Applied Probability Trust (20 December 2013)

NON-NORMAL SMALL JUMP APPROXIMATION OF INFINITELY DIVISIBLE DISTRIBUTIONS

ZHIYI CHI,* *University of Connecticut*

Abstract

We study a type of non-Normal small jump approximation of infinitely divisible distributions. By incorporating compound Poisson, Gamma, and Normal distributions, the approximation has a higher order of cumulant matching than its Normal counterpart, and hence in many cases a higher rate of approximation error decay as the cut-off for jump size tends to 0. The parameters of the approximation are easy to fix, and its random sampling has the same order of computational complexity as the Normal one. An error bound of the approximation in terms of total variation distance is derived. Simulations show empirically the non-Normal approximation can have significantly smaller error than its Normal counterpart.

Keywords: Infinitely divisible; Normal approximation; compound Poisson approximation; Gamma approximation; cumulant matching; sampling

2010 Mathematics Subject Classification: Primary 60E07
Secondary 60G51

1. Introduction

Simulation of infinitely divisible (i.d.) random variables has many applications. In most cases, since exact simulation is either unavailable or computationally costly, good approximation methods are desired. The Normal approximation of i.d. distributions, which was studied in [24] and later developed in [1] under the framework of small jump approximation, has received much attention in the literature [2, 11–13, 17, 20, 22, 23, 34, 35].

Let X be a real-valued i.d. random variable and λ its Lévy measure. Without loss of generality, we will always assume that X has no Normal component. The Normal (small jump) approximation starts with the decomposition $X = X_r + \Delta_r$ given $r > 0$, where X_r and Δ_r are independent and i.d. with Lévy measures $\lambda_r(\mathrm{d}x) = \mathbf{1}\{|x| < r\}\,\lambda(\mathrm{d}x)$ and $\lambda - \lambda_r$, respectively. As Δ_r is compound Poisson, its sampling is standard. The key is to approximate X_r by a Normal random variable with the same mean and variance [1]. Thus one can regard the approximation as relying on second-order cumulant matching. By certain measure, the error of the approximation is bounded by

$$C|\kappa|_{3,X_r}/\sigma^3_{X_r}, \tag{1}$$

where C is a universal constant, σ_{X_r} the standard deviation of X_r, and for $j \ge 2$, $|\kappa|_{j,X_r} = \int |x|^j \lambda_r(\mathrm{d}x)$ [1]. Currently, the smallest available C seems to be 0.4748 [33]. In symmetric cases, since the third cumulants of X_r and the Normal random variable are 0, the bound can be more or less replaced with $C|\kappa|_{4,X_r}/\sigma^4_{X_r}$ [1]. The pattern of the bound suggests that, if X_r

* Postal address: Department of Statistics, Glenbrook Road, U-4120, Storrs, CT 06269, USA.
* Email address: zhiyi.chi@uconn.edu

and some Y_r have the same cumulants of order 1, ..., $q-1$ with $q \ge 5$, then X_r might be approximated by Y_r with the error being bounded by

$$C(r)(|\kappa|_{q,X_r} + |\kappa|_{q,Y_r})/\sigma^q_{X_r},$$

where $C(r)$ is nearly constant, at least for small r, or most ideally is a small universal constant. Since the q^{th} cumulant of a Normal random variable is 0, the bound is consistent with the one for the Normal approximation.

Even by rough Fourier analysis, there is good reason to expect the above bound to be true, and elementary calculations indicate that in many cases it vanishes at a strictly higher rate than the bound for the Normal approximation as $r \to 0$. However, perhaps one need first ask if an approximation based on higher-order cumulant matching can possibly be implemented easily. The meaning of the question is twofold. First, the approximating distribution should be easy to identify and preferably i.d. Second, the approximating distribution should be easy to sample; preferably, the computational complexity of the sampling is of the same order as the Normal approximation. If the answer to the question is positive, then another question is how large q can be. It can be anticipated that the larger q is, the faster the error of approximation vanishes as $r \to 0$. If both questions are answered, then a wide range of available techniques can potentially be adapted to establish error bounds.

Clearly, cumulant matching is equivalent to moment matching. In fact, our proof of the above type of bound ultimately relies on moment matching. However, thanks to the Lévy-Khintchine representation, it is much more convenient to work on cumulants than moments. In Section 2.1, we present a simple way to construct approximating i.d. distributions with matching cumulants up to at least the fourth order, i.e., $q \ge 5$. In many important cases, we get $q = 6$, and in symmetric cases, $q = 10$. Each approximating distribution is a convolution of compound Poisson and Normal distributions, with the former made from Gamma variables. Importantly, using standard algorithms [14, 21], the computational complexity of sampling for the approximation is universally bounded, and hence is of the same order as for the Normal approximation.

We refer to the approximation as Poisson-Gamma-Normal (PGN) approximation, although a term like "compound Poisson-Normal small jump approximation with Gamma summands and higher order of cumulant matching" should be more accurate. In Section 2.2, we bound its error in terms of total variation distance. The bound is non-asymptotic and of the desired type. Section 2.3 gives some examples of the PGN approximation. The examples show that the bound yields substantially higher rate of precision than the Normal approximation as $r \to 0$. However, they also indicate that the bound may be far from being optimal or even practically useless. Therefore, in Section 3, we conduct large scale simulations to show that empirically, the PGN approximation can have significantly smaller error than the Normal one.

In Section 4, we prove the error bound by combining Fourier analysis, Lindeberg method, and a device in [1]. Of course, on modern treatments of Poisson, compound Poisson, and Normal approximations, there is now an extensive literature, and on Gamma and other types of approximations, there is also a growing literature; see [3, 4, 7, 8, 18, 25, 26, 30] and references therein. However, it appears that there has been little work on using convolutions of different types of simple distributions to improve approximation. We remark that while this paper only concerns the univariate case, accompanying results on the multivariate case have been reported elsewhere; see [10].

The rest of the section fixes notations and recalls basic facts. A Borel measure λ on $\mathbb{R}$ is the Lévy measure of an i.d. distribution $\iff \lambda(\{0\}) = 0$ and $\int(u^2 \wedge 1)\,\lambda(\mathrm{d}u) < \infty$ [32, Theorem

8.1]. Denote by $\mathrm{sppt}(\lambda)$ the support of λ. If X is i.d. with Lévy measure λ, denote by $\Psi_X(t)$ and $\psi_X(t) = \exp\{-\Psi_X(t)\} = \mathrm{E}[e^{\mathrm{i}tX}]$ its characteristic exponent and characteristic function, respectively, and

$$\kappa_{j,X} = \frac{\mathrm{d}^j}{\mathrm{d}t^j} \ln \mathrm{E}[e^{tX}]\bigg|_{t=0}, \quad |\kappa|_{j,X} = \int |u|^j\, \lambda(\mathrm{d}u), \quad j \in \mathbb{N},$$

and $\sigma_X = \kappa_{2,X}^{1/2}$. $\kappa_{j,X}$ is known as the j^{th} cumulant of X. It is well-defined if $\mathrm{E}[e^{tX}] < \infty$ for all t in a neighborhood of 0. We refer to $|\kappa|_{j,X}$ as the j^{th} absolute cumulant of X. For $a > 0$, $\mathrm{E}|X|^a < \infty \iff \int \mathbf{1}\{|u| > 1\}\, |u|^a\, \lambda(\mathrm{d}u) < \infty$ ([32], Theorem 25.3 & Proposition 25.4). If $\mathrm{sppt}(\lambda)$ is bounded, then $\mathrm{E}[e^{tX}] < \infty$ for all t [32, Theorem 25.17]. Since X has no Normal component, $\kappa_{j,X} = \int u^j\, \lambda(\mathrm{d}u)$, $j > 1$, and $\mathrm{Var}(X) = \kappa_{2,X}$. If in addition $\Psi_X(t) = \int (1 + \mathrm{i}tu - e^{\mathrm{i}tu})\, \lambda(\mathrm{d}u)$, then $\kappa_{1,X} = \mathrm{E}X = 0$. If j is even or $\mathrm{sppt}(\lambda) \subset \mathbb{R}_+ := [0, \infty)$, then $\kappa_{j,X} = |\kappa|_{j,X}$. Denote the total variation distance of X and random variable Y by $d_{\mathrm{TV}}(X, Y) = \sup\{|\mathrm{P}\{X \in A\} - \mathrm{P}\{Y \in A\}| : A \text{ measurable}\}$ and their Kolmogorov-Smirnov (KS) distance by $d_{\mathrm{KS}}(X, Y) = \sup\{|\mathrm{P}\{X \le x\} - \mathrm{P}\{Y \le x\}| : x \in \mathbb{R}\}$.

Henceforth, we set X_r such that

$$\Psi_{X_r}(t) = \int (1 + \mathrm{i}tu - e^{\mathrm{i}tu})\, \lambda_r(\mathrm{d}u). \tag{2}$$

Consequently, $\mathrm{E}X_r = 0$ and $\mathrm{E}|X_r|^p < \infty$ for all $p > 0$. Let $Z \sim N(0, 1)$ be independent of Δ_r. Then Δ_r and $\sigma_{X_r} Z + \Delta_r$ are known as the compound Poisson (CP) and Normal approximations of X, respectively [1]. We refer to r as the cut-off for jump size for the approximations.

2. PGN approximation

2.1. Cumulant matching

In general, one can decompose $X = X_+ - X_- + X_s$, where $X_\pm$ and X_s are independent and i.d., with the Lévy measures of $X_\pm$ being supported on $\mathbb{R}_+$, and X_s being symmetric, i.e., $X_s \sim -X_s$. Indeed, the Lévy measure of X_s can be any symmetric Borel measure λ_s such that for $A \subset \mathbb{R}_+$, $\lambda_s(A) \le \min(\lambda(A), \lambda(-A))$, and the Lévy measures $\lambda_\pm$ of $X_\pm$ are defined by $\lambda_\pm(A) = \lambda(\pm A) - \lambda_s(A)$. Although one can always set $\lambda_s = 0$, as seen below, it is useful to exploit $\lambda_s \ne 0$.

Thus, without loss of generality, we will only consider the asymmetric case where $\mathrm{sppt}(\lambda) \subset \mathbb{R}_+$ and the symmetric case. First, let $\mathrm{sppt}(\lambda) \subset \mathbb{R}_+$. Given $r > 0$ and $p \ge -1$, let Y_r be an i.d. random variable independent of Δ_r such that

$$\Psi_{Y_r}(t) = \int_0^\infty (1 + \mathrm{i}tu - e^{\mathrm{i}tu})\, \gamma_r(\mathrm{d}u) \quad \text{with } \gamma_r(\mathrm{d}u) = m(r) u^p e^{-u/s(r)}\, \mathrm{d}u, \tag{3}$$

where $m(r) > 0$ and $s(r) > 0$ are constants to be determined. Then let

$$T_r = Y_r + \sigma(r) Z \quad \text{with } Z \sim N(0, 1) \text{ independent of } (Y_r, \Delta_r), \tag{4}$$

where $\sigma(r) > 0$ is a constant that also needs to be determined.

To use $T_r + \Delta_r$ to approximate X, first a comment on the sampling of T_r, which boils down to that of Y_r. Since $Y_r = U - \mathrm{E}U$, where $U \ge 0$ is i.d. with Lévy density $m(r)\mathbf{1}\{u > 0\}\, u^p e^{-u/s(r)}$ and $\mathrm{E}U = \Gamma(p+2) m(r) s(r)^{p+2}$, the sampling of Y_r is reduced to that of U. If $p = -1$, then $U \sim \mathrm{Gamma}(m(r), s(r))$, the Gamma distribution with shape parameter $m(r)$ and scale parameter

$s(r)$, whose sampling has universally bounded complexity ([14], p. 407–420). If $p > -1$, then $U \sim \sum_{i=1}^N \xi_i$, where $N \sim \text{Poisson}(a)$ with $a = \int_0^\infty m(r) u^p e^{-u/s(r)}\, \mathrm{d}u = \Gamma(p+1) m(r) s(r)^{p+1}$, and ξ_i are i.i.d. $\text{Gamma}(p+1, s(r))$ random variables independent of N. The sampling of $\text{Poisson}(a)$ has universally bounded complexity ([15] or [21], p. 228–241). Then, because conditional on N, $U \sim \text{Gamma}(N(p+1), s(r))$, the sampling of U, and hence that of T_r, has the same order of complexity as the sampling of $N(0,1)$.

Due to its Lévy-Khintchine representation, we refer to $T_r + \Delta_r$ as the PGN approximation of X with cut-off r. To match the cumulants of X_r and T_r, note that $\mathrm{E} T_r = \mathrm{E} Y_r = 0$ and

$$\begin{cases} \kappa_{j,Y_r} = \Gamma(j+p+1) m(r) s(r)^{j+p+1}, \\ \kappa_{j,T_r} = \kappa_{j,Y_r} + \mathbf{1}\{j=2\}\, \sigma(r)^2. \end{cases} \qquad j \ge 2. \tag{5}$$

We next show two results. The first one allows $r = \infty$ and hence applies to any i.d. random variable with finite 4^{th} cumulant, subject to a mild constraint. However, it only attains 4^{th}-order matching. The second result asserts that one can obtain 5^{th}-order cumulant matching provided the existence of a slowly varying Lévy density at 0+.

Proposition 2.1. (Fourth-order cumulant matching.) *Let $0 < r \le \infty$ and $0 < \kappa_{4,X_r} < \infty$. Suppose λ_r is not concentrated on a single point. Then*

$$\frac{p+4}{p+3} < \frac{\kappa_{2,X_r} \kappa_{4,X_r}}{\kappa_{3,X_r}^2} \quad \text{for all large } p. \tag{6}$$

For any $p \ge -1$ satisfying (6), if

$$s(r) = \frac{\kappa_{4,X_r}}{(p+4)\kappa_{3,X_r}}, \quad m(r) = \frac{\kappa_{3,X_r}}{\Gamma(p+4) s(r)^{p+4}}, \tag{7}$$

and if Y_r is defined by (3), then $\kappa_{2,X_r} > \kappa_{2,Y_r}$, and by setting

$$\sigma(r) = (\kappa_{2,X_r} - \kappa_{2,Y_r})^{1/2}, \tag{8}$$

$\kappa_{j,X_r} = \kappa_{j,T_r}$ for $2 \le j \le 4$.

Remark. If $\lambda(\mathbb{R}_+) = \infty$, then for any $r > 0$, $\lambda_r \ne 0$ and is not concentrated on a single point.

Proof. Since λ_r is not concentrated on a single point and $0 < \kappa_{4,X_r} < \infty$, by Hölder inequality, $0 < \kappa_{3,X_r}^2 < \kappa_{2,X_r}\kappa_{4,X_r} < \infty$, which implies (6). From (5), by setting $s(r)$ and $m(r)$ as in (7), $\kappa_{j,X_r} = \kappa_{j,Y_r}$ for $j = 3, 4$ and

$$\kappa_{2,Y_r} = \Gamma(p+3) m(r) s(r)^{p+3} = \Gamma(p+3) \frac{\kappa_{3,X_r}}{\Gamma(p+4) s(r)} = \frac{(p+4)\kappa_{3,X_r}^2}{(p+3)\kappa_{4,X_r}}.$$

Then for $p \ge -1$ satisfying (6), $\kappa_{2,Y_r} < \kappa_{2,X_r}$. The rest of the result is then clear.

Proposition 2.2. (Fifth-order cumulant matching.) *Let $\lambda(\mathrm{d}u) = \mathbf{1}\{u > 0\}\, u^{-\alpha-1} \ell(u)\, \mathrm{d}u$, with $\alpha \in (0,2)$ and $\ell(u)$ slowly varying at 0+. Let $p = p(r)$ be defined by the equation*

$$\frac{1}{p+4} = \frac{\kappa_{3,X_r}\kappa_{5,X_r}}{\kappa_{4,X_r}^2} - 1.$$

Then for all small $r > 0$, $p > -1$ and satisfies (6). For any $r > 0$ with such p, set $s(r)$, $m(r)$ by (7) and $\sigma(r)$ by (8). Then $\kappa_{j,X_r} = \kappa_{j,T_r}$, $2 \le j \le 5$.

Proof. For $j \geq 3$, as $r \to 0+$, $\kappa_{j,X_r} = \int_0^r u^{j-1-\alpha}\ell(u)\,\mathrm{d}u \sim r^{j-\alpha}\ell(r)/(j-\alpha)$ [5, Theorem 1.5.11], so

$$\frac{1}{p+4} \sim \frac{(4-\alpha)^2}{(3-\alpha)(5-\alpha)} - 1 = \frac{1}{(3-\alpha)(5-\alpha)}$$
$$\implies p \sim \alpha^2 - 8\alpha + 11 > -1.$$

As a result, for all small $r > 0$, $p > -1$. Furthermore, as

$$\frac{\kappa_{2,X_r}\kappa_{4,X_r}}{\kappa_{3,X_r}^2} \sim \frac{(3-\alpha)^2}{(2-\alpha)(4-\alpha)} = 1 + \frac{1}{\alpha^2 - 6\alpha + 8},$$

combining with the previous display, it is not hard to get (6). By Proposition 2.1, it only remains to show given $r > 0$ such that $p > -1$ and satisfies (6), $\kappa_{5,X_r} = \kappa_{5,Y_r}$. However, this follows from $\kappa_{3,X_r}\kappa_{5,X_r}/\kappa_{4,X_r}^2 = (p+5)/(p+4) = \kappa_{3,Y_r}\kappa_{5,Y_r}/\kappa_{4,Y_r}^2$, where the second equality is due to (5).

Now consider the symmetric case. Let $X = X^{(1)} - X^{(2)}$, where $X^{(i)}$ are i.i.d. with Lévy measure λ supported in $\mathbb{R}_+$. Let $X_r = X_r^{(1)} - X_r^{(2)}$ and approximate it by $T_r = T_r^{(1)} - T_r^{(2)}$, where $T_r^{(i)}$ are i.i.d. defined in (4). Since all the odd-ordered cumulants of X_r and T_r are 0, it suffices to match their even-ordered cumulants. The next result asserts that in general one can match their cumulants up to the 7^{th} order, and in some important cases up to the 9^{th} order.

Proposition 2.3. (Symmetric case.) *1) Let $0 < r \leq \infty$ and $0 < \kappa_{4,X_r} < \infty$. Suppose λ_r is not concentrated on a single point. Then*

$$\frac{(p+5)(p+6)}{(p+3)(p+4)} < \frac{\kappa_{2,X_r}\kappa_{6,X_r}}{\kappa_{4,X_r}^2} \quad \textit{for all large } p. \tag{9}$$

For any $p \geq -1$ satisfying (9), if

$$s(r) = \sqrt{\frac{\kappa_{6,X_r}}{(p+5)(p+6)\kappa_{4,X_r}}}, \quad m(r) = \frac{\kappa_{4,X_r}}{2\Gamma(p+5)s(r)^{p+5}}, \tag{10}$$

and $Y_r = Y_r^{(1)} - Y_r^{(2)}$ with $Y_r^{(i)}$ i.i.d. defined by (3), then $\kappa_{2,X_r} > \kappa_{2,Y_r}$ and by setting $\sigma(r)$ by (8), $\kappa_{j,X_r} = \kappa_{j,T_r}$ for $2 \leq j \leq 7$.

2) Let $\lambda(\mathrm{d}u) = \mathbf{1}\{u > 0\}\, u^{-\alpha-1}\ell(u)\,\mathrm{d}u$, with $\alpha \in (0,2)$ and $\ell(u)$ slowly varying at $0+$. Then for all small $r > 0$, there is a unique $p = p(r) > 0$ satisfying (9) and

$$\frac{(p+7)(p+8)}{(p+5)(p+6)} = \frac{\kappa_{4,X_r}\kappa_{8,X_r}}{\kappa_{6,X_r}^2}. \tag{11}$$

Given $r > 0$ with such p, set $s(r)$, $m(r)$ by (10) and $\sigma(r)$ by (8). Then $\kappa_{j,X_r} = \kappa_{j,T_r}$, $2 \leq j \leq 9$.

Proof. 1) By the assumption and Hölder inequality, $0 < \kappa_{4,X_r}^2 < \kappa_{2,X_r}\kappa_{6,X_r} < \infty$, so (9) holds for all large p. Since for even j, $\kappa_{j,Y_r} = 2\kappa_{j,Y_r^{(1)}} = 2\Gamma(j+p+1)m(r)s(r)^{j+p+1}$, it is easy to see $\kappa_{4,X_r} = \kappa_{4,Y_r}$ and $\kappa_{6,X_r} = \kappa_{6,Y_r}$. On the other hand, for all odd j, $\kappa_{j,X_r} = \kappa_{j,Y_r} = 0$. Finally, by similar argument for Proposition 2.1, $\kappa_{2,Y_r} < \kappa_{2,X_r}$, leading to $\kappa_{j,X_r} = \kappa_{j,T_r}$ for $2 \leq j \leq 7$.

2) Following the proof of Proposition 2.2, as $r \to 0+$,

$$\frac{\kappa_{4,X_r}\kappa_{8,X_r}}{\kappa^2_{6,X_r}} \sim \frac{(6-\alpha)^2}{(4-\alpha)(8-\alpha)} = 1 + \frac{4}{(4-\alpha)(8-\alpha)} := h(\alpha).$$

The function h is strictly increasing on $(0, 2)$. On the other hand,

$$g(p) := \frac{(p+7)(p+8)}{(p+5)(p+6)} = \left(1 + \frac{2}{p+5}\right)\left(1 + \frac{2}{p+6}\right)$$

is strictly decreasing on $(-1, \infty)$, with $g(0) > h(2) > h(\alpha) > h(0) > 1 = g(\infty)$. Therefore, there is a unique $p > 0$ satisfying (11). We have to show that for this $p = p(r)$, (9) is satisfied for all small $r > 0$. By continuity, it suffices to show that for $p > 0$,

$$\frac{(p+7)(p+8)}{(p+5)(p+6)} = \frac{(6-\alpha)^2}{(4-\alpha)(8-\alpha)} \implies \frac{(p+5)(p+6)}{(p+3)(p+4)} < \frac{(4-\alpha)^2}{(2-\alpha)(6-\alpha)}.$$

By calculation, the equality is equivalent to $2p^2 = 2p(\alpha^2 - 12\alpha + 21) + 13\alpha^2 - 156\alpha + 356$, while the inequality is equivalent to $2p^2 > 2p(\alpha^2 - 8\alpha + 5) + 9\alpha^2 - 72\alpha + 84$. Then, by $p > 0$ and $0 < \alpha < 2$, it is not hard to see the implication holds. The rest of the proof then follows the proof for 1).

Propositions 2.2 and 2.3 immediately lead to the next result on i.d. distributions with truncated stable Lévy measures. It should be pointed out that simple exact sampling methods for stable distributions are well-known [14] and i.d. distributions with truncated stable Lévy measures with $\alpha \in (0, 1)$ can be sampled exactly but with high computational complexity [9].

Corollary 2.1. *Let* $\lambda(\mathrm{d}u) = c\mathbf{1}\{0 < u < r_0\}\, u^{-\alpha-1}\,\mathrm{d}u$, *where* $c > 0$, $0 < r_0 \le \infty$, *and* $\alpha \in (0, 2)$.

1) Let X *have Lévy measure* λ. *If* $p = \alpha^2 - 8\alpha + 11$, *then* $p > -1$ *and for all* $0 < r < r_0$, *by setting* $(s(r), m(r))$ *by (7),* $\kappa_{2,X_r} > \kappa_{2,Y_r}$, *and by setting* $\sigma(r)$ *by (8),* $\kappa_{j,X_r} = \kappa_{j,T_r}$ *for* $2 \le j \le 5$.

2) Let $X = X^{(1)} - X^{(2)}$, *with* $X^{(i)}$ *i.i.d. with Lévy measure* λ. *If* p *is the (unique) solution to*

$$\frac{(p+7)(p+8)}{(p+5)(p+6)} = \frac{(6-\alpha)^2}{(4-\alpha)(8-\alpha)}, \quad p \in (0, \infty)$$

then for all $0 < r \le r_0$, *by setting* $(s(r), m(r))$ *by (10),* $\kappa_{2,X_r} > \kappa_{2,Y_r}$, *and by setting* $\sigma(r)$ *by (8),* $\kappa_{j,X_r} = \kappa_{j,T_r}$ *for* $2 \le j \le 9$.

2.2. Error bound for PGN approximation

Denote $C_0 = (\sin 1)^2/2 = 0.354...$ Observe that by $|\kappa|_{j+1,X_r} \le r|\kappa|_{j,X_r}$, $j \ge 2$, for $s(r)$ defined in (7) or (10), $s(r) < r/(p+3)$. The main result is the following.

Theorem 2.1. *Fix* $r \in (0, \infty)$ *and* $q \ge 5$. *Let* T_r *be defined by (3)–(4) in the asymmetric case, or* $T_r = T_r^{(1)} - T_r^{(2)}$ *in the symmetric case, where* $T_r^{(i)}$ *are i.i.d. defined by (3)–(4). Suppose* $s(r) < r/(p+3)$ *and* $\sigma(r) > 0$. *For* $j \ge 1$, *define*

$$Q_j(r) = \left[\frac{\Gamma(j+1/2)}{2C_0^{j+1/2}} + \sigma_{X_r}^{2j+1}\int_{1/r}^{\infty} t^{2j} e^{-2L(t,r)}\,\mathrm{d}t\right]^{1/2},$$

where $L(t,r) = t^2 \min\{C_0 \kappa_{2,X_{1/|t|}},\ \sigma(r)^2/2\}$. *Suppose* $\kappa_{j,X_r} = \kappa_{j,T_r}$ *for* $2 \le j < q$. *Then*

$$d_{\mathrm{TV}}(X, T_r + \Delta_r) \le \frac{|\kappa|_{q,X_r} + |\kappa|_{q,Y_r}}{q!\sigma^q_{X_r}}[qQ_{q-1}(r) + Q_q(r) + Q_{q+1}(r)]. \tag{12}$$

Remark.

1. The bound in (12) is on d_{TV} instead of the more commonly used d_{KS} [1, 24]. However, we have not been able to derive a Berry-Esseen type of bound $C(|\kappa|_{q,X_r} + |\kappa|_{q,Y_r})/\sigma^q_{X_r}$, with C a universal constant only depending on q. It appears that some key ingredients for the proof of the Berry-Esseen bound for the Normal approximation cannot hold for higher order approximations. Also, as seen later, the bound sometimes is quite suboptimal.
2. The bound will be proved by combining Fourier analysis, the Lindeberg method (cf. [6] for a modern application of it), and a device in [1] (cf. the proof of Theorem 25.18 in [32]). Although a bound on d_{KS} may be established solely based on Fourier analysis [24], our proof seems to be more transparent and suitable for generalization.
3. As seen later, in order for the right hand side (RHS) of (12) to be finite, X_r must have a density, in particular, $\lambda(\mathbb{R}) = \infty$. The last condition implies that X is not compound Poisson and has no atom [32, Theorem 27.4]. It also excludes lattice distributions, to which the Poisson-Charlier approximation applies [3, 24]. Although the condition $\lambda(\mathbb{R}) = \infty$ is necessary for X to have a valid Normal small jump approximation, it is not sufficient, which is a fact with several interesting and important implications [1, 12].

To evaluate the RHS of (12), we need to evaluate $\sigma(r)^2$ and $|\kappa|_{q,Y_r}$. Both can be expressed in terms of the cumulants of X_r. For the asymmetric case, the proof of Proposition 2.1 shows $\kappa_{2,Y_r} = (p+4)\kappa^2_{3,X_r}/[(p+3)\kappa_{4,X_r}]$. Then by (8),

$$\sigma(r)^2 = \kappa_{2,X_r} - \frac{(p+4)\kappa^2_{3,X_r}}{(p+3)\kappa_{4,X_r}}.$$

Similarly, for the symmetric case, using (5) and (10),

$$\sigma(r)^2 = \kappa_{2,X_r} - \frac{(p+5)(p+6)\kappa^2_{4,X_r}}{(p+3)(p+4)\kappa_{6,X_r}}.$$

Furthermore, if $\kappa_{j,X_r} = \kappa_{j,T_r}$, for $2 \le j < q$, then $|\kappa|_{q,Y_r}$ can also be expressed in terms of the absolute cumulants of X_r. In the asymmetric case, by (5),

$$|\kappa|_{q,Y_r} = \kappa_{q,Y_r} = (q+p)s(r)\kappa_{q-1,Y_r} = (q+p)s(r)\kappa_{q-1,X_r}.$$

Similarly, in the symmetric case, if q is even, then

$$\begin{aligned} |\kappa|_{q,Y_r} = \kappa_{q,Y_r} &= (q+p)(q+p-1)s(r)^2\kappa_{q-2,Y_r} \\ &= (q+p)(q+p-1)s(r)^2\kappa_{q-2,X_r}. \end{aligned}$$

In the bound (12), the $Q_j(r)$'s look rather technical. The next result gives their asymptotics as $r \to 0+$.

Proposition 2.4. *For* $b \in (0,1)$ *and* $q \ge 3$, *there is* $M = M(b,q) > 0$, *such that if*

$$\limsup_{s\to 0} \frac{\kappa_{2,Y_s}}{\kappa_{2,X_s}} < b, \quad \liminf_{s\to 0} \frac{\kappa_{2,X_s}}{s^2\ln(1/s)} > M, \tag{13}$$

then for any $2 \le j \le q+1$,

$$Q_j(r)^2 = \frac{\Gamma(j+1/2)}{2C_0^{j+1/2}} + o(1), \quad r \to 0.$$

Here is a short proof. By (13), for small $r > 0$, $\sigma(r)^2 = \kappa_{2,X_r} - \kappa_{2,Y_r} > (1-b)\kappa_{2,X_r}$. Then, from the increasing monotonicity of κ_{2,X_r} in r, there is a constant $c = c(b) > 0$, such that for $t \ge 1/r$, $L(t,r) \ge ct^2\kappa_{2,X_{1/t}}$. Consequently, if $M \ge (q+2)/c$, then by (13), for $t \ge 1/r$, $L(t,r) \ge Mc \ln t \ge (q+2)\ln t$, and hence for all $2 \le j \le q+1$,

$$\int_{1/r}^{\infty} t^{2j} e^{-2L(t,r)}\,\mathrm{d}t \le \int_{1/r}^{\infty} t^{2(q+1)-2Mc}\,\mathrm{d}t = o(1), \quad r \to 0.$$

Since $\sigma_{X_r} = o(1)$ as $r \to 0$, the proof is complete.

2.3. Examples

Example 2.1. (*Truncated stable Lévy measure.*) Let $\lambda(\mathrm{d}u) = c\mathbf{1}\{0 < u < r_0\}\, u^{-\alpha-1}\,\mathrm{d}u$ with $c > 0$, $0 < r_0 \le \infty$, and $\alpha \in (0,2)$. By Corollary 2.1, given $r \in (0, r_0)$, if $p = \alpha^2 - 8\alpha + 11$, and $s(r)$, $m(r)$ and $\sigma(r)$ are set by (7)–(8), then $\kappa_{j,X_r} = \kappa_{j,T_r}$ for $2 \le j < q = 6$. To apply (12), we need to know κ_{2,X_r}, κ_{6,X_r}, and κ_{6,Y_r}. For $j \ge 2$, $\kappa_{j,X_r} = cr^{j-\alpha}/(j-\alpha)$. Then

$$s(r) = \frac{\kappa_{4,X_r}}{(p+4)\kappa_{3,X_r}} = \frac{(3-\alpha)r}{(p+4)(4-\alpha)} = \frac{r}{(4-\alpha)(5-\alpha)},$$

$$\kappa_{2,Y_r} = \frac{\kappa_{3,Y_r}}{(p+3)s(r)} = \frac{\kappa_{3,X_r}}{(p+3)s(r)} = \frac{c(4-\alpha)(5-\alpha)r^{2-\alpha}}{(3-\alpha)(\alpha^2-8\alpha+14)},$$

and $\kappa_{6,Y_r} = (6+p)s(r)\kappa_{5,Y_r} = (6+p)s(r)\kappa_{5,X_r} = cA(\alpha)r^{6-\alpha}$, with $A(\alpha) = (\alpha^2-8\alpha+17)/(4-\alpha)(5-\alpha)^2$. Therefore, by Theorem 2.1,

$$d_{\mathrm{TV}}(X, T_r + \Delta_r) \le (2-\alpha)^3 \left[\frac{1}{6-\alpha} + A(\alpha)\right] \times \frac{6Q_5(r) + Q_6(r) + Q_7(r)}{6!} \times (r^\alpha/c)^2. \tag{14}$$

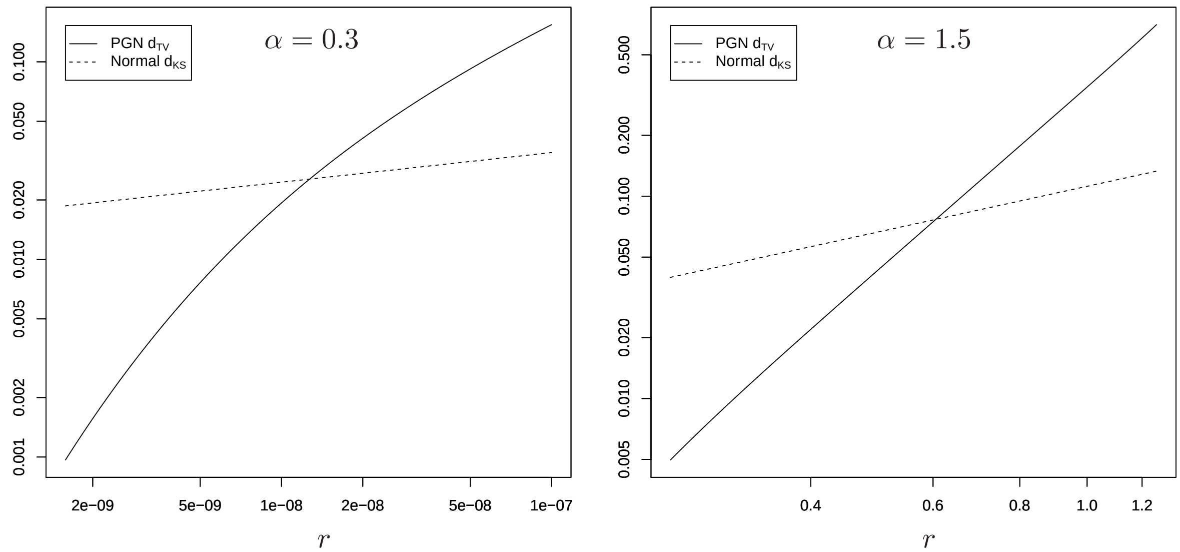
Since $0 < \kappa_{2,Y_r}/\kappa_{2,X_r} < 1$ is a constant independent of r, and λ satisfies Orey's condition $\liminf_{s\to 0} \kappa_{2,X_s}/s^{2-\alpha} > 0$ ([28]; also see [32], Proposition 28.3), the conditions in (13) are satisfied no matter the value of M. Then by Proposition 2.4, $d_{\mathrm{TV}}(X, T_r + \Delta_r) = O(r^{2\alpha})$. This may be compared to the Normal approximation in [1, 24], where d_{KS} between X and its Normal approximation is $O(r^{\alpha/2})$ when X is asymmetric. Specifically, by (1),

$$d_{\mathrm{KS}}(X, \sigma_{X_r} Z + \Delta_r) \le \frac{C(2-\alpha)^{3/2}}{(3-\alpha)} \times (r^\alpha/c)^{1/2}, \quad C = 0.4748. \tag{15}$$

Furthermore, if $X = X^{(1)} - X^{(2)}$ is symmetric, where $X^{(i)}$ are i.i.d. with Lévy measure λ, then by 2) of Corollary 2.1, it can be seen that we can set $q = 10$ and get $d_{\mathrm{TV}}(X, T_r + \Delta_r) = O(r^{4\alpha})$, whereas the d_{KS} between X and its Normal approximation in this case is $O(r^\alpha)$ [1].

Although the bound in (14) vanishes at a higher rate than the one in (15) as $r \to 0$, the asymptotic result says little about how the bounds compare if r is not too small. This is especially the case when $\alpha < 1$. In Fig. 1, for $c = 1$ and $r_0 = \infty$, the bounds are plotted as functions of r. The bound in (14) is evaluated numerically; see Appendix for detail. As seen from the plots, for $\alpha = 1.5$, the bound in (14) is smaller than the one in (15) once $r < 0.6$, whereas for $\alpha = 0.3$, this happens only if $r < 2 \times 10^{-8}$. Therefore, (14) may provide little evidence on whether the PGN approximation is better than the Normal approximation in practice. To address this issue, we resort to numerical simulation in next section.

FIGURE 1: The bounds in (14) and (15) as functions of r, for $\alpha = 0.3$ and 1.5, respectively, with $\lambda(\mathrm{d}x) = \mathbf{1}\{u > 0\}\, u^{-\alpha-1}\,\mathrm{d}x$. The axes are plotted in logarithmic scale with base 10.



Example 2.2. (*Tempered stable.*) Let $\lambda(\mathrm{d}u) = \mathbf{1}\{u > 0\}\, u^{-\alpha-1}\exp(-u^\theta)\,\mathrm{d}u$, where $\alpha \in (0,2)$ and $\theta > 0$. Then for $j \geq 2$,

$$\kappa_{j,X_r} = \int_0^r u^{j-\alpha-1}\exp(-u^\theta)\,\mathrm{d}u = \frac{1}{\theta}\int_0^{r^\theta} u^{(j-\alpha)/\theta-1}e^{-u}\,\mathrm{d}u.$$

The last integral can be numerically evaluated as an incomplete Gamma function [27]. However, it has no closed formulas. The following method can be used if one wants to avoid the problem. Recall that for any odd $n \geq 1$, $e^{-u} \geq f_n(u)$ for $u \geq 0$, where $f_n(u) = \sum_{i=0}^n (-u)^i/i!$. Let $n = 2\lfloor\alpha/(2\theta)\rfloor + 1$, which is the smallest odd number greater than $\alpha/\theta - 1$. Let $F(u) = \mathbf{1}\{0 < u < r_0\}\, f_n(u^\theta)$, with $r_0 = \sup\{r > 0 : f_n(u) > 0 \text{ for all } 0 \leq u < r^\theta\}$. Decompose $\lambda = \lambda_1 + \lambda_2$, where $\lambda_1(\mathrm{d}u) = \mathbf{1}\{u > 0\}\, u^{-\alpha-1}F(u)\,\mathrm{d}u$. It is easy to evaluate $\int_0^r u^j \lambda_1(\mathrm{d}u)$. Then we can apply the PGN approximation to λ_1, with all parameters set in closed form. Meanwhile, since $u^{-\alpha-1}[\exp(-u^\theta) - F(u)] = O(u^{(n+1)\theta-\alpha-1})$ as $u \to 0+$, λ_2 has finite mass, and hence corresponds to a compound Poisson random variable that can be sampled exactly. If X, X', and X'' denote i.d. random variables with Lévy measures λ, λ_1, and λ_2, respectively, and Δ_r and T_r the i.d. random variables from the PGN approximation to X', then by Proposition 2.2, $d_{\mathrm{TV}}(X, T_r + \Delta_r + X'') \leq d_{\mathrm{TV}}(X', T_r + \Delta_r) = O(r^{2\alpha})$.

Example 2.3. Let $\lambda(\mathrm{d}u) = c\mathbf{1}\{0 < u < 1\}\, u^{-1}\ln(1/u)\,\mathrm{d}u$. Since

$$\int_{u<r} u^2\,\lambda(\mathrm{d}u) = c\int_0^r u\ln(1/u)\,\mathrm{d}u = \frac{cr^2[2\ln(1/r) + 1]}{4},$$

by Proposition 2.1 in [1], the Normal approximation is valid in the sense that its error in terms of d_{KS} tends to 0 as $r \to 0$. However, since for $|t| \gg 1$, $L(t,r) = C_0 t^2 \int_{u<1/|t|} u^2\,\lambda(\mathrm{d}u) \sim cC_0 \ln|t|$, it can be seen that the RHS of (12) is finite only when c is large enough, and even in that case the RHS of (12) decreases to 0 very slowly as $r \to 0$.

3. A numerical study

As seen in Section 2.3, the bound (12) sometimes is a poor indicator on the precision of the PGN approximation in practice. To address this issue, we conducted a simulation study to compare empirically the errors of the PGN and Normal approximations in terms of the KS distance. We exclusively considered the approximations to stable distributions for two reasons. First, besides Gamma distributions, they are the only class of non-Normal i.d. distributions whose Lévy measures and distribution functions are both known (semi-)explicitly. Second, unlike Gamma distributions, they have valid Normal approximations [1]. The simulations were implemented in R language [29], with its `stabledist` package used for all computations involving stable distributions.

The Lévy measure of a stable distribution with exponent $\alpha \in (0,2)$ was parametrized as

$$\lambda(\mathrm{d}u) = [M_+ \mathbf{1}\{u>0\} + M_- \mathbf{1}\{u<0\}]|u|^{-1-\alpha}\,\mathrm{d}u,$$

where $M_\pm \ge 0$ such that $M := M_+ + M_- > 0$. For simplicity, the stable distribution to be approximated ("target distribution") was centered in the sense that

$$\Psi_X(t) = \begin{cases} \gamma^\alpha |t|^\alpha \,[1 - \mathrm{i}\,\mathrm{sgn}(t)\,\beta \tan(\pi\alpha/2)] & \alpha \ne 1 \\ \gamma|t|\,[1 + \mathrm{i}\,\mathrm{sgn}(t)\,(2/\pi)\beta \ln|t|] & \alpha = 1, \end{cases}$$

where $\gamma > 0$ and $\beta \in [-1,1]$ ([31], p. 5). Then by [16], p. 568–570

$$\Psi_X(t) = \int \phi(t,u)\lambda(\mathrm{d}u) \quad \text{with} \quad \phi(t,u) = \begin{cases} e^{\mathrm{i}tu} - 1 & 0<\alpha<1 \\ e^{\mathrm{i}tu} - 1 - \mathrm{i}tu & 1<\alpha<2 \\ e^{\mathrm{i}tu} - 1 - \mathrm{i}t\sin u & \alpha = 1 \end{cases}$$

and

$$\gamma = \left[\frac{M\pi}{2\Gamma(\alpha+1)\sin(\pi\alpha/2)}\right]^{1/\alpha}, \quad \beta = (M_+ - M_-)/M.$$

Given $r > 0$, Example 2.1 provides all the parameters for the PGN and Normal approximations to X_r. The PGN approximation was sampled according to the description following Eq. (4). On the other hand, Δ_r was sampled by the representation

$$\Delta_r \sim d(r) + r\sum_{i=1}^N \epsilon_i U_i^{-1/\alpha},$$

where $d(r)$ is a constant and N, $\epsilon_i \in \{\pm 1\}$, $U_i \sim \mathrm{Unif}(0,1)$, $i = 1,2,\ldots$ are mutually independent, with N being Poisson distributed with mean $\int_{|u|\ge r}\lambda(\mathrm{d}u) = M\alpha^{-1}r^{-\alpha}$ and $\mathrm{P}\{\epsilon_i = 1\} = M_+/M$. To determine $d(r)$, use $\Psi_X(t) = \Psi_{X_r}(t) + \Psi_{\Delta_r}(t)$. It follows that

$$\mathrm{i}d(r)t = \int \phi(t,u)\lambda(\mathrm{d}u) - \int_{-r}^{r}(e^{\mathrm{i}tu} - 1 - \mathrm{i}tu)\lambda(\mathrm{d}u) - \int_{|u|\ge r}(e^{\mathrm{i}tu} - 1)\lambda(\mathrm{d}u),$$

giving $d(r) = (M_+ - M_-)d_0(r)$, where

$$d_0(r) = \begin{cases} r^{1-\alpha}/(1-\alpha) & \alpha \ne 1, \\ \displaystyle\int_0^\infty u^{-2}(u\mathbf{1}\{u<r\} - \sin u)\,\mathrm{d}u & \alpha = 1. \end{cases}$$

As in last section, $M_- = 0$ or $M_- = M_+$. Equivalently, $\beta = 1$ or $\beta = 0$. The simulations became quite numerically unstable for small α, so we started with $\alpha = 0.2$. Also, for $\alpha = 1.0$, the R routines provided by `stabledist` package appeared to become numerically unstable in the asymmetric case but showed no serious problems in the symmetric case. Therefore, in the simulations for the asymmetric case, we chose $\alpha = 1.01$ instead of $\alpha = 1.0$, while for the symmetric case, we chose $\alpha = 1.0$. With this exception, we let α range from 0.2 to 1.8 at step size 0.2.

TABLE 1: Errors of approximations in terms of KS distance from the target distribution at cut-off $r = 5$.

$\beta = 1(0)$, $M_+ = 1$				
α	r	PGN d_{KS}	Norm d_{KS}	CP d_{KS}
0.2	5(5)	3.84e-02(8.71e-04)	3.99e-02(8.72e-04)	5.72e-02(9.34e-04)
0.3	5(5)	2.82e-03(8.31e-04)	8.75e-03(8.53e-04)	7.00e-02(8.49e-03)
0.4	5(5)	5.14e-03(8.39e-04)	2.08e-02(1.64e-03)	1.50e-01(3.63e-02)
0.5	5(5)	7.45e-03(8.44e-04)	3.33e-02(3.54e-03)	2.35e-01(8.36e-02)
0.6	5(5)	8.96e-03(8.66e-04)	4.47e-02(6.13e-03)	3.05e-01(1.41e-01)
0.8	5(5)	9.60e-03(9.23e-04)	6.14e-02(1.16e-02)	4.10e-01(2.51e-01)
1*	5(5)	7.79e-03(9.22e-04)	6.88e-02(1.54e-02)	4.75e-01(3.35e-01)
1.2	5(5)	5.26e-03(8.72e-04)	6.72e-02(1.61e-02)	5.10e-01(3.93e-01)
1.4	5(5)	2.87e-03(8.22e-04)	5.79e-02(1.38e-02)	5.20e-01(4.30e-01)
1.5	5(5)	2.02e-03(8.32e-04)	5.05e-02(1.17e-02)	5.20e-01(4.44e-01)
1.6	5(5)	1.42e-03(8.38e-04)	4.15e-02(9.19e-03)	5.15e-01(4.55e-01)
1.8	5(5)	8.74e-04(8.23e-04)	1.98e-02(3.70e-03)	5.00e-01(4.70e-01)
α	P-N d_{KS} Ratio	$\text{CL}_{0.99}(R_{\text{P-N}})$	N-C d_{KS} Ratio	$\text{CL}_{0.99}(R_{\text{N-C}})$
0.2	9.62e-01(9.99e-01)	9.63e-01(1.00e+00)	6.97e-01(9.42e-01)	6.98e-01(9.47e-01)
0.3	3.23e-01(9.77e-01)	3.25e-01(9.80e-01)	1.25e-01(1.01e-01)	1.25e-01(1.03e-01)
0.4	2.47e-01(5.10e-01)	2.47e-01(5.16e-01)	1.39e-01(4.52e-02)	1.39e-01(4.56e-02)
0.6	2.00e-01(1.41e-01)	2.01e-01(1.42e-01)	1.47e-01(4.36e-02)	1.47e-01(4.37e-02)
0.8	1.56e-01(7.92e-02)	1.57e-01(8.03e-02)	1.50e-01(4.64e-02)	1.50e-01(4.65e-02)
1*	1.13e-01(6.00e-02)	1.14e-01(6.08e-02)	1.45e-01(4.58e-02)	1.45e-01(4.59e-02)
1.2	7.83e-02(5.41e-02)	7.86e-02(5.49e-02)	1.32e-01(4.10e-02)	1.32e-01(4.11e-02)
1.4	4.96e-02(5.95e-02)	4.99e-02(6.04e-02)	1.11e-01(3.21e-02)	1.11e-01(3.22e-02)
1.5	4.01e-02(7.11e-02)	4.04e-02(7.21e-02)	9.71e-02(2.64e-02)	9.72e-02(2.64e-02)
1.6	3.42e-02(9.13e-02)	3.46e-02(9.27e-02)	8.06e-02(2.02e-02)	8.06e-02(2.02e-02)
1.8	4.42e-02(2.24e-01)	4.48e-02(2.27e-01)	3.96e-02(7.88e-03)	3.96e-02(7.91e-03)

* $\alpha = 1.01$ for $\beta = 1$

For each value of α, at a given value of r, we sampled 10^6 triplets $(T_r, \sigma_{X_r} Z, \Delta_r)$. The paired sums $T_r + \Delta_r$ and $\sigma_{X_r} Z + \Delta_r$ formed samples from the PGN and Normal approximations to X, respectively. Meanwhile, Δ_r formed a sample from the CP approximation of X. Denote by $\widehat{F}_{\text{PGN}}$, $\widehat{F}_{\text{Norm}}$, and $\widehat{F}_{\text{CP}}$ the corresponding empirical distributions. For $\theta \in (0, 1)$, let x_θ be the (unique) quantile of X such that $\text{P}\{X \le x_\theta\} = \theta$. The empirical KS distance between the PGN approximation and the target distribution was defined as $\widehat{D}_{\text{PGN}} = \max_\theta |\widehat{F}_{\text{PGN}}(x_\theta) - \theta|$ with $\theta \in \{i/200 : i = 1, \dots, 199\}$. Likewise, $\widehat{D}_{\text{Norm}}$ and $\widehat{D}_{\text{CP}}$ were calculated for the Normal and CP approximations, respectively. This step was repeated 2,500 times. The resulting 2,500 triplets $(\widehat{D}_{\text{PGN}}, \widehat{D}_{\text{Norm}}, \widehat{D}_{\text{CP}})$ were used to estimate $d_{\text{KS}}(X, T_r + \Delta_r)$, $d_{\text{KS}}(X, \sigma_{X_r} Z + \Delta_r)$, and $d_{\text{KS}}(X, \Delta_r)$, respectively, and their pairwise ratios. The focus here was the ratio of $d_{\text{KS}}(X, T_r + \Delta_r)$ to $d_{\text{KS}}(X, \sigma_{X_r} Z + \Delta_r)$. However, to make sure that our implementation of the Normal approximation was correct, we also estimated the ratio of $d_{\text{KS}}(X, \sigma_{X_r} Z + \Delta_r)$ to $d_{\text{KS}}(X, \Delta_r)$.

In the following, all errors are in terms of KS distance from target distribution.

TABLE 2: Errors of approximations, with $M_+ = 1$ and r set so that the empirical KS distance between the Normal approximation and the target distribution was roughly 1%.

$\beta = 1(0)$, $M_+ = 1$				
α	r	PGN d_{KS}	Norm d_{KS}	CP d_{KS}
0.2	4.38e+01(2.00e+03)	9.33e-03(4.38e-03)	9.78e-03(1.00e-02)	5.50e-02(5.61e-02)
0.4	2.34e+00(2.03e+01)	1.70e-03(2.03e-03)	9.55e-03(9.70e-03)	9.00e-02(1.12e-01)
0.6	1.37e+00(7.03e+00)	8.55e-04(1.12e-03)	1.02e-02(1.03e-02)	1.40e-01(1.78e-01)
0.8	1.12e+00(4.69e+00)	8.35e-04(8.74e-04)	1.01e-02(1.05e-02)	1.80e-01(2.42e-01)
1*	1.07e+00(3.91e+00)	1.47e-03(8.37e-04)	1.06e-02(1.01e-02)	2.25e-01(3.00e-01)
1.2	1.03e+00(3.71e+00)	8.21e-04(8.29e-04)	9.83e-03(9.67e-03)	2.55e-01(3.54e-01)
1.4	1.12e+00(4.10e+00)	8.24e-04(8.32e-04)	1.02e-02(9.86e-03)	3.10e-01(4.10e-01)
1.6	1.37e+00(5.08e+00)	8.24e-04(8.32e-04)	1.00e-02(9.42e-03)	3.75e-01(4.56e-01)
1.8	2.54e+00(9.38e+00)	8.36e-04(8.38e-04)	1.01e-02(9.70e-03)	4.65e-01(4.90e-01)
α	P-N d_{KS} Ratio	$\text{CL}_{0.99}(R_{\text{P-N}})$	N-C d_{KS} Ratio	$\text{CL}_{0.99}(R_{\text{N-C}})$
0.2	9.54e-01(4.37e-01)	1.05e+00(4.81e-01)	1.78e-01(1.78e-01)	1.89e-01(1.87e-01)
0.4	1.78e-01(2.09e-01)	1.79e-01(2.11e-01)	1.06e-01(8.69e-02)	1.06e-01(8.70e-02)
0.6	8.41e-02(1.08e-01)	8.52e-02(1.10e-01)	7.27e-02(5.78e-02)	7.28e-02(5.79e-02)
0.8	8.27e-02(8.34e-02)	8.39e-02(8.46e-02)	5.62e-02(4.33e-02)	5.63e-02(4.33e-02)
1*	1.39e-01(8.26e-02)	1.40e-01(8.38e-02)	4.70e-02(3.38e-02)	4.71e-02(3.38e-02)
1.2	8.36e-02(8.58e-02)	8.49e-02(8.71e-02)	3.86e-02(2.73e-02)	3.86e-02(2.74e-02)
1.4	8.10e-02(8.44e-02)	8.22e-02(8.57e-02)	3.29e-02(2.40e-02)	3.29e-02(2.41e-02)
1.6	8.25e-02(8.83e-02)	8.37e-02(8.96e-02)	2.67e-02(2.07e-02)	2.68e-02(2.07e-02)
1.8	8.31e-02(8.64e-02)	8.44e-02(8.77e-02)	2.17e-02(1.98e-02)	2.17e-02(1.98e-02)

* $\alpha = 1.01$ for $\beta = 1$

We first compared the approximations with $M_+ = 1$ and the cut-off r fixed at 5. To compare with Example 2.1, we included $\alpha = 0.3$ and 1.5 in the simulations. The results are summarized in Table 1. The top panel of the table displays the sample means of $\widehat{D}_{\text{PGN}}$ ("PGN d_{KS}"), $\widehat{D}_{\text{Norm}}$ ("Norm d_{KS}"), and $\widehat{D}_{\text{CP}}$ ("CP d_{KS}"), respectively. The bottom panel of the table displays $\widehat{D}_{\text{PGN}}/\widehat{D}_{\text{Norm}}$ ("P-N d_{KS} Ratio"), the upper 99% t-confidence limit of $\text{E}[\widehat{D}_{\text{PGN}}/\widehat{D}_{\text{Norm}}]$ ("$\text{CL}_{0.99}(R_{\text{P-N}})$"), $\widehat{D}_{\text{Norm}}/\widehat{D}_{\text{CP}}$ ("N-C d_{KS} Ratio"), and the upper 99% t-confidence limit of $\text{E}[\widehat{D}_{\text{Norm}}/\widehat{D}_{\text{CP}}]$ ("$\text{CL}_{0.99}(R_{\text{N-C}})$"), respectively. Since all of the standard errors are less than 1% of the corresponding sample means, they are omitted for brevity. To compare the performances of the approximations when the distribution is asymmetric ($\beta = 1$), and when the distribution is symmetric ($\beta = 0$), the sample means under these two conditions are displayed in pair, with the results under the symmetric condition put between parentheses. The table shows that, generally speaking, except for small α, the error of the PGN approximation is significantly smaller than that of the Normal approximation. For example, in the asymmetric case, for $\alpha = 0.3$ and $r = 5$, the sample mean of $\widehat{D}_{\text{PGN}}$ is about 1/3 of that of $\widehat{D}_{\text{Norm}}$. This may be compared with Figure 1, which shows that the bound in (14) for the PGN approximation is smaller than the one in (15) for the Normal approximation only if r is extremely small. The results for $\alpha = 1.5$ in the table can also be compared with Figure 1. This indicates that the bound in (14) is quite conservative.

Table 1 also confirms that the Normal approximation has significantly smaller error than the CP approximation. In fact, from the confidence limits shown in the table, the ratio of reduction of error by the Normal approximation as compared to the CP approximation is greater than that by the PGN approximation as compared to the Normal approximation.

Also, the bound in (15) for the Normal approximation is quite conservative as compared to the numerical results. For example, in the asymmetric case, for $\alpha = 0.8$, the sample mean of $\widehat{D}_{\text{Norm}}$ is about .06, whereas the bound in (15) gives .54. In the other sets of simulations, the greater ratio of reduction of error by the Normal approximation and the conservativeness of the bound in (15) were observed as well.

TABLE 3: Errors of approximations, with $M_+ = 0.1$ and r set so that the empirical KS distance between the Normal approximation and the target distribution was roughly 1%, or as low as possible if this could not be attained numerically.

$\beta = 1(0)$, $M_+ = 0.1$				
α	r	PGN d_{KS}	Norm d_{KS}	CP d_{KS}
0.2	2.44e-01(1.71e-01)	1.47e-01(2.38e-02)	1.73e-01(3.19e-02)	3.70e-01(1.45e-01)
0.4	2.29e-02(6.71e-02)	1.26e-02(2.37e-03)	3.12e-02(1.00e-02)	1.86e-01(1.17e-01)
0.6	2.90e-02(1.46e-01)	8.51e-04(1.08e-03)	9.95e-03(9.79e-03)	1.35e-01(1.74e-01)
0.8	6.10e-02(2.56e-01)	8.39e-04(8.71e-04)	9.56e-03(1.00e-02)	1.70e-01(2.38e-01)
1*	1.04e-01(3.91e-01)	1.47e-03(8.37e-04)	9.60e-03(1.01e-02)	2.15e-01(3.00e-01)
1.2	1.53e-01(5.37e-01)	8.35e-04(8.33e-04)	1.00e-02(9.42e-03)	2.55e-01(3.52e-01)
1.4	2.20e-01(7.81e-01)	8.38e-04(8.26e-04)	1.04e-02(9.64e-03)	3.10e-01(4.09e-01)
1.6	3.17e-01(1.27e+00)	8.28e-04(8.28e-04)	9.75e-03(1.03e-02)	3.70e-01(4.59e-01)
1.8	6.84e-01(2.73e+00)	8.33e-04(8.29e-04)	9.74e-03(1.04e-02)	4.60e-01(4.91e-01)
α	P-N d_{KS} Ratio	$\text{CL}_{0.99}(R_{\text{P-N}})$	N-C d_{KS} Ratio	$\text{CL}_{0.99}(R_{\text{N-C}})$
0.2	8.53e-01(7.32e-01)	8.54e-01(7.39e-01)	4.66e-01(2.19e-01)	4.66e-01(2.20e-01)
0.4	3.60e-01(2.39e-01)	3.69e-01(2.43e-01)	1.66e-01(8.59e-02)	1.68e-01(8.62e-02)
0.6	8.56e-02(1.10e-01)	8.67e-02(1.11e-01)	7.37e-02(5.63e-02)	7.38e-02(5.64e-02)
0.8	8.79e-02(8.69e-02)	8.92e-02(8.82e-02)	5.62e-02(4.20e-02)	5.63e-02(4.21e-02)
1*	1.53e-01(8.26e-02)	1.54e-01(8.38e-02)	4.47e-02(3.38e-02)	4.47e-02(3.38e-02)
1.2	8.35e-02(8.84e-02)	8.47e-02(8.97e-02)	3.93e-02(2.68e-02)	3.94e-02(2.68e-02)
1.4	8.07e-02(8.58e-02)	8.19e-02(8.71e-02)	3.36e-02(2.36e-02)	3.36e-02(2.36e-02)
1.6	8.51e-02(8.07e-02)	8.64e-02(8.19e-02)	2.64e-02(2.24e-02)	2.64e-02(2.24e-02)
1.8	8.58e-02(8.01e-02)	8.70e-02(8.13e-02)	2.12e-02(2.11e-02)	2.12e-02(2.11e-02)

* $\alpha = 1.01$ for $\beta = 1$

At cut-off $r = 5$, the error of the Normal approximation varies with α. One question is how the PGN compares to the Normal approximation when the error of the latter is fixed at a specified level. In the second set of simulations, we let r vary according to α, such that the empirical KS distance between the Normal approximation and the target distribution was roughly 1%. The value of r was selected as follows. For each r, ten values of $\widehat{D}_{\text{Norm}}$ were sampled each based on 10^6 observations from the Normal approximation at cut-off r. Starting with a large r, we reduced r by half if the average of the ten sample values of $\widehat{D}_{\text{Norm}}$ was greater 1.05%. We kept doing this until the average was within $(0.95\%, 1.05\%)$ or was less than 0.95%. In the former case r was selected. In the latter case we got two values of r, one giving an average greater than 1.05%, the other giving an average smaller than 0.95%. Then a bisection search was used to get a value of r with the corresponding average of $\widehat{D}_{\text{Norm}}$ within $(0.95\%, 1.05\%)$.

After a value of r was selected, the simulations preceded as the ones for Table 1. The results are summarized in Table 2, which also reports the selected values of r. The mean values of $\widehat{D}_{\text{Norm}}$ realized by the simulations are included to make sure the values of r were selected appropriately. In general, the mean value of $\widehat{D}_{\text{Norm}}$ fell into the interval $(0.95\%, 1.05\%)$. However, due to random fluctuations, the mean value could fall outside of $(0.95\%, 1.05\%)$, even though during the selection of r, the average of the ten sampled values of $\widehat{D}_{\mathrm{Norm}}$ fell into the interval. Similar to Table 1, except for small values of α, the error of the PGN approximation is significantly smaller than that of the Normal approximation.

In the above simulations, $M_+ = 1$. In the last set of simulations, we set $M_+ = 0.1$ to see how the approximations performed. The results are summarized in Table 3. Again, we attempted to set r so that the empirical KS distance between the Normal approximation and the target distribution was roughly 1%. However, although theoretically the approximation error vanishes as $r \to 0$, in our simulations, the numerical precision of the approximation deteriorated for small M_+, especially when α was small as well, and the minimum empirical KS distance for the Normal approximation could be quite larger than 1%. In this case, we set r so that the empirical KS distance was as small as possible. Table 3 shows that for $\alpha \le 0.4$, the empirical KS distance for the Normal approximation sometimes could not reach 1%. When this happened, the empirical KS distance for the PGN approximation did not reach 1% either. Since in our simulation, each pair $(T_r + \Delta_r, \sigma_{X_r} Z + \Delta_r)$ shared the same sampled value of Δ_r, this suggests that the deterioration of the numerical precision might be largely due to the error in Δ_r. However, a thorough solution to the issue is beyond the scope of the article. On the other hand, regardless of this issue, Table 3 again shows that the error of the PGN approximation can be significantly smaller than that of the Normal approximation.

4. Technical details

4.1. Proof of Theorem 2.1

Denote by $\mathscr{S}$ the space of smooth and rapidly decreasing functions on $\mathbb{R}$. It is classical that the Fourier transform $h \to \widehat{h}(t) = \int e^{\mathrm{i}tx} h(x)\,\mathrm{d}x$ is an homeomorphism of $\mathscr{S}$ onto itself ([19], p. 103). Let f_X be the probability density of X. If it exists, then $\psi_X = \widehat{f}_X$. Let

$$\int_0^\infty t^{2(q+1)} e^{-2L(t,r)}\,\mathrm{d}t < \infty. \tag{16}$$

Otherwise, $Q_{q+1} = \infty$ and (12) is trivial. We need two lemmas. Note that the second one does not require matching of cumulants.

Lemma 4.1. *1) Let ξ be i.d. with $\Psi_\xi(t) = \int (1 + \mathrm{i}tu - e^{\mathrm{i}tu})\,\nu(\mathrm{d}u)$ and $\mathrm{E}|\xi|^j < \infty$ for all $j \ge 1$. Given $\epsilon > 0$, let $Z \sim N(0, \epsilon^2)$ be independent of ξ. Then $\psi_{\xi+Z} \in \mathscr{S}$. 2) Under condition (16), $f_{X_r} \in C^q(\mathbb{R})$, and for $0 \le j \le q$, $f^{(j)}_{X_r}(x) \to 0$ as $|x| \to \infty$.*

Lemma 4.2. *Let T_r be defined as in Theorem 2.1 with $s(r) < 1/(p+3)$ and $\sigma(r) > 0$. Fix $\epsilon > 0$. Given A, $B \ge 0$ with $A+B = 1$, let W be i.d. with $\Psi_W(t) = A\Psi_{X_r}(t) + B\Psi_{T_r}(t) + \epsilon^2 t^2/2$. Let $\xi = W/\nu$, where $\nu = \sqrt{A\kappa_{2,X_r} + B\kappa_{2,T_r}}$. Then $f_\xi \in \mathscr{S}$ and for $j \ge 1$,*

$$\int |f^{(j)}_\xi(x)|\,\mathrm{d}x \le jI_{j-1}(r) + I_j(r) + (1 + \epsilon^2/\nu^2) I_{j+1}(r),$$

where for $j \ge 0$,

$$I_j(r) = \nu^{j+1/2} \left[\frac{\Gamma(j+1/2)}{2D(r)^{2j+1}} + \int_{1/r}^\infty t^{2j} e^{-2H(t,r)}\,\mathrm{d}t \right]^{1/2},$$

with $D(r) = \sqrt{2AC_0\kappa_{2,X_r} + B(C_0\kappa_{2,Y_r} + \sigma(r)^2)}$ and

$$H(t,r) = AC_0 t^2 \int_{u<1/|t|} u^2\,\lambda(\mathrm{d}u) + \frac{B\sigma(r)^2 t^2}{2}.$$

Assume the lemmas are true for now. Since $d_{\mathrm{TV}}(X, T_r + \Delta_r) = d_{\mathrm{TV}}(X_r + \Delta_r, T_r + \Delta_r) \le d_{\mathrm{TV}}(X_r, T_r)$, to show Theorem 2.1, it suffices to show that for any measurable $A \subset \mathbb{R}$,

$$\varrho(A) \le \frac{M}{q!}(|\kappa|_{q,X_r} + |\kappa|_{q,Y_r}) \tag{17}$$

where $\varrho(A) = |\mathrm{P}\{X_r \in A\} - \mathrm{P}\{T_r \in A\}|$ and $M = \sigma_{X_r}^{-q}[qQ_{q-1}(r) + Q_q(r) + Q_{q+1}(r)]$.

We start with smoothing X_r and T_r while maintaining the same order of cumulant matching. Let Z, Z' be i.i.d. $\sim N(0,1)$ and independent of (X_r, T_r). Fix $\epsilon > 0$. Let h be a measurable function with $\|h\|_\infty \le 1$. The goal now is to bound

$$\Delta_\epsilon = \mathrm{E}[h(X_r + \epsilon Z) - h(T_r + \epsilon Z')].$$

For $n \ge 2$, let $U_i = U_{i,n}$ and $V_j = V_{j,n}$, $i, j = 1, \ldots, n+1$, be independent and i.d. with

$$\Psi_{U_i}(t) = n^{-1}\Psi_{X_r+\epsilon Z}(t), \quad \Psi_{V_i}(t) = n^{-1}\Psi_{T_r+\epsilon Z'}(t).$$

For $k = 1, \ldots, n+1$, let

$$W_k = \sum_{1\le j<k} V_j + \sum_{k<j\le n+1} U_j, \quad g_k(x) = \mathrm{E}h(W_k + x).$$

Since $X_r + \epsilon Z \sim W_1$ and $T_r + \epsilon Z' \sim W_{n+1}$, then $\Delta_\epsilon = g_1(0) - g_{n+1}(0)$, giving

$$|\Delta_\epsilon| \le |\mathrm{E}[g_1(U_1) - g_{n+1}(V_{n+1})]| + |\mathrm{E}[g_1(U_1) - g_1(0)]| + |\mathrm{E}[g_{n+1}(V_{n+1}) - g_{n+1}(0)]|. \tag{18}$$

We bound the expectations on the RHS separately. By $W_k + V_k = W_{k+1} + U_{k+1}$,

$$h(W_1 + U_1) - h(W_{n+1} + V_{n+1}) = \sum_{k=1}^{n+1}[h(W_k + U_k) - h(W_k + V_k)].$$

For each k, since W_k, U_k, and V_k are independent, by conditioning, $\mathrm{E}h(W_k + U_k) = \mathrm{E}g_k(U_k)$ and $\mathrm{E}h(W_k + V_k) = \mathrm{E}g_k(V_k)$. Then taking expectation on both sides of the displayed identity yields

$$\mathrm{E}[g_1(U_1) - g_{n+1}(V_{n+1})] = \sum_{k=1}^{n+1}\mathrm{E}[g_k(U_k) - g_k(V_k)]. \tag{19}$$

Denote $\nu = \sigma_{X_r}$. Let $\xi_k = W_k/\nu$. By Lemma 4.1, $f_{\xi_k} \in \mathscr{S}$. As a result,

$$g_k(x) = \mathrm{E}[h(\nu\xi_k + x)] = \int h(\nu u) f_{\xi_k}(u - x/\nu)\,\mathrm{d}u \tag{20}$$

is smooth. By Taylor expansion around 0,

$$g_k(U_k) - g_k(V_k) = \sum_{j=1}^{q-1}\frac{g_k^{(j)}(0)}{j!}(U_k^j - V_k^j) + \frac{1}{q!}[g_k^{(q)}(\theta(U_k)U_k)U_k^q - g_k^{(q)}(\theta(V_k)V_k)V_k^q],$$

where $\theta(x) \in [0,1]$. By assumption, $\kappa_{j,X_r} = \kappa_{j,T_r}$ for $1 \le j < q$. Since $\kappa_{j,U_k} = n^{-1}\kappa_{j,X_r+\epsilon Z} = n^{-1}(\kappa_{j,X_r} + \epsilon^2 \mathbf{1}\{j=2\})$, and likewise $\kappa_{j,V_k} = n^{-1}(\kappa_{j,T_r} + \epsilon^2 \mathbf{1}\{j=2\})$, then $\kappa_{j,U_k} = \kappa_{j,V_k}$ for $1 \le j < q$. As a result, $\mathrm{E} U_k^j = \mathrm{E} V_k^j$ for $1 \le j < q$ and hence

$$\mathrm{E}[g_k(U_k) - g_k(V_k)] = \frac{1}{q!}\mathrm{E}[g_k^{(q)}(\theta(U_k)V_k)U_k^q - g_k^{(q)}(\theta(V_k)V_k)V_k^q],$$

$$\Longrightarrow \ |\mathrm{E}[g_k(U_k) - g_k(V_k)]| \le \frac{\|g_k^{(q)}\|_\infty}{q!}[\mathrm{E}|U_k|^q + \mathrm{E}|V_k|^q]. \tag{21}$$

Since by (20) we have $g_k^{(q)}(x) = (-\nu)^{-q}\int h(\nu u) f_{\xi_k}^{(q)}(u - x/\nu)\,\mathrm{d}u$, then

$$\|g_k^{(q)}\|_\infty \le \nu^{-q}\int |f_{\xi_k}^{(q)}(u)|\,\mathrm{d}u < \infty. \tag{22}$$

By $\Psi_{W_k}(t) = (k-1)\Psi_{V_1}(t) + (n+1-k)\Psi_{U_1}(t)$,

$$\Psi_{W_k}(t) = \frac{n+1-k}{n}\Psi_{X_r}(t) + \frac{k-1}{n}\Psi_{T_r}(t) + \frac{\epsilon^2 t^2}{2}.$$

Then we can apply Lemma 4.2 with $\nu^2 = \kappa_{2,X_r} = \kappa_{2,T_r}$, $A = (n+1-k)/n$ and $B = (k-1)/n$ therein. By definition of $D(r)$ and $H(t,r)$ in Lemma 4.2,

$$D(r)^2 = 2AC_0\nu^2 + B(C_0\kappa_{2,Y_r} + \sigma(r)^2) \ge C_0\nu^2$$

and

$$\begin{aligned} H(t,r) &= AC_0 t^2 \int_{u<1/|t|} u^2\,\lambda(\mathrm{d}u) + \frac{B\sigma(r)^2 t^2}{2} \\ &\ge (A+B)t^2 \min\left\{C_0 \int_{u<1/|t|} u^2\,\lambda(\mathrm{d}u),\ \sigma(r)^2/2\right\} = L(t,r). \end{aligned}$$

By definition of $Q_j(r)$ in Theorem 2.1 and definition of $I_j(r)$ in Lemma 4.2, $I_j(r) \le Q_j(r)$. By condition (16), $Q_j(r) < \infty$ for $0 \le j \le q+1$. Thus (22) and Lemma 4.2 give

$$\|g_k^{(q)}\|_\infty \le \nu^{-q}\left[qQ_{q-1}(r) + Q_q(r) + (1+\epsilon^2/\nu^2)Q_{q+1}(r)\right] := M_\epsilon < \infty.$$

Since M_ϵ is independent of k, by (19) and (21),

$$|\mathrm{E}g_1(U_1) - \mathrm{E}g_{n+1}(V_{n+1})| \le \frac{M_\epsilon}{q!}\sum_{k=1}^{n+1}(\mathrm{E}|U_k|^q + \mathrm{E}|V_k|^q).$$

Since the Lévy measure of X_r has bounded support, $\mathrm{E}|X_r + \epsilon Z|^q < \infty$. Meanwhile, from (5), $\mathrm{E}|Y_r + \epsilon Z|^q < \infty$. Then by Lemma 3.1 in [1],

$$\sum_{k=1}^{n+1}\mathrm{E}|U_k|^q \to |\kappa|_{q,X_r+\epsilon Z} = |\kappa|_{q,X_r}, \quad \sum_{k=1}^{n+1}\mathrm{E}|V_k|^q \to |\kappa|_{q,T_r+\epsilon Z'} = |\kappa|_{q,Y_r}.$$

Thus, for the first term on the RHS of (18),

$$\limsup_{n\to\infty}|\mathrm{E}g_1(U_1) - \mathrm{E}g_{n+1}(V_{n+1})| \le \frac{M_\epsilon}{q!}(|\kappa|_{q,X_r} + |\kappa|_{q,Y_r}). \tag{23}$$

To bound the other terms on the RHS of (18), first note $|\mathrm{E}[g_1(U_1) - g_1(0)]| \le \|g_1'\|_\infty \mathrm{E}|U_1|$. As in (22), $\|g_1'\|_\infty < \infty$. Since $g_1(x) = \mathrm{E}h(X_r + \epsilon Z + x)$, $\|g_1'\|_\infty$ is independent of n. On the other hand, by $\mathrm{E}U_1 = 0$ and Cauchy-Schwartz inequality, $\mathrm{E}|U_1| \le \sigma_{U_1} = \sigma_{X_r+\epsilon Z}/\sqrt{n}$, so $\mathrm{E}[g_1(U_1) - g_1(0)] \to 0$ as $n \to \infty$. Likewise, $\mathrm{E}[g_{n+1}(V_{n+1}) - g_{n+1}(0)] \to 0$. Together with (18) and (23), this implies

$$|\mathrm{E}h(X_r + \epsilon Z) - \mathrm{E}h(T_r + \epsilon Z')| \le \frac{M_\epsilon}{q!}(|\kappa|_{q,X_r} + |\kappa|_{q,Y_r}).$$

Let G be the union of a finite number of (a_i, b_i) and $h(x) = \mathbf{1}\{x \in G\}$. By 2) of Lemma 4.1, $\mathrm{P}\{X_r = a_i \text{ or } b_i, \text{ some } i\} = 0$. Let $\epsilon \to 0$. Then $h(X_r + \epsilon Z) - h(X_r) \to 0$ a.s. On the other hand, since T_r is the sum of Y_r and an independent non-zero Normal random variable, by 1) of Lemma 4.1, $f_{T_r} \in \mathscr{S}$. As a result, $h(T_r + \epsilon Z') - h(T_r) \to 0$ a.s. Finally, $M_\epsilon \to M$. Then by dominated convergence, (17) holds for G.

Let A be measurable. Given $\delta > 0$, fix $R > 0$ such that $\mathrm{P}\{|X_r| \ge R\} + \mathrm{P}\{|T_r| \ge R\} < \delta$. Let $B = A \cap (-R, R)$. Then $\varrho(A) \le \varrho(B) + \delta$. There is an open $G \supset B$ such that $\ell(G \setminus B) < \delta$, where ℓ is the Lebesgue measure. G is the union of at most countably many disjoint open intervals (a_i, b_i). Let $G_k = \cup_{i=1}^k (a_i, b_i)$. Then $\varrho(B) \le \varrho(G_k) + \mathrm{P}\{X_r \in B \triangle G_k\} + \mathrm{P}\{T_r \in B \triangle G_k\}$. From last paragraph, (17) holds for G_k. Next, $B \triangle G_k \subset (G \setminus G_k) \cup (G \setminus B)$, and $\mathrm{P}\{X_r \in G \setminus B\} \le \|f_{X_r}\|_\infty \ell(G \setminus B)$ and a similar inequality holds for T_r. Then

$$\varrho(B) \le \frac{M}{q!}(|\kappa|_{q,X_r} + |\kappa|_{q,Y_r}) + \mathrm{P}\{X_r \in G \setminus G_k\} + \mathrm{P}\{T_r \in G \setminus G_k\} + (\|f_{X_r}\|_\infty + \|f_{T_r}\|_\infty)\delta.$$

By Lemma 4.1, $\|f_{X_r}\|_\infty + \|f_{T_r}\|_\infty < \infty$. Letting $k \to \infty$ and then $\delta \to 0$, it is seen (17) holds for A.

4.2. Proofs of Lemmas 4.1 and 4.2

We need an elementary result for the proofs.

Lemma 4.3. *1)* $1 - \cos x \ge C_0 x^2$ *for* $|x| \le 1$. *2)* $\inf_{p>0} \frac{1}{\Gamma(p)} \int_0^p u^{p-1} e^{-u}\,\mathrm{d}u = 1/2$.

Proof. 1) For $x \in [0, 1]$, since $\sin x$ is concave, $\sin x \ge x \sin 1$, giving $1 - \cos x = 2[\sin(x/2)]^2 \ge 2[(x/2)\sin 1]^2 = C_0 x^2$. For $x \in [-1, 0]$, the proof follows from symmetry.

2) The inequality can be written as $\inf_{p>0} \mathrm{P}\{\xi_p \le p\} = 1/2$, where $\xi_p \sim \mathrm{Gamma}(p, 1)$. By Central Limit Theorem, $\mathrm{P}\{\xi_p \le p\} \to 1/2$ as $p \to \infty$. Therefore, it suffices to show that for every $p > 0$, $\mathrm{P}\{\xi_p \le p\} > 1/2$, or equivalently, $\int_0^p u^{p-1} e^{-u}\,\mathrm{d}u > \int_p^\infty u^{p-1} e^{-u}\,\mathrm{d}u$. Applying change of variable $u \leftarrow pu$ to the first integral and $u \leftarrow p/u$ to the second one, the inequality is equivalent to $\int_0^1 u^{-p-1}[u^{2p} e^{-pu} - e^{-p/u}]\,\mathrm{d}u > 0$, which holds if for all $u \in (0, 1)$, $u^{2p} e^{-pu} > e^{-p/u}$, or equivalently, $2 \ln u + 1/u - u > 0$. The last inequality follows directly from calculus.

Proof of Lemma 4.1. 1) From the assumption, $\int |u|^j \lambda(\mathrm{d}u) < \infty$ for all $j \ge 2$. Then by dominated convergence, $\Psi_\xi \in C^\infty(\mathbb{R})$ with $\Psi_\xi^{(j)}(t) = \int (\mathbf{1}\{j = 1\} - e^{\mathrm{i}tu})(\mathrm{i}u)^j \nu(\mathrm{d}u)$ for $j \ge 1$. By $|1 - e^{\mathrm{i}x}| \le |x|$ for $x \in \mathbb{R}$, $|\Psi_\xi'(t)| \le \kappa_{2,\xi}|t|$. Clearly, $|\Psi_\xi^{(j)}(t)| \le |\kappa|_{j,\xi}$ for $j \ge 2$. Since $\psi_{\xi+Z}(t) = \exp(-\Psi_\xi(t) - \epsilon^2 t^2/2)$, then for $j \ge 0$, $\psi_{\xi+Z}^{(j)}(t) = P_j(\Psi_\xi'(t), \ldots, \Psi_\xi^{(j)}(t), t)\,\psi_\xi(t) \exp(-\epsilon^2 t^2/2)$, where $P_j(z)$ is a multivariate polynomial in $z = (z_1, \ldots, z_{j+1})$ of order j. It follows that $|\psi_{\xi+Z}^{(j)}(t)| = O(|t|^j e^{-\epsilon^2 t^2/2})$ and hence for any $p \ge 0$, $|t|^p |\psi_{\xi+Z}^{(j)}(t)| \to 0$ as $|t| \to \infty$, which yields the proof.

2) For $|t| \ge 1/r$,

$$\mathrm{Re}[\Psi_{X_r}(t)] = \int_{|u|<r} (1-\cos tu)\,\lambda(\mathrm{d}u) \ge \int_{|u|<1/|t|} (1-\cos tu)\,\lambda(\mathrm{d}u).$$

Then by Lemma 4.3, $1-\cos tu \ge C_0 t^2 u^2$ for $0 \le u < 1/|t|$ and hence

$$\mathrm{Re}[\Psi_{X_r}(t)] \ge C_0 t^2 \int_{u<1/|t|} u^2\,\lambda(\mathrm{d}u) \ge L(t,r).$$

On the other hand, by Cauchy-Schwartz inequality,

$$\begin{aligned}\int_{|t|\ge 1/r} |t|^q |\psi_{X_r}(t)|\,\mathrm{d}t &\le \left(\int \frac{\mathrm{d}t}{1+t^2}\right)^{1/2} \left(\int_{|t|\ge 1/r} (1+t^2)t^{2q}|\psi_{X_r}(t)|^2\,\mathrm{d}t\right)^{1/2} \\ &\le \sqrt{\pi}\left(\int (1+t^2)t^{2q}e^{-2L(t,r)}\,\mathrm{d}t\right)^{1/2}\end{aligned}$$

Then by (16), $|t|^q|\psi_{X_r}(t)| \in L^1(\mathbb{R})$ and the proof follows from Proposition 28.1 of [32].

To prove Lemma 4.2, we need a type of inequalities known in the literature (cf. [4], Lemma 11.6). Since the expression of $(\widehat{f})^{(j)}$ becomes complicated rapidly as j increases, the following specific form is used to reduce the maximum order of derivative involved.

Lemma 4.4. *Let $f \in \mathscr{S}$ and $\psi = \widehat{f}$. Then for $j \ge 1$,*

$$\int |f^{(j)}| \le \frac{1}{\sqrt{2}}\left[\left(\int |t^j\psi(t)|^2\,\mathrm{d}t\right)^{1/2} + j\left(\int |t^{j-1}\psi(t)|^2\,\mathrm{d}t\right)^{1/2} + \left(\int |t^j\psi'(t)|^2\,\mathrm{d}t\right)^{1/2}\right].$$

Proof. By Cauchy-Schwartz and Minkowski inequalities,

$$\begin{aligned}\int |f^{(j)}| &\le \left(\int \frac{\mathrm{d}x}{1+x^2}\right)^{1/2} \left(\int |f^{(j)}(x)|^2(1+x^2)\,\mathrm{d}x\right)^{1/2} \\ &\le \sqrt{\pi}\left[\left(\int |f^{(j)}(x)|^2\,\mathrm{d}x\right)^{1/2} + \left(\int |xf^{(j)}(x)|^2\,\mathrm{d}x\right)^{1/2}\right].\end{aligned}$$

Then by Plancherel theorem and the fact that the Fourier transforms of $f^{(j)}(x)$ and $x^j f(x)$ are $(-it)^j\psi(t)$ and $(-i)^j\psi^{(j)}(t)$, respectively ([19], p. 100-102),

$$\int |f^{(j)}| \le \frac{1}{\sqrt{2}}\left[\left(\int |t^j\psi(t)|^2\,\mathrm{d}t\right)^{1/2} + \left(\int |(t^j\psi(t))'|^2\,\mathrm{d}t\right)^{1/2}\right].$$

The proof is complete by applying Minkowski inequality to the last integral.

Proof of Lemma 4.2. We only consider the case where $\mathrm{sppt}(\lambda) \subset \mathbb{R}_+$. The proof for the symmetric case is similar. For brevity, write $f = f_\xi$, $\psi = \psi_\xi$, and $\Psi = \Psi_\xi$. By Lemma 4.1, f, $\psi \in \mathscr{S}$. Write $M = \epsilon^2 + B\sigma(r)^2$. Then

$$\mathrm{Re}[\Psi(t)] = \mathrm{Re}[\Psi_W(t/\nu)] = \int (1-\cos tu/\nu)[A\lambda_r(\mathrm{d}u) + B\gamma_r(\mathrm{d}u)] + \frac{Mt^2}{2\nu^2}.$$

If $|t| \le \nu/r$, then $|tu|/\nu \le 1$ for $0 \le u < r$, so by Lemma 4.3, $1 - \cos tu/\nu \ge C_0 t^2 u^2/\nu^2$. Consequently,

$$\begin{aligned}\mathrm{Re}[\Psi(t)] &\ge \frac{C_0 t^2}{\nu^2}\int_0^r u^2[A\lambda_r(\mathrm{d}u) + B\gamma_r(\mathrm{d}u)] + \frac{Mt^2}{2\nu^2} \\ &= \frac{AC_0\kappa_{2,X_r}t^2}{\nu^2} + \frac{BC_0 m(r)s(r)^{p+3}t^2}{\nu^2}\int_0^{r/s(r)} u^{p+2}e^{-u}\,\mathrm{d}u + \frac{Mt^2}{2\nu^2}.\end{aligned}$$

Since $s(r) < r/(p+3)$, by 2) of Lemma 4.3,

$$\int_0^{r/s(r)} u^{p+2}e^{-u}\,\mathrm{d}u \ge \int_0^{p+3} u^{p+2}e^{-u}\,\mathrm{d}u \ge \Gamma(p+3)/2.$$

Then by $\Gamma(p+3)m(r)s(r)^{p+3} = \kappa_{2,Y_r}$,

$$\begin{aligned}\mathrm{Re}[\Psi(t)] &\ge \frac{AC_0\kappa_{2,X_r}t^2}{\nu^2} + \frac{BC_0 m(r)s(r)^{p+3}\Gamma(p+3)t^2}{2\nu^2} + \frac{Mt^2}{2\nu^2} \\ &\ge \frac{AC_0\kappa_{2,X_r}t^2}{\nu^2} + \frac{BC_0\kappa_{2,Y_r}t^2}{2\nu^2} + \frac{B\sigma(r)^2t^2}{2\nu^2} = \frac{D(r)^2t^2}{2\nu^2}.\end{aligned}$$

If $|t| > \nu/r$, then $r > \nu/|t|$ and

$$\mathrm{Re}[\Psi(t)] \ge \frac{AC_0t^2}{\nu^2}\int_{u<\nu/|t|} u^2\lambda(\mathrm{d}u) + \frac{B\sigma(r)^2t^2}{2\nu^2} = H(t/\nu, r).$$

Therefore, for $j \ge 0$,

$$\begin{aligned}\int |t^j\psi(t)|^2\,\mathrm{d}t &= 2\int_0^\infty t^{2j}e^{-2\mathrm{Re}[\Psi(t)]}\,\mathrm{d}t \\ &\le 2\int_0^{\nu/r} t^{2j}e^{-D(r)^2t^2/\nu^2}\,\mathrm{d}t + 2\int_{\nu/r}^\infty t^{2j}e^{-2H(t/\nu,r)}\,\mathrm{d}t \\ &\le 2\int_0^\infty t^{2j}e^{-D(r)^2t^2/\nu^2}\,\mathrm{d}t + 2\nu^{2j+1}\int_{1/r}^\infty t^{2j}e^{-2H(t,r)}\,\mathrm{d}t \\ &\le \frac{\nu^{2j+1}\Gamma(j+1/2)}{D(r)^{2j+1}} + 2\nu^{2j+1}\int_{1/r}^\infty t^{2j}e^{-2H(t,r)}\,\mathrm{d}t = 2I_j(r)^2. \qquad (24)\end{aligned}$$

Next, $\psi'(t) = -\Psi'(t)\psi(t)$, with $\Psi'(t) = (\mathrm{i}/\nu)\int(1-e^{\mathrm{i}tu/\nu})u\,[A\lambda_r(\mathrm{d}u) + B\gamma_r(\mathrm{d}u)] + Mt/\nu^2$. By $|1-e^{\mathrm{i}x}| \le |x|$ for all $x \in \mathbb{R}$,

$$\begin{aligned}|\Psi'(t)| &\le \frac{t}{\nu^2}\int u^2\,[A\lambda_r(\mathrm{d}u) + B\gamma_r(\mathrm{d}u)] + \frac{Mt}{\nu^2} \\ &= \frac{A\kappa_{2,X_r}t}{\nu^2} + \frac{B\kappa_{2,Y_r}t}{\nu^2} + \frac{(\epsilon^2 + B\sigma(r)^2)t}{\nu^2} \\ &= \frac{(A\kappa_{2,X_r} + B\kappa_{2,T_r})t}{\nu^2} + \frac{\epsilon^2t}{\nu^2} = (1+\epsilon^2/\nu^2)t.\end{aligned}$$

As a result,

$$\begin{aligned}\int |t^j\psi'(t)|^2\,\mathrm{d}t &= \int |t^j\Psi'(t)\psi(t)|^2\,\mathrm{d}t \\ &\le (1+\epsilon^2/\nu^2)^2\int |t^{j+1}\psi(t)|^2\,\mathrm{d}t \le 2(1+\epsilon^2/\nu^2)^2 I_{j+1}(r)^2. \qquad (25)\end{aligned}$$

The proof is complete by combining Lemma 4.4, (24) and (25).

Appendix A.

To evaluate the RHS of (14), we need to evaluate $Q_j(r)$, $j = 5, 6, 7$, which involves the integral of $t^{2j}e^{-2L(t,r)}$ over $t \in [1/r, \infty)$. To get good numerical precision, one way is to employ incomplete Gamma functions [27]. For $\lambda(\mathrm{d}u) = c\mathbf{1}\{u > 0\}\, u^{-\alpha-1}\,\mathrm{d}u$, $\kappa_{2,X_s} = \int_0^s u^2\lambda(\mathrm{d}u) = c\int_0^s u^{1-\alpha}\,\mathrm{d}u = cs^{2-\alpha}/(2-\alpha)$, $s > 0$. Then it is not hard to get $2L(t,r) = \min\{At^2, B|t|^\alpha\}$, where $A = \sigma(r)^2$ and $B = 2cC_0/(2-\alpha)$. Let $t_0 = (B/A)^{1/(2-\alpha)}$. Then $2L(t,r) = At^2\mathbf{1}\{|t| \le t_0\} + B|t|^\alpha\mathbf{1}\{|t| > t_0\}$, and hence, letting $t_1 = \max\{1/r, t_0\}$,

$$\begin{aligned}\int_{1/r}^\infty t^{2j}e^{-2L(t,r)}\,\mathrm{d}t &= \int_{1/r}^{t_1} t^{2j}e^{-At^2}\,\mathrm{d}t + \int_{t_1}^\infty t^{2j}e^{-Bt^\alpha}\,\mathrm{d}t \\ &= \frac{1}{2A^{j+1/2}}\int_{A/r^2}^{At_1^2} u^{j-1/2}e^{-u}\,\mathrm{d}u + \frac{1}{\alpha B^{(2j+1)/\alpha}}\int_{Bt_1^\alpha}^\infty u^{(2j+1)/\alpha-1}e^{-u}\,\mathrm{d}u.\end{aligned}$$

The integrals on the last line can be expressed as incomplete Gamma functions. For symmetric $\lambda(\mathrm{d}u) = c\mathbf{1}\{|u| > 0\}\,|u|^{-\alpha-1}\,\mathrm{d}u$, the formula is the same, except that $B = 4cC_0/(2-\alpha)$.

Acknowledgements

The author would like to thank an anonymous referee for helpful suggestions.

References

[1] Asmussen, S. and Rosiński, J. (2001). Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.* **38,** 482–493.

[2] Baeumer, B. and Kovács, M. (2012). Approximating multivariate tempered stable processes. *J. Appl. Probab.* **49,** 167–183.

[3] Barbour, A. D. and Chen, L. H. Y. (2005). The permutation distribution of matrix correlation statistics. In *Stein's method and applications*. vol. 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* Singapore Univ. Press, Singapore pp. 223–245.

[4] Bhattacharya, R. N. and Ranga Rao, R. (1976). *Normal approximation and asymptotic expansions.* John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.

[5] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular variation* vol. 27 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge.

[6] Chatterjee, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34,** 2061–2076.

[7] Chatterjee, S. and Shao, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie-Weiss model. *Ann. Appl. Probab.* **21,** 464–483.

[8] Chen, L. H. Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal approximation by Stein's method.* Probability and its Applications (New York). Springer, Heidelberg.

[9] Chi, Z. (2012). On exact sampling of nonnegative infinitely divisible random variables. *Adv. Appl. Probab.* **44,** 842–873.

[10] Chi, Z. (2013). Nonnormal small jump approximation of infinitely divisible distributions. *Technical report* 13. Department of Statistics, University of Connecticut. Available on the web at `http://arxiv.org/abs/1304.6326`.

[11] Cohen, S. and Rosiński, J. (2007). Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli* **13,** 195–210.

[12] Covo, S. (2009). On approximations of small jumps of subordinators with particular emphasis on a Dickman-type limit. *J. Appl. Probab.* **46,**.

[13] Dereich, S. (2011). Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.* **21,** 283–311.

[14] Devroye, L. (1986). *Nonuniform random variate generation*. Springer-Verlag, New York.

[15] Devroye, L. (1987). A simple generator for discrete log-concave distributions. *Computing* **39,** 87–91.

[16] Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York.

[17] Fournier, N. (2011). Simulation and approximation of Lévy-driven stochastic differential equations. *ESAIM Probab. Stat.* **15,** 233–248.

[18] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19,** 724–739.

[19] Grafakos, L. (2008). *Classical Fourier analysis* 2nd ed. vol. 249 of *Graduate Texts in Mathematics*. Springer, New York.

[20] Hilber, N., Reich, N., Schwab, C. and Winter, C. (2009). Numerical methods for Lévy processes. *Finance Stoch.* **13,** 471–500.

[21] Hörmann, W., Leydold, J. and Derflinger, G. (2004). *Automatic nonuniform random variate generation*. Statistics and Computing. Springer-Verlag, Berlin.

[22] Kawai, R. and Masuda, H. (2011). On simulation of tempered stable random variates. *J. Comput. Appl. Math.* **235,** 2873–2887.

[23] Kohatsu-Higa, A. and Tankov, P. (2010). Jump-adapted discretization schemes for Lévy-driven SDEs. *Stochastic Processes Appl.* **120,** 2258–2285.

[24] Lorz, U. and Heinrich, L. (1991). Normal and Poisson approximation of infinitely divisible distribution functions. *Statistics* **22,** 627–649.

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

[26] Nourdin, I. and Peccati, G. (2012). *Normal approximations with Malliavin calculus* vol. 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge. From Stein's method to universality.

[27] Olver, F. W. J., Lozier, D. W., Boisvert, R. F. and Clark, C. W., Eds. (2010). *NIST handbook of mathematical functions*. U.S. Department of Commerce National Institute of Standards and Technology, Washington, DC.

[28] Orey, S. (1968). On continuity properties of infinitely divisible distribution functions. *Ann. Math. Stat.* **39,** 936–937.

[29] R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. `http://www.R-project.org`.

[30] Reinert, G. (2005). Three general approaches to Stein's method. In *An introduction to Stein's method*. vol. 4 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* Singapore Univ. Press, Singapore pp. 183–221.

[31] Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York. Stochastic models with infinite variance.

[32] Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions* vol. 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Translated from the 1990 Japanese original, Revised by the author.

[33] Shevtsova, I. On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands 2011. Available at http://arxiv.org/abs/1111.6554.

[34] Signahl, M. (2003). On error rates in normal approximations and simulation schemes for Lévy processes. *Stoch. Models* **19,** 287–298.

[35] Veillette, M. S. and Taqqu, M. S. (2012). Berry-Esseen and Edgeworth approximations for the normalized tail of an infinite sum of independent weighted gamma random variables. *Stochastic Processes Appl.* **122,** 885–909.