

Geometric sums, size biasing and zero biasing

Qingwei Liu* Aihua Xia†

Abstract

The geometric sum plays a significant role in risk theory and reliability theory [Kalashnikov (1997)] and a prototypical example of the geometric sum is Rényi's theorem [Rényi (1956)] saying a sequence of suitably parameterised geometric sums converges to the exponential distribution. There is extensive study of the accuracy of exponential distribution approximation to the geometric sum [Sugakova (1995), Kalashnikov (1997), Peköz & Röllin (2011)] but there is little study on its natural counterpart of gamma distribution approximation to negative binomial sums. In this note, we show that a nonnegative random variable follows a gamma distribution if and only if its size biasing equals its zero biasing. We combine this characterisation with Stein's method to establish simple bounds for gamma distribution approximation to the sum of nonnegative independent random variables, a class of compound Poisson distributions and the negative binomial sum of random variables.

Keywords: Stein's method; size-biasing; zero-biasing; gamma distribution; geometric sum.

MSC2020 subject classifications: 60F05; 60E15.

Submitted to ECP on September 16, 2021, final version accepted on March 13, 2022.

1 Introduction

Rényi's theorem [Rényi (1956)] states that $\mathcal{L}(p \sum_{i=1}^N X_i) \rightarrow \text{Exp}(1)$ as $p \rightarrow 0$, where \mathcal{L} denotes the distribution, $\{X_i\}$ is a sequence of independent and identically distributed (*i.i.d.*) random variables with $\mathbb{E} X_1 = 1$, N is a geometric random variable with distribution $\mathbb{P}(N = k) = p(1-p)^k$ for $k \in \mathbb{Z}_+ := \{0, 1, 2, \dots\}$, denoted by $N \sim \text{Ge}(p)$, and N is independent of $\{X_i\}$. Geometric sums are a natural object in risk theory and reliability theory [Kalashnikov (1997)] and, under mild conditions, they are asymptotically close to the exponential distribution (see [Sugakova (1995), Kalashnikov (1997)] and references therein). The accuracy of exponential distribution approximation can be estimated through renewal techniques hinged on the memoryless property of the geometric distribution [Sugakova (1995), Kalashnikov (1997)] but these techniques seem to be less efficient in tackling its natural counterpart of gamma distribution approximation to negative binomial sums. Stein's method related to gamma distribution approximation [Diaconis & Zabell (1991), Luk (1994), Peköz & Röllin (2011), Gaunt, Pickett & Reinert (2017), Gaunt (2019), Slepov (2021)] is more flexible than renewal techniques and gamma distribution approximation of random variables in a fixed Wiener

*School of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia.

E-mail: qingweil@student.unimelb.edu.au

†School of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia.

E-mail: aihuaxia@unimelb.edu.au

chaos of a general Gaussian process using the Malliavin-Stein method has been investigated by [Eichelsbacher & Thäle (2015), Ledoux, Nourdin & Peccati (2015), Nourdin & Peccati (2009)]. The key to the success of using Stein’s method is to find suitable distributional transformations of the random object under consideration [Goldstein & Reinert (2005)], hence the first obstacle we need to overcome is to find such distributional transformations characterising the gamma distribution. To this end, let us recall the two most commonly used distributional transformations, namely, size biasing and zero biasing.

For any nonnegative random variable V with finite mean μ , we say that V^s has V -size biased distribution if

$$\mathbb{E}[Vf(V)] = \mu \mathbb{E}f(V^s) \tag{1.1}$$

for all functions f such that $\mathbb{E}|Vf(V)| < \infty$. For the geometric sum $S := \sum_{i=1}^N X_i$ mentioned above [Rényi (1956)], it is a routine exercise to verify that $S^s \stackrel{d}{=} S + S' + X_1^s$, where $\stackrel{d}{=}$ stands for “equal in distribution”, $S' \stackrel{d}{=} S$, and S, S', X_1^s are independent. This form of size biasing does not seem promising to study gamma distribution approximation of the geometric sum using the size biasing only.

For a random variable W with mean μ and variance $\sigma^2 \in (0, \infty)$, we say W^z has the W -zero biased distribution if for all differentiable function f with $\mathbb{E}|Wf(W)| < \infty$,

$$\mathbb{E}[(W - \mu)f(W)] = \sigma^2 \mathbb{E}f'(W^z). \tag{1.2}$$

The zero-bias transformation dates back at least to [Lukacs (1970), Theorem 12.2.5 (a)] and it was first systematically studied by [Goldstein & Reinert (1997)]. [Stein (1972)] appears to be the first to observe that the normal distribution is the fixed point of the zero-bias transformation: Z is a normal random variable with zero mean and variance σ^2 if and only if for all absolutely continuous f with $\mathbb{E}|Zf(Z)| < \infty$,

$$\mathbb{E}[Zf(Z)] = \sigma^2 \mathbb{E}f'(Z).$$

The discrete version of zero biasing was introduced in [Goldstein & Xia (2006)] and it is slightly different from the zero biasing defined above. It is an elementary exercise to verify that zero biasing satisfies $W^z \stackrel{d}{=} (W - a)^z + a$ for all $a \in \mathbb{R}$.

As observed in [Goldstein & Reinert (2005)], both size biasing and zero biasing are special cases of the transformation that can be used to construct approximation theory based on a distribution which is the fixed point of the transformation. Both of the biasing transformations play significant roles in a wide range of distributional approximations including normal [Chen, Goldstein & Shao (2010)], Poisson [Barbour, Holst & Janson (1992)] and exponential [Peköz & Röllin (2011)].

In Section 2, we show that the gamma distribution is uniquely characterised by the property that its size biased distribution is the same as its zero biased distribution. We then combine this characterisation with Stein’s method in Section 3 to establish simple bounds for gamma distribution approximation with application to the sum of independent nonnegative random variables. As the gamma distribution is in the family of infinitely divisible distributions, we present the direct relationship between size biasing and zero biasing for infinitely divisible distributions on $\mathbb{R}_+ := [0, \infty)$ in Lemma 4.2. In the remaining part of Section 4, we consider gamma distribution approximation to a class of compound Poisson distributions and the negative binomial sum of random variables. In particular, our result provides an intuitive explanation why the gamma distribution is a continuous counterpart of the negative binomial distribution and why the gamma process is a pure jump increasing process while the gamma distribution is continuous.

2 Gamma distribution: the intersection of size biasing and zero biasing

The gamma distribution $\Gamma(r, \alpha)$ we consider has the probability density function $\frac{1}{\Gamma(r)}\alpha^r x^{r-1}e^{-\alpha x}$, $x > 0$. The following theorem states that the unique distribution having the same size biased distribution and zero biased distribution is a gamma distribution. A well-known fact is that for $W \sim \Gamma(r, \alpha)$, $W^s \sim \Gamma(r + 1, \alpha)$ [Arratia, Goldstein & Kochman (2019)].

Theorem 2.1. For a random variable $W \geq 0$ with mean μ and variance $\sigma^2 \in (0, \infty)$, we have $W \sim \Gamma(r, \alpha)$ with $\alpha := \frac{\mu}{\sigma^2}$, $r := \frac{\mu^2}{\sigma^2}$ if and only if

$$W^s \stackrel{d}{=} W^z. \tag{2.1}$$

Proof The proof relies on Stein’s identity of the gamma distribution [Luk (1994)]:

$$xf''(x) + (r - \alpha x)f'(x) = h(x) - \Gamma_{r,\alpha}h, \tag{2.2}$$

where $\Gamma_{r,\alpha}h := \mathbb{E}h(Z)$ for $Z \sim \Gamma(r, \alpha)$. Therefore, the random variable $W \sim \Gamma(r, \alpha)$ if and only if

$$\mathbb{E}[Wf''(W) + (r - \alpha W)f'(W)] = 0$$

for all twice differentiable functions f such that the expectations $\mathbb{E}|Wf''(W)|$ and $\mathbb{E}|Wf'(W)|$ are finite.

For the necessity part, assume that $W \sim \Gamma(r, \alpha)$ for some $r, \alpha > 0$, hence

$$\mu = \frac{r}{\alpha}, \quad \sigma^2 = \frac{r}{\alpha^2}.$$

For a differentiable h such that $\mathbb{E}|Wh(W)| < \infty$,

$$\begin{aligned} \sigma^{-2} \mathbb{E}[(W - \mu)h(W)] &= \alpha \mathbb{E}[h(W^s) - h(W)] \\ &= \alpha \int_0^\infty h(x) \left[\frac{1}{\Gamma(r+1)}\alpha^{r+1}x^r e^{-\alpha x} - \frac{1}{\Gamma(r)}\alpha^r x^{r-1} e^{-\alpha x} \right] dx \\ &= - \int_0^\infty h(x) d \left[\frac{1}{\Gamma(r+1)}\alpha^{r+1}x^r e^{-\alpha x} \right] \\ &= \int_0^\infty h'(x) \frac{1}{\Gamma(r+1)}\alpha^{r+1}x^r e^{-\alpha x} dx \\ &= \Gamma_{r+1,\alpha}h', \end{aligned}$$

which ensures that $W^z \sim \Gamma(r + 1, \alpha)$. Since $W^s \sim \Gamma(r + 1, \alpha)$, (2.1) follows. Conversely, if (2.1) holds, since $\mu = \alpha\sigma^2$, for all twice differentiable f such that the following expectations exist, we have

$$\begin{aligned} &\mathbb{E}[Wf''(W) + (r - \alpha W)f'(W)] \\ &= \mathbb{E}[Wf''(W)] - \alpha \mathbb{E}[(W - \mu)f'(W)] \\ &= \mu \mathbb{E}f''(W^s) - \alpha\sigma^2 \mathbb{E}f''(W^z) \\ &= 0. \end{aligned}$$

This ensures that $W \sim \Gamma(r, \alpha)$ and the proof is complete. □

3 Main results

In this section, we bound the errors of gamma distribution approximation in terms of the Wasserstein distance d_W and the Kolmogorov distance d_K defined as

$$d_W(\mathcal{L}(X), \mathcal{L}(Y)) := \sup_{f \in \mathcal{F}_W} |\mathbb{E} f(X) - \mathbb{E} f(Y)|, \tag{3.1}$$

$$d_K(\mathcal{L}(X), \mathcal{L}(Y)) := \sup_{f \in \mathcal{F}_K} |\mathbb{E} f(X) - \mathbb{E} f(Y)|,$$

where $\mathcal{F}_W := \{f : \mathbb{R} \rightarrow \mathbb{R}, |f(x) - f(y)| \leq |x - y|, \forall x, y \in \mathbb{R}\}$ and $\mathcal{F}_K := \{\mathbf{1}_{\{\cdot \leq z\}}, \forall z \in \mathbb{R}\}$.

Theorem 3.1. *Let W be a nonnegative random variable with mean μ and variance $\sigma^2 \in (0, \infty)$. For $\alpha = \mu/\sigma^2, r = \mu^2/\sigma^2$,*

$$d_W(\mathcal{L}(W), \Gamma(r, \alpha)) \leq 8\sqrt{\frac{3r}{\alpha(r+2)}} \Theta^{1/2} + \frac{8r}{r+2} \Theta, \tag{3.2}$$

$$d_K(\mathcal{L}(W), \Gamma(r, \alpha)) \leq a_{r,\alpha} \Theta^{\frac{r\wedge 1}{r\wedge 1+2}} + b_{r,\alpha} \Theta^{\frac{r\wedge 1+1}{r\wedge 1+2}}, \tag{3.3}$$

where $\Theta := d_W(\mathcal{L}(W^s), \mathcal{L}(W^z))$ and

$$a_{r,\alpha} = \begin{cases} 0.5(48\alpha)^{r/(r+2)} \left(\frac{r+2}{\Gamma(r+1)}\right)^{2/(r+2)} \leq 1.15(48\alpha)^{r/(r+2)}, & 0 < r < 1, \\ 3\left(\frac{6\alpha r}{r+2}\right)^{1/3} \left[\frac{1}{\Gamma(r)} \left(\frac{r-1}{e}\right)^{r-1}\right]^{2/3} \leq 5.46\left(\frac{\alpha}{2r+1}\right)^{1/3}, & r \geq 1; \end{cases} \tag{3.4}$$

$$b_{r,\alpha} = \begin{cases} 8\left(\frac{\alpha r}{r+2}\right)^{\frac{r+1}{r+2}} \left[\frac{1}{48\Gamma(r)}\right]^{\frac{1}{r+2}} \leq 1.11\alpha^{(r+1)/(r+2)}, & 0 < r < 1, \\ \frac{4\alpha^{2/3}r}{(r+2)^{2/3}} \left[\frac{1}{6\Gamma(r+1)} \left(\frac{r-1}{e}\right)^{r-1}\right]^{1/3} \leq 4.24\alpha^{2/3}r^{-1/6}, & r \geq 1. \end{cases} \tag{3.5}$$

Remark 3.2. (i) Due to the generality of the theorem, we don't know the optimal orders of the bounds in terms of Θ . Tractable examples indicate that the powers of Θ seem to be the artefacts of the proofs.

(ii) In respect of r as $r \rightarrow \infty$, (3.2) is of optimal order but the optimal orders of $a_{r,\alpha}$ and $b_{r,\alpha}$ in (3.3) seem to be $O(r^{-1/2})$. To check this, take W as a sum of i.i.d. random variables having finite third moment, then both W and $\Gamma(r, \alpha)$ can be well approximated by the normal distribution with the same mean and variance and we can use the order of normal approximation as the benchmark. Corollary 3.3 ensures that Θ does not depend on r , hence the optimal order of $d_W(\mathcal{L}(W), \Gamma(r, \alpha))$ is $O(1)$ [Esseen (1958), Zolotarev (1964)] or [Chen, Goldstein & Shao (2010), Theorem 4.2] and $d_K(\mathcal{L}(W), \Gamma(r, \alpha))$ is of order $O(r^{-1/2})$ [Berry (1941), Esseen (1942)] or [Chen, Goldstein & Shao (2010), Theorem 3.6].

Proof of Theorem 3.1 For any $h \in \mathcal{F}_W$ and $\delta > 0$, we can construct an interpolating spline function \tilde{h} as follows. For $i \in \mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$, let $x_i = -i\delta, (x_i, y_i) := (x_i, h(x_i))$ and

$$\tilde{h}(x) := y_i + (y_{i+1} - y_i)\phi\left(\frac{x - x_i}{x_{i+1} - x_i}\right), \quad x \in [x_i, x_{i+1}),$$

where

$$\phi(t) = \begin{cases} 0, & t < 0, \\ 2t^2, & 0 \leq t < 1/2, \\ 1 - 2(1-t)^2, & 1/2 \leq t \leq 1, \\ 1, & t > 1. \end{cases}$$

It is easy to verify that \tilde{h} is twice differentiable and

$$\|\tilde{h}'\| := \sup_{x \in \mathbb{R}} |\tilde{h}'(x)| \leq 2, \quad \|h - \tilde{h}\| \leq \delta, \quad \|\tilde{h}''\| \leq 4\delta^{-1},$$

since $\left| \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right| \leq 1$. Let $\tilde{f} := f_{\tilde{h}}$ be the solution of the Stein equation (2.2) with \tilde{h} in place of h , then

$$\|\tilde{f}'''\| \leq \frac{2}{r+2} (3\|\tilde{h}''\| + 2\alpha\|\tilde{h}'\|) \leq \frac{8}{r+2} (3\delta^{-1} + \alpha), \tag{3.6}$$

see [Gaunt, Pickett & Reinert (2017), Theorem 2.1]. By virtue of (2.2), (1.1) and (1.2), we have

$$\begin{aligned} \mathbb{E} \tilde{h}(W) - \Gamma_{r,\alpha} \tilde{h} &= \mathbb{E}[W \tilde{f}''(W) + (r - \alpha W) \tilde{f}'(W)] \\ &= \mu \mathbb{E} \tilde{f}''(W^s) - \alpha \mathbb{E}[(W - \mu) \tilde{f}'(W)] \\ &= \mu \mathbb{E}[\tilde{f}''(W^s) - \tilde{f}''(W^z)], \end{aligned}$$

which, together with (3.6), implies

$$|\mathbb{E} \tilde{h}(W) - \Gamma_{r,\alpha} \tilde{h}| \leq \frac{8\mu}{r+2} (3\delta^{-1} + \alpha) \Theta. \tag{3.7}$$

By the triangle inequality, we have

$$\begin{aligned} |\mathbb{E} h(W) - \Gamma_{r,\alpha} h| &\leq |\mathbb{E} h(W) - \mathbb{E} \tilde{h}(W)| + |\mathbb{E} \tilde{h}(W) - \Gamma_{r,\alpha} \tilde{h}| + |\Gamma_{r,\alpha} \tilde{h} - \Gamma_{r,\alpha} h| \\ &\leq 2\delta + \frac{8\mu}{r+2} (3\delta^{-1} + \alpha) \Theta, \end{aligned}$$

and (3.2) follows from (3.1) and $\delta = \sqrt{\frac{12\mu}{r+2}} \Theta^{1/2}$.

The proof of (3.3) relies on the following concentration inequality of $Z \sim \Gamma(r, \alpha)$: for $\delta > 0$ and $z \geq 0$,

$$\mathbb{P}(z < Z \leq z + \delta) \leq \epsilon(\delta) := \begin{cases} \frac{\alpha^r \delta^r}{\Gamma(r+1)}, & 0 < r < 1, \\ M(r, \alpha) \delta, & r \geq 1, \end{cases}$$

where $M(r, \alpha) := \frac{\alpha}{\Gamma(r)} \left(\frac{r-1}{e}\right)^{r-1}$ is the maximum of the density function of $\Gamma(r, \alpha)$. In fact, for $r \geq 1$, the bound is obvious, and for $r \in (0, 1)$, the bound follows from the fact that $g(\delta) := \int_z^{z+\delta} x^{r-1} e^{-\alpha x} dx - \delta^r/r$ is decreasing in δ and $g(0) = 0$.

Assume now $h_z(\cdot) = \mathbf{1}_{\{\cdot \leq z\}} \in \mathcal{F}_K$, $z \geq 0$. Denote

$$\tilde{h}_z(x) := 1 - \phi\left(\frac{x-z}{\delta}\right),$$

and note that $\|\tilde{h}'_z\| \leq 2\delta^{-1}$ and $\|\tilde{h}''_z\| \leq 4\delta^{-2}$. As

$$\begin{aligned} \mathbb{P}(W \leq z) - \mathbb{P}(Z \leq z) &\leq \mathbb{E} \tilde{h}_z(W) - \mathbb{E} \tilde{h}_z(Z) + \mathbb{E} \tilde{h}_z(Z) - \mathbb{P}(Z \leq z) \\ &\leq \mathbb{E} \tilde{h}_z(W) - \mathbb{E} \tilde{h}_z(Z) + \mathbb{P}(z < Z \leq z + \delta) \\ &\leq |\mathbb{E} \tilde{h}_z(W) - \Gamma_{r,\alpha} \tilde{h}_z| + \epsilon(\delta), \end{aligned}$$

and, following the same argument for (3.7), we have

$$|\mathbb{E} \tilde{h}_z(W) - \Gamma_{r,\alpha} \tilde{h}_z| \leq \mu \|\tilde{f}'''_z\| \Theta \leq \frac{8\mu}{r+2} (3\delta^{-2} + \alpha\delta^{-1}) \Theta =: e(\delta).$$

Analogously,

$$\begin{aligned} \mathbb{P}(W \leq z) - \mathbb{P}(Z \leq z) &\geq \mathbb{E} \tilde{h}_{z-\delta}(W) - \mathbb{E} \tilde{h}_{z-\delta}(Z) - \mathbb{P}(z - \delta < Z \leq z) \\ &\geq -e(\delta) - \epsilon(\delta). \end{aligned}$$

Therefore, for $0 < r < 1$, we have

$$d_K(\mathcal{L}(W), \Gamma(r, \alpha)) \leq e(\delta) + \frac{\alpha^r \delta^r}{\Gamma(r+1)} = \frac{8\mu}{r+2} (3\delta^{-2} + \alpha\delta^{-1})\Theta + \frac{\alpha^r \delta^r}{\Gamma(r+1)},$$

which implies (3.3) with $\delta = \left[\frac{48\mu\Theta\Gamma(r)}{(r+2)\alpha^r} \right]^{1/(r+2)}$. Similarly, for $r \geq 1$,

$$d_K(\mathcal{L}(W), \Gamma(r, \alpha)) \leq \frac{8\mu}{r+2} (3\delta^{-2} + \alpha\delta^{-1})\Theta + M(r, \alpha)\delta,$$

which, together with $\delta = \left(\frac{48\mu\Theta}{(r+2)M(r,\alpha)} \right)^{1/3}$ and $\mu = r/\alpha$, ensures (3.3). In terms of the upper bounds in (3.4), if $0 < r < 1$, we use the fact that $\Gamma(r+1) \geq 0.8856$ and $\left(\frac{r+2}{0.8856} \right)^{2/(r+2)}$ attains its maximum at $r_0 = 0.8856e - 2$ to obtain the upper bound; if $r \geq 1$, we use [Batir (2008), Theorem 1.5] to get $\Gamma(r) \geq \sqrt{2e} \left(\frac{r-1/2}{e} \right)^{r-1/2}$ and the fact that $\frac{r}{(r+2)(r-1/2)} \leq \frac{1}{r+1/2}$ to obtain

$$\sqrt{\frac{r}{r+2}} \left(\frac{r-1}{e} \right)^{r-1} \frac{1}{\Gamma(r)} \leq \sqrt{\frac{r}{2(r+2)(r-1/2)}} \left(\frac{r-1}{r-1/2} \right)^{r-1} \leq (2r+1)^{-1/2}.$$

Similarly, we can establish the upper bounds in (3.5). For $0 < r < 1$,

$$\begin{aligned} \left(\frac{r}{r+2} \right)^{\frac{r+1}{r+2}} \left[\frac{1}{48\Gamma(r)} \right]^{\frac{1}{r+2}} &= \frac{r}{(r+2)^{\frac{r+1}{r+2}}} \left[\frac{1}{48\Gamma(r+1)} \right]^{\frac{1}{r+2}} \\ &\leq \frac{r}{(r+2)^{\frac{r+1}{r+2}}} \left[\frac{1}{48 \times 0.8856} \right]^{\frac{1}{r+2}} =: \tilde{g}(r). \end{aligned}$$

It is easy to verify that $\tilde{g}(r)$ is increasing in r , so $\tilde{g}(r) \leq \tilde{g}(1)$ and the upper bound follows immediately. For $r \geq 1$, we apply the bound [Batir (2008), Theorem 1.5] again to get $\Gamma(r+1) \geq \sqrt{2e} \left(\frac{r+1/2}{e} \right)^{r+1/2}$ so

$$\begin{aligned} \frac{4r}{(r+2)^{2/3}} \left[\frac{1}{6\Gamma(r+1)} \left(\frac{r-1}{e} \right)^{r-1} \right]^{1/3} &\leq \frac{4er}{6^{1/3}\sqrt{2}(r+2)^{2/3}} \left(\frac{(r-1)^{r-1}}{(r+1/2)^{r+1/2}} \right)^{1/3} \\ &\leq 4.24r^{-1/6}. \quad \square \end{aligned}$$

Theorem 3.1 says that the accuracy of gamma distribution approximation with respect to d_W and d_K is determined by $\Theta = d_W(\mathcal{L}(W^s), \mathcal{L}(W^z))$. The next corollary says that for the sum of nonnegative independent random variables, Θ can be easily bounded.

Corollary 3.3. *Let $\{X_i : 1 \leq i \leq n\}$ be nonnegative independent random variables with positive finite variances and $W = \sum_{i=1}^n X_i$, then (3.2) and (3.3) hold with*

$$\Theta = \mathbb{E} |X_{I_1}^s + X_{I_2} - X_{I_1} - X_{I_2}^z|,$$

where X_i^s and X_i^z have the size-biased distribution and zero-biased distribution of X_i , respectively, I_1, I_2 are random indices, independent of X_1, \dots, X_n , with distributions

$$\mathbb{P}(I_1 = i) = \frac{\mathbb{E} X_i}{\mathbb{E} W}, \quad \mathbb{P}(I_2 = i) = \frac{\text{Var}(X_i)}{\text{Var}(W)}, \quad (3.8)$$

for $i = 1, \dots, n$. In particular, when X_1, \dots, X_n are i.i.d. random variables, we have (3.2) and (3.3) with $\Theta = \mathbb{E} |X_1^s - X_1^z|$.

Proof According to [Goldstein & Rinott (2005)] and [Goldstein & Reinert (1997)], we can set $W^s := \sum_{j \neq I_1} X_j + X_{I_1}^s$ and $W^z := \sum_{j \neq I_2} X_j + X_{I_2}^z$, hence

$$\mathbb{E} |W^s - W^z| = \mathbb{E} |X_{I_1}^s + X_{I_2} - X_{I_1} - X_{I_2}^z|. \tag{3.9}$$

When X_1, \dots, X_n are i.i.d. random variables, we may take $I_1 = I_2$, which is uniformly distributed on $\{1, \dots, n\}$. Hence, (3.9) can be reduced to

$$\mathbb{E} |X_{I_1}^s - X_{I_1}^z| = \mathbb{E} |X_1^s - X_1^z|,$$

as claimed. □

Convolutions of gamma distributions commonly arise in queuing type problems. For examples, the total waiting time of a system with independent exponentially distributed component lifetimes and the total amount of water collected in a dam from independent excess flows at different occasions [Moschopoulos (1985)]. In fact, as a more general form of the exponential distribution, convolutions of gamma distributions may appear in applications where random events in space or time occur in tandem with gaps of events following gamma distributions, see [Stewart et al. (2007), Vellaisamy & Upadhye (2009)] for more discussion of such applications. There are extensive theoretical studies of the problem and we refer interested readers to [Vellaisamy & Upadhye (2009), Covo & Elalouf (2014)] for further reading in this area. In particular, it is well known that weighted sums of independent Chi-square random variables can be written as a convolution of gamma random variables [Jensen & Solomon (1972)]. The convolution of gamma distributions is generally intractable but can be well approximated by a gamma distribution with the same first and second moments [Moschopoulos (1985)] and the next corollary provides the error estimates of the gamma distribution approximation to such convolutions. Before we state the corollary, we mention that other parameterisations of the approximating gamma distribution may be possible [Covo & Elalouf (2014)].

Corollary 3.4. *If $X_i \sim \Gamma(r_i, \alpha_i)$, $1 \leq i \leq n$, are independent gamma distributed random variables, with $\mu = \sum_{i=1}^n r_i/\alpha_i$, $\sigma^2 = \sum_{i=1}^n r_i/\alpha_i^2$, $\alpha = \mu/\sigma^2$, $r = \alpha\mu$, then (3.2) and (3.3) hold with $\Theta = \mathbb{E} |Y_{I_1} - Y_{I_2}|$, where I_1 and I_2 are random indices satisfying*

$$\mathbb{P}(I_1 = i) = \frac{r_i \alpha_i^{-1}}{\sum_{j=1}^n r_j \alpha_j^{-1}}, \quad \mathbb{P}(I_2 = i) = \frac{r_i \alpha_i^{-2}}{\sum_{j=1}^n r_j \alpha_j^{-2}}, \tag{3.10}$$

$Y_i \sim \Gamma(1, \alpha_i)$, $i = 1, \dots, n$, $\{Y_i\}$ are independent and are independent of $\{I_1, I_2\}$.

Proof By Corollary 3.3, it suffices to bound $\mathbb{E} |X_{I_1}^s + X_{I_2} - X_{I_1} - X_{I_2}^z|$. Recalling that $X_i^s \stackrel{d}{=} X_i^z \sim \Gamma(r_i + 1, \alpha_i)$, we can set $X_i^s := X_i + Y_i$, $X_i^z = X_i^s$, where $Y_i \sim \Gamma(1, \alpha_i)$ is independent of $\{X_j : 1 \leq j \leq n\}$, $1 \leq i \leq n$. The distributions of I_1, I_2 in (3.8) are reduced to (3.10) and $\mathbb{E} |X_{I_1}^s + X_{I_2} - X_{I_1} - X_{I_2}^z| = \mathbb{E} |Y_{I_1} - Y_{I_2}|$. □

4 Applications

Before we consider applications, it is handy to have the following lemma bounding the Wasserstein distance between two gamma distributions.

Lemma 4.1. *For $r_1, r_2, \alpha_1, \alpha_2 > 0$,*

$$d_W(\Gamma(r_1, \alpha_1), \Gamma(r_2, \alpha_2)) \leq \frac{|r_1 - r_2|}{\alpha_1 \vee \alpha_2} + (r_1 \vee r_2) \left| \frac{1}{\alpha_1} - \frac{1}{\alpha_2} \right|.$$

Proof Without loss, we assume $\alpha_1 < \alpha_2$, then by the triangle inequality,

$$\begin{aligned} d_W(\Gamma(r_1, \alpha_1), \Gamma(r_2, \alpha_2)) &\leq d_W(\Gamma(r_1, \alpha_1), \Gamma(r_1, \alpha_2)) + d_W(\Gamma(r_1, \alpha_2), \Gamma(r_2, \alpha_2)) \\ &= r_1 \left| \frac{1}{\alpha_1} - \frac{1}{\alpha_2} \right| + \frac{|r_1 - r_2|}{\alpha_2}, \end{aligned}$$

where the last equality is due to the facts that $\Gamma(r, \alpha)$ is stochastically increasing in r and stochastically decreasing in α , and for two distributions F_1 and F_2 such that F_2 is stochastically smaller than F_1 , i.e., $F_1(t) \leq F_2(t)$ for all $t \in \mathbb{R}$, then the main theorem of [Vallender (1972)] ensures

$$d_w(F_1, F_2) = \int t dF_1(t) - \int t dF_2(t). \quad \square \tag{4.1}$$

We will also need size biasing and zero biasing of infinitely divisible distributions on \mathbb{R}_+ , see [Arratia, Goldstein & Kochman (2019), Theorem 11.2] and [Arras & Houdré (2019), Proposition 3.8] for more details. The direct relationship between the two biasings seems to be not noted anywhere in the literature so we give a proof for the relationship.

Lemma 4.2. *Suppose X is a nonnegative random variable with $\text{Var}(X) > 0$.*

- (a) $\mathcal{L}(X)$ is infinitely divisible if and only if there exists a random variable $X' \geq 0$ a.s. independent of X such that the size biased distribution of X satisfies

$$X^s \stackrel{d}{=} X + X'. \tag{4.2}$$

- (b) $\mathcal{L}(X)$ is infinitely divisible if and only if there exists a random variable X'' independent of X such that the zero biased distribution of X satisfies

$$X^z \stackrel{d}{=} X + X''. \tag{4.3}$$

The distribution of X'' has the density function

$$\frac{1}{\mathbb{E} X'} \mathbb{P}(X' \geq x), \quad x \geq 0, \tag{4.4}$$

where X' is uniquely determined in (4.2).

Proof of (4.4) The proof is a simple application of the Laplace transform. For $\theta > 0$, denote $\phi_V(\theta) := \mathbb{E} e^{-\theta V}$ for some non-negative random variable V . For simplicity, we denote $\mu := \mathbb{E} X$, and $\sigma^2 := \text{Var}(X)$. By taking $V = X$ and $f(x) = e^{-\theta x}$ in (1.2), together with $\mathbb{E} X' = \mathbb{E} X^s - \mathbb{E} X = \frac{\sigma^2}{\mu}$, we have

$$\begin{aligned} \phi_{X^s}(\theta) &= -\frac{1}{\sigma^2 \theta} \mathbb{E} [(X - \mathbb{E} X) e^{-\theta X}] \\ &= -\frac{\mu}{\sigma^2 \theta} [\phi_{X^s}(\theta) - \phi_X(\theta)] \\ &= \phi_X(\theta) \frac{\mu}{\sigma^2 \theta} \mathbb{E}[1 - e^{-\theta X'}] \\ &= \phi_X(\theta) \frac{\mu}{\sigma^2} \int_0^\infty e^{-\theta x} [1 - \mathbb{P}(X' \leq x)] dx \\ &= \phi_X(\theta) \int_0^\infty e^{-\theta x} \frac{1 - \mathbb{P}(X' \leq x)}{\mathbb{E} X'} dx, \end{aligned}$$

where the third equality is due to (4.2), and the fourth one is from the integration by parts. This is equivalent to (4.3). \square

We note that the distribution of X'' is also called the *equilibrium distribution* with respect to X' in [Peköz & Röllin (2011)]. [Arras & Houdré (2019)] state that the only probability measure that has an additive exponential size biased distribution is the gamma distribution and [Peköz & Röllin (2011)] say that the exponential distribution is the unique fixed point under the equilibrium transformation, their observations confirm Theorem 2.1 in the case of infinitely divisible distributions on \mathbb{R}_+ .

The first application we consider is to estimate the difference between a compound Poisson distribution and a gamma distribution having the same mean and variance. Recall that a compound Poisson distribution, denoted by $\text{CP}(\lambda, \mathcal{L}(X))$, is the distribution of $W = \sum_{i=1}^N X_i$, where $\{X, X_i, i \geq 1\}$ are i.i.d. random variables independent of $N \sim \text{Pn}(\lambda)$.

Proposition 4.3. Assume that $X \geq 0$ and $\text{Var}(X) \in (0, \infty)$ such that both X^s and X^z exist. Let $W \sim \text{CP}(\lambda, \mathcal{L}(X))$ with $\mu = \mathbb{E}W$ and $\sigma^2 = \text{Var}(W)$. Taking $\alpha = \mu/\sigma^2$, $r = \mu^2/\sigma^2$, then (3.2) and (3.3) hold with $\Theta = d_W(\mathcal{L}(X^s), \mathcal{L}(\tilde{X}))$, where the density function of \tilde{X} is given by

$$f_{\tilde{X}}(y) = \frac{1}{\mathbb{E}(X^2)} \mathbb{E}[X \mathbf{1}_{\{X \geq y\}}], \quad y \geq 0.$$

Proof In this case, $\mathcal{L}(W)$ is finitely divisible and $W^s = W + X^s$ [Arratia, Goldstein & Kochman (2019), p. 7], Lemma 4.2 ensures $W^z = W + \tilde{X}$, the claim is an immediate consequence of Theorem 3.1. \square

The intriguing phenomenon that the gamma process is a pure-jump increasing process while the gamma distribution is continuous can be well explained by the following bound. The bound is the result of direct calculations from Proposition 4.3 and (4.1), so we omit the details. The example tells us that the bounds in Proposition 4.3 have some room for improvement. Recall that the Lévy measure of the gamma process is $\gamma(dy) = e^{-y}/y \, dy$.

Example 4.4. Let $W \sim \text{CP}(\lambda, \mathcal{L}(X))$ with $\lambda = \int_{\delta}^{\infty} \frac{e^{-x}}{x} dx$ and the density of $\mathcal{L}(X)$ given by $\frac{1}{\lambda} \frac{e^{-x}}{x} \mathbf{1}_{\{x \geq \delta\}}$ for some $\delta > 0$, then

$$d_W(\mathcal{L}(W), \Gamma(1, 1)) \leq 8\sqrt{\delta} + \frac{17}{3}\delta.$$

Remark 4.5. The bound in Example 4.4 is not of optimal order. In fact, let Ξ be a Poisson point process on $(0, \infty)$ with intensity measure $\mu(dx) = \frac{e^{-x}}{x} dx$, and $\tilde{\Xi} := \Xi|_{(\delta, \infty)}$ be the restriction of Ξ to (δ, ∞) . Let $Z := \sum_{\xi \in \Xi} \xi$ and $\tilde{W} := \sum_{\xi \in \tilde{\Xi}} \xi$. It is easy to verify that $\tilde{W} \stackrel{d}{=} W$ and $Z \sim \Gamma(1, 1)$, which ensure that $\mathcal{L}(Z)$ is stochastically bigger than $\mathcal{L}(W)$, hence

$$d_W(\mathcal{L}(W), \Gamma(1, 1)) = \mathbb{E}Z - \mathbb{E}W = 1 - e^{-\delta} \leq \delta.$$

It is tempting to ask whether for $W \sim \text{CP}(\lambda, \mathcal{L}(X))$ with $\lambda \rightarrow \infty, \mathbb{E}W \rightarrow \frac{a}{b}$ and $\text{Var}(W) \rightarrow \frac{a}{b^2}$ for some $a, b > 0$ are sufficient to ensure that W converges in distribution to $\Gamma(a, b)$. The following example gives a negative answer to this question, indicating that $d_W(\mathcal{L}(X^s), \mathcal{L}(\tilde{X})) \rightarrow 0$ in Proposition 4.3 is also necessary for the compound Poisson distribution to be close to the gamma distribution.

Counterexample 4.6. Let $\mathbb{P}(X = 1) = 1/\lambda = 1 - \mathbb{P}(X = 0)$, then $W \sim \text{CP}(\lambda, \mathcal{L}(X)) = \text{Pn}(1)$, $\mathbb{E}W = \text{Var}(W) = 1$, but W does not converge to $\Gamma(1, 1)$ as $\lambda \rightarrow \infty$.

For $\kappa > 0$ and $0 < p < 1$, we write $V \sim \text{NB}(\kappa, p)$ if

$$\mathbb{P}(V = i) = \frac{\Gamma(\kappa + i)}{\Gamma(\kappa)i!} p^{\kappa}(1 - p)^i, \quad i = 0, 1, \dots$$

Hence, $\mathbb{E}V = \kappa(1 - p)/p$ and $\text{Var}(V) = \kappa(1 - p)/p^2$.

To estimate gamma distribution approximation to negative binomial sums, we first bound the difference between a rescaled negative binomial distribution and a gamma distribution.

Proposition 4.7. *Let $T_p \sim \text{NB}(\kappa, p)$, $W_p := pT_p$. Then*

$$d_W(\mathcal{L}(W_p), \Gamma(\kappa(1-p), 1)) \leq 4\sqrt{\frac{6\kappa(1-p)p}{\kappa(1-p)+2}} + \frac{4\kappa(1-p)p}{\kappa(1-p)+2},$$

$$d_K(\mathcal{L}(W_p), \Gamma(\kappa(1-p), 1)) \leq a_{r,\alpha}(0.5p)^{\frac{r\wedge 1}{r\wedge 1+2}} + b_{r,\alpha}(0.5p)^{\frac{r\wedge 1+1}{r\wedge 1+2}},$$

where $a_{r,\alpha}$ and $b_{r,\alpha}$ are given in (3.4) and (3.5).

Proof We write $T_p := \sum_{i=1}^N X_i$, where $\{X, X_1, X_2, \dots\}$ are i.i.d. random variables with the logarithmic distribution

$$\mathbb{P}(X = i) = -\frac{1}{\ln(p)} \frac{(1-p)^i}{i}, \quad i \geq 1, \tag{4.5}$$

and $N \sim \text{Pn}(-\kappa \ln(p))$, independent of X_i 's. It is obvious to see that $\mathbb{E} X = -\frac{1-p}{p \ln(p)}$, and $\mathbb{E}(X^2) = -\frac{1-p}{p^2 \ln(p)}$, giving

$$\mathbb{E} W_p = \text{Var}(W_p) = \kappa(1-p).$$

From Lemma 4.2, we know that $W_p = p \sum_{i=1}^N X_i$ has the size biased distribution

$$W_p^s \stackrel{d}{=} W_p + (pX)^s,$$

where W_p and $(pX)^s$ are independent. For any $i \in \mathbb{N}$, $x = pi$, we have

$$\mathbb{P}((pX)^s = x) = \frac{x \mathbb{P}(pX = x)}{\mathbb{E}(pX)} = \frac{i \mathbb{P}(X = i)}{\mathbb{E} X} = \mathbb{P}(X^s = i). \tag{4.6}$$

From (4.5), we can derive the size biased distribution of X : for $i \geq 1$

$$\mathbb{P}(X^s = i) = \frac{i \mathbb{P}(X = i)}{\mathbb{E} X} = p(1-p)^{i-1}. \tag{4.7}$$

Likewise, $W_p^z \stackrel{d}{=} W_p + \tilde{X}$, where \tilde{X} is independent of W_p , having density

$$f_{\tilde{X}}(y) = \frac{\mathbb{E}[(pX)\mathbf{1}_{\{pX \geq y\}}]}{\mathbb{E}[(pX)^2]} = \frac{\mathbb{E}[X\mathbf{1}_{\{X \geq y/p\}}]}{p \mathbb{E}(X^2)}.$$

Noting that for $p(i-1) < y \leq pi$, we have $i-1 < y/p \leq i$, thus

$$\begin{aligned} f_{\tilde{X}}(y) &= \frac{\mathbb{E} X \mathbb{P}(X^s \geq y/p)}{p \mathbb{E}(X^2)} = \mathbb{P}(X^s \geq i) \\ &= (1-p)^{i-1}. \end{aligned}$$

$$\begin{aligned} d_W(\mathcal{L}((pX)^s), \mathcal{L}(\tilde{X})) &= \int_0^\infty \left| \mathbb{P}((pX)^s \geq x) - \mathbb{P}(\tilde{X} \geq x) \right| dx \\ &= \sum_{i=1}^\infty \int_{p(i-1)}^{pi} \left| \mathbb{P}((pX)^s \geq x) - \mathbb{P}(\tilde{X} \geq x) \right| dx. \end{aligned}$$

For $p(i-1) < x \leq pi$, from (4.6) and (4.7), we have

$$\mathbb{P}((pX)^s \geq x) = \sum_{k=i}^\infty \mathbb{P}(X^s = k) = (1-p)^{i-1}.$$

On the other hand,

$$\begin{aligned} \mathbb{P}(\tilde{X} \geq x) &= \int_x^\infty f_{\tilde{X}}(y)dy = (pi - x)(1 - p)^{i-1} + \sum_{k=i}^\infty p(1 - p)^k \\ &= (pi - x)(1 - p)^{i-1} + (1 - p)^i < (1 - p)^{i-1} = \mathbb{P}((pX)^s \geq x). \end{aligned}$$

Therefore,

$$\begin{aligned} d_W(\mathcal{L}((pX)^s), \mathcal{L}(\tilde{X})) &= \int_0^\infty (\mathbb{P}((pX)^s \geq x) - \mathbb{P}(\tilde{X} \geq x)) dx \\ &= \mathbb{E}(pX)^s - \mathbb{E}\tilde{X} \\ &= \frac{\mathbb{E}(pX)^2}{\mathbb{E}(pX)} - \sum_{i=1}^\infty \int_{p^{i-1}}^{p^i} \mathbb{P}(\tilde{X} \geq x)dx \\ &= 1 - \sum_{i=1}^\infty (p(1 - p)^i + 0.5p^2(1 - p)^{i-1}) \\ &= 0.5p. \end{aligned}$$

This, together with Proposition 4.3, completes the proof. □

Corollary 4.8. *Let $T_p \sim \text{NB}(\kappa, p)$ and $\{X_i\}$ be a sequence of random variables, define $S_n = \sum_{i=1}^n X_i$. Assume that $\mathbb{E}(X_i|T_p) = 1$ and $\text{Var}(S_i|T_p = i) = i\nu^2$ for all $i \geq 1$, then $W_p = p \sum_{i=1}^{T_p} X_i$ satisfies*

$$d_W(\mathcal{L}(W_p), \Gamma(\kappa, 1)) \leq 4\sqrt{\frac{6\kappa(1 - p)p}{\kappa(1 - p) + 2}} + \frac{4\kappa(1 - p)p}{\kappa(1 - p) + 2} + \nu\sqrt{\kappa(1 - p)p} + p\kappa.$$

Proof By the triangle inequality,

$$\begin{aligned} &d_W(\mathcal{L}(W_p), \Gamma(\kappa, 1)) \\ &\leq d_W(\mathcal{L}(W_p), \mathcal{L}(pT_p)) + d_W(\mathcal{L}(pT_p), \Gamma(\kappa(1 - p), 1)) + d_W(\Gamma(\kappa(1 - p), 1), \Gamma(\kappa, 1)) \\ &\leq \mathbb{E}|W_p - pT_p| + 4\sqrt{\frac{6\kappa p(1 - p)}{\kappa(1 - p) + 2}} + \frac{4\kappa p(1 - p)}{\kappa(1 - p) + 2} + p\kappa, \end{aligned}$$

where the last inequality comes from Proposition 4.7 and Lemma 4.1 using the fact that $\Gamma(r, 1)$ is stochastically increasing in r . The remaining part $\mathbb{E}|W_p - pT_p|$ is bounded by the Cauchy-Schwarz inequality:

$$\begin{aligned} \mathbb{E}|W_p - pT_p| &= p \mathbb{E} \left| \sum_{i=1}^{T_p} (X_i - 1) \right| \leq p \sqrt{\text{Var} \left(\sum_{i=1}^{T_p} (X_i - 1) \right)} \\ &= p \sqrt{\mathbb{E} \left[\text{Var} \left(\sum_{i=1}^{T_p} (X_i - 1) \middle| T_p \right) \right]} = p \sqrt{\mathbb{E}[T_p \nu^2]} = \nu \sqrt{\kappa(1 - p)p}. \quad \square \end{aligned}$$

References

- [Arras & Houdré (2019)] Arras, B. & Houdré, C. (2019). On Stein’s Method for Infinitely Divisible Laws with Finite First Moment. *Springer International Publishing*. MR3931309
- [Arratia, Goldstein & Kochman (2019)] Arratia, R., Goldstein, L. & Kochman, F. (2019). Size bias for one and all. *Probab. Surveys* **16**, 1-61. MR3896143 MR3896143

- [Barbour, Holst & Janson (1992)] Barbour, A. D., Holst, L. & Janson, S. (1992). Poisson Approximation. *The Clarendon Press Oxford University Press*. MR1163825
- [Batir (2008)] Batir, N. (2008). Inequalities for the gamma function. *Arch. Math.* **91**, 554–563. MR2465874
- [Berry (1941)] Berry, A. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society* **49**, 122–136. MR0003498
- [Chen, Goldstein & Shao (2010)] Chen, L. H. Y., Goldstein, L. & Shao, Q. M. (2010). Normal approximation by Stein’s method. *Springer Science & Business Media*. MR2732624
- [Covo & Elalouf (2014)] Covo, S. & Elalouf, A. (2014). A novel single-gamma approximation to the sum of independent gamma variables, and a generalization to infinitely divisible distributions. *Electron. J. Stat.* **8**, 894–926. MR3229102
- [Diaconis & Zabell (1991)] Diaconis, P. & Zabell, S. (1991). Closed Form Summation for Classical Distributions: Variations on a Theme of De Moivre. *Statist. Sci.* **6**, 284–302. MR1144242
- [Eichelsbacher & Thäle (2015)] Eichelsbacher, P. & Thäle, C. (2015). Malliavin-Stein method for Variance-Gamma approximation on Wiener space. *Electron. J. Probab.* **20**, 1–28. MR3425543
- [Esseen (1942)] Esseen, C. G. (1942). On the Liapounoff limit of error in the theory of probability. *Arkiv För Matematik, Astronomi Och Fysik*, 28A, 19 pp. MR0011909
- [Esseen (1958)] Esseen, C. G. (1958). On mean central limit theorems. *Kungl. Tekn. Högsk. Handl. Stockholm* **121**, 31. MR0097111
- [Gaunt (2019)] Gaunt, R. E. (2019). New error bounds for Laplace approximation via Stein’s method. arXiv:1911.03574 MR4291370
- [Gaunt, Pickett & Reinert (2017)] Gaunt, R. E., Pickett, A. M. & Reinert, G. (2017). Chi-square approximation by Stein’s method with application to Pearson’s statistic. *Ann. Appl. Probab.* **27**, 720–756. MR3655852
- [Goldstein & Reinert (1997)] Goldstein, L. & Reinert, G. (1997). Stein’s method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935–952. MR1484792
- [Goldstein & Reinert (2005)] Goldstein, L. & Reinert, G. (2005). Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18**, 237–260. MR2132278
- [Goldstein & Rinott (2005)] Goldstein, L. & Rinott, Y. (2005). Multivariate normal approximations by Stein’s method and size bias couplings. *J. Appl. Probab.* **33**, 1–17. MR1371949
- [Goldstein & Xia (2006)] Goldstein, L. and Xia, A. (2006). Zero biasing and a discrete central limit theorem. *Ann. Probab.* **34**, 1782–1806. MR2271482
- [Jensen & Solomon (1972)] Jensen, D. R. & Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form. *J. Amer. Statist. Assoc.* **67**, 898–902. MR1294073
- [Kalashnikov (1997)] Kalashnikov, V. V. (1997). Geometric Sums: Bounds for Rare Events with Applications: Risk Analysis, Reliability, Queueing. *Mathematics and Its Applications* **413**. Kluwer, Dordrecht. MR1471479
- [Ledoux, Nourdin & Peccati (2015)] Ledoux, M., Nourdin, I. & Peccati, G. (2015). Stein’s method, logarithmic Sobolev and transport inequalities. *Geometric and Functional Analysis* **25**, 256–306. MR3320893
- [Luk (1994)] Luk, H. M. (1994). Stein’s method for the gamma distribution and related statistical applications (Doctoral dissertation, University of Southern California). MR2693204
- [Lukacs (1970)] Lukacs, E. (1970). Characteristic functions. Griffin, London. MR0346874
- [Moschopoulos (1985)] Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.* **37**, Part A, 541–544. MR0818052
- [Nourdin & Peccati (2009)] Nourdin, I. & Peccati, G. (2009). Stein’s method on Wiener chaos. *Probab. Theory Related Fields* **145**, 75–118. MR2520122
- [Peköz & Röllin (2011)] Peköz, E. A. & Röllin, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.* **39**, 587–608. MR2789507

- [Rényi (1956)] Rényi, A. (1956). A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató. Int. Közl.* **1**, 519–527. MR0094861
- [Slepov (2021)] Slepov, N. A. (2021). Convergence Rate of Random Geometric Sum Distributions to the Laplace Law. *Theory Probab. Appl.* **66**, 121–141. MR4213093
- [Stein (1972)] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 2: Probability Theory. The Regents of the University of California. MR0402873
- [Stewart et al. (2007)] Stewart, T., Strijbosch, L. W. G., Moors, H. & Batenburg, P. (2007). A Simple Approximation to the Convolution of Gamma Distributions. *Center Discussion Paper No. 2007-70*, <http://dx.doi.org/10.2139/ssrn.900109>.
- [Sugakova (1995)] Sugakova, E. V. (1995). Estimates in the Rényi theorem for differently distributed terms. *Ukrainian Mathematical Journal* **47**, 1128–1134. MR1367955
- [Vallender (1972)] Vallender, S. S. (1972). Calculation of the Wasserstein distance between probability distributions on the line. *Theory Probab. Appl.* **18**, 784–786. MR0328982
- [Vellaisamy & Upadhye (2009)] Vellaisamy, P. & Upadhye, N. S. (2009). On the sums of compound negative binomial and gamma random variables. *J. Appl. Probab.* **46**, 272–283. MR2508518
- [Zolotarev (1964)] Zolotarev, V. M. (1964). On asymptotically best constants in refinements of mean central limit theorems. *Theory Probab. Appl.* **9**, 268–276. MR0163338

Acknowledgments. We thank Nathan Ross for suggesting the direct coupling proof in Remark 4.5 and in the context of infinitely divisible distributions on \mathbb{R}_+ , Theorem 2.1 is indirectly confirmed in the literature. We also thank a referee for bringing to our attention that the zero-bias transformation dates back at least to [Lukacs (1970), Theorem 12.2.5 (a)].

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>