STEIN'S METHOD AND THE LAPLACE DISTRIBUTION

John Pike
Department of Mathematics
Cornell University
jpike@cornell.edu

Haining Ren
Department of Mathematics
University of Southern California
hren@usc.edu

Abstract. Using Stein's method techniques, we develop a framework which allows one to bound the error terms arising from approximation by the Laplace distribution and apply it to the study of random sums of mean zero random variables. As a corollary, we deduce a Berry-Esseen type theorem for the convergence of certain geometric sums. Our results make use of a second order characterizing equation and a distributional transformation which is related to zero-biasing.

1. Background and Introduction

Beginning with the publication of Charles Stein's seminal paper [16], probabilists and statisticians have developed a wide range of techniques based on characterizing equations for bounding the distance between the distribution of a random variable X and that of a random variable Z having some specified target distribution. The metrics for which these techniques are applicable are of the form $d_{\mathcal{H}}(\mathscr{L}(X), \mathscr{L}(Z)) = \sup_{h\in\mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|$ for some suitable class of functions $\mathcal{H}$, and include as special cases the Wasserstein, Kolmogorov, and total variation distances. (The Kolmogorov distance gives the L^∞ distance between the associated distribution functions, so $\mathcal{H} = \{1_{(-\infty,a]}(x) : a \in \mathbb{R}\}$. The total variation and Wasserstein distances correspond to letting $\mathcal{H}$ consist of indicators of Borel sets and 1-Lipschitz functions, respectively.) The basic idea is to find an operator $\mathcal{A}$ such that $\mathbb{E}[(\mathcal{A}f)(X)] = 0$ for all f belonging to some sufficiently large class of functions $\mathcal{F}$ if and only if $\mathscr{L}(X) = \mathscr{L}(Z)$. For example, Stein showed that $Z \sim \mathcal{N}(0, \sigma^2)$ if and only if $\mathbb{E}[(\mathcal{A}_N f)(Z)] = \mathbb{E}[Zf(Z) - \sigma^2 f'(Z)] = 0$ for all absolutely continuous functions f with $\|f'\|_\infty < \infty$ [16], and shortly thereafter Louis Chen proved that $Z \sim \text{Poisson}(\lambda)$ if and only if $\mathbb{E}[(\mathcal{A}_P f)(Z)] = \mathbb{E}[Zf(Z) - \lambda f(Z+1)] = 0$ for all functions f for which the expectations exist [2]. Similar characterizing operators have since been worked out for several other common distributions (e.g. [6, 11, 1, 13]).

2010 *Mathematics Subject Classification.* 60F05.
Key words and phrases. Stein's Method, Laplace distribution, geometric summation.

Given a Stein operator $\mathcal{A}$ for $\mathscr{L}(Z)$, one then shows that for every $h \in \mathcal{H}$, the equation $(\mathcal{A}f)(x) = h(x) - \mathbb{E}[h(Z)]$ has solution $f_h \in \mathcal{F}$. Taking expectations, absolute values, and suprema yields

$$d_{\mathcal{H}}(\mathscr{L}(X), \mathscr{L}(Z)) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]| = \sup_{h \in \mathcal{H}} |\mathbb{E}[(\mathcal{A}f_h)(X)]| .$$

(In practice, this is usually how one proves that $\mathbb{E}[(\mathcal{A}f)(X)] = 0$ for $f \in \mathcal{F}$ is sufficient for $\mathscr{L}(X) = \mathscr{L}(Z)$.)

The intuition is that since $\mathbb{E}[(\mathcal{A}f)(Z)] = 0$ for $f \in \mathcal{F}$, the distribution of X should be close to that of Z when $\mathbb{E}[(\mathcal{A}f)(X)]$ is close to zero. Remarkably, it is often easier to work with the right-hand side of the above equation, and the tools for analyzing distances between distributions in this manner are collectively known as Stein's method. For more on this rich and fascinating subject, the authors recommend [17, 15, 3, 4].

In this paper, we apply the above ideas to the Laplace distribution. For $a \in \mathbb{R}$, $b > 0$, a random variable $W \sim \text{Laplace}(a, b)$ has distribution function

$$F_W(w; a, b) = \begin{cases} \frac{1}{2} e^{\frac{w-a}{b}}, & w \leq a \\ 1 - \frac{1}{2} e^{-\frac{w-a}{b}}, & w \geq a \end{cases}$$

and density

$$f_W(w; a, b) = \frac{1}{2b} e^{-\frac{|w-a|}{b}}, \ w \in \mathbb{R}.$$

If $W \sim \text{Laplace}(0, b)$, then its moments are given by

$$\mathbb{E}[W^k] = \begin{cases} 0, & k \text{ is odd} \\ b^k k!, & k \text{ is even} \end{cases},$$

and its characteristic function is

$$\varphi_W(t) = \frac{1}{1 + b^2 t^2}.$$

This distribution was introduced by P.S. Laplace in 1774, four years prior to his proposal of the "second law of errors," now known as the normal distribution. Though nowhere near as ubiquitous as its younger sibling, the Laplace distribution appears in numerous applications, including image and speech compression, options pricing, and modeling sizes of sand particles, diamonds, and beans. For more properties and applications of the Laplace distribution, the reader is referred to the text [10].

Our interest in the Laplace distribution was sparked by the fact that if $X_1, X_2, \ldots$ is a sequence of random variables (satisfying certain technical assumptions) and $N_p \sim \text{Geometric}(p)$ is independent of the X_i's, then the sum $p^{\frac{1}{2}} \sum_{i=1}^{N_p} X_i$ converges weakly to the Laplace distribution as $p \to 0$ [10]. Such geometric sums arise in a variety of settings [9], and the general setup (distributional convergence of sums of random variables) is exactly the type of problem for which one expects Stein's method computations to yield useful results. Indeed, Erol Peköz and Adrian Röllin have applied Stein's method arguments to generalize a theorem due to Rényi concerning the convergence of sums of a random number of positive random variables to the exponential distribution [12]. By an analogous line of reasoning, we are able to carry out a similar program for convergence of random sums of certain mean zero random variables to the Laplace distribution.

We begin in Section 2 by introducing a Stein operator which we show completely characterizes the mean zero Laplace distribution. Specifically, we prove

Theorem 1.1. *Let $W \sim \text{Laplace}(0, b)$ and define the operator $\mathcal{A}$ by*

$$(\mathcal{A}f)(x) = f(x) - f(0) - b^2 f''(x).$$

Then $\mathbb{E}[(\mathcal{A}g)(W)] = 0$ for every function g such that g and g' are locally absolutely continuous and $\mathbb{E}\left|g'(W)\right|, \mathbb{E}\left|g''(W)\right| < \infty$.

Conversely, if X is any random variable such that $\mathbb{E}[(\mathcal{A}g)(X)] = 0$ for every twice-differentiable function g with $\|g\|_\infty, \|g'\|_\infty, \|g''\|_\infty < \infty$, then $X \sim \text{Laplace}(0, b)$.

In Section 3, we use this characterization to bound the distance to the Laplace distribution. For technical reasons, we work in the bounded Lipschitz metric, d_{BL}, which is defined in terms of 1-Lipschitz test functions with sup norm 1. We begin by defining the centered equilibrium transformation $X \mapsto X^L$ in terms of the functional equation $\mathbb{E}[f(X)] - f(0) = \frac{1}{2}\mathbb{E}[X^2]\mathbb{E}[f''(X^L)]$ for all twice-differentiable functions f such that f, f', and f'' are bounded. After establishing that X^L exists whenever X has mean zero and finite nonzero variance, we derive the following coupling bound.

Theorem 1.2. *Suppose that X is a random variable with $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = 2b^2 \in (0, \infty)$, and let X^L have the centered equilibrium distribution for X. Then*

$$d_{BL}(\mathscr{L}(X), \text{Laplace}(0, b)) \leq \frac{b+2}{b}\mathbb{E}\left|X - X^L\right|.$$

Finally, in Section 4 we apply these tools to the study of random sums of mean zero random variables. As a special case, we show

Theorem 1.3. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 2b^2$, $\sup_{i \geq 1} \mathbb{E}\left[|X_i|^3\right] = \rho < \infty$, and let $N \sim \text{Geometric}(p)$ (with strictly positive support) be independent of the $X_i's$. Then*

$$d_{BL}\left(\mathscr{L}\left(p^{\frac{1}{2}} \sum_{i=1}^{N} X_i\right), \text{Laplace}\left(0, b\right)\right) \leq p^{\frac{1}{2}} \frac{b+2}{b}\left(b\sqrt{2} + \frac{\rho}{6b^2}\right)$$

for all $p \in (0, 1)$.

2. Characterizing the Laplace Distribution

Our first order of business is to establish a characterizing operator for the Laplace distribution. As is typical in Stein's method constructions, we split the proof of Theorem 1.1 into two parts. We begin with

Lemma 2.1. *Suppose that $W \sim \text{Laplace}(0, b)$. If g and g' are locally absolutely continuous with $\mathbb{E}\left|g'(W)\right|, \mathbb{E}\left|g''(W)\right| < \infty$, then*

$$\mathbb{E}[g(W)] - g(0) = b^2\mathbb{E}[g''(W)].$$

Proof. Applying Fubini's theorem twice shows that

$$\begin{aligned}
\int_0^\infty g''(x)e^{-\frac{x}{b}}dx &= \int_0^\infty g''(x)\left(\int_x^\infty \frac{1}{b}e^{-\frac{y}{b}}dy\right)dx \\
&= \frac{1}{b}\int_0^\infty \int_0^y g''(x)e^{-\frac{y}{b}}dxdy \\
&= \frac{1}{b}\int_0^\infty g'(y)e^{-\frac{y}{b}}dy - \frac{g'(0)}{b}\int_0^\infty e^{-\frac{y}{b}}dy \\
&= \frac{1}{b}\int_0^\infty g'(y)\left(\int_y^\infty \frac{1}{b}e^{-\frac{z}{b}}dz\right)dy - g'(0) \\
&= \frac{1}{b^2}\int_0^\infty \int_0^z g'(y)e^{-\frac{z}{b}}dydz - g'(0) \\
&= \frac{1}{b^2}\int_0^\infty g(z)e^{-\frac{z}{b}}dz - \frac{g(0)}{b^2}\int_0^\infty e^{-\frac{z}{b}}dz - g'(0) \\
&= \frac{1}{b^2}\int_0^\infty g(z)e^{-\frac{z}{b}}dz - \frac{g(0)}{b} - g'(0).
\end{aligned}$$

Setting $h(y) = g(-y)$, it follows from the previous calculation that

$$\begin{aligned}
\int_{-\infty}^0 g''(x)e^{\frac{x}{b}}dx = \int_0^\infty g''(-y)e^{-\frac{y}{b}}dy &= \int_0^\infty h''(y)e^{-\frac{y}{b}}dy \\
&= \frac{1}{b^2}\int_0^\infty h(z)e^{-\frac{z}{b}}dz - \frac{h(0)}{b} - h'(0) \\
&= \frac{1}{b^2}\int_0^\infty g(-z)e^{-\frac{z}{b}}dz - \frac{g(0)}{b} + g'(0) \\
&= \frac{1}{b^2}\int_{-\infty}^0 g(z)e^{\frac{z}{b}}dz - \frac{g(0)}{b} + g'(0).
\end{aligned}$$

Summing the above expressions gives

$$\begin{aligned}
\mathbb{E}[g''(W)] &= \frac{1}{2b}\int_{-\infty}^0 g''(x)e^{\frac{x}{b}}dx + \frac{1}{2b}\int_0^\infty g''(x)e^{-\frac{x}{b}}dx \\
&= \frac{1}{2b}\left[\left(\frac{1}{b^2}\int_{-\infty}^0 g(z)e^{\frac{z}{b}}dz - \frac{g(0)}{b} + g'(0)\right)\right. \\
&\qquad \left. + \left(\frac{1}{b^2}\int_0^\infty g(z)e^{-\frac{z}{b}}dz - \frac{g(0)}{b} - g'(0)\right)\right] \\
&= \frac{1}{2b}\left(\frac{1}{b^2}\int_{-\infty}^\infty g(z)e^{-\frac{|z|}{b}}dz - 2\frac{g(0)}{b}\right) = \frac{1}{b^2}\left(\mathbb{E}[g(W)] - g(0)\right).
\end{aligned}$$

□

Note that since the density of a Laplace$(0, b)$ random variable is given by $f_W(w) = \frac{1}{2b}e^{-\frac{|w|}{b}}$, the density method [3] suggests the following characterizing equation for the Laplace distribution:

$$g'(w) - \frac{1}{b}\mathrm{sgn}(w)g(w) = g'(w) + \frac{f_W'(w)}{f_W(w)}g(w) = 0,$$

and indeed one can verify that if $W \sim \text{Laplace}(0, b)$, then

$$\mathbb{E}[g'(W)] = \frac{1}{b}\mathbb{E}[\text{sgn}(W)g(W)]$$

for all absolutely continuous g for which these expectations exist. Thus if g' is such a function as well, setting $G(w) = \text{sgn}(w)\,(g(w) - g(0))$ gives

$$\begin{aligned}\mathbb{E}[g''(W)] &= \frac{1}{b}\mathbb{E}[\text{sgn}(W)g'(W)] = \frac{1}{b}\mathbb{E}[G'(W)] \\ &= \frac{1}{b^2}\mathbb{E}[\text{sgn}(W)G(W)] = \frac{1}{b^2}\left(\mathbb{E}[g(W)] - g(0)\right),\end{aligned}$$

so the general form of the equation in Theorem 1.1 can be ascertained by iterating the density method.

Alternatively, it is known [15] that if $Z \sim \text{Exponential}(1)$, then $\mathbb{E}[g'(Z)] = \mathbb{E}[g(Z)] - g(0)$ for all absolutely continuous g with $\mathbb{E}\,|g'(Z)| < \infty$. Thus if g' is also absolutely continuous and $\mathbb{E}\,|g''(Z)| < \infty$, then

$$\mathbb{E}[g''(Z)] = \mathbb{E}[g'(Z)] - g'(0) = \mathbb{E}[g(Z)] - g(0) - g'(0).$$

Using this observation, one can derive the equation in Lemma 2.1 by noting that if J is independent of Z with $\mathbb{P}(J = 1) = \mathbb{P}(J = -1) = \frac{1}{2}$, then $W = bJZ$ has the $\text{Laplace}(0, b)$ distribution.

We include each of these approaches because their analogues may be variously applicable in different situations involving the construction of characterizing equations. Observe that there is an iterative step involved in each case. By manipulating the integral defining $\mathbb{E}[g'(W)]$ or using the usual Stein equation for the exponential along with the representation $W = bJZ$, one arrives at the first order equation $\mathbb{E}[g'(W)] = \frac{1}{b}\mathbb{E}[\text{sgn}(W)g(W)]$ suggested by one application of the density method. However, we were not able to get much mileage out of the operator $(\widetilde{\mathcal{A}}g)(x) = g'(x) - \frac{1}{b}\text{sgn}(x)g(x)$, while the second-order operator $(\mathcal{A}g)(x) = g(x) - g(0) - b^2 g''(x)$ turned out to be quite effective. This idea of iterating more traditional procedures to obtain higher order characterizing equations which are simpler to work with is one of the key insights of this paper.

Now, in order to establish the second part of Theorem 1.1, we will show that any X satisfying the hypotheses has

$$d_{BL}\left(\mathscr{L}(X), \text{Laplace}(0, b)\right) = 0$$

where d_{BL} denotes the bounded Lipschitz distance given by

$$d_{BL}\left(\mathscr{L}(X), \mathscr{L}(Y)\right) = \sup_{h \in \mathcal{H}_{BL}} \left|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]\right|,$$

$$\mathcal{H}_{BL} = \{h : \|h\|_\infty \leq 1 \text{ and } |h(x) - h(y)| \leq |x - y| \text{ for all } x, y \in \mathbb{R}\}.$$

The claim will follow since d_{BL} is a metric on the space of Borel probability measures on $\mathbb{R}$ [19].

In keeping with the general strategy laid out in the introduction, we consider the initial value problem

$$g(x) - b^2 g''(x) = h(x) - Wh,\ g(0) = 0$$

where $h \in \mathcal{H}_{BL}$ and $Wh := \mathbb{E}[h(W)]$, $W \sim \text{Laplace}(0, b)$.

Lemma 2.2. *For $h \in \mathcal{H}_{BL}$, $W \sim \text{Laplace}(0,b)$, $\widetilde{h}(x) = h(x) - Wh$, a bounded, twice-differentiable solution to the initial value problem*

$$g(x) - b^2 g''(x) = \widetilde{h}(x),\ g(0) = 0$$

is given by

$$g_h(x) = \frac{1}{2b}\left(e^{\frac{x}{b}}\int_x^\infty e^{-\frac{y}{b}}\widetilde{h}(y)dy + e^{-\frac{x}{b}}\int_{-\infty}^x e^{\frac{y}{b}}\widetilde{h}(y)dy\right).$$

This solution satisfies $\|g_h\|_\infty \le 2$, $\|g_h'\|_\infty \le \frac{2}{b}$, $\|g_h''\|_\infty \le \frac{4}{b^2}$ and $\|g_h'''\|_\infty \le \frac{b+2}{b^3}$.

Proof. The general solution to the homogeneous equation $g''(x) - b^{-2}g(x) = 0$ is given by $g_0(x) = C_1 e^{\frac{x}{b}} + C_2 e^{-\frac{x}{b}}$, so, since the associated Wronskian is nonzero, the variation of parameters method suggests that a solution to the inhomogeneous equation $g''(x) - b^{-2}g(x) = -b^{-2}\widetilde{h}(x)$ is

$$g_h(x) = u_h(x)e^{\frac{x}{b}} + v_h(x)e^{-\frac{x}{b}}$$

where

$$u_h(x) = \frac{1}{2b}\int_x^\infty e^{-\frac{y}{b}}\widetilde{h}(y)dy, \qquad v_h(x) = \frac{1}{2b}\int_{-\infty}^x e^{\frac{y}{b}}\widetilde{h}(y)dy.$$

Differentiation gives

$$g_h'(x) = -\frac{1}{2b}\widetilde{h}(x) + \frac{1}{b}u_h(x)e^{\frac{x}{b}} + \frac{1}{2b}\widetilde{h}(x) - \frac{1}{b}v_h(x)e^{-\frac{x}{b}} = \frac{1}{b}\left(u_h(x)e^{\frac{x}{b}} - v_h(x)e^{-\frac{x}{b}}\right),$$

and thus

$$g_h''(x) = \frac{1}{b^2}\left(-\widetilde{h}(x) + u_h(x)e^{\frac{x}{b}} + v_h(x)e^{-\frac{x}{b}}\right) = \frac{1}{b^2}\left(g_h(x) - \widetilde{h}(x)\right),$$

so g_h is indeed a solution.

To see that the initial condition is satisfied, we observe that

$$\begin{aligned} g_h(0) = u_h(0) + v_h(0) &= \frac{1}{2b}\int_{-\infty}^\infty e^{-\frac{|y|}{b}}\widetilde{h}(y)dy = \int_{-\infty}^\infty f_W(y)\left(h(y) - Wh\right)dy \\ &= \int_{-\infty}^\infty h(y)f_W(y)dy - Wh\int_{-\infty}^\infty f_W(y)dy = Wh - Wh = 0. \end{aligned}$$

Moreover, since $\|h\|_\infty \le 1$,

$$|Wh| = \left|\frac{1}{2b}\int_{-\infty}^\infty h(x)e^{-\frac{|x|}{b}}dx\right| \le \frac{1}{2b}\int_{-\infty}^\infty e^{-\frac{|x|}{b}}dx = 1,$$

and thus $\left|\tilde{h}(x)\right| \le |h(x)| + |Wh| \le 2$. Consequently,

$$\left|u_h(x)e^{\frac{x}{b}}\right| \le \frac{1}{2b}e^{\frac{x}{b}}\int_x^\infty 2e^{-\frac{y}{b}}dy = 1$$

and

$$\left|v_h(x)e^{-\frac{x}{b}}\right| \le \frac{1}{2b}e^{-\frac{x}{b}}\int_{-\infty}^x 2e^{\frac{y}{b}}dy = 1,$$

for all $x \in \mathbb{R}$, and the bounds on $\|g_h\|_\infty$ and $\|g_h'\|_\infty$ follow.

Noting that $g_h''(x) = \frac{1}{b^2}\left(g_h(x) - \widetilde{h}(x)\right)$ and thus $g_h'''(x) = \frac{1}{b^3}g_h'(x) - \frac{1}{b^2}h'(x)$ completes the proof since $\left\|\tilde{h}\right\|_\infty \le 2$ and $\|h'\|_\infty \le 1$. □

With the preceding result in hand, we can finish of the proof of Theorem 1.1 via

Lemma 2.3. *If X is a random variable such*

$$\mathbb{E}[g(X)] - g(0) = b^2\mathbb{E}[g''(X)]$$

for every twice-differentiable function g with $\|g\|_\infty, \|g'\|_\infty, \|g''\|_\infty < \infty$, then $X \sim$ Laplace$(0, b)$.

Proof. Let $W \sim$ Laplace$(0, b)$ and, for $h \in \mathcal{H}_{BL}$, let g_h be as in Lemma 2.2. Because $g_h(0) = 0$ and g_h, g_h', g_h'' are bounded, it follows from the above assumptions that

$$\mathbb{E}[h(X)] - \mathbb{E}[h(W)] = \mathbb{E}[g_h(X) - b^2 g_h''(X)] = 0.$$

Taking the supremum over $h \in \mathcal{H}_{BL}$ shows that $d_{BL}(\mathscr{L}(X), \mathscr{L}(W)) = 0$. □

Before moving on, we observe that the reason we are working with the bounded Lipschitz distance is that the bounds on g_h and its derivatives depended on both h and h' having finite sup norm. As d_{BL} is not especially common (at least explicitly) in the Stein's method literature, we conclude this section with a proposition relating it to the more familiar Kolmogorov distance

$$d_K(\mathscr{L}(X), \mathscr{L}(Y)) = \sup_{x\in\mathbb{R}} |\mathbb{P}\{X \le x\} - \mathbb{P}\{Y \le x\}|.$$

Proposition 2.4. *If Z is an absolutely continuous random variable whose density, f_Z, is uniformly bounded by a constant $C < \infty$, then for any random variable X,*

$$d_K(\mathscr{L}(X), \mathscr{L}(Z)) \le \frac{C+2}{2}\sqrt{d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))}.$$

Proof. We first note that the inequality holds trivially if $d_{BL}(\mathscr{L}(X), \mathscr{L}(Z)) = 0$ as d_{BL} and d_K are metrics. Also, since $d_K(P, Q) \le 1$ for all probability measures P and Q, $\frac{C+2}{2} \ge 1$, and $d_{BL}(\mathscr{L}(X), \mathscr{L}(Z)) \ge 1$ implies $\sqrt{d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))} \ge 1$, we have

$$d_K(\mathscr{L}(X), \mathscr{L}(Z)) \le 1 \le \frac{C+2}{2}\sqrt{d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))}$$

whenever $d_{BL}(\mathscr{L}(X), \mathscr{L}(Z)) \ge 1$. Thus it suffices to consider the case where $d_{BL}(\mathscr{L}(X), \mathscr{L}(Z)) \in (0, 1)$.

Now, for $x \in \mathbb{R}$, $\varepsilon > 0$, write

$$h_x(z) = 1_{(-\infty,x]}(z) = \begin{cases} 1, & z \le x \\ 0, & z > x \end{cases} \text{ and } h_{x,\varepsilon}(z) = \begin{cases} 1, & z \le x \\ 1 - \frac{z-x}{\varepsilon}, & z \in (x, x+\varepsilon] \\ 0, & z > x+\varepsilon \end{cases}.$$

Then for all $x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E}[h_x(X) - h_x(Z)] &= \mathbb{E}[h_x(X)] - \mathbb{E}[h_{x,\varepsilon}(Z)] + \mathbb{E}[h_{x,\varepsilon}(Z)] - \mathbb{E}[h_x(Z)] \\ &\le (\mathbb{E}[h_{x,\varepsilon}(X)] - \mathbb{E}[h_{x,\varepsilon}(Z)]) + \int_x^{x+\varepsilon}\left(1 - \frac{z-x}{\varepsilon}\right) f_Z(z)dz \\ &\le |\mathbb{E}[h_{x,\varepsilon}(X)] - \mathbb{E}[h_{x,\varepsilon}(Z)]| + \frac{C\varepsilon}{2}.\end{aligned}$$

Since $d_{BL}(\mathscr{L}(X), \mathscr{L}(Z)) \in (0,1)$, if we take $\varepsilon = \sqrt{d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))} \in (0,1)$, then $\varepsilon h_{x,\varepsilon} \in \mathcal{H}_{BL}$ and thus

$$\begin{aligned}\mathbb{E}[h_x(X) - h_x(Z)] &\leq \frac{1}{\varepsilon}\left|\mathbb{E}[\varepsilon h_{x,\varepsilon}(X)] - \mathbb{E}[\varepsilon h_{x,\varepsilon}(Z)]\right| + \frac{C\varepsilon}{2} \\ &\leq \frac{1}{\varepsilon} d_{BL}(\mathscr{L}(X), \mathscr{L}(Z)) + \frac{C\varepsilon}{2} \\ &= \frac{C+2}{2}\sqrt{d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))}.\end{aligned}$$

A similar argument using

$$\begin{aligned}\mathbb{E}[h_x(Z) - h_x(X)] &= \mathbb{E}[h_x(Z)] - \mathbb{E}[h_{x-\varepsilon,\varepsilon}(Z)] + \mathbb{E}[h_{x-\varepsilon,\varepsilon}(Z)] - \mathbb{E}[h_x(X)] \\ &\leq \frac{C\varepsilon}{2} + (\mathbb{E}[h_{x-\varepsilon,\varepsilon}(Z)] - \mathbb{E}[h_{x-\varepsilon,\varepsilon}(X)])\end{aligned}$$

shows that

$$\left|\mathbb{E}[h_x(X)] - \mathbb{E}[h_x(Z)]\right| \leq \frac{C+2}{2}\sqrt{d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))}$$

for all $x \in \mathbb{R}$, and the proposition follows by taking suprema. □

Remark. When $C \geq 1$, we can take $\varepsilon = \sqrt{\frac{1}{C} d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))}$ in the above argument to obtain an improved bound of

$$d_K(\mathscr{L}(X), \mathscr{L}(Z)) \leq \frac{3}{2}\sqrt{C d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))}.$$

To the best of the authors' knowledge, the above proposition is original, though the proof follows the same basic line of reasoning as the well-known bound on the Kolmogorov distance by the Wasserstein distance (see Proposition 1.2 in [15]). It seems that the primary reason for using the Wasserstein metric, d_W, is that it enables one to work with smoother test functions while still implying convergence in the more natural Kolmogorov distance. Proposition 2.4 shows that d_{BL} also upper-bounds d_K while enjoying all of the resulting smoothness of Wasserstein test functions and with additional boundedness properties to boot. Moreover, the Wasserstein distance is not always well-defined (e.g. if one of the distributions does not have a first moment), whereas d_{BL} always exists. Finally, d_{BL} is a fairly natural measure of distance since it metrizes weak convergence [19]. However, we always have $d_W(\mathscr{L}(X), \mathscr{L}(Z)) \geq d_{BL}(\mathscr{L}(X), \mathscr{L}(Z))$, and it is possible for a sequence to converge in d_{BL} but not in d_W or d_K. Furthermore, the bounded Lipschitz metric does not scale as nicely as the Kolmogorov or Wasserstein distances when the associated random variables are multiplied by a positive constant. For the remainder of this paper, we will state our results in terms of d_{BL} with the corresponding Kolmogorov bound being implicit therein, though one should note that, as with the Wasserstein bound on d_K, Kolmogorov bounds obtained in this fashion are not necessarily optimal, often giving the root of the true rate.

3. The Centered Equilibrium Transformation

Our next task is to use the characterization in Theorem 1.1 to obtain bounds on the error terms resulting from approximation by the Laplace distribution. To this end, we introduce the following definition.

Definition 3.1. For any nondegenerate random variable X with mean zero and finite variance, we say that the random variable X^L has the *centered equilibrium distribution* with respect to X if

$$\mathbb{E}[f(X)] - f(0) = \frac{1}{2}\mathbb{E}[X^2]\mathbb{E}[f''(X^L)]$$

for all twice-differentiable functions f such that f, f', and f'' are bounded. We call the map $X \mapsto X^L$ the *centered equilibrium transformation.*

Note that the centered equilibrium distribution is uniquely defined because $\mathbb{E}[f''(X)] = \mathbb{E}[f''(Y)]$ for all twice continuously differentiable functions with compact support implies $X =_d Y$. The nomenclature is in reference to the equilibrium distribution from renewal theory, which was used in a similar manner in [12] for a related problem involving the exponential distribution.

Since the characterizing equation for the Laplace distribution involves second derivatives and a variance term, one expects some kind of relation between X^L and the zero bias distribution for X (defined by $\mathbb{E}[Xf(X)] = \mathbb{E}[X^2]\mathbb{E}[f'(X^z)]$ for all absolutely continuous f for which $\mathbb{E}\,|Xf(X)| < \infty$ [7]) in much the same way as the equilibrium distribution is related to the size bias distribution [12]. The following theorem shows that this is indeed the case. Moreover, since the zero bias distribution is defined for any X with $\mathbb{E}[X] = 0$ and $\mathrm{Var}(X) \in (0, \infty)$, it will follow that every such random variable has a centered equilibrium distribution.

Theorem 3.2. *Suppose that X has mean zero and variance $\sigma^2 \in (0, \infty)$. Let X^z have the zero bias distribution with respect to X and let $B \sim \mathrm{Beta}(2, 1)$ be independent of X^z. Then $X^L := BX^z$ satisfies*

$$\mathbb{E}[f(X)] - f(0) = \frac{\sigma^2}{2}\mathbb{E}[f''(X^L)]$$

for all twice-differentiable f with $\|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty < \infty$.

Proof. Applying the fundamental theorem of calculus, Fubini's theorem, the definition of X^z, and the fact that B has density $p(x) = 2x1_{[0,1]}(x)$ gives

$$\begin{aligned}\mathbb{E}[f(X)] - f(0) &= \mathbb{E}\left[\int_0^1 Xf'(uX)du\right] = \int_0^1 \mathbb{E}\left[Xf'(uX)\right]du \\ &= \sigma^2 \int_0^1 u\mathbb{E}[f''(uX^z)]du = \sigma^2\mathbb{E}\left[\int_0^1 uf''(uX^z)du\right] \\ &= \frac{\sigma^2}{2}\mathbb{E}\left[f''(BX^z\right]\end{aligned}$$

The assumptions ensure that all of the functions are integrable. □

Corollary 3.3. *If X has variance $\mathbb{E}[X^2] = 2b^2$ and X^L has the centered equilibrium distribution with respect to X, then X^L is absolutely continuous with density*

$$f_{X^L}(x) = \frac{1}{b^2}\int_0^1 \mathbb{E}\left[X; X > \frac{x}{v}\right]dv.$$

Proof. $X^L =_d BX^z$ with B and X^z as in Theorem 3.2, so the claim follows from the fact [7] that X^z is absolutely continuous with density $f_{X^z}(x) = \frac{1}{2b^2}\mathbb{E}[X; X > x]$ by the usual method of computing the density of a product. □

Remark. In an earlier version of this paper, we established the existence of the centered equilibrium distribution by showing that for certain random variables X, X^L can be obtained by iterating the $X-P$ bias transformation from [8] with $P(x)=\operatorname{sgn}(x)$. Though there may be some merit to such a strategy and it provides another example of how results for higher order Stein operators may be obtained by iterating more traditional techniques, in our case it required the rather artificial assumption that the variates in the domain of the transformation have median zero. Those interested in the iterated $X-P$ bias approach are referred to the article [5] by Christian Döbler, which contains the essential technical details of our original argument and suggests certain improvements.

Lemma 2.3 shows that, up to scaling, the mean zero Laplace distribution is the unique fixed point of the centered equilibrium transformation. Thus one expects that if a random variable is close to its centered equilibrium transform, then its distribution is close to the Laplace law. Theorem 1.2 formalizes this intuition.

Proof of Theorem 1.2. If X is a random variable with $\mathbb{E}[X^2]=2b^2<\infty$ and X^L has the centered equilibrium distribution for X, then for all $h\in\mathcal{H}_{BL}$, taking g_h as in Lemma 2.2, we see that

$$\begin{aligned}|Wh-\mathbb{E}[h(X)]| &= \left|\mathbb{E}[g_h(X)-b^2g_h''(X)]\right| = \left|\mathbb{E}[b^2g_h''(X^L)-b^2g_h''(X)]\right| \\ &\le b^2\mathbb{E}\left|g_h''(X^L)-g_h''(X)\right| \le b^2\left\|g_h'''\right\|\mathbb{E}\left|X^L-X\right| \\ &= \frac{b+2}{b}\mathbb{E}\left|X-X^L\right|.\end{aligned}$$

□

For the example in Section 4, we will also need the following complementary result.

Proposition 3.4. *If Y^L has the centered equilibrium distribution for Y and $\mathbb{E}[Y^2]=2b^2$, then*

$$\mathbb{E}\left|Y-Y^L\right| \le \mathbb{E}\left|Y\right| + \frac{1}{6b^2}\mathbb{E}[|Y|^3].$$

Proof. We may assume that $\mathbb{E}[|Y|^3]<\infty$ as the inequality is trivial otherwise. The result will follow immediately from the triangle inequality if we can show that $\mathbb{E}\left|Y^L\right| = \frac{1}{6b^2}\mathbb{E}\left|Y^3\right|$, which is what one would obtain by formally plugging $f(y)=|y|^3$ into the definition of the transformation $Y\mapsto Y^L$.

Of course, neither f, f', nor f'' is bounded, so we must proceed by approximation. To this end, define

$$f_n(x)=\begin{cases}|x|^3, & |x|\le n\\ n^3+3n^2(|x|-n)-\frac{3}{2}(|x|-n)^2, & n<|x|\le n^2+n\\ n^3+3n^2(2n^2+n-|x|)-\frac{3}{2}(2n^2+n-|x|)^2, & n^2+n<|x|\le 2n^2+n\\ (2n^2+2n-|x|)^3, & 2n^2+n<|x|\le 2n^2+2n\\ 0, & |x|>2n^2+2n\end{cases}$$

By construction, f_n is smooth with compact support and satisfies $\left|f_n^{(i)}\right|\le\left|f^{(i)}\right|$ for all $n\in\mathbb{N}$, $i=0,1,2$, thus fulfilling the conditions in the definition of the centered

equilibrium transformation when $\mathbb{E}[|Y|^3] < \infty$. Moreover, $f_n^{(i)} \to f^{(i)}$ pointwise for $i = 0, 1, 2$, so it follows from the dominated convergence theorem that

$$\mathbb{E}[f(Y)] = \lim_{n\to\infty} \mathbb{E}[f_n(Y)] = \lim_{n\to\infty} b^2\mathbb{E}[f_n''(Y^L)].$$

Fatou's lemma shows that $f''(Y^L)$ is integrable since

$$\mathbb{E}\left[f''(Y^L)\right] = \mathbb{E}\left[\liminf_{n\to\infty} f_n''(Y^L)\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[f_n''(Y^L)\right] = \frac{1}{b^2}\mathbb{E}[f(Y)] < \infty,$$

so another application of dominated convergence gives

$$\mathbb{E}[f(Y)] = \lim_{n\to\infty} b^2\mathbb{E}[f_n''(Y^L)] = b^2\mathbb{E}[f''(Y^L)].$$

□

The same general argument shows that if Y^L has the centered equilibrium distribution for Y and $\mathbb{E}\left[|Y|^n\right] < \infty$, then $\mathbb{E}[q(Y)] - q(0) = \frac{1}{2}\mathbb{E}[Y^2]\mathbb{E}[q''(Y^L)]$ for any polynomial q of degree at most n. As is often the case with distributional transformations defined in terms of functional equations, we find it more convenient to define $X \mapsto X^L$ in terms of a relatively small class of test functions and then argue by approximation when we want to apply the relation more generally.

4. Random Sums

The p-geometric summation of a sequence of random variables $X_1, X_2, ...$ is defined as $S_p = X_1 + X_2 + ... + X_{N_p}$ where N_p is geometric with success probability p - that is, $\mathbb{P}\{N_p = n\} = p(1-p)^{n-1}$, $n \in \mathbb{N}$ - and is independent of all else. A result due to Rényi [14] states that if $X_1, X_2, ...$ are i.i.d., positive, nondegenerate random variables with $\mathbb{E}[X_i] = 1$, then $\mathscr{L}(pS_p) \to \text{Exponential}(1)$ as $p \to 0$. In fact, just as the normal law is the only nondegenerate distribution with finite variance that is stable under ordinary summation (in the sense that if $X, X_1, X_2, ...$ are i.i.d. nondegenerate random variables with finite variance, then for every $n \in \mathbb{N}$, there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that $X =_d a_n\left(X_1 + ... + X_n\right) + b_n$), it can be shown that if $X, X_1, X_2, ...$ are i.i.d., positive, and nondegenerate with finite variance, then there exists $a_p > 0$ such that $a_p\left(X_1 + ... + X_{N_p}\right) =_d X$ for all $p \in (0,1)$ if and only if X has an exponential distribution. Similarly, if we assume that $Y, Y_1, Y_2, ...$ are i.i.d., symmetric, and nondegenerate with finite variance, then there exists $a_p > 0$ such that $a_p\left(Y_1 + ... + Y_{N_p}\right) =_d Y$ for all $p \in (0,1)$ if and only if Y has a Laplace distribution. Moreover, it must be the case that $a_p = p^{\frac{1}{2}}$. In addition, we have an analog of Rényi's theorem [10]:

Theorem 4.1. *Suppose that $X_1, X_2, ...$ are i.i.d., symmetric, and nondegenerate random variables with finite variance σ^2, and let $N_p \sim$ Geometric(p) be independent of the $X_i's$. If*

$$a_p \sum_{i=1}^{N_p} X_i \to_d X \text{ as } p \to 0,$$

then there exists $\gamma > 0$ such that $a_p = p^{\frac{1}{2}}\gamma + o(p^{\frac{1}{2}})$ and X has the Laplace distribution with mean 0 and variance $\sigma^2\gamma^2$.

A recent theorem due to Alexis Toda [18] gives the following Lindeberg-type conditions for the existence of the distributional limit in Theorem 4.1.

Theorem 4.2 (Toda)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent (but not necessarily identically distributed) random variables such that $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = \sigma_i^2$, and let $N_p \sim \mathrm{Geometric}(p)$ independent of the $X_i's$. Suppose that*

$$\lim_{n\to\infty} n^{-\alpha}\sigma_n^2 = 0 \textit{ for some } 0 < \alpha < 1,$$

$$\sigma^2 := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 > 0 \textit{ exists},$$

and for all $\epsilon > 0$,

$$\lim_{p\to 0} \sum_{i=1}^{\infty} (1-p)^{i-1} p\mathbb{E}[X_i^2; |X_i| \geq \epsilon p^{-\frac{1}{2}}] = 0.$$

Then as $p \to 0$, the sum $p^{\frac{1}{2}} \sum_{i=1}^{N_p} X_i$ converges weakly to the Laplace distribution with mean 0 and variance σ^2.

Remark. The original statement of Toda's theorem is slightly more general, allowing for convergence to a possibly asymmetric Laplace distribution.

In 2011, Peköz and Röllin were able to generalize Rényi's theorem by using a distributional transformation inspired by Stein's method considerations [12]. Specifically, for a nonnegative random variable X with $\mathbb{E}[X] < \infty$, they say that X^e has the equilibrium distribution with respect to X if $\mathbb{E}[f(X)] - f(0) = \mathbb{E}[X]\mathbb{E}[f'(X^e)]$ for all Lipschitz f and use this to bound the Wasserstein and Kolmogorov distances to the Exponential(1) distribution. The equilibrium distribution arises in renewal theory, but its utility in analyzing convergence to the exponential distribution comes from the fact that a Stein operator for the exponential distribution with mean one is given by $(\mathcal{A}_E f)(x) = f'(x) - f(x) + f(0)$, so $X \sim \mathrm{Exponential}(1)$ is the unique fixed point of the equilibrium transformation [15]. This transformation and the similarity between our characterization of the Laplace and the above characterization of the exponential inspired our construction of the centered equilibrium transformation, and the fact that both distributions are stable under geometric summation led us to parallel their argument for bounding the distance between p-geometric sums of positive random variables and the exponential distribution in order to obtain corresponding results for the Laplace case. Our results are summarized in the following theorem.

Theorem 4.3. *Let N be any $\mathbb{N}$-valued random variable with $\mu = \mathbb{E}[N] < \infty$ and let $X_1, X_2, \ldots$ be a sequence of independent random variables, independent of N, with $\mathbb{E}[X_i] = 0$, and $\mathbb{E}[X_i^2] = \sigma_i^2 \in (0, \infty)$. Set $\sigma^2 = \mathbb{E}\left[\left(\sum_{i=1}^{N} X_i\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{N} \sigma_i^2\right]$ and let M be any $\mathbb{N}$-valued random variable, independent of the $X_i's$ and defined on the same space as N, satisfying*

$$\mathbb{P}\{M = m\} = \frac{\sigma_m^2}{\sigma^2}\mathbb{P}\{N \geq m\}.$$

Then

$$d_{BL}\left(\mathscr{L}\left(\mu^{-\frac{1}{2}}\sum_{i=1}^{N}X_i\right), \text{Laplace}\left(0, \frac{\sigma}{\sqrt{2\mu}}\right)\right)$$
$$\leq \left(\mu^{-\frac{1}{2}} + \frac{\sqrt{8}}{\sigma}\right)\left(\mathbb{E}\left|X_M - X_M^L\right| + \sup_{i\geq 1}\sigma_i \mathbb{E}\left[|N-M|^{\frac{1}{2}}\right]\right).$$

Proof. We first note that

$$\sigma^2 = \sum_{m=1}^{\infty}\mathbb{P}\{N=m\}\sum_{i=1}^{m}\sigma_i^2 = \sum_{m=1}^{\infty}\mathbb{P}\{N\geq m\}\sigma_m^2,$$

so M is well-defined.

Now, taking $V = \mu^{-\frac{1}{2}}\sum_{i=1}^{N}X_i$, we claim that $V^L = \mu^{-\frac{1}{2}}\left(\sum_{i=1}^{M-1}X_i + X_M^L\right)$ has the centered equilibrium distribution with respect to V. (Throughout, X_m^L is taken to be independent of M, N, and X_k for $k \neq m$.) To see that this is so, let f be any function satisfying the assumptions in Definition 3.1. Then, using the notation

$$\mathbb{X} = \{X_i\}_{i\geq 1},\ g_m(\mathbb{X}) = f\left(\mu^{-\frac{1}{2}}\sum_{i=1}^{m}X_i\right),\ f_s(x) = f\left(\mu^{-\frac{1}{2}}s + \mu^{-\frac{1}{2}}x\right),$$

letting ν_m denote the distribution of

$$S_{m-1} = \sum_{i=1}^{m-1}X_i,$$

and observing that, by independence,

$$\mathbb{E}[h''(X_m^L)|S_{m-1}=s] = \mathbb{E}[h''(X_m^L)] = \frac{2}{\sigma_m^2}\left(\mathbb{E}[h(X_m)-h(0)]\right)$$
$$= \frac{2}{\sigma_m^2}\mathbb{E}[h(X_m)-h(0)|S_{m-1}=s]$$

for all $s\in\mathbb{R}$ and all twice differentiable h with h, h', h'' bounded, we see that

$$\mathbb{E}\left[f''\left(\mu^{-\frac{1}{2}}\sum_{i=1}^{m-1}X_i + \mu^{-\frac{1}{2}}X_m^L\right)\right] = \int \mathbb{E}\left[f''(\mu^{-\frac{1}{2}}s + \mu^{-\frac{1}{2}}X_m^L)\,|S_{m-1}=s\right]d\nu_m(s)$$
$$= \int \mathbb{E}\left[\mu f_s''(X_m^L)\,|S_{m-1}=s\right]d\nu_m(s)$$
$$= \frac{2\mu}{\sigma_m^2}\int \mathbb{E}\left[f_s(X_m) - f_s(0)\,|S_{m-1}=s\right]d\nu_m(s)$$
$$= \frac{2\mu}{\sigma_m^2}\int \mathbb{E}\Big[f(\mu^{-\frac{1}{2}}s + \mu^{-\frac{1}{2}}X_m)$$
$$-f(\mu^{-\frac{1}{2}}s)\,|S_{m-1}=s\Big]d\nu_m(s)$$
$$= \frac{2\mu}{\sigma_m^2}\mathbb{E}\left[f\left(\mu^{-\frac{1}{2}}\sum_{i=1}^{m}X_i\right) - f\left(\mu^{-\frac{1}{2}}\sum_{i=1}^{m-1}X_i\right)\right]$$
$$= \frac{2\mu}{\sigma_m^2}\mathbb{E}\left[g_m(\mathbb{X}) - g_{m-1}(\mathbb{X})\right]$$

for all $m \in \mathbb{N}$, hence

$$\begin{aligned}
\mathbb{E}[f''(V^L)] &= \sum_{m=1}^{\infty} \mathbb{P}\{M=m\}\mathbb{E}\left[f''\left(\mu^{-\frac{1}{2}}\sum_{i=1}^{m-1} X_i + \mu^{-\frac{1}{2}} X_m^L\right)\right] \\
&= \frac{2\mu}{\sigma^2}\sum_{m=1}^{\infty}\frac{\sigma^2}{\sigma_m^2}\mathbb{P}\{M=m\}\mathbb{E}\left[g_m(\mathbb{X}) - g_{m-1}(\mathbb{X})\right] \\
&= \frac{2}{\mathbb{E}[V^2]}\mathbb{E}\left[\sum_{m=1}^{\infty}\mathbb{P}\{N \geq m\}\left(g_m(\mathbb{X}) - g_{m-1}(\mathbb{X})\right)\right] \\
&= \frac{2}{\mathbb{E}[V^2]}\left(\mathbb{E}[g_N(\mathbb{X})] - g_0(\mathbb{X})\right) = \frac{2}{\mathbb{E}[V^2]}\left(\mathbb{E}[f(V)] - f(0)\right).
\end{aligned}$$

Having shown that V^L does in fact have the centered equilibrium distribution for V, we can apply Theorem 1.2 with $2B^2 = \mathbb{E}[V^2] = \frac{\sigma^2}{\mu}$ to obtain

$$\begin{aligned}
d_{BL}(\mathscr{L}(V), \mathrm{Laplace}(0,B)) &\leq \left(1 + \frac{2}{B}\right)\mathbb{E}\left|V - V^L\right| \\
&= \left(\mu^{-\frac{1}{2}} + \frac{\sqrt{8}}{\sigma}\right)\mathbb{E}\left|\left(X_M - X_M^L\right) + \mathrm{sgn}(N-M)\sum_{i=(N\wedge M)+1}^{N\vee M} X_i\right| \\
&\leq \left(\mu^{-\frac{1}{2}} + \frac{\sqrt{8}}{\sigma}\right)\left(\mathbb{E}\left|X_M - X_M^L\right| + \mathbb{E}\left|\sum_{i=(N\wedge M)+1}^{N\vee M} X_i\right|\right).
\end{aligned}$$

Finally, since the $X_i's$ are independent with mean zero, the Cauchy-Schwarz inequality gives

$$\begin{aligned}
\mathbb{E}\left|\sum_{i=(N\wedge M)+1}^{N\vee M} X_i\right| &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbb{P}\{N\wedge M = j, |N-M| = k\}\mathbb{E}\left|\sum_{i=j+1}^{j+k} X_i\right| \\
&\leq \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbb{P}\{N\wedge M = j, |N-M| = k\}\mathbb{E}\left[\sum_{i=j+1}^{j+k} X_i^2\right]^{\frac{1}{2}} \\
&\leq \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbb{P}\{N\wedge M = j, |N-M| = k\}k^{\frac{1}{2}}\sup_{i\geq 1}\sigma_i \\
&= \sup_{i\geq 1}\sigma_i\sum_{k=1}^{\infty}k^{\frac{1}{2}}\mathbb{P}\{|N-M| = k\} = \sup_{i\geq 1}\sigma_i\mathbb{E}\left[|N-M|^{\frac{1}{2}}\right],
\end{aligned}$$

and the proof is complete. □

Remark. When the summands in Theorem 4.3 have common variance σ_1^2, we have $\sigma^2 = \mathbb{E}\left[\sum_{i=1}^N \sigma_1^2\right] = \mu\sigma_1^2$, so M is defined by

$$\mathbb{P}\{M=m\} = \frac{\sigma_m^2}{\sigma^2}\mathbb{P}\{N \geq m\} = \frac{1}{\mu}\mathbb{P}\{N \geq m\}.$$

In other words, M is the discrete equilibrium transformation of N defined in [13] for use in geometric approximation. Thus the $\mathbb{E}[|N-M|^{\frac{1}{2}}]$ term in the bound

from Theorem 4.3 may be regarded as measuring how close $\mathscr{L}(N)$ is to a geometric distribution. If the summands are i.i.d., then $X_M =_d X_1$ (as M is assumed to be independent of all else), so Theorem 1.2 shows that the $\mathbb{E}\left|X_M - X_M^L\right|$ term measures how close the distribution of the summands is to the Laplace. To put this into context, recall that if the $X_i's$ are i.i.d. Laplace$(0,b)$ and $N \sim$ Geometric(p), then $p^{\frac{1}{2}}\sum_{i=1}^N X_i \sim$ Laplace$(0,b)$.

We conclude our discussion with a proof of Theorem 1.3, which gives sufficient conditions for weak convergence in the setting of Theorem 4.1. Though it requires that the $X_i's$ have uniformly bounded third absolute moments, the condition of symmetry is dropped and the identical distribution assumption is reduced to the requirement that the $X_i's$ have common variance. This result is not quite as general as Theorem 4.2, but it does provide bounds on the error terms.

Proof of Theorem 1.3. We are trying to bound the distance between the Laplace$(0,b)$ distribution and that of $p^{\frac{1}{2}}\sum_{i=1}^N X_i$ where $X_1, X_2, ...$ are independent mean zero random variables with common variance $\mathbb{E}[X_i^2] = 2b^2$ and uniformly bounded third absolute moments $\sup_{i\geq 1}\mathbb{E}\left[|X_i|^3\right] = \rho < \infty$, and $N \sim$ Geometric(p) is independent of the $X_i's$.

In the language of Theorem 4.3, we have $\mu = \frac{1}{p}$, $\sigma = b\sqrt{2\mu}$, and

$$\frac{\sigma_m^2}{\sigma^2}\mathbb{P}\{N \geq m\} = p\mathbb{P}\{N \geq m\} = p^2\sum_{i=m}^{\infty}(1-p)^{m-1} = p(1-p)^{m-1} = \mathbb{P}\{N = m\},$$

so we can take $M = N$ to obtain

$$d_{BL}\left(\mathscr{L}\left(p^{\frac{1}{2}}\sum_{i=1}^N X_i\right), \text{Laplace}(0,b)\right) \leq \left(p^{\frac{1}{2}} + \frac{2p^{\frac{1}{2}}}{b}\right)\mathbb{E}\left|X_N - X_N^L\right|.$$

Applying Proposition 3.4 gives

$$\begin{aligned}\mathbb{E}\left|X_N - X_N^L\right| &= \sum_{n=1}^{\infty}\mathbb{P}\{N = n\}\mathbb{E}\left|X_n - X_n^L\right| \\ &\leq \sum_{n=1}^{\infty}\mathbb{P}\{N = n\}\left(\mathbb{E}\left|X_n\right| + \frac{1}{6b^2}\mathbb{E}\left[|X_n|^3\right]\right) \\ &\leq \sum_{n=1}^{\infty}\mathbb{P}\{N = n\}\left(\mathbb{E}\left[X_n^2\right]^{\frac{1}{2}} + \frac{\rho}{6b^2}\right) = b\sqrt{2} + \frac{\rho}{6b^2},\end{aligned}$$

and the result follows □

Acknowledgements

The authors would like to thank Larry Goldstein for suggesting the use of Stein's method to get convergence rates in the general setting of Theorem 4.2 and for many helpful comments throughout the preparation of this paper. Thanks also to Alex Rozinov for several illuminating conversations concerning the material in Section 4 and to the anonymous referees whose careful notes were of great help to us.

References

[1] Timothy C. Brown and M. J. Phillips. Negative binomial approximation with Stein's method. *Methodol. Comput. Appl. Probab.*, 1(4):407–421, 1999. MR1770372.

[2] Louis H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probability*, 3(3):534–545, 1975. MR0428387.

[3] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg, 2011. MR2732624.

[4] Persi Diaconis and Susan Holmes, editors. *Stein's method: expository lectures and applications. Papers from the Workshop on Stein's Method held at Stanford University, Stanford, CA, 1998*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 46. Institute of Mathematical Statistics, Beachwood, OH, 2004. MR2118599.

[5] Christian Döbler. Distributional transformations without orthogonality relations. *ArXiv e-prints*, 2013. arXiv:1312.6093.

[6] Werner Ehm. Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.*, 11(1):7–16, 1991. MR1093412.

[7] Larry Goldstein and Gesine Reinert. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, 7(4):935–952, 1997. MR1484792.

[8] Larry Goldstein and Gesine Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.*, 18(1):237–260, 2005. MR2132278.

[9] Vladimir Kalashnikov. *Geometric sums: bounds for rare events with applications. Risk analysis, reliability, queueing*, volume 413 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1997. MR1471479.

[10] Samuel Kotz, Tomasz J. Kozubowski, and Krzysztof Podgórski. *The Laplace distribution and generalizations. A revisit with applications to communications, economics, engineering, and finance*. Birkhäuser Boston, Inc., Boston, MA, 2001. MR1935481.

[11] Ho Ming Luk. *Stein's method for the Gamma distribution and related statistical applications. Thesis (Ph.D.)–University of Southern California*. ProQuest LLC, Ann Arbor, MI, 1994. MR2693204.

[12] Erol A. Peköz and Adrian Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*, 39(2):587–608, 2011. MR2789507.

[13] Erol A. Peköz, Adrian Röllin, and Nathan Ross. Total variation error bounds for geometric approximation. *Bernoulli*, 19(2):610–632, 2013. MR3037166.

[14] Alfréd Rényi. A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 1:519–527 (1957), 1957. MR0094861.

[15] Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011. MR2861132.

[16] Charles Stein. 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 (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, Calif., 1972. MR0402873.

[17] Charles Stein. *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986. MR882007.

[18] Alexis A. Toda. Weak limit of the geometric sum of independent but not identically distributed random variables. *ArXiv e-prints*, 2011. arXiv:1111.1786.

[19] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes. With applications to statistics*. Springer Series in Statistics. Springer-Verlag, New York, 1996. MR1385671.