The Annals of Probability
2011, Vol. 39, No. 2, 587–608
DOI: 10.1214/10-AOP559

NEW RATES FOR EXPONENTIAL APPROXIMATION AND THE THEOREMS OF RÉNYI AND YAGLOM[1]

BY EROL A. PEKÖZ AND ADRIAN RÖLLIN[2]

Boston University and National University of Singapore

We introduce two abstract theorems that reduce a variety of complex exponential distributional approximation problems to the construction of couplings. These are applied to obtain new rates of convergence with respect to the Wasserstein and Kolmogorov metrics for the theorem of Rényi on random sums and generalizations of it, hitting times for Markov chains, and to obtain a new rate for the classical theorem of Yaglom on the exponential asymptotic behavior of a critical Galton–Watson process conditioned on nonextinction. The primary tools are an adaptation of Stein's method, Stein couplings, as well as the equilibrium distributional transformation from renewal theory.

1. Introduction.

The exponential distribution arises as an asymptotic limit in a wide variety of settings involving rare events, extremes, waiting times, and quasi-stationary distributions. As discussed in the preface of Aldous (1989), the tremendous difficulty in obtaining explicit bounds on the error of the exponential approximation in more than the most elementary of settings apparently has left a gap in the literature. The classical theorem of Yaglom (1947) describing the asymptotic exponential behavior of a critical Galton–Watson process conditioned on nonextinction, for example, has a large literature of extensions and embellishments [see Lalley and Zheng (2011), e.g.] but the complex dependencies between offspring have apparently not previously allowed for obtaining explicit error bounds. Stein's method, introduced in Stein (1972), is now a well-established method for obtaining explicit bounds in distributional approximation problems in settings with dependence [see Ross and Peköz (2007) for an introduction]. Results for the normal and Poisson approximation, in particular, are extensive but also are currently very actively being further developed; see, for example, Chatterjee (2008) and Chen and Röllin (2009).

There have been a few attempts to apply Stein's method to exponential approximation. Weinberg (2005) sketches a few potential applications but only tackles simple examples thoroughly, and Bon (2006) only considers geometric convolutions. Chatterjee, Fulman and Röllin (2006) breaks new ground by applying the

Received March 2010.

[1]Supported by the Institute for Mathematical Sciences at the National University of Singapore.

[2]Supported in part by NUS Grant R-155-000-098-133.

AMS 2000 subject classifications. Primary 60F05; secondary 60J10, 60J80.

Key words and phrases. Exponential approximation, geometric convolution, first passage times, critical Galton–Watson branching process, Stein's method, equilibrium and size-biased distribution.

method to a challenging problem in spectral graph theory using exchangeable pairs, but the calculations involved are application-specific and far from elementary. In this article, in contrast, we develop a general framework that more conveniently reduces a broad variety of complex exponential distributional approximation problems to the construction of couplings. We provide evidence that our approach can be fruitfully applied to nontrivial applications and in settings with dependence—settings where Stein's method typically is expected to shine.

The article is organized as follows. In Section 2, we present two abstract theorems formulated in terms of couplings. We introduce a distributional transformation (the "equilibrium distribution" from renewal theory) which has not yet been extensively explored using Stein's method. We also make use of Stein couplings similar to those introduced in Chen and Röllin (2009). In Section 3, we give applications using these couplings to obtain exponential approximation rates for the theorem of Rényi on random sums and hitting times for Markov chains; our approach yields generalizations of these results not previously available in the literature. Furthermore, we consider the rate of convergence in the classical theorem of Yaglom on the exponential asymptotic behavior of a critical Galton–Watson process conditioned on nonextinction; this is the first place this latter result has appeared in the literature. In Section 4, we then give the postponed proofs for the main theorems.

2. Main results. In this section, we present the framework in abstract form that will subsequently be used in concrete applications in Section 3. This framework is comprised of two approaches that we will describe here and then prove in Section 4.

To define the probability metrics used in this article, we need the sets of test functions

$$\begin{aligned}\mathcal{F}_{\mathrm{K}} &= \{\mathrm{I}[\cdot \le z] | z \in \mathbb{R}\},\\ \mathcal{F}_{\mathrm{W}} &= \{h : \mathbb{R} \to \mathbb{R} | h \text{ is Lipschitz}, \|h'\| \le 1\},\\ \mathcal{F}_{\mathrm{BW}} &= \{h : \mathbb{R} \to \mathbb{R} | h \text{ is Lipschitz}, \|h\| \le 1 \text{ and } \|h'\| \le 1\}\end{aligned}$$

and then the distance between two probability measures P and Q with respect to $\mathcal{F}$ is defined as

$$d_{\mathcal{F}}(P, Q) := \sup_{f \in \mathcal{F}} \left| \int_{\mathbb{R}} f \, dP - \int_{\mathbb{R}} f \, dQ \right| \tag{2.1}$$

if the corresponding integrals are well-defined. Denote by d_{K}, d_{W} and d_{BW} the respective distances corresponding to the sets $\mathcal{F}_{\mathrm{K}}$, $\mathcal{F}_{\mathrm{W}}$ and $\mathcal{F}_{\mathrm{BW}}$. The subscripts respectively denote the Kolmogorov, Wasserstein and bounded Wasserstein distances. We can use the following two relations:

$$d_{\mathrm{BW}} \le d_{\mathrm{W}}, \qquad d_{\mathrm{K}}(P, \mathrm{Exp}(1)) \le 1.74\sqrt{d_{\mathrm{W}}(P, \mathrm{Exp}(1))}. \tag{2.2}$$

The first relation is clear, as $\mathcal{F}_{\mathrm{BW}} \subset \mathcal{F}_{\mathrm{W}}$, and we refer to Gibbs and Su (2002) for the second relation. It is worthwhile noting that the second inequality can yield optimal bounds with respect to the d_{K} metric. This is in contrast to normal approximation where in fact d_{W} and d_{K} often exhibit the same order of convergence and hence the corresponding equivalent of (2.2) for the normal distribution does not yield optimal bounds on d_{K}; cf. Corollary 3.4.

Our first approach involves a coupling with the *equilibrium distribution* from renewal theory, and is related to the zero-bias coupling from Goldstein and Reinert (1997) used for normal approximation [see also Bon (2006), Lemma 6, Goldstein (2005, 2007) and Ghosh (2009)].

DEFINITION 2.1. Let X be a nonnegative random variable with finite mean. We say that a random variable X^e has the *equilibrium distribution w.r.t. X* if for all Lipschitz f

$$\mathbb{E}f(X) - f(0) = \mathbb{E}X\mathbb{E}f'(X^e). \tag{2.3}$$

It is straightforward that this implies

$$\mathbb{P}(X^e \le x) = \frac{1}{\mathbb{E}X}\int_0^x \mathbb{P}[X > y]\,dy \tag{2.4}$$

and our first result below can be thought of as formalizing the notion that when $\mathscr{L}(W)$ and $\mathscr{L}(W^e)$ are approximately equal then W has approximately an exponential distribution.

THEOREM 2.1. *Let W be a nonnegative random variable with $\mathbb{E}W = 1$ and let W^e have the equilibrium distribution w.r.t. W. Then, for any $\beta > 0$,*

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le 12\beta + 2\mathbb{P}[|W^e - W| > \beta] \tag{2.5}$$

and

$$d_{\mathrm{K}}(\mathscr{L}(W^e), \mathrm{Exp}(1)) \le \beta + \mathbb{P}[|W^e - W| > \beta]. \tag{2.6}$$

If in addition W has finite second moment, then

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le 2\mathbb{E}|W^e - W| \tag{2.7}$$

and

$$d_{\mathrm{K}}(\mathscr{L}(W^e), \mathrm{Exp}(1)) \le \mathbb{E}|W^e - W|; \tag{2.8}$$

bound (2.8) *also holds for $d_{\mathrm{W}}(\mathscr{L}(W^e), \mathrm{Exp}(1))$.*

Our second approach involves an adaptation of the *linear Stein couplings* introduced in Chen and Röllin (2009).

DEFINITION 2.2. A triple (W, W', G) of random variables is called a *constant Stein coupling* if

$$\mathbb{E}\{Gf(W') - Gf(W)\} = \mathbb{E}f(W) \tag{2.9}$$

for all f with $f(0) = 0$ and for which the expectations exist.

Let

$$r_1(\mathcal{F}) = \sup_{\substack{f\in\mathcal{F},\\ f(0)=0}} |\mathbb{E}\{Gf(W') - Gf(W) - f(W)\}|$$

and $r_2 = \mathbb{E}|\mathbb{E}^{W''}(GD) - 1|$, where here and in the rest of the article $D := W' - W$. The random variable W'' is defined on the same probability space as (W, W', G) and can be used to simplify the bounds (it is typically chosen so that $r_2 = 0$); let $D' := W'' - W$. At first reading one may simply set $W'' = W$ (in which case typically $r_2 \neq 0$); we refer to Chen and Röllin (2009) for a more detailed discussion of Stein couplings. Our next result applies to general random variables, but useful bounds can only be expected if they are coupled together so that $r_1(\mathcal{F}_{\mathrm{W}})$ is small.

THEOREM 2.2. *Let W, W', W'' and G be random variables with finite first moments such that also $\mathbb{E}|GD| < \infty$ and $\mathbb{E}|GD'| < \infty$. Then with the above definitions*,

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le r_1(\mathcal{F}_{\mathrm{W}}) + r_2 + 2r_3 + 2r_3' + 2r_4 + 2r_4', \tag{2.10}$$

where

$$r_3 = \mathbb{E}\big|GD\mathrm{I}[|D| > 1]\big|, \qquad r_3' = \mathbb{E}\big|(GD-1)\mathrm{I}[|D'| > 1]\big|,$$
$$r_4 = \mathbb{E}|G(D^2 \wedge 1)|, \qquad r_4' = \mathbb{E}\big|(GD-1)(|D'| \wedge 1)\big|.$$

The same bound holds for d_{BW} with $r_1(\mathcal{F}_{\mathrm{W}})$ replaced by $r_1(\mathcal{F}_{\mathrm{BW}})$. Furthermore, for any α, β and β',

$$\begin{aligned} &d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \\ &\quad \le 2r_1(\mathcal{F}_{\mathrm{BW}}) + 2r_2 + 2r_5 + 2r_5' + 22(\alpha\beta + 1)\beta' + 12\alpha\beta^2, \end{aligned} \tag{2.11}$$

where

$$r_5 = \mathbb{E}\big|GD\mathrm{I}[|G| > \alpha \text{ or } |D| > \beta]\big|,$$
$$r_5' = \mathbb{E}\big|(1 - GD)\mathrm{I}[|G| > \alpha \text{ or } |D| > \beta \text{ or } |D'| > \beta']\big|.$$

2.1. *Couplings*. In this section, we present a way to construct the equilibrium distribution more explicitly and also discuss a few constant Stein couplings.

2.1.1. *Equilibrium distribution via size biasing.* Assume that $\mathbb{E}W = 1$ and let W^s have the size bias distribution of W, that is,

$$\mathbb{E}\{Wf(W)\} = \mathbb{E}f(W^s)$$

for all f for which the expectation exist. Then, if U has the uniform distribution on [0, 1] independent of all else, $W^e := UW^s$ has the equilibrium distribution w.r.t. W. Indeed, for any Lipschitz f with $f(0) = 0$ we have

$$\mathbb{E}f(W) = \mathbb{E}f(W) - f(0) = \mathbb{E}\{Wf'(UW)\} = \mathbb{E}f'(UW^s) = \mathbb{E}f'(W^e).$$

We note that this construction was also considered by Goldstein (2009) and it has been observed by Pakes and Khattree (1992) that for a nonnegative random variable W with $\mathbb{E}W < \infty$, we have that $\mathscr{L}(W) = \mathscr{L}(UW^s)$ if and only if W has exponential distribution.

2.1.2. *Exchangeable pairs.* Let (W, W') be an exchangeable pair. Assume that

$$\mathbb{E}^W(W' - W) = -\lambda + \lambda R \qquad \text{on } \{W > 0\}.$$

Then, if we set $G = (W' - W)/(2\lambda)$, we have $r_1(\mathcal{F}_{\mathrm{BW}}) \le \mathbb{E}|R|$ and $r_1(\mathcal{F}_{\mathrm{W}}) \le \mathbb{E}|RW|$.

This coupling was used by Chatterjee, Fulman and Röllin (2006) to obtain an exponential approximation for the spectrum of the Bernoulli–Laplace Markov chain. In order to obtain optimal rates, Chatterjee, Fulman and Röllin (2006) develop more application specific theorems than ours.

2.1.3. *Conditional distribution of W given E^c.* Let E be an event and let $p = \mathbb{P}[E]$, where p is small. Assume that W' and Y are defined on the same probability space and that $\mathscr{L}(W') = \mathscr{L}(W|E^c)$ and $\mathscr{L}(Y) = \mathscr{L}(W|E)$. Then, for any Lipschitz f with $f(0) = 0$, and with $G = (1-p)/p$,

$$\begin{aligned}
&\mathbb{E}\{Gf(W') - Gf(W)\}\\
&\quad = \frac{1-p}{p}\mathbb{E}f(W') - \frac{1}{p}\mathbb{E}f(W) + \mathbb{E}f(W)\\
&\quad = \frac{1-p}{p}\mathbb{E}f(W') - \frac{1-p}{p}\mathbb{E}(f(W)|E^c) - \mathbb{E}(f(W)|E) + \mathbb{E}f(W)\\
&\quad = \frac{1-p}{p}\mathbb{E}f(W') - \frac{1-p}{p}\mathbb{E}f(W') - \mathbb{E}f(Y) + \mathbb{E}f(W)\\
&\quad = \mathbb{E}f(W) - \mathbb{E}\{Yf'(UY)\},
\end{aligned}$$

so that $r_1(\mathcal{F}_{\mathrm{W}}) \le \mathbb{E}Y$. This coupling is used by Peköz (1996) for geometric approximation in total variation. The Stein operator used there is a discrete version of the Stein operator used in this article. Clearly, one will typically aim for an event $E \supset \{W = 0\}$ in order to have $Y = 0$.

REMARK 2.1. The roles of W and W' from the previous coupling can be reversed. Let E and p be as before. However, assume now that $\mathscr{L}(W) = \mathscr{L}(W'|E^c)$ and $\mathscr{L}(Y) = \mathscr{L}(W'|E)$. Then, it is again straightforward to see that $(W, W', -1/p)$ is a constant Stein coupling.

3. Applications.

3.1. *Random sums.* A classical result of Rényi (1957) states that $\mathscr{L}(p \times \sum_{i=1}^{N} X_i) \to \mathrm{Exp}(1)$ as $p \to 0$ when N has the $\mathrm{Ge}(p)$ distribution (independent of all else) and X_i are i.i.d. with $\mathbb{E}X_i = 1$. There have been some generalizations [see Brown (1990), Kalashnikov (1997) and the references therein]. Sugakova (1995), in particular, gives uniform error bounds for independent but nonidentically distributed summands with identical means. Our next result can be viewed as generalizing this to dependent summands and to nongeometric N. For a random variable X, we denote by F_X its distribution function and by F_X^{-1} its generalized inverse. We adopt the standard convention that $\sum_a^b = 0$ if $b < a$.

THEOREM 3.1. *Let $X = (X_1, X_2, \ldots)$ be a sequence of square integrable, nonnegative random variables, independent of all else, such that, for all $i \geq 1$,*

$$\mathbb{E}(X_i | X_1, \ldots, X_{i-1}) = \mu_i < \infty \qquad \textit{almost surely.} \tag{3.1}$$

Let N be a positive, integer valued random variable with $\mathbb{E}N < \infty$ and let M be a random variable satisfying

$$P(M = m) = \mu_m P(N \geq m)/\mu, \qquad m = 1, 2, \ldots \tag{3.2}$$

with

$$\mu = \mathbb{E}\sum_{i=1}^{N} X_i = \sum_{m \geq 1} \mu_m P(N \geq m).$$

Then, with $W = \mu^{-1}\sum_{i=1}^{N} X_i$, we have

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \leq 2\mu^{-1}\Big(\mathbb{E}|X_M - X_M^e| + \sup_{i \geq 1} \mu_i \mathbb{E}|N - M|\Big), \tag{3.3}$$

where each X_i^e is a random variable having the equilibrium distribution w.r.t. X_i given $X_1, \ldots, X_{i-1}$. If, in addition, $X_i \leq C$ for all i and $|N - M| \leq K$, then

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \leq 12\mu^{-1}\Big\{\sup_{i \geq 1} \|F_{X_i}^{-1} - F_{X_i^e}^{-1}\| + CK\Big\}; \tag{3.4}$$

if $K = 0$, the same bound also holds for unbounded X_i.

PROOF. We first show that $W^e := \mu^{-1}(\sum_{i=1}^{M-1} X_i + X_M^e)$ has the equilibrium distribution w.r.t. W. For a given Lipschitz f, we write $g(m) = f(\mu^{-1}\sum_{i=1}^{m} X_i)$ and we have

$$\mu\mathbb{E}\left[\frac{g(M)}{\mu_M} - \frac{g(M-1)}{\mu_M}\right] = \sum_{m\geq 0} P(N \geq m)\big(g(m) - g(m-1)\big) = \mathbb{E}g(N)$$

and, for any integer $m > 0$,

$$\mathbb{E}f'\left(\mu^{-1}\sum_{i=1}^{m-1} X_i + \mu^{-1}X_m^e\right) = \frac{\mu}{\mu_m}\mathbb{E}[g(m) - g(m-1)]$$

[using (2.3), (3.1) and the assumptions on X_i^e] that together give $\mathbb{E}f'(W^e) = \mathbb{E}f(W)$. Then using

$$W^e - W = \mu^{-1}\left\{(X_M^e - X_M) + \operatorname{sgn}(M-N)\sum_{i=(M\wedge N)+1}^{N\vee M} X_i\right\} \tag{3.5}$$

we obtain (3.3) from (2.7). Letting $\beta = \mu^{-1}\{\sup_{i\geq 1}\|F_{X_i}^{-1} - F_{X_i^e}^{-1}\| + CK\}$, and using Strassen's theorem we obtain (3.4) from (2.5); the remark after (3.4) follows similarly. □

REMARK 3.1. Let $N \sim \mathrm{Ge}(p)$ and assume that the μ_i are bounded from above and bounded away from 0. This implies in particular that $\mu \asymp 1/p$ as $p \to 0$. Using

$$d_{\mathrm{W}}(\mathscr{L}(N), \mathscr{L}(M)) = \inf_{(N,M)} \mathbb{E}|N - M| \tag{3.6}$$

from Kantorovič and Rubinšteĭn (1958) [see also Vallander (1973)], where the infimum ranges over all possible couplings of N and M, we can replace $\mathbb{E}|N-M|$ in (3.3) by the left-hand side of (3.6). To bound this quantity note first that from (3.2) we have $\mathbb{E}h(M) = \mathbb{E}\{\frac{\mu_N}{\mu p}h(N)\}$ for every function h for which the expectations exist. Note also that $\mathbb{E}(\mu_N) = \mu p$. Let h now be Lipschitz with Lipschitz constant 1 and assume without loss of generality that $h(0) = 0$, so that $|h(N)| \leq N$. Then

$$\begin{aligned}|\mathbb{E}\{h(M) - h(N)\}| &= \left|\mathbb{E}\left\{\left(\frac{\mu_N}{\mu p} - 1\right)h(N)\right\}\right| \leq \mathbb{E}\left|\left\{\frac{\mu_N}{\mu p} - 1\right\}N\right| \\ &\leq \frac{\sqrt{\operatorname{Var}(\mu_N)\mathbb{E}N^2}}{\mu p} \leq \frac{\sqrt{2\operatorname{Var}(\mu_N)}}{\mu p^2}.\end{aligned}$$

Hence, under the assumptions of this remark, $\mu^{-1}\sup_i \mu_i \mathbb{E}|N - M|$ is at most of order $\operatorname{Var}(\mu_N)$ as $p \to 0$.

Next, we have an immediate corollary by coupling stochastically ordered random variables.

COROLLARY 3.1. *In the setting in Theorem 3.1, assume either $N \le_{\mathrm{st}} M$ or $N \ge_{\mathrm{st}} M$ holds as well as that the X_i are independent and, for each i, we have $\mathbb{E}X_i = 1$ and either $X_i \le_{\mathrm{st}} X_i^e$ or $X_i \ge_{\mathrm{st}} X_i^e$. Then*

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le 2\mu^{-1} \sup_{i\ge 1}\left|\frac{1}{2}\mathbb{E}X_i^2 - 1\right| + 2\left|\frac{\mathbb{E}N^2}{2\mu^2} + \frac{1}{2\mu} - 1\right| \tag{3.7}$$

and, furthermore, if N has a Ge(p) *distribution then*

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le 2.47\left(p \sup_{i\ge 1}\left|\frac{1}{2}\mathbb{E}X_i^2 - 1\right|\right)^{1/2}. \tag{3.8}$$

REMARK 3.2. A nonnegative random variable X with finite mean is said to be NBUE (new better than used in expectation) if $X^e \le_{\mathrm{st}} X$ or NWUE (new worse than used in expectation) if $X^e \ge_{\mathrm{st}} X$ [see Shaked and Shanthikumar (2007) and Sengupta, Chatterjee and Chakraborty (1995) for other sufficient conditions]. A result similar to (3.8) appears as Theorem 6.1 in Brown and Ge (1984) with a larger constant, though Brown (1990) and Daley (1988) subsequently derived significant improvements.

EXAMPLE 3.1 (Geometric convolution of i.i.d. random variables). Assume that $N \sim \mathrm{Ge}(p)$ and that $\mathbb{E}X_1 = 1$. Since $\mathscr{L}(M) = \mathscr{L}(N)$, we can set $M = N$. Denote by $\delta(\mathcal{F})$ the distance between $\mathscr{L}(X_1)$ and Exp(1) as defined in (2.1) with respect to the set of test functions $\mathcal{F}$; define $\delta^e(\mathcal{F})$ analogously but between $\mathscr{L}(X)$ and $\mathscr{L}(X^e)$. In this case, the estimates of Theorem 3.1 reduce to

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le 2p\delta^e(\mathcal{F}_{\mathrm{W}}), \tag{3.9}$$

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le 12p\|F_{X_1}^{-1} - F_{X_1^e}^{-1}\| \tag{3.10}$$

which can be compared with the (slightly simplified)

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le p\delta(\mathcal{F}_{\mathrm{W}}) + 2p\delta(\mathcal{F}_2), \tag{3.11}$$

where

$$\mathcal{F}_2 = \{f \in C^1(\mathbb{R}) | f' \in \mathcal{F}_{\mathrm{W}}\},$$

from Kalashnikov (1997), Theorem 3.1 for $s = 2$, page 151.

Noting $\delta(\mathcal{F}_{\mathrm{W}}) \le 2\delta^e(\mathcal{F}_{\mathrm{W}})$ (using the Kantorovich–Rubinstein theorem), let $Z \sim \mathrm{Exp}(1)$ and let h be a differentiable function with $h(0) = 0$. Then, recalling that $\mathscr{L}(Z^e) = \mathscr{L}(Z)$, we have from (2.3) that $\mathbb{E}h(Z) = \mathbb{E}h'(Z)$ and, using again (2.3) for X and X^e,

$$\mathbb{E}h(X) - \mathbb{E}h(Z) = \mathbb{E}h'(X^e) - \mathbb{E}h'(Z). \tag{3.12}$$

This implies

$$\delta(\mathcal{F}_2) = d_{\mathrm{W}}(\mathscr{L}(X_1^e), \mathrm{Exp}(1)) \tag{3.13}$$

and hence, from (2.8), we have $\delta(\mathcal{F}_2) \le \delta^e(\mathcal{F}_{\mathrm{W}})$, so that (3.11) gives a bound which is not as good as (3.9) if the bound is to be expressed in terms of $\delta^e(\mathcal{F}_{\mathrm{W}})$.

On the other hand, from (3.13) and the triangle inequality,

$$\delta^e(\mathcal{F}_{\mathrm{W}}) \le \delta(\mathcal{F}_{\mathrm{W}}) + d_{\mathrm{W}}(\mathscr{L}(X_1^e), \mathrm{Exp}(1)) = \delta(\mathcal{F}_{\mathrm{W}}) + \delta(\mathcal{F}_2).$$

Hence, although much broader in applicability, our Theorem 3.1 yields results comparable to those in the literature when specialized to the setting of geometric convolutions.

THEOREM 3.2. *Let $X = (X_1, X_2, \dots)$ be a sequence of random variables with $\mathbb{E}X_i = \mu_i$ and $\mathbb{E}X_i^2 < \infty$. Let N, N' and N'' be nonnegative, square integrable, integer valued random variables independent of the sequence X. Assume that*

$$p := \mathbb{P}[N = 0] > 0, \qquad \mathscr{L}(N') = \mathscr{L}(N | N > 0), \qquad N'' \le N \le N'.$$

Define $S(k, l) := X_{k+1} + \cdots + X_l$ for $k < l$ and $S(k, l) = 0$ for $k \ge l$. Let $\mu = \mathbb{E}S(0, N)$ and $W = S(0, N)/\mu$. Then,

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le \frac{qs}{p\mu} + \frac{4q\mathbb{E}\{S(N, N')(1 + S(N'', N))\}}{p\mu^2} + \frac{4\mathbb{E}S(N'', N)}{\mu},$$

where $q = 1 - p$, $s^2 = \operatorname{Var}\mathbb{E}(S(N, N')|\mathcal{F}_{N''})$ and $\mathcal{F}_k := \sigma(X_1, \dots, X_k)$. If, in addition,

$$X_i \le C, \qquad N' - N \le K_1, \qquad N - N'' \le K_2, \tag{3.14}$$

for positive constants C, K_1 and K_2, then

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le \frac{qs}{p\mu} + \frac{22CK_2}{\mu} + \frac{2C^2K_1(11K_2 + 6K_1)}{p\mu^2}. \tag{3.15}$$

PROOF. We make use of the coupling construction from Section 2.1.3. Let $E = \{N = 0\}$, let $Y = 0$, let $W' = \mu^{-1}\sum_{i=1}^{N'} X_i$ and likewise $W'' = \mu^{-1}\sum_{i=1}^{N''} X_i$. Then the conditions of Section 2.1.3 are satisfied with $G = q/p$ and we can apply Theorem 2.2, in particular (2.11). We have $r_1(\mathcal{F}_{\mathrm{BW}}) = 0$ as proved in Section 2.1.3. Note now that $D = S(N, N')$ and $D' = S(N'', N)$. Hence, $r_2 = \mathbb{E}|1 - q(p\mu)^{-1}\mathbb{E}(S(N, N')|\mathcal{F}_{N''})|$. As (2.9) implies that $\mathbb{E}(GD) = \mathbb{E}W = 1$, the variance bound of r_2 follows. The d_{W}-bound follows from (2.10), using the rough estimates $r_3 + r_4 \le \mathbb{E}|GD^2|$ and $r_3' + r_4' \le 2\mathbb{E}|D'| + 2\mathbb{E}|GDD'|$ as we assume bounded second moments. To obtain the d_{K}-bound choose $\alpha = G = q/p$, $\beta = CK_1/\mu$ and $\beta' = CK_2/\mu$; then $r_5 = r_5' = 0$. Hence, (2.11) yields

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le r_2 + 22(\alpha\beta + 1)\beta' + 12\alpha\beta^2.$$

Plugging in the value for r_2 and the constants, the theorem is proved. □

EXAMPLE 3.2 (Geometric convolution under local dependence). If $N+1 \sim \mathrm{Ge}(p)$ (that is, N is a geometric distribution starting at 0) we can choose $N' = N+1$, as $\mathscr{L}(N|N>0) = \mathscr{L}(N+1)$ due to the well-known lack-of-memory property; hence $K_1 = 1$. Assume now there is a nonnegative integer m such that, for each i, $(X_1, \dots, X_i)$ is independent of $(X_{i+m+1}, X_{i+m+2}, \dots)$. We can set $N'' = \max(N-m, 0)$, hence $s^2 \le \operatorname{Var} \mu_{N+1}$, where $\mu_i := \mathbb{E}X_i$. Assume also that $\mu_i \ge \mu_0$ for some $\mu_0 > 0$, so that $\mu \ge \mu_0/p$. Hence, Theorem 3.2 yields

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \le \frac{\sqrt{\operatorname{Var}(\mu_{N+1})}}{\mu_0} + \frac{22Cpm}{\mu_0} + \frac{2C^2 p(11m+6)}{\mu_0^2}.$$

Again, convergence is obtained if $\operatorname{Var}(\mu_{N+1}) \to 0$ as $p \to 0$; cf. Remark 3.1.

3.2. *First passage times*. Approximately exponential hitting times for Markov chains have been widely studied; see Aldous (1989), Aldous and Brown (1992) and Aldous and Brown (1993) for entry points to this literature. Let $X_0, X_1, \dots$ be a stationary ergodic Markov chain with a countable state space $\mathcal{X}$, transition probability matrix $P = (P_{i,j})_{i,j\in\mathcal{X}}$ and stationary distribution $\pi = (\pi_i)_{i\in\mathcal{X}}$ and let

$$\mathscr{L}(T_{\pi,i}) = \mathscr{L}(\inf\{t \ge 0 : X_t = i\}) \qquad \text{starting with } \mathscr{L}(X_0) = \pi$$

be the hitting time on state i started according to the stationary distribution π and let

$$\mathscr{L}(T_{i,j}) = \mathscr{L}(\inf\{t > 0 : X_t = j\}) \qquad \text{starting with } X_0 = i$$

be the hitting time on state j starting from state i. We also say a stopping time $T_{i,\pi}$ is a stationary time starting from state i if $\mathscr{L}(X_{T_{i,\pi}} | X_0 = i) = \pi$.

COROLLARY 3.2. *With the above definitions, we have*

$$d_{\mathrm{K}}(\mathscr{L}(\pi_i T_{\pi,i}), \mathrm{Exp}(1)) \le 2\pi_i + \min\{\pi_i \mathbb{E}|T_{\pi,i} - T_{i,i}|, \mathbb{P}(T_{\pi,i} \ne T_{i,i})\}. \tag{3.16}$$

PROOF. Using a renewal argument to obtain $\mathbb{P}(T_{\pi,i} = k) = \pi_i \mathbb{P}(T_{i,i} > k)$, it is then straightforward to see that $\mathscr{L}(T_{i,i}^e) = \mathscr{L}(T_{\pi,i} + U)$ when U is a uniform random variable on [0, 1], independent of all else: with $f(0) = 0$ and using (2.3) we have

$$\begin{aligned}
\mathbb{E}f'(T_{\pi,i} + U) &= \mathbb{E}f(T_{\pi,i} + 1) - f(T_{\pi,i}) \\
&= \pi_i \sum_k \mathbb{P}(T_{i,i} > k)\bigl(f(k+1) - f(k)\bigr) \\
&= \pi_i \sum_k \sum_{j>k} \mathbb{P}(T_{i,i} = j)\bigl(f(k+1) - f(k)\bigr) \\
&= \pi_i \sum_j \sum_{0\le k<j} \mathbb{P}(T_{i,i} = j)\bigl(f(k+1) - f(k)\bigr) \\
&= \pi_i \sum_j \mathbb{P}(T_{i,i} = j) f(j) = \pi_i \mathbb{E}f(T_{i,i}).
\end{aligned}$$

We then have

$$
\begin{aligned}
d_{\mathrm{K}}(\mathscr{L}(\pi_i T_{\pi,i}), \mathrm{Exp}(1)) &\le \pi_i + d_{\mathrm{K}}\big(\mathscr{L}\big(\pi_i(T_{\pi,i}+U)\big), \mathrm{Exp}(1)\big) \\
&= \pi_i + d_{\mathrm{K}}(\mathscr{L}(\pi_i T^e_{i,i}), \mathrm{Exp}(1)),
\end{aligned}
\tag{3.17}
$$

where we use $d_{\mathrm{K}}(\mathscr{L}(T_{\pi,i}), \mathscr{L}(T_{\pi,i}+U)) \le \pi_i$ in the first line and $\mathscr{L}(T^e_{i,i}) = \mathscr{L}(T_{\pi,i}+U)$ in the second line. We obtain inequality (3.16) from (3.17) and (2.8), and then using

$$
\mathbb{E}|T_{\pi,i}+U-T_{i,i}| \le \mathbb{E}U + \mathbb{E}|T_{\pi,i}-T_{i,i}| \le 0.5 + \mathbb{E}|T_{\pi,i}-T_{i,i}|
$$

along with (3.17) and (2.6) using $\beta = \pi_i$ (since $\{|T_{\pi,i}+U-T_{i,i}| > 1\}$ implies $\{T_{\pi,i} \ne T_{i,i}\}$). $\square$

Below, whenever $T_{i,i}$ and $T_{i,\pi}$ are used together in an expression it assumed they are both based on a single copy of the Markov chain.

COROLLARY 3.3. *With the above definitions and* $\rho = \mathbb{P}[T_{i,i} < T_{i,\pi}]$,

$$
\begin{aligned}
&d_{\mathrm{K}}(\mathscr{L}(\pi_i T_{\pi,i}), \mathrm{Exp}(1)) \\
&\quad \le 2\pi_i + \min\left\{\pi_i\Big(\mathbb{E}T_{i,\pi} + \rho \sup_j \mathbb{E}T_{j,i}\Big), \sum_{n=1}^{\infty} |P^{(n)}_{i,i} - \pi_i|\right\}.
\end{aligned}
\tag{3.18}
$$

PROOF. Letting $X_0 = i$, $T_{\pi,i} = \inf\{t \ge 0 : X_{T_{i,\pi}+t} = i\}$, $T_{i,i} = \inf\{t > 0 : X_t = i\}$ and $A = \{T_{i,i} < T_{i,\pi}\}$ we have

$$
|T_{\pi,i} - T_{i,i}| \le (T_{\pi,i} + T_{i,\pi}) I_A + T_{i,\pi} I_{A^c} \le T_{i,\pi} + T_{\pi,i} I_A
$$

and the first argument in the minimum of (3.18) follows from (3.16) after noting $\mathbb{E}[T_{\pi,i}|A] \le \sup_j \mathbb{E}T_{j,i}$.

For the second argument in the minimum, let $X_1, X_2, \ldots$ be the stationary Markov chain and let $Y_0, Y_1, \ldots$ be a coupled copy of the Markov chain started in state i at time 0, but let $Y_1, Y_2, \ldots$ be coupled with $X_1, X_2, \ldots$ according to the maximal coupling of Griffeath (1974/75) so that we have $\mathbb{P}(X_n = Y_n = i) = \pi_i \wedge P^{(n)}_{i,i}$. Let $T_{\pi,i}$ and $T_{i,i}$ be hitting times respectively defined on these two Markov chains. Then

$$
\mathbb{P}(T_{\pi,i} \ne T_{i,i}) \le \sum_n \mathbb{P}(X_n = i, Y_n \ne i) + \mathbb{P}(Y_n = i, X_n \ne i)
$$

and since

$$
\begin{aligned}
\mathbb{P}(X_n = i, Y_n \ne i) &= \pi_i - \mathbb{P}(X_n = i, Y_n = i) \\
&= \pi_i - \pi_i \wedge P^{(n)}_{i,i} \\
&= [\pi_i - P^{(n)}_{i,i}]^+,
\end{aligned}
$$

and a similar calculation yields $\mathbb{P}(Y_n = i, X_n \neq i) = [P_{i,i}^{(n)} - \pi_i]^+$, and then we obtain (3.18). □

EXAMPLE 3.3. With the above definitions and further assuming X_n is an m-dependent Markov chain, we can let $T_{i,\pi} = m$ and we thus have

$$d_{\mathrm{K}}(\mathscr{L}(\pi_i T_{\pi,i}), \mathrm{Exp}(1)) \\ \leq 2\pi_i + \min\left\{\pi_i\Bigl(m + \mathbb{P}(T_{i,i} < m) \sup_j \mathbb{E}T_{j,i}\Bigr), \sum_{n=1}^{m-1} \bigl|P_{i,i}^{(n)} - \pi_i\bigr|\right\}.$$

If we consider flipping a biased coin repeatedly, let T be the number of flips required until the beginning of a given pattern (that cannot overlap with itself) of heads and tails of length k first appears as a run. The current run of k flips can be encoded in the state space of a k-dependent Markov chain and then applying the second result above we obtain

$$d_{\mathrm{K}}(\mathscr{L}(\pi_i T_{\pi,i}), \mathrm{Exp}(1)) \leq \pi_i (k + 1).$$

Using the "de-clumping" trick of counting the flips T preceding the first appearance of tails followed by k heads in row we have

$$d_{\mathrm{K}}(\mathscr{L}(q p^k T_{\pi,i}), \mathrm{Exp}(1)) \leq (k + 2) p^k,$$

where $p = 1 - q$ is the probability of heads. Similar results are obtained using Poisson and geometric approximations respectively in Barbour, Holst and Janson [(1992), page 164] and Peköz (1996).

Recall the definitions of NBUE and NWUE from Remark 3.2 and, as discussed in Aldous and Fill (2010), that stationary reversible continuous-time Markov chain hitting times are NWUE. The next results are immediate consequences of Theorem 2.1 and (2.2). While (3.21) appears to be new, inequality (3.19) appears in Brown (1990), Lemma 2.3. Inequality (3.20) with a larger constant of 3.119 appears in Brown and Ge [(1984), Theorem 3.6] for the NBUE case and in Brown and Ge [(1984), equation (5.3)] for the NWUE case; this constant was later improved in both cases to 1.41 for small ρ in Daley (1988), equation (1.7).

COROLLARY 3.4. *If W is either NBUE or NWUE with $\mathbb{E}W = 1$, finite second moment and letting $\rho = |\frac{1}{2}\mathbb{E}W^2 - 1|$, we have*

$$d_{\mathrm{K}}(\mathscr{L}(W^e), \mathrm{Exp}(1)) \leq \rho, \tag{3.19}$$

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1)) \leq 2.47\rho^{1/2} \tag{3.20}$$

and

$$d_{\mathrm{W}}(\mathscr{L}(W^e), \mathrm{Exp}(1)) \leq \rho, \qquad d_{\mathrm{W}}(\mathscr{L}(W), \mathrm{Exp}(1)) \leq 2\rho. \tag{3.21}$$

3.3. *Critical Galton–Watson branching process.* Let $Z_0 = 1, Z_1, Z_2, \dots$ be a Galton–Watson branching process with offspring distribution $\nu = \mathscr{L}(Z_1)$. A theorem due to Yaglom (1947) states that, if $\mathbb{E}Z_1 = 1$ and $\operatorname{Var} Z_1 = \sigma^2 < \infty$, then $\mathscr{L}(n^{-1}Z_n|Z_n > 0)$ converges to an exponential distribution with mean $\sigma^2/2$. We give a rate of convergence for this asymptotic under finite third moment of the offspring distribution using the idea from Section 2.1.1. Though exponential limits in this context are an active area of research [see, e.g., Lalley and Zheng (2011)], the question of rates does not appear to have been previously studied in the literature. To this end, we make use the of construction from Lyons, Pemantle and Peres (1995); we refer to that article for more details on the construction and only present what is needed for our purpose.

THEOREM 3.3. *For a critical Galton–Watson branching process with offspring distribution $\nu = \mathscr{L}(Z_1)$ such that $\mathbb{E}Z_1^3 < \infty$ we have*

$$d_{\mathrm{W}}\bigl(\mathscr{L}\bigl(2Z_n/(\sigma^2 n)|Z_n > 0\bigr), \operatorname{Exp}(1)\bigr) = \mathrm{O}\biggl(\frac{\log n}{n}\biggr).$$

PROOF. First, we construct a size-biased branching tree as in Lyons, Pemantle and Peres (1995). We assume that this tree is labeled and ordered, in the sense that, if w and v are vertices in the tree from the same generation and w is to the left of v, then the offspring of w is to the left of the offspring of v, too. Start in generation 0 with one vertex v_0 and let it have a number of offspring distributed according to the size-bias distribution of ν. Pick one of the offspring of v_0 uniformly at random and call it v_1. To each of the siblings of v_1, attach an independent Galton–Watson branching process with offspring distribution ν. For v_1 proceed as for v_0, that is, give it a size-biased number of offspring, pick one at uniformly at random, call it v_2, attach independent Galton–Watson branching process to the siblings of v_2 and so on. It is clear that this will always give an infinite tree as the "spine" $v_0, v_1, v_2, \dots$ of the tree will never die out.

We next need some notation. Denote by S_n the total number of particles in generation n. Denote by L_n and R_n, respectively, the number of particles to the left (exclusive v_n) and to the right (inclusive v_n), respectively, of vertex v_n, so that $S_n = L_n + R_n$. We can describe these particles in more detail, according to the generation at which they split off from the spine. Denote by $S_{n,j}$ the number of particles in generation n that stem from any of the siblings of v_j (but not v_j itself). Clearly, $S_n = 1 + \sum_{j=1}^n S_{n,j}$, where the summands are independent. Likewise, let $L_{n,j}$ and $R_{n,j}$, respectively, be the number of particles in generation n that stem from the siblings to the left and right, respectively, of v_j (note that $L_{n,n}$ and $R_{n,n}$ are just the number of siblings of v_n to the left and to the right, respectively). We have the relations $L_n = \sum_{j=1}^n L_{n,j}$ and $R_n = 1 + \sum_{j=1}^n R_{n,j}$. Note that, for fixed j, $L_{n,j}$ and $R_{n,j}$ are in general not independent, as they are linked through the offspring size of v_{j-1}.

Now let $R'_{n,j}$ be independent random variables such that

$$\mathscr{L}(R'_{n,j}) = \mathscr{L}(R_{n,j} | L_{n,j} = 0)$$

and, with $A_{n,j} = \{L_{n,j} = 0\}$, define

$$R^*_{n,j} = R_{n,j} I_{A_{n,j}} + R'_{n,j} I_{A^c_{n,j}} = R_{n,j} + (R'_{n,j} - R_{n,j}) I_{A^c_{n,j}}. \tag{3.22}$$

Define also $R^*_n = 1 + \sum_{j=1}^n R^*_{n,j}$. Let us collect a few facts which we will then use to give the proof of the theorem:

(i) for any nonnegative random variable X the size-biased distribution of $\mathscr{L}(X)$ is the same as the size-biased distribution of $\mathscr{L}(X | X > 0)$;
(ii) S_n has the size-biased distribution of Z_n;
(iii) given S_n, the vertex v_n is uniformly distributed among the particles of the nth generation;
(iv) $\mathscr{L}(R^*_n) = \mathscr{L}(Z_n | Z_n > 0)$;
(v) $\mathbb{E}\{R'_{n,j} I_{A^c_{n,j}}\} \le \sigma^2 \mathbb{P}[A^c_{n,j}]$;
(vi) $\mathbb{E}\{R_{n,j} I_{A^c_{n,j}}\} \le \gamma \mathbb{P}[A^c_{n,j}]$, where $\gamma = \mathbb{E} Z_1^3$;
(vii) $\mathbb{P}[A^c_{n,j}] \le \sigma^2 \mathbb{P}[Z_{n-j} > 0] \le C(\nu)/(n - j + 1)$ for some absolute constant $C(\nu)$.

Statement (i) is easy to verify, (ii) follows from Lyons, Pemantle and Peres (1995), equation (2.2), (iii) follows from Lyons, Pemantle and Peres (1995), comment after (2.2), (iv) follows from Lyons, Pemantle and Peres (1995), proof of Theorem C(i). Using independence,

$$\mathbb{E}\{R'_{n,j} I_{A^c_{n,j}}\} = \mathbb{E} R'_{n,j} \mathbb{P}[A^c_{n,j}] \le \sigma^2 \mathbb{P}[A^c_{n,j}],$$

where the second inequality is due to Lyons, Pemantle and Peres (1995), proof of Theorem C(i), which proves (v). If X_j denotes the number of siblings of v_j, having the size bias distribution of Z_1 minus 1, we have

$$\begin{aligned}
\mathbb{E}\{R_{n,j} I_{A^c_{n,j}}\} \le \mathbb{E}\{X_j I_{A^c_{n,j}}\} &\le \sum_k k \mathbb{P}[X_j = k, A^c_{n,j}] \\
&\le \sum_k k \mathbb{P}[X_j = k] \mathbb{P}[A^c_{n,j} | X_j = k] \\
&\le \sum_k k^2 \mathbb{P}[X_j = k] \mathbb{P}[A^c_{n,j}] \le \gamma \mathbb{P}[A^c_{n,j}],
\end{aligned}$$

hence (vi). Finally,

$$\mathbb{P}[A^c_{n,j}] = \mathbb{E}\{\mathbb{P}[A^c_{n,j} | X_j]\} \le \mathbb{E}\{X_j \mathbb{P}[Z_{n-j} > 0]\} \le \sigma^2 \mathbb{P}[Z_{n-j} > 0].$$

Using Kolmogorov's estimate [see Lyons, Pemantle and Peres (1995), Theorem C(i)], we have $\lim_{n\to\infty} n\mathbb{P}[Z_n > 0] = 2/\sigma^2$, which implies (vii).

We are now in the position to prove the theorem using (2.5) of Theorem 2.1. Let $c = 2/\sigma^2$. Due to (iv) we can set $W = cR_n^*/n$. Due to (i) and (ii), S_n has the size bias distribution of R_n^*. Let U be an independent and uniform random variable on [0, 1]. Now, $R_n - U$ is a continuous random variable taking values on $[0, S_n]$ and, due to (iii), has distribution $\mathscr{L}(US_n)$; hence we can set $W^e = c(R_n - U)/n$. It remains to bound $\mathbb{E}|W - W^e|$. From (3.22) and using (v)–(vii), we have

$$nc^{-1}\mathbb{E}|W - W^e| \le \mathbb{E}U + \mathbb{E}|R_n^* - R_n| \le 1 + \sum_{j=1}^{n} \mathbb{E}\{R'_{n,j} I_{A_{n,j}^c} + R_{n,j} I_{A_{n,j}^c}\}$$
$$\le 1 + C(\nu) \sum_{j=1}^{n} \frac{\sigma^2 + \gamma}{n - j + 1} \le 1 + C(\nu)(\sigma^2 + \gamma)(1 + \log n).$$

Hence, for a possibly different constant $C(\nu)$,

$$\mathbb{E}|W - W^e| \le \frac{C(\nu)\log n}{n}.$$

Plugging this into (2.7) yields the final bound. □

4. Proofs of main results. Our results are based on the Stein operator

$$Af(x) = f'(x) - f(x) \tag{4.1}$$

and the corresponding Stein equation

$$f'(w) - f(w) = h(w) - \mathbb{E}h(Z), \qquad w \ge 0 \tag{4.2}$$

previously studied (independently of each other and, in the case of the first two, independent of the present work) by Weinberg (2005), Bon (2006) and Chatterjee, Fulman and Röllin (2006). It is straightforward that the solution f to (4.2) can be written as

$$f(w) = -e^w \int_w^\infty \big(h(x) - \mathbb{E}h(Z)\big)e^{-x}\, dx. \tag{4.3}$$

We next need some properties of the solution (4.3). Some preliminary results can be found in Weinberg (2005), Bon (2006), Chatterjee, Fulman and Röllin (2006) and Daly (2008). We give self-contained proofs of the following bounds.

LEMMA 4.1 (Properties of the solution to the Stein equation). *Let f be the solution to* (4.2). *If h is bounded, we have*

$$\|f\| \le \|h\|, \qquad \|f'\| \le 2\|h\|. \tag{4.4}$$

If h is Lipschitz, we have

$$|f(w)| \le (1 + w)\|h'\|, \qquad \|f'\| \le \|h'\|, \qquad \|f''\| \le 2\|h'\|. \tag{4.5}$$

For any $a > 0$ *and any* $\varepsilon > 0$, *let*

$$h_{a,\varepsilon}(x) := \varepsilon^{-1} \int_0^\varepsilon \mathrm{I}[x + s \le a]\, ds. \tag{4.6}$$

Define $f_{a,\varepsilon}$ *as in* (4.3) *with respect to* $h_{a,\varepsilon}$. *Define* $h_{a,0}(x) = \mathrm{I}[x \le a]$ *and* $f_{a,0}$ *accordingly. Then, for all* $\varepsilon \ge 0$,

$$\|f_{a,\varepsilon}\| \le 1, \qquad \|f'_{a,\varepsilon}\| \le 1, \tag{4.7}$$

$$|f_{a,\varepsilon}(w+t) - f_{a,\varepsilon}(w)| \le 1, \qquad |f'_{a,\varepsilon}(w+t) - f'_{a,\varepsilon}(w)| \le 1 \tag{4.8}$$

and, for all $\varepsilon > 0$,

$$|f'_{a,\varepsilon}(w+t) - f'_{a,\varepsilon}(w)| \le (|t| \wedge 1) + \varepsilon^{-1} \int_{t\wedge 0}^{t\vee 0} \mathrm{I}[a - \varepsilon \le w + u \le a]\, du. \tag{4.9}$$

PROOF. Write $\tilde{h}(w) = h(w) - \mathbb{E}h(Z)$. Assume now that h is bounded. Then

$$|f(w)| \le e^w \int_w^\infty |\tilde{h}(x)| e^{-x}\, dx \le \|h\|.$$

Rearranging (4.2) we have $f'(w) = f(w) + \tilde{h}(w)$, hence

$$|f'(w)| \le |f(w)| + |\tilde{h}(w)| \le 2\|h\|.$$

This proves (4.4). Assume now that h is Lipschitz. We can further assume without loss of generality that $h(0) = 0$ as f will not change under shift; hence we may assume that $|h(x)| \le x\|h'\|$. Thus,

$$|f(w)| \le e^w \int_w^\infty x\|h'\| e^{-x}\, dx = (1 + w)\|h'\|,$$

which is the first bound of (4.5). Now, differentiate both sides of (4.2) to obtain

$$f''(w) - f'(w) = h'(w), \tag{4.10}$$

hence, analogous to (4.3), we have

$$f'(w) = -e^w \int_w^\infty h'(x) e^{-x}\, dx.$$

The same arguments as before lead to the second and third bound of (4.5).

We now look at the properties of $f_{a,\varepsilon}$. It is easy to check that

$$f_{a,0}(x) = (e^{x-a} \wedge 1) - e^{-a}, \qquad f'_{a,0}(x) = e^{x-a}\mathrm{I}[x \le a] \tag{4.11}$$

is the explicit solution to (4.10) with respect to $h_{a,0}$. Now, it is not difficult to see that, for $\varepsilon > 0$, we can write

$$f_{a,\varepsilon}(x) = \varepsilon^{-1} \int_0^\varepsilon f_{a,0}(x + s)\, ds$$

and this $f_{a,\varepsilon}$ satisfies (4.2). These representations immediately lead to the bounds (4.7) and (4.8) for $\varepsilon \geq 0$ from the explicit formulas (4.11). Now let $\varepsilon > 0$; observe that, from (4.10),

$$f'(x+t) - f'(x) = \big(f(x+t) - f(x)\big) + \big(h(x+t) - h(x)\big).$$

Again from (4.11), we deduce that $|f_{a,\varepsilon}(x+t) - f_{a,\varepsilon}(x)| \leq (|t| \wedge 1)$, which yields the first part of the bound (4.9). For the second part, assume that $t > 0$ and write

$$h_{a,\varepsilon}(x+t) - h_{a,\varepsilon}(x) = \int_0^t h'_{a,\varepsilon}(x+s)\,ds = -\varepsilon^{-1}\int_0^t \mathrm{I}[a-\varepsilon \leq x+u \leq a]\,du.$$

Taking the absolute value this gives the second part of the bound (4.9) for $t > 0$; a similar argument yields the same bound for $t < 0$. □

The following lemmas are straightforward and hence given without proof.

LEMMA 4.2 (Smoothing lemma). *For any* $\varepsilon > 0$

$$d_{\mathrm{K}}(\mathscr{L}(W), \mathscr{L}(Z)) \leq \varepsilon + \sup_{a>0} |\mathbb{E}h_{a,\varepsilon}(W) - \mathbb{E}h_{a,\varepsilon}(Z)|,$$

where $h_{a,\varepsilon}$ *are defined as in Lemma* 4.1.

LEMMA 4.3 (Concentration inequality). *For any random variable* V,

$$\mathbb{P}[a \leq V \leq b] \leq (b-a) + 2d_{\mathrm{K}}(\mathscr{L}(V), \mathrm{Exp}(1)).$$

For the rest of the article, write $\kappa = d_{\mathrm{K}}(\mathscr{L}(W), \mathrm{Exp}(1))$.

PROOF OF THEOREM 2.1. Let $\Delta := W - W^e$. Define $I_1 := \mathrm{I}[|\Delta| \leq \beta]$; note that W^e may not have finite first moment. With f as in (4.2) with respect to (4.6), the quantity $\mathbb{E}f'(W^e)$ is well defined as $\|f'\| < \infty$, and we have

$$\begin{aligned}
&\mathbb{E}\{f'(W) - f(W)\}\\
&\qquad = \mathbb{E}\big\{I_1\big(f'(W) - f'(W^e)\big)\big\} + \mathbb{E}\big\{(1-I_1)\big(f'(W) - f'(W^e)\big)\big\} =: J_1 + J_2.
\end{aligned}$$

Using (4.7), $|J_2| \leq \mathbb{P}[|\Delta| > \beta]$. Now, using (4.10) and in the last step Lemma 4.3,

$$\begin{aligned}
J_1 &= \mathbb{E}\left\{I_1 \int_0^\Delta f''(W+t)\,dt\right\}\\
&= \mathbb{E}\left\{I_1 \int_0^\Delta \big(f'(W+t) - \varepsilon^{-1}\mathrm{I}[a-\varepsilon \leq W+t \leq a]\big)\,dt\right\}\\
&\leq \mathbb{E}|I_1\Delta| + \int_{-\beta}^0 \mathbb{P}[a-\varepsilon \leq W+t \leq a]\,dt \leq 2\beta + 2\beta\varepsilon^{-1}\kappa.
\end{aligned}$$

Similarly,

$$J_1 \geq -\mathbb{E}|I_1\Delta| - \int_0^\beta \mathbb{P}[a-\varepsilon \leq W+t \leq a]\,dt \geq -2\beta - 2\beta\varepsilon^{-1}\kappa,$$

hence $|J_1| \leq 2\beta + 2\beta\varepsilon^{-1}\kappa$. Using Lemma 4.2 and choosing $\varepsilon = 4\beta$,

$$\kappa \leq \varepsilon + \mathbb{P}[|\Delta| > \beta] + 2\beta + 2\beta\varepsilon^{-1}\kappa \leq \mathbb{P}[|\Delta| > \beta] + 6\beta + 0.5\kappa.$$

Solving for κ proves (2.5).

To obtain (2.6), write

$$\begin{aligned}\mathbb{E}\{f'(W^e) - f(W^e)\} &= \mathbb{E}\{f(W) - f(W^e)\}\\ &= \mathbb{E}\big\{I_1\big(f(W) - f(W^e)\big)\big\} + \mathbb{E}\big\{(1-I_1)\big(f(W) - f(W^e)\big)\big\}.\end{aligned}$$

Hence, using Taylor's expansion along with the bounds (4.7) for $\varepsilon = 0$,

$$|\mathbb{E}f'(W^e) - f(W^e)| \leq \|f'\|\mathbb{E}|I_1\Delta| + \mathbb{P}[|\Delta| > \beta] \leq \beta + \mathbb{P}[|\Delta| > \beta],$$

which gives (2.6).

Assume now in addition that W has finite variance so that W^e has finite mean. Then

$$|\mathbb{E}\{f'(W) - f(W)\}| = |\mathbb{E}\{f'(W) - f'(W^e)\}| \leq \|f''\|\mathbb{E}|\Delta|.$$

From the bound (4.5), (2.7) follows. Also,

$$|\mathbb{E}\{f'(W^e) - f(W^e)\}| \leq \|f'\|\mathbb{E}|\Delta|$$

which yields (2.8) from (4.7) with $\varepsilon = 0$; the remark after (2.8) follows from (4.5). □

PROOF THEOREM 2.2. Let f be the solution (4.2) to (4.3), hence $f(0) = 0$, and assume that f is Lipschitz. From the fundamental theorem of calculus, we have

$$f(W') - f(W) = \int_0^D f'(W+t)\,dt.$$

Multiplying both sides by G and comparing it with the left-hand side of (4.2), we have

$$\begin{aligned}f'(W) - f(W) = {}& Gf(W') - Gf(W) - f(W)\\ &+ (1-GD)f'(W'')\\ &+ (1-GD)\big(f'(W) - f'(W'')\big)\\ &- G\int_0^D \big(f'(W+t) - f'(W)\big)\,dt.\end{aligned}$$

Note that we can take expectation component-wise due to the moment assumptions. Hence,

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = R_1(f) + R_2(f) + R_3(f) - R_4(f),$$

where

$$\begin{aligned}
R_1(f) &= \mathbb{E}\{Gf(W') - Gf(W) - f(W)\},\\
R_2(f) &= \mathbb{E}\{(1 - GD)f'(W'')\},\\
R_3(f) &= \mathbb{E}\{(1 - GD)\big(f'(W) - f'(W'')\big)\},\\
R_4(f) &= \mathbb{E}\bigg\{G\int_0^D \big(f'(W+t) - f'(W)\big)\,dt\bigg\}.
\end{aligned}$$

Assume now that $h \in \mathcal{F}_{\mathrm{BW}}$ and f the solution to (4.2). Then from (4.4) and (4.5) we obtain $\|f\| \le 1$, $\|f'\| \le 1$ and $\|f''\| \le 2$. Hence, $f \in \mathcal{F}_{\mathrm{BW}}$, $|R_1(f)| \le r_1(\mathcal{F}_{\mathrm{BW}})$ and $|R_2(f)| \le r_2$. Furthermore,

$$\begin{aligned}
|R_3(f)| &\le \mathbb{E}\big|(1 - GD)\big(f'(W'') - f'(W)\big)\big|\\
&\le 2\mathbb{E}\{|1 - GD|\mathrm{I}[|D'| > 1]\} + 2\mathbb{E}\{|1 - GD|(|D'| \wedge 1)\}\\
&= 2r_3' + 2r_4'
\end{aligned}$$

and

$$\begin{aligned}
|R_4(f)| &\le \mathbb{E}\bigg|G\int_0^D \big(f'(W+t) - f'(W)\big)\,dt\bigg|\\
&\le 2\mathbb{E}\big|GD\mathrm{I}[|D| > 1]\big| + 2\mathbb{E}|G(D^2 \wedge 1)|\\
&= 2r_3 + 2r_4.
\end{aligned}$$

This yields the d_{BW} results. Now let $h \in \mathcal{F}_{\mathrm{W}}$ and f the solution to (4.2). Then, from (4.4) and (4.5), we have $|f(x)| \le (1 + x)$, $\|f'\| \le 1$ and $\|f''\| \le 2$, hence the bounds on $R_2(f)$, $R_3(f)$ and $R_4(f)$ remain, whereas now $f \in \mathcal{F}_{\mathrm{W}}$ and, thus, $|R_1(f)| \le r_1(\mathcal{F}_{\mathrm{W}})$. This proves the d_{W} estimate.

Now let f be the solution to (4.2) with respect to $h_{a,\varepsilon}$ as in (4.6). Then, from (4.7), we have $\|f\| \le 1$ and $\|f'\| \le 1$, hence $f \in \mathcal{F}_{\mathrm{BW}}$, $|R_1(f)| \le r_1(\mathcal{F}_{\mathrm{BW}})$ and $|R_2(f)| \le r_2$. Let $I_1 = \mathrm{I}[|G| \le \alpha, |D| \le \beta', |D'| \le \beta']$. Write

$$\begin{aligned}
R_3(f) &= \mathbb{E}\{(1 - I_1)(1 - GD)\big(f'(W'') - f'(W)\big)\}\\
&\qquad + \mathbb{E}\{I_1(1 - GD)\big(f'(W'') - f'(W)\big)\} =: J_1 + J_2.
\end{aligned}$$

Using (4.7), $|J_1| \le r_5'$ is immediate. Using (4.9) and Lemma 4.3,

$$\begin{aligned}
|J_2| &\le \mathbb{E}\big|(GD - 1)I_1\big(f'(W'') - f'(W)\big)\big|\\
&\le (\alpha\beta + 1)\beta' + (\alpha\beta + 1)\varepsilon^{-1}\int_{-\beta'}^{\beta'} \mathbb{P}[a - \varepsilon \le W + u \le a]\,du
\end{aligned}$$

$$\leq (\alpha\beta+1)\beta' + (\alpha\beta+1)\varepsilon^{-1}\int_{-\beta'}^{\beta'}(\varepsilon+2\kappa)\,du$$
$$= 3(\alpha\beta+1)\beta' + 4(\alpha\beta+1)\beta'\varepsilon^{-1}\kappa.$$

Similarly, let $I_2 = \mathrm{I}[|G| \leq \alpha, |D| \leq \beta]$ and write

$$R_4(f) = \mathbb{E}\biggl\{G(1-I_2)\int_0^D \bigl(f'(W+t) - f'(W)\bigr)\,dt\biggr\} + \mathbb{E}\biggl\{GI_2\int_0^D \bigl(f'(W+t) - f'(W)\bigr)\,dt\biggr\} =: J_3 + J_4.$$

By (4.7), $|J_3| \leq r_5$. Using again (4.9) and Lemma 4.3,

$$|J_4| \leq \mathbb{E}\biggl\{GI_2\int_{D\wedge 0}^{D\vee 0} |f'(W+t) - f'(W)|\,dt\biggr\}$$
$$\leq \alpha\mathbb{E}\biggl\{\int_{-\beta}^{\beta}\biggl[(|t|\wedge 1) + \varepsilon^{-1}\int_{t\wedge 0}^{t\vee 0}\mathrm{I}[a-\varepsilon \leq W+u \leq a]\,du\biggr]dt\biggr\}$$
$$\leq \alpha\beta^2 + \alpha\varepsilon^{-1}\mathbb{E}\biggl\{\int_{-\beta}^{\beta}\int_{t\wedge 0}^{t\vee 0}(\varepsilon+2\kappa)\,du\,dt\biggr\} = 2\alpha\beta^2 + 2\alpha\beta^2\varepsilon^{-1}\kappa.$$

Using Lemma 4.2 and collecting the bounds above, we obtain

$$\kappa \leq \varepsilon + r_1(\mathcal{F}_{\mathrm{BW}}) + r_2 + |J_1| + |J_2| + |J_3| + |J_4|$$
$$\leq \varepsilon + r_1(\mathcal{F}_{\mathrm{BW}}) + r_2 + r_5 + r_5' + 3(\alpha\beta+1)\beta' + 2\alpha\beta^2 + \bigl(4(\alpha\beta+1)\beta' + 2\alpha\beta^2\bigr)\varepsilon^{-1}\kappa$$

so that, setting $\varepsilon = 8(\alpha\beta+1)\beta' + 4\alpha\beta^2$,

$$\kappa \leq \varepsilon + r_1(\mathcal{F}_{\mathrm{BW}}) + r_2 + r_5 + r_5' + 11(\alpha\beta+1)\beta' + 6\alpha\beta^2 + 0.5\kappa.$$

Solving for κ yields the final bound. □

Acknowledgments. The authors would like to express gratitude for the gracious hospitality of Louis Chen and Andrew Barbour during a visit to the National University of Singapore in January 2009, where a portion of this work was completed. We also thank the referees for their helpful comments. We are indebted to Nathan Ross for many valuable suggestions and for pointing out an error (and its elegant solution) in an early version of this paper. We also thank Mark Brown, Fraser Daly and Larry Goldstein for inspiring discussions.

REFERENCES

ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic. Applied Mathematical Sciences* **77**. Springer, New York. MR969362

ALDOUS, D. J. and BROWN, M. (1992). Inequalities for rare events in time-reversible Markov chains. I. In *Stochastic Inequalities* (*Seattle*, *WA*, 1991). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **22** 1–16. IMS, Hayward, CA. MR1228050

ALDOUS, D. J. and BROWN, M. (1993). Inequalities for rare events in time-reversible Markov chains. II. *Stochastic Process*. *Appl*. **44** 15–25. MR1198660

ALDOUS, D. J. and FILL, J. A. (2010). *Reversible Markov Chains and Random Walks on Graphs*. Preprint. Available at stat-www.berkeley.edu/users/aldous/RWG/book.html.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. *Oxford Studies in Probability* **2**. Oxford Univ. Press, New York. MR1163825

BON, J. L. (2006). Error bounds for exponential approximation of large-system reliability. *J. Math*. *Sci*. (*N. Y.*) **138** 5366–5376. MR2261586

BROWN, M. (1990). Error bounds for exponential approximations of geometric convolutions. *Ann*. *Probab*. **18** 1388–1402. MR1062073

BROWN, M. and GE, G. P. (1984). Exponential approximations for two classes of aging distributions. *Ann*. *Probab*. **12** 869–875. MR744241

CHATTERJEE, S. (2008). A new method of normal approximation. *Ann*. *Probab*. **36** 1584–1610. MR2435859

CHATTERJEE, S., FULMAN, J. and RÖLLIN, A. (2006). Exponential approximation by exchangeable pairs and spectral graph theory. Preprint. Available at http://arxiv.org/abs/math/0605552.

CHEN, L. H. Y. and RÖLLIN, A. (2009). Stein couplings for normal approximation. Preprint. Available at http://arxiv.org/abs/1003.6039.

DALEY, D. J. (1988). Tight bounds on the exponential approximation of some aging distributions. *Ann*. *Probab*. **16** 414–423. MR920282

DALY, F. (2008). Upper bounds for Stein-type operators. *Electron*. *J*. *Probab*. **13** 566–587. MR2399291

GHOSH, S. (2009). L^p bounds for a combinatorial central limit theorem with involutions. Preprint.

GIBBS, A. L. and SU, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review / Revue Internationale de Statistique* **70** 419–435.

GOLDSTEIN, L. (2005). Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J*. *Appl*. *Probab*. **42** 661–683. MR2157512

GOLDSTEIN, L. (2007). L^1 bounds in normal approximation. *Ann*. *Probab*. **35** 1888–1930. MR2349578

GOLDSTEIN, L. (2009). Personal communication and unpublished notes. Stein workshop, January 2009, Singapore.

GOLDSTEIN, L. and REINERT, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann*. *Appl*. *Probab*. **7** 935–952. MR1484792

GRIFFEATH, D. (1974/75). A maximal coupling for Markov chains. *Z*. *Wahrsch*. *Verw*. *Gebiete* **31** 95–106. MR0370771

KALASHNIKOV, V. (1997). *Geometric Sums*: *Bounds for Rare Events with Applications*: *Risk Analysis*, *Reliability*, *Queueing*. *Mathematics and Its Applications* **413**. Kluwer, Dordrecht. MR1471479

KANTOROVIČ, L. V. and RUBINŠTEĬN, G. Š. (1958). On a space of completely additive functions. *Vestnik Leningrad*. *Univ*. **13** 52–59. MR0102006

LALLEY, S. P. and ZHENG, X. (2011). Occupation statistics of critical branching random walks in two or higher dimensions. *Ann*. *Probab*. **39** 327–368.

LYONS, R., PEMANTLE, R. and PERES, Y. (1995). Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann*. *Probab*. **23** 1125–1138. MR1349164

PAKES, A. G. and KHATTREE, R. (1992). Length-biasing, characterizations of laws and the moment problem. *Austral*. *J*. *Statist*. **34** 307–322. MR1193781

PEKÖZ, E. A. (1996). Stein's method for geometric approximation. *J*. *Appl*. *Probab*. **33** 707–713. MR1401468

RÉNYI, A. (1957). A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **1** 519–527. MR0094861

ROSS, S. and PEKÖZ, E. (2007). *A Second Course in Probability*. Available at www.ProbabilityBookstore.com.

SENGUPTA, D., CHATTERJEE, A. and CHAKRABORTY, B. (1995). Reliability bounds and other inequalities for discrete life distributions. *Microelectronics and Reliability* **35** 1473–1478.

SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. Springer, New York. MR2265633

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab. Vol. II: Probability Theory* 583–602. Univ. California Press, Berkeley, CA. MR0402873

SUGAKOVA, E. V. (1995). Estimates in the Rényi theorem for differently distributed summands. *Ukraïn. Mat. Zh.* **47** 984–989. MR1367955

VALLANDER, S. (1973). Calculations of the Wasserstein distance between probability distributions on the line. *Teory Probab. Appl.* **18** 784–786. MR0328982

WEINBERG, G. V. (2005). Stein's method and its application in radar signal processing. DSTO Formal Reports (DSTO-TR-1735). Electronic Warfare & Radar Division, Dept. Defence (Australian Goverment). Available at dspace.dsto.defence.gov.au/dspace/handle/1947/4064.

YAGLOM, A. M. (1947). Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR* (*N.S.*) **56** 795–798. MR0022045

SCHOOL OF MANAGEMENT
BOSTON UNIVERSITY
595 COMMONWEALTH AVENUE
BOSTON, MASSACHUSETTS 02215
E-MAIL: pekoz@bu.edu

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
6 SCIENCE DRIVE 2
SINGAPORE 117546
E-MAIL: staar@nus.edu.sg