Bernoulli **19**(2), 2013, 610–632
DOI: 10.3150/11-BEJ406

Total variation error bounds for geometric approximation

EROL A. PEKÖZ[1], ADRIAN RÖLLIN[2] and NATHAN ROSS[3]

[1]*School of Management, Boston University, 595 Commonwealth Avenue Boston, MA 02215, USA. E-mail: pekoz@bu.edu*
[2]*Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore. E-mail: adrian.roellin@nus.edu.sg*
[3]*Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, USA. E-mail: ross@stat.berkeley.edu*

We develop a new formulation of Stein's method to obtain computable upper bounds on the total variation distance between the geometric distribution and a distribution of interest. Our framework reduces the problem to the construction of a coupling between the original distribution and the "discrete equilibrium" distribution from renewal theory. We illustrate the approach in four non-trivial examples: the geometric sum of independent, non-negative, integer-valued random variables having common mean, the generation size of the critical Galton–Watson process conditioned on non-extinction, the in-degree of a randomly chosen node in the uniform attachment random graph model and the total degree of both a fixed and randomly chosen node in the preferential attachment random graph model.

Keywords: discrete equilibrium distribution; geometric distribution; preferential attachment model; Stein's method; Yaglom's theorem

1. Introduction

The exponential and geometric distributions are convenient and accurate approximations in a wide variety of complex settings involving rare events, extremes and waiting times. The difficulty in obtaining explicit error bounds for these approximations beyond elementary settings is discussed in the preface of [1], where the author also points out a lack of such results. Recently, Peköz and Röllin [18] developed a framework to obtain error bounds for the Kolmogorov and Wasserstein distance metrics between the exponential distribution and a distribution of interest. The main ingredients there are Stein's method (see [23,24] for introductions) along with the equilibrium distribution from renewal theory. Due to the flexibility of Stein's method and the close connection between the exponential and geometric distributions, it is natural to attempt to use similar techniques to obtain bounds for the stronger total variation distance metric between the geometric distribution and an integer supported distribution.

Our formulation rests on the idea that a positive, integer-valued random variable W will be approximately geometrically distributed with parameter $p = 1/\mathbb{E}W$ if $\mathscr{L}(W) \approx \mathscr{L}(W^e)$, where W^e has the (discrete) equilibrium distribution with respect to W, defined by

$$\mathbb{P}(W^e = k) = \frac{1}{\mathbb{E}W}\mathbb{P}(W \geq k), \qquad k = 1, 2, \ldots. \tag{1.1}$$

This distribution arises in discrete-time renewal theory as the time until the next renewal when the process is stationary, and the transformation which maps a distribution to its equilibrium distribution has the geometric distribution with positive support as its unique fixed point. Our main result is an upper bound on the variation distance between the distribution of W and a geometric distribution with parameter $(\mathbb{E}W)^{-1}$, in terms of a coupling between the random variables W^e and W.

This setup is closely related to the exponential approximation formulation of [18] and also [11], which is also related to the zero-bias transformation of [12]. A difficulty in pushing the results of [18] through to the stronger total variation metric is that the support of the distribution to be approximated may not match the support of the geometric distribution well enough. This issue is typical in bounding the total variation distance between integer-valued random variables and can be handled by introducing a term into the bound that quantifies the "smoothness" of the distribution of interest; see, for example, [2,21,22]. To illustrate this point, we apply our abstract formulation to obtain total variation error bounds in two of the examples treated in [18].

It is also important to note that geometric approximation can be appropriate in situations where exponential approximation is not; for example if a sequence of random variables has the geometric distribution with fixed parameter as its distributional limit. Thus, we will also apply our theory in two examples (discussed in more detail immediately below) that fall into this category. In these applications, the "smoothness" term in our bounds does not play a part. The same effect can be observed when comparing translated Poisson and Poisson approximation; see [21] and [4].

The first application in this article is a bound on the total variation distance between the geometric distribution and the sum of a geometrically distributed number of independent, non-negative, integer-valued random variables with common mean. The distribution of such geometric convolutions have been considered in many places in the literature in the setting of exponential approximation and convergence; the book-length treatment is given in [13]. The second application is a variation on the classical theorem of Yaglom [25], describing the asymptotic behavior of the generation size of a critical Galton–Watson process conditioned on non-extinction. This theorem has a large literature of extensions and embellishments; see, for example, [14]. Peköz and Röllin [18] obtained a rate of convergence for the Kolmogorov distance between the generation size of a critical Galton–Watson process conditioned on non-extinction and the exponential distribution. Here we obtain an analogous bound for the geometric distribution in total variation distance. The third application is to the in-degree of a randomly chosen node in the uniform attachment random graph discussed in [5], and the final application is to the total degree of both a fixed and a randomly chosen node in the preferential attachment random graph discussed in [5]. As mentioned before, these examples do not derive from an exponential approximation result.

Finally, we mention that there are other formulations of geometric approximation using Stein's method. For example, Peköz [17] and Barbour and Grübel [3] use the intuition that a positive, integer-valued random variable W approximately has a geometric distribution with parameter $p = \mathbb{P}(W = 1)$ if

$$\mathscr{L}(W) \approx \mathscr{L}(W - 1|W > 1).$$

Other approaches can be found in [20] and [9].

The organization of this article is as follows. In Section 2 we present our main theorems, and Sections 3, 4, 5, and 6, respectively, contain applications to geometric sums, the critical Galton–

Watson process conditioned on non-extinction, the uniform attachment random graph model and the preferential attachment random graph model.

2. Main results

A typical issue when discussing the geometric distribution is whether to have the support begin at zero or one. Denote by $\mathrm{Ge}(p)$ the geometric distribution with positive support; that is, $\mathscr{L}(Z) = \mathrm{Ge}(p)$ if $\mathbb{P}(Z = k) = (1 - p)^{k-1}p$ for positive integers k. Alternatively, denote by $\mathrm{Ge}^0(p)$ the geometric distribution $\mathrm{Ge}(p)$ shifted by minus one, that is, "starting at 0." Since $\mathscr{L}(Z) = \mathrm{Ge}(p)$ implies $\mathscr{L}(Z - 1) = \mathrm{Ge}^0(p)$, it is typical that results for one of $\mathrm{Ge}(p)$ or $\mathrm{Ge}^0(p)$ easily pass to the other. Unfortunately, our methods do not appear to trivially transfer between these two distributions, so we are forced to develop our theory for both cases in parallel.

First, we give an alternate definition of the equilibrium distribution that we will use in the proof of our main result.

Definition 2.1. *Let X be a positive, integer-valued random variable with finite mean. We say that an integer-valued random variable X^e has the discrete equilibrium distribution w.r.t. X if, for all bounded f and $\nabla f(x) = f(x) - f(x - 1)$, we have*

$$\mathbb{E}f(X) - f(0) = \mathbb{E}X\mathbb{E}\nabla f(X^e). \tag{2.1}$$

Remark 2.2. To see how (2.1) is equivalent to (1.1), note that we have

$$\mathbb{E}f(X) - f(0) = \mathbb{E}\sum_{i=1}^{X}\nabla f(i) = \sum_{i=1}^{\infty}\nabla f(i)\mathbb{P}(X \geq i) = \mathbb{E}X\mathbb{E}\nabla f(X^e).$$

Note that Definition 2.1 and (1.1) do not actually require positive support of W so that we can define W^e for a non-negative random variable which is not identically zero (and our results below still hold in this case). However, we will use W^e to compare the distribution of W to a geometric distribution with positive support, so the assumption that $W > 0$ is not restrictive in most cases of interest. In order to handle geometric approximation with support on the non-negative integers, we could shift the non-negative random variable W by 1 and then consider geometric approximation on the positive integers, but this strategy is not practical since $(W + 1)^e$ is typically inconvenient to work with. Fortunately, developing an analogous theory for non-negative random variables with mass at zero is no more difficult than that for positive random variables.

Definition 2.3. *If X is a non-negative, integer-valued random variable with $\mathbb{P}(X = 0) > 0$, we say that an integer-valued random variable X^{e_0} has the discrete equilibrium distribution w.r.t. X if, for all bounded f and with $\Delta f(x) = f(x + 1) - f(x)$, we have*

$$\mathbb{E}f(X) - f(0) = \mathbb{E}X\mathbb{E}\Delta f(X^{e_0}).$$

It is not difficult the see that $W^{e_0} \stackrel{\mathscr{D}}{=} W^e - 1$. Note that we are defining the term "discrete equilibrium distribution" in both of the previous definitions, but this should not cause confusion as the support of the base distribution dictates the meaning of the terminology.

Besides the total variation metric, we will also give bounds on the local metric

$$d_{\mathrm{loc}}(\mathscr{L}(U), \mathscr{L}(V)) := \sup_{m \in \mathbb{Z}} |\mathbb{P}(U = m) - \mathbb{P}(V = m)|.$$

It is clear that d_{loc} will be less than or equal to $\sup_m [\mathbb{P}(U = m) \vee \mathbb{P}(V = m)]$, so that typically better rates need to be obtained in order to provide useful information in this metric.

As a final bit of notation, before the statement of our main results, for a function g with domain $\mathbb{Z}$, let $\|g\| = \sup_{k \in \mathbb{Z}} |g(k)|$, and for any integer-valued random variable W and any σ-algebra $\mathcal{F}$, define the conditional smoothness

$$S_1(W|\mathcal{F}) = \sup_{\|g\| \le 1} |\mathbb{E}\{\Delta g(W)|\mathcal{F}\}| = 2 d_{\mathrm{TV}}\big(\mathscr{L}(W+1|\mathcal{F}), \mathscr{L}(W|\mathcal{F})\big), \tag{2.2}$$

and the second order conditional smoothness

$$S_2(W|\mathcal{F}) = \sup_{\|g\| \le 1} |\mathbb{E}\{\Delta^2 g(W)|\mathcal{F}\}|,$$

where $\Delta^2 g(k) = \Delta g(k+1) - \Delta g(k)$. In order to simplify the presentation of the main theorems, we let $d_1 = d_{\mathrm{TV}}$ and $d_2 = d_{\mathrm{loc}}$. Also, let I_A denote the indicator random variable of an event A.

Theorem 2.1. *Let W be a positive, integer-valued random variable with $\mathbb{E}W = 1/p$ for some $0 < p \le 1$, and let W^e have the discrete equilibrium distribution w.r.t. W. Then, with $D = W - W^e$, any σ-algebra $\mathcal{F} \supseteq \sigma(D)$ and event $A \in \mathcal{F}$, we have*

$$d_l(\mathscr{L}(W), \mathrm{Ge}(p)) \le \mathbb{E}\{|D| S_l(W|\mathcal{F}) I_A\} + 2\mathbb{P}(A^c) \tag{2.3}$$

for $l = 1, 2$, and

$$d_{\mathrm{TV}}(\mathscr{L}(W^e), \mathrm{Ge}(p)) \le p\mathbb{E}|D|, \tag{2.4}$$

$$d_{\mathrm{loc}}(\mathscr{L}(W^e), \mathrm{Ge}(p)) \le p\mathbb{E}\{|D| S_1(W|\mathcal{F})\}; \tag{2.5}$$

on the RHS of (2.3) and (2.5), $S_l(W|\mathcal{F})$ can be replaced by $S_l(W^e|\mathcal{F})$.

Theorem 2.2. *Let W be a non-negative, integer-valued random variable with $\mathbb{P}(W = 0) > 0$, $\mathbb{E}W = (1-p)/p$ for some $0 < p \le 1$, and let W^{e_0} have the discrete equilibrium distribution w.r.t. W. Then, with $D = W - W^{e_0}$, any σ-algebra $\mathcal{F} \supseteq \sigma(D)$ and event $A \in \mathcal{F}$, we have*

$$d_l(\mathscr{L}(W), \mathrm{Ge}^0(p)) \le (1-p)\mathbb{E}\{|D| S_l(W|\mathcal{F}) I_A\} + 2(1-p)\mathbb{P}(A^c) \tag{2.6}$$

for $l = 1, 2$, and

$$d_{\mathrm{TV}}(\mathscr{L}(W^{e_0}), \mathrm{Ge}^0(p)) \le p\mathbb{E}|D|, \tag{2.7}$$

$$d_{\mathrm{loc}}(\mathscr{L}(W^{e_0}), \mathrm{Ge}^0(p)) \le p\mathbb{E}\{|D| S_1(W|\mathcal{F})\}, \tag{2.8}$$

on the RHS of (2.6) *and* (2.8), $S_l(W|\mathcal{F})$ *can be replaced by* $S_l(W^{e_0}|\mathcal{F})$.

Before we prove Theorems 2.1 and 2.2, we make a few remarks related to these results.

Remark 2.4. It is easy to see that the a random variable W with law equal to $\mathrm{Ge}(p)$ has the property that $\mathscr{L}(W) = \mathscr{L}(W^e)$, so that W^e can be taken to be W and the theorem yields the correct error term in this case. The analogous statement is true for $\mathrm{Ge}^0(p)$ and W^{e_0}.

Remark 2.5. By choosing $A = \{W = W^{e_0}\}$ in (2.6) with $l = 1$, we find

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Ge}^0(p)) \le 2(1-p)\mathbb{P}(W \ne W^{e_0}), \tag{2.9}$$

and an analogous corollary holds for Theorem 2.1.

In order to use the theorem we need to be able to construct random variables with the discrete equilibrium distribution. The next proposition provides such a construction for a non-negative, integer-valued random variable W. We say W^s has the size-bias distribution of W, if

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

for all f for which the expectation exists.

Proposition 2.3. *Let W be an integer-valued random variable, and let W^s have the size-bias distribution of W.*

1. *If $W > 0$ and we define the random variable W^e such that conditional on W^s, W^e has the uniform distribution on the integers $\{1, 2, \dots, W^s\}$, then W^e has the discrete equilibrium distribution w.r.t. W.*
2. *If $W \ge 0$ with $\mathbb{P}(W = 0) > 0$, and we define the random variable W^{e_0} such that conditional on W^s, W^{e_0} has the uniform distribution on the integers $\{0, 1, \dots, W^s - 1\}$, then W^{e_0} has the discrete equilibrium distribution w.r.t. W.*

Proof. For any bounded f we have

$$\mathbb{E}f(W) - f(0) = \mathbb{E}\sum_{i=1}^{W} \nabla f(i) = \mathbb{E}W\mathbb{E}\left\{\frac{1}{W^s}\sum_{i=1}^{W^s}\nabla f(i)\right\} = \mathbb{E}W\mathbb{E}\nabla f(W^e),$$

which implies Item 1. The second item is proved analogously. □

As mentioned in the Introduction, there can be considerable technical difficulty in ensuring the support of the distribution to be approximated is smooth. In Theorems 2.1 and 2.2 this issue is accounted for in the term $S_1(W|\mathcal{F})$. Typically, our strategy to bound this term will be to write W (or W^e) as a sum of terms which are independent given $\mathcal{F}$ and then apply the following lemma from [16], Corollary 1.6.

Lemma 2.4 ([16], Corollary 1.6). *If $X_1, \dots, X_n$ are independent, integer-valued random variables and*

$$u_i = 1 - d_{\mathrm{TV}}\big(\mathscr{L}(X_i), \mathscr{L}(X_i + 1)\big),$$

then

$$d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{i=1}^{n} X_i\right), \mathscr{L}\left(1 + \sum_{i=1}^{n} X_i\right)\right) \le \sqrt{\frac{2}{\pi}}\left(\frac{1}{4} + \sum_{i=1}^{n} u_i\right)^{-1/2}.$$

Before we present the proof of Theorems 2.1 and 2.2, we must first develop the Stein method machinery we will need. As in [17], for any subset B of the integers and any $p = 1 - q$, we construct the function $f = f_{B,p}$ defined by $f(0) = 0$ and for $k \ge 1$,

$$qf(k) - f(k-1) = I_{k \in B} - \mathrm{Ge}(p)\{B\}, \tag{2.10}$$

where $\mathrm{Ge}(p)\{B\} = \sum_{i \in B}(1-p)^{i-1}p$ is the chance that a positive, geometric random variable with parameter p takes a value in the set B. It can be easily verified that the solution of (2.10) is given by

$$f(k) = \sum_{i \in B} q^{i-1} - \sum_{i \in B, i \ge k+1} q^{i-k-1}. \tag{2.11}$$

Equivalently, for $k \ge 0$,

$$qf(k+1) - f(k) = I_{k \in B-1} - \mathrm{Ge}^0(p)\{B-1\},$$

where we define $\mathrm{Ge}^0(p)\{B\}$ analogously to $\mathrm{Ge}(p)\{B\}$.

Peköz [17] and Daly [8] study properties of these solutions, but we need the following additional lemma to obtain our main result.

Lemma 2.5. *For f as above, we have*

$$\sup_{k \ge 1} |\nabla f(k)| = \sup_{k \ge 0} |\Delta f(k)| \le 1. \tag{2.12}$$

If, in addition, $B = \{m\}$ for some $m \in \mathbb{Z}$, then

$$\sup_{k \ge 0} |f(k)| \le 1. \tag{2.13}$$

Proof. To show (2.12), note that

$$\begin{aligned}\nabla f(k) &= \sum_{i \in B, i \ge k} q^{i-k} - \sum_{i \in B, i \ge k+1} q^{i-k-1} \\ &= I_{k \in B} + \sum_{i \in B, i \ge k+1} (q^{i-k} - q^{i-k-1}) = I_{k \in B} - p \sum_{i \in B, i \ge k+1} q^{i-k-1},\end{aligned}$$

thus $-1 \le \nabla f(k) \le 1$. If now $B = \{m\}$, (2.13) is immediate from (2.11). □

Proof of Theorem 2.1. Given any positive, integer-valued random variable W with $\mathbb{E}W = 1/p$ and $D = W - W^e$ we have, using (2.10), Definition 2.1, and Lemma 2.5 in the two inequalities,

$$\begin{aligned}
&\mathbb{P}(W \in B) - \mathrm{Ge}(p)\{B\} \\
&\quad = \mathbb{E}\{qf(W) - f(W-1)\} \\
&\quad = \mathbb{E}\{\nabla f(W) - pf(W)\} \\
&\quad = \mathbb{E}\{\nabla f(W) - \nabla f(W^e)\} \\
&\quad \le \mathbb{E}\{I_A(\nabla f(W) - \nabla f(W^e))\} + 2\mathbb{P}(A^c) \\
&\quad = \mathbb{E}\left\{I_A I_{D>0} \sum_{i=0}^{D-1} \mathbb{E}(\nabla f(W^e + i + 1) - \nabla f(W^e + i)|\mathcal{F})\right\} \\
&\qquad + \mathbb{E}\left\{I_A I_{D<0} \sum_{i=0}^{-D-1} \mathbb{E}(\nabla f(W^e - i - 1) - \nabla f(W^e - i)|\mathcal{F})\right\} + 2\mathbb{P}(A^c) \\
&\quad \le \mathbb{E}\{|D| S_1(W^e|\mathcal{F}) I_A\} + 2\mathbb{P}(A^c),
\end{aligned}$$

which is (2.3) for $l = 1$; analogously, one can obtain (2.3) with $S_1(W|\mathcal{F})$ in place of $S_1(W^e|\mathcal{F})$ on the RHS. In the case of $B = \{m\}$ we can make use of (2.13) to obtain

$$|\mathbb{P}(W = m) - \mathrm{Ge}(p)\{m\}| \le \mathbb{E}\{|D| S_2(W^e|\mathcal{F})\mathrm{I}_A\} + 2\mathbb{P}(A^c),$$

instead, which proves (2.3) for $l = 2$. For (2.4), we have that

$$\begin{aligned}
\mathbb{P}(W^e \in B) - \mathrm{Ge}(p)\{B\} &= \mathbb{E}\{qf(W^e) - f(W^e - 1)\} \\
&= \mathbb{E}\{\nabla f(W^e) - pf(W^e)\} = p\mathbb{E}\{f(W) - f(W^e)\} \le p\mathbb{E}|D|,
\end{aligned}$$

where the last line follows by writing $f(W) - f(W^e)$ as a telescoping sum of $|D|$ terms no greater than $\|\nabla f\|$, which can be bounded using (2.12); (2.5) is straightforward using (2.13) and (2.2). □

Proof of Theorem 2.2. Let W be a non-negative, integer-valued random variable with $\mathbb{E}W = (1-p)/p$ and W^{e_0} as in the theorem. If we define I to be independent of all else and such that $\mathbb{P}(I = 1) = 1 - \mathbb{P}(I = 0) = p$, then a short calculation shows that the variable defined by

$$[(W+1)^e | I = 1] = W + 1 \quad \text{and} \quad [(W+1)^e | I = 0] = W^{e_0} + 1$$

has the positive discrete equilibrium transform with respect to $W + 1$. Equation (2.6) now follows after noting that

$$d_l(\mathscr{L}(W), \mathrm{Ge}^0(p)) = d_l(\mathscr{L}(W+1), \mathrm{Ge}(p)),$$

and then applying the following stronger variation of (2.3) which is easily read from the proof of Theorem 2.1:

$$\begin{aligned} &d_l\big(\mathscr{L}(W+1), \mathrm{Ge}(p)\big) \\ &\quad \le \mathbb{E}\{|W+1-(W+1)^e|S_l(W|\mathcal{F})I_A\} + 2\mathbb{P}\big((W+1)^e \neq W+1\big)\mathbb{P}(A^c). \end{aligned}$$

Equations (2.7) and (2.8) can be proved in a manner similar to their analogs in Theorem 2.1. □

3. Applications to geometric sums

In this section we apply the results above to a sum of the geometric number of independent but not necessarily identically distributed random variables. As in our theory above, we will have separate results for the two cases where the sum is strictly positive and the case where it can take on the value zero with positive probability. We reiterate that although there are a variety of exponential approximation results in the literature for this example, there do not appear to be bounds available for the analogous geometric approximation in the total variation metric.

Theorem 3.1. *Let $X_1, X_2, \dots$ be a sequence of independent, square integrable, positive and integer-valued random variables, such that, for some $u > 0$, we have, for all $i \ge 1$, $\mathbb{E}X_i = \mu$ and $u \le 1 - d_{\mathrm{TV}}(\mathscr{L}(X_i), \mathscr{L}(X_i+1))$. Let $\mathscr{L}(N) = \mathrm{Ge}(a)$ for some $0 < a \le 1$ and $W = \sum_{i=1}^N X_i$. Then with $p = 1-q = a/\mu$, we have*

$$d_l(\mathscr{L}(W), \mathrm{Ge}(p)) \le C_l \sup_{i\ge 1} \mathbb{E}|X_i - X_i^e| \le C_l\left(\mu_2/2 + \frac{1}{2} + \mu\right) \tag{3.1}$$

for $l = 1, 2$, where $\mu_2 := \sup_i \mathbb{E}X_i^2$ and

$$\begin{aligned} C_1 &= \min\left\{1, a\left[1 + \left(-\frac{2}{u\log(1-a)}\right)^{1/2}\right]\right\}, \\ C_2 &= \min\left\{1, a\left[1 - \frac{6\log(a)}{\pi u}\right]\right\}. \end{aligned}$$

Theorem 3.2. *Let $X_1, X_2, \dots$ be a sequence of independent, square integrable, non-negative and integer-valued random variables, such that for some $u > 0$ we have, for all $i \ge 1$, $\mathbb{E}X_i = \mu$ and $u \le 1 - d_{\mathrm{TV}}(\mathscr{L}(X_i), \mathscr{L}(X_i+1))$. Let $\mathscr{L}(M) = \mathrm{Ge}^0(a)$ for some $0 < a \le 1$ and $W = \sum_{i=1}^M X_i$. Then with $p = 1-q = a/(a+\mu(1-a))$, we have*

$$d_l(\mathscr{L}(W), \mathrm{Ge}^0(p)) \le C_l \sup_{i\ge 1} \mathbb{E}X_i^{e_0} \le C_l\left(\mu_2/(2\mu) - \frac{1}{2}\right) \tag{3.2}$$

for $l = 1, 2$, where $\mu_2 := \sup_i \mathbb{E}X_i^2$ and the C_l are as in Theorem 3.1.

Remark 3.1. The first inequality in (3.1) yields the correct bound of zero when X_i is geometric, since in this case we would have $X_i = X_i^e$; see Remark 2.4 following Theorem 2.1. Similarly, in the case where the X_i have a Bernoulli distribution with expectation μ, we have that $X^{e_0} = 0$ so that the left-hand side of (3.2) is zero. That is, if $M \sim \mathrm{Ge}^0(a)$ and conditional on M, W has the binomial distribution with parameters M and μ for some $0 \le \mu \le 1$, then $W \sim \mathrm{Ge}^0(a/(a + \mu(1-a)))$.

Remark 3.2. In the case where X_i are i.i.d. but not necessarily integer valued and $0 < a \le \frac{1}{2}$, Brown [6], Theorem 2.1, obtains the exponential approximation result

$$d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathrm{Exp}(1/p)\bigr) \le \frac{a\mu_2}{\mu} \tag{3.3}$$

for the weaker Kolmogorov metric. To compare (3.3) with (3.1) for small a, we observe that bound (3.1) is linear in a whereas (3.3), within a constant factor, behaves like $a(-\log(1-a))^{-1/2} \sim \sqrt{a}$. Therefore bound (3.3) is better, but (3.1) applies to non-i.i.d. random variables (albeit having identical means) and to the stronger total variation metric.

Proof of Theorem 3.1. First, let us prove that $W^e := \sum_{i=1}^{N-1} X_i + X_N^e$ has the discrete equilibrium distribution w.r.t. W, where, for each $i \ge 1$, X_i^e is a random variable having the equilibrium distribution w.r.t. X_i, independent of all else. Note first that we have, for bounded f and every m,

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

Note also that since N is geometric, for any bounded function g with $g(0) = 0$, we have $\mathbb{E}\{g(N) - g(N-1)\} = a\mathbb{E}g(N)$. We now assume that $f(0) = 0$. Hence, using the above two facts and independence between N and the sequence $X_1, X_2, \ldots,$ we have

$$\begin{aligned}\mathbb{E}W\mathbb{E}\nabla f(W^e) &= \frac{\mu}{a}\mathbb{E}\nabla f\left(\sum_{i=1}^{N-1} X_i + X_N^e\right)\\ &= \frac{1}{a}\mathbb{E}f\left[\left(\sum_{i=1}^{N} X_i\right) - f\left(\sum_{i=1}^{N-1} X_i\right)\right] = \mathbb{E}f\left(\sum_{i=1}^{N} X_i\right) = \mathbb{E}f(W).\end{aligned}$$

Now, $D = W - W^e = X_N - X_N^e$ and setting $\mathcal{F} = \sigma(N, X_N^e, X_N)$, we have

$$\begin{aligned}S_1(W^e|\mathcal{F}) = S_1\left(\sum_{i=1}^{N-1} X_i \middle| \mathcal{F}\right) &\le 1 \wedge \left(\frac{2}{\pi(0.25 + (N-1)u)}\right)^{1/2}\\ &\le 1 \wedge \left(\frac{2}{\pi(N-1)u}\right)^{1/2},\end{aligned} \tag{3.4}$$

where we have used Lemma 2.4 and the fact that $S_1(W^e|\mathcal{F})$ is almost surely bounded by one. We now have

$$\mathbb{E}[|D|S_1(W^e|\mathcal{F})] \le \mathbb{E}\left[\left(1 \wedge \left(\frac{2}{\pi(N-1)u}\right)^{1/2}\right)\mathbb{E}^N|X_N - X_N^e|\right]. \tag{3.5}$$

From here, we can obtain the first inequality in (3.1) by applying Theorem 2.1, after noting that

$$\mathbb{E}^N|X_N - X_N^e| \le \sup_{i\ge 1}\mathbb{E}|X_i - X_i^e|, \tag{3.6}$$

and

$$\begin{aligned}
\mathbb{E}\left(1 \wedge \left(\frac{2}{\pi(N-1)u}\right)^{1/2}\right) &\le 1 \wedge \left(a + \left(\frac{2}{\pi u}\right)^{1/2}\sum_{i\ge 1}\frac{a(1-a)^i}{i^{1/2}}\right)\\
&\le 1 \wedge \left(a + a\left(\frac{2}{\pi u}\right)^{1/2}\left(-\frac{\pi}{\log(1-a)}\right)^{1/2}\right)\\
&= 1 \wedge \left(a\left[1 + \left(-\frac{2}{\log(1-a)u}\right)^{1/2}\right]\right),
\end{aligned}$$

where we have used

$$\sum_{i\ge 1}\frac{a(1-a)^{i-1}}{i^{1/2}} \le \frac{a}{1-a}\int_0^\infty \frac{(1-a)^x}{x^{1/2}}\,\mathrm{d}x = \frac{a}{1-a}\left(-\frac{\pi}{\log(1-a)}\right)^{1/2}.$$

The second inequality in (3.1) follows from Theorem 2.1 and the fact (from the definition of the transformation X^e) that $\mathbb{E}^N|X_N - X_N^e| \le \mu_2/2 + \frac{1}{2} + \mu$.

To obtain the local limit result, note that, if $V = X + Y$ is the sum of two independent random variables, then $S_2(V) \le S_1(X)S_1(Y)$. Hence,

$$S_2(W^e|\mathcal{F}) \le 1 \wedge \frac{2}{\pi(0.25 + (N/2-1)_+ u)} \le 1 \wedge \frac{6}{\pi(N-1)u}.$$

From here we have

$$\mathbb{E}S_2(W^e|\mathcal{F}) \le 1 \wedge \left(a + \frac{6}{\pi u}\sum_{i\ge 1}\frac{a(1-a)^i}{i}\right) = 1 \wedge \left(a - \frac{6a\log(a)}{\pi u}\right).$$

□

Proof of Theorem 3.2. It is straightforward to check that $W^{e_0} := \sum_{i=1}^M X_i + X_{M+1}^{e_0}$ has the equilibrium distribution with respect to W. Now, $D = W - W^{e_0} = -X_{M+1}^{e_0}$ and setting $\mathcal{F} = \sigma(M, X_{M+1}^{e_0})$, we have

$$S_1(W^{e_0}|\mathcal{F}) = S_1\left(\sum_{i=1}^M X_i \middle| \mathcal{F}\right),$$

which can be bounded above by (3.4), as in the proof of Theorem 3.1. The remainder of the proof follows closely to that of Theorem 3.1. For example, the expression analogous to (3.5) is

$$\mathbb{E}[|D|S_1(W^{e_0}|\mathcal{F})] \le \mathbb{E}\biggl[\min\biggl\{1, \frac{\sqrt{2}}{(\pi M u)^{1/2}}\biggr\}\mathbb{E}^M X_M^{e_0}\biggr],$$

and the definition of the transform X^{e_0} implies that

$$\mathbb{E}X_i^{e_0} = \frac{\mathbb{E}X_i^2}{2\mu} - \frac{1}{2}. \qquad \square$$

4. Application to the critical Galton–Watson branching process

Let $Z_0 = 1, Z_1, Z_2, \ldots$ be a Galton–Watson branching process with offspring distribution $\mathscr{L}(Z_1)$. A theorem due to Yaglom [25] 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 $\sigma^2/2$. The recent article [18] is the first to give an explicit bound on the rate of convergence for this asymptotic result. Using ideas from there, we give a convergence rate for the total variation error of a geometric approximation to Z_n under finite third moment of the offspring distribution and the natural periodicity requirement that

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(Z_1), \mathscr{L}(Z_1+1)\bigr) < 1. \tag{4.1}$$

This type of smoothness condition is typical in the context of Stein's method for approximation by a discrete distribution; see, for example, [2] and [22].

For the proof of the following theorem, we make use the of construction of Lyons *et al.* [15]; we refer to that article for more details on the construction and only present what is needed for our purpose.

Theorem 4.1. *For a critical Galton–Watson branching process with offspring distribution $\mathscr{L}(Z_1)$, such that $\mathbb{E}Z_1^3 < \infty$ and (4.1) hold, we have*

$$d_{\mathrm{TV}}\biggl(\mathscr{L}(Z_n|Z_n > 0), \mathrm{Ge}\biggl(\frac{2}{\sigma^2 n}\biggr)\biggr) \le \frac{C \log n}{n^{1/4}} \tag{4.2}$$

for some constant C which is independent of n.

Remark 4.1. From [18], Theorem 4.1, we have

$$d_{\mathrm{K}}\bigl(\mathscr{L}\bigl(2Z_n/(\sigma^2 n)|Z_n > 0\bigr), \mathrm{Exp}(1)\bigr) \le C\biggl(\frac{\log n}{n}\biggr)^{1/2} \tag{4.3}$$

without condition (4.1) for the weaker Kolmogorov metric. It can be seen that the bound in (4.2) is not as good as the bound in (4.3) for large n, but (4.2) applies to the stronger total variation metric.

Proof of Theorem 4.1. First we construct a size-biased branching tree as in [15]. 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. 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 $\mathscr{L}(Z_1)$. 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 $\mathscr{L}(Z_1)$. For v_1 proceed as for v_0, that is, give it a size-biased number of offspring, pick one 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" $v_0, v_1, v_2, \ldots$ is an infinite sequence where v_i is an individual (or particle) in generation i.

We now need some notation. Denote by S_n the total number of particles in generation n. Denote by L_n and R_n the number of particles to the left (excluding v_n) and to the right (including v_n), 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 $R_{n,j}$, be the number of particles in generation n that stem from the siblings to the left and right 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}.

Let now $R'_{n,j}$ 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

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

Define also $R^*_n = 1 + \sum_{j=1}^{n} R^*_{n,j}$. Let us collect a few facts from [18] which we will then use to give the proof of the theorem (here and in the rest of the proof, C shall denote a constant which is independent of n, but may depend on $\mathscr{L}(Z_1)$ and may also be different from formula to formula):

(i) $\mathscr{L}(R^*_n) = \mathscr{L}(Z_n | Z_n > 0)$;
(ii) S_n has the size-biased distribution of Z_n, and v_n is equally likely to be any of the S_n particles;
(iii) $\mathbb{E}\{R'_{n,j} I_{A^c_{n,j}}\} \leq \sigma^2 \mathbb{P}[A^c_{n,j}]$;
(iv) $\mathbb{E}\{R_{n,j} I_{A^c_{n,j}}\} \leq \gamma \mathbb{P}[A^c_{n,j}]$, and $\mathbb{E}\{R_{n-1,j} I_{A^c_{n,j}}\} \leq \gamma \mathbb{P}[A^c_{n,j}]$, where $\gamma = \mathbb{E} Z_1^3$;
(v) $\mathbb{P}[A^c_{n,j}] \leq \sigma^2 \mathbb{P}[Z_{n-j} > 0] \leq C/(n - j + 1)$ for some $C > 0$.

In light of (i) and (ii) (and then using the construction in Proposition 2.3 to see that R_n has the discrete equilibrium distribution w.r.t $\mathscr{L}(R^*_n)$) we can let $W = R^*_n$, $W^e = R_n$ and

$$D = R^*_n - R_n = \sum_{j=1}^{n} (R'_{n,j} - R_{n,j}) I_{A^c_{n,j}}.$$

Also let

$$N = \sum_{j=1}^{n-1} R_{n-1,j} I_{A^c_{n,j}} \quad \text{and} \quad M = \sum_{j=1}^{n} R_{n,j} I_{A^c_{n,j}},$$

and note that (iii)–(v) give

$$\mathbb{E}|D| \le C \log n \quad \text{and} \quad \mathbb{E}N \le C \log n. \tag{4.4}$$

Next with $\mathcal{F} = \sigma(N, D, R_{n-1}, M, R_{n,n} I_{A_{n,n}})$ and, letting Z_1^i, $i = 1, 2, \dots,$ be i.i.d. copies of Z_1, we have

$$\mathscr{L}(R_n - M - R_{n,n} I_{A_{n,n}} - 1|\mathcal{F}) = \mathscr{L}\left(\sum_{i=1}^{R_{n-1}-N} Z_1^i \,\middle|\, R_{n-1}, N\right),$$

which follows since $R_n - M = 1 + \sum_{i=1}^n R_{n,j} I_{A_{n,j}}$, and the particles counted by $R_{n-1} - N$ will be parents of the particles counted by $R_n - M - 1 + R_{n,n} I_{A_{n,n}}$.

Then we use Lemma 2.4 to obtain

$$S_1(W^e|\mathcal{F}) = S_1(R_n - M - R_{n,n} I_{A_{n,n}} - 1|\mathcal{F}) \le \frac{0.8}{(0.25 + (R_{n-1} - N)u)^{1/2}}. \tag{4.5}$$

As a direct corollary of (2.4), for any bounded function f, we have

$$\mathbb{E}f(W^e) \le \mathbb{E}f(X_p) + p\|f\|\mathbb{E}|W^e - W|, \tag{4.6}$$

where $X_p \sim \mathrm{Ge}(p)$. Fix $q = 1/\mathbb{E}[Z_{n-1}|Z_{n-1} > 0]$, $k = q^{-1/4}$, and let

$$A = \{N \le k, |D| \le k, R_{n-1} > 2k\},$$

and

$$f(x) = (x-k)^{-1/2} I_{x \ge 2k+1}.$$

Using (4.4), (4.5), (4.6) and the fact that $\|f\| \le k^{-1/2}$, we find

$$\mathbb{E}[f(X_q)] \le q \sum_{j=1}^{\infty} \frac{(1-q)^j}{j^{1/2}} \le (q\pi)^{1/2}$$

to obtain

$$\begin{aligned}
\mathbb{E}[|D| S_1(W^e|\mathcal{F}) I_A] &\le k u^{-1/2} \mathbb{E}f(R_{n-1}) \\
&\le k u^{-1/2} \big(\mathbb{E}f(X_q) + q k^{-1/2} \mathbb{E}|D_{n-1}|\big) \\
&\le C q^{1/4} \log n,
\end{aligned}$$

where $D_{n-1} = R_{n-1} - R^*_{n-1}$. Now, applying (2.4) yields

$$\mathbb{P}(R_{n-1} \leq 2k) \leq 1 - (1-q)^{2k} + q\mathbb{E}|D_{n-1}| \leq q(2k + \mathbb{E}|D_{n-1}|),$$

and, by Markov's inequality and (4.4), we finally obtain

$$\mathbb{P}(A^c) \leq k^{-1}(\mathbb{E}N + \mathbb{E}|D|) + q(2k + \mathbb{E}|D|) \leq Cq^{1/4}\log n.$$

The theorem follows after using (v) and $\mathbb{E}Z_n = 1$ to obtain $\mathbb{E}[Z_n|Z_n > 0] \leq Cn$. □

5. Application to the uniform attachment random graph model

Let G_n be a directed random graph on n nodes defined by the following recursive construction. Initially the graph starts with one node with a single loop, where one end of the loop contributes to the "in-degree" and the other to the "out-degree." Now, for $2 \leq m \leq n$, given the graph with $m-1$ nodes, add node m along with an edge directed from m to a node chosen uniformly at random among the m nodes present. Note that this model allows edges connecting a node with itself. This random graph model is referred to as uniform attachment.

This model has been well studied, and it was shown in [5] that if W is equal to the in-degree of a node chosen uniformly at random from G_n, then W converges to a geometric distribution (starting at 0) with parameter $1/2$ as $n \to \infty$. We will give an explicit bound on the total variation distance between the distribution of W and the geometric distribution that yields this asymptotic.

The same result, but with a larger constant, was obtained in [10], where the author uses Stein's method an ad hoc analysis of the model.

Theorem 5.1. *If W is the in-degree of a node chosen uniformly at random from the random graph G_n generated according to uniform attachment, then*

$$d_{\mathrm{TV}}\left(\mathscr{L}(W), \mathrm{Ge}^0\left(\frac{1}{2}\right)\right) \leq \frac{1}{n}.$$

Proof. Let X_i have a Bernoulli distribution, independent of all else, with parameter $\mu_i := (n - i + 1)^{-1}$, and let N be an independent random variable that is uniform on the integers $1, 2, \ldots, n$. If we imagine that node $n + 1 - N$ is the randomly selected node, then it's easy to see that we can write $W := \sum_{i=1}^{N} X_i$.

Next, let us prove that $\sum_{i=1}^{N-1} X_i$ has the discrete equilibrium distribution w.r.t. W. First note that we have, for bounded f and every m,

$$\mu_m \mathbb{E}\Delta f\left(\sum_{i=1}^{m-1} X_i\right) = \mathbb{E}\left[f\left(\sum_{i=1}^{m} X_i\right) - f\left(\sum_{i=1}^{m-1} X_i\right)\right],$$

where we use

$$\mathbb{E}f(X_m) - f(0) = \mathbb{E}X_m \mathbb{E}\Delta f(0)$$

and thus the fact that we can write $(X_m)^{e_0} \equiv 0$. Note also that, for any bounded function g with $g(0) = 0$, we have

$$\mathbb{E}\left(\frac{g(N)}{\mu_N} - \frac{g(N-1)}{\mu_N}\right) = \mathbb{E}g(N).$$

We now assume that $f(0) = 0$. Hence, using the above two facts and independence between N and the sequence $X_1, X_2, \ldots,$ we have

$$\mathbb{E}W\mathbb{E}\Delta f\left(\sum_{i=1}^{N-1} X_i\right) = \mathbb{E}f(W).$$

Now, let

$$N' = \begin{cases} N & \text{if } 1 \le N < n, \\ 0 & \text{if } N = n. \end{cases}$$

We have that $\mathscr{L}(N') = \mathscr{L}(N-1)$ so that $W^{e_0} := \sum_{i=1}^{N'} X_i$ has the equilibrium distribution with respect to W, and it is plain that

$$\mathbb{P}[W \neq W^{e_0}] \le \mathbb{P}[N = n] = \frac{1}{n}.$$

Applying (2.9) of Remark 2.5 yields the theorem. □

6. Application to the preferential attachment random graph model

Define the directed graph G_n on n nodes by the following recursive construction. Initially the graph starts with one node with a single loop where one end of the loop contributes to the "in-degree" and the other to the "out-degree." Now, for $2 \le m \le n$, given the graph with $m - 1$ nodes, add node m along with an edge directed from m to a random node chosen proportionally to the total degree of the node. Note that at step m, the chance that node m connects to itself is $1/(2m - 1)$, since we consider the added vertex m as immediately having out-degree equal to one. This random graph model is referred to as preferential attachment.

This model has been well studied, and it was shown in [5] that if W is equal to the in-degree of a node chosen uniformly at random from G_n, then W converges to the Yule–Simon distribution (defined below). We will give a rate of convergence in the total variation distance for this asymptotic, a result that cannot be read from the main results in [5]. Some rates of convergence in this and related random graph models can be found in the thesis [10], but the techniques and results there do not appear to overlap with ours below. Detailed asymptotics for the individual degrees, along with rates, can be found in [19].

We say the random variable Z has the Yule–Simon distribution if

$$\mathbb{P}(Z = k) = \frac{4}{k(k+1)(k+2)}, \qquad k = 1, 2, \ldots.$$

The following is our main result.

Theorem 6.1. *Let $W_{n,i}$ be the total degree of vertex i in the preferential attachment graph on n vertices, and let I uniform on $\{1,\dots,n\}$ independent of $W_{n,i}$. If Z has the Yule–Simon distribution, then*

$$d_{\mathrm{TV}}(\mathscr{L}(W_{n,I}),\mathscr{L}(Z)) \le \frac{C\log n}{n}$$

for some constant C independent of n.

Remark 6.1. The notation $\mathscr{L}(W_{n,I})$ in the statement of Theorem 6.1 should be interpreted as

$$\mathscr{L}(W_{n,I}|I=i)=\mathscr{L}(W_{n,i}).$$

We will use similar notation in what follows without further mention.

Proposition 6.2. *If U has the uniform distribution on* $(0, 1)$, *and given U, we define Z such that $\mathscr{L}(Z)=\mathrm{Ge}(\sqrt{U})$, then Z has the Yule–Simon distribution.*

Our strategy to prove Theorem 6.1 will be to show that, for I uniform on $\{1,\dots,n\}$ and U uniform on $(0, 1)$, we have:

(1) $d_{\mathrm{TV}}(\mathscr{L}(W_{n,I}),\mathrm{Ge}(\mathbb{E}[W_{n,I}|I]^{-1}))\le C\log(n)/n$,
(2) $d_{\mathrm{TV}}(\mathrm{Ge}(\mathbb{E}[W_{n,I}|I]^{-1}),\mathrm{Ge}(\sqrt{I/n}))\le C\log(n)/n$,
(3) $d_{\mathrm{TV}}(\mathrm{Ge}(\sqrt{I/n}),\mathrm{Ge}(\sqrt{U}))\le C\log(n)/n$,

where, here and in what follows, we use the letter C as a generic constant which may differ from line to line. From this point, Theorem 6.1 follows from the triangle inequality and Proposition 6.2.

Item (1) will follow from our framework above; in particular we will use the following result, which may be of independent interest. We postpone the proof to the end of the section.

Theorem 6.3. *Retaining the notation and definitions above, we have*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W_{n,i}),\mathrm{Ge}\big(1/\mathbb{E}(W_{n,i})\big)\big)\le\frac{C}{i}$$

for some constant C independent of n and i.

To show Items (2) and (3) we will need the following lemma. The first statement is found in [5], page 283, and the second follows easily from the first.

Lemma 6.4 (Bollobás *et al.* [5]). *Retaining the notation and definitions above, for all* $1\le i\le n$,

$$\left|\mathbb{E}W_{n,i}-\sqrt{\frac{n}{i}}\right|\le C\sqrt{\frac{n}{i^3}}\quad\textit{and}\quad\left|\frac{1}{\mathbb{E}W_{n,i}}-\sqrt{\frac{i}{n}}\right|\le\frac{C}{\sqrt{ni}}.$$

Our final general lemma is useful for handling total variation distance for conditionally defined random variables.

Lemma 6.5. *Let W and V be random variables, and let X be a random element defined on the same probability space. Then*

$$d_{\mathrm{TV}}(\mathscr{L}(W),\mathscr{L}(V)) \le \mathbb{E}d_{\mathrm{TV}}(\mathscr{L}(W|X),\mathscr{L}(V|X)).$$

Proof. If $f:\mathbb{R}\to[0,1]$, then

$$|\mathbb{E}[f(W)-f(V)]| \le \mathbb{E}|\mathbb{E}[f(W)-f(V)|X]| \le \mathbb{E}d_{\mathrm{TV}}(\mathscr{L}(W|X),\mathscr{L}(V|X)). \qquad \square$$

Proof of Theorem 6.1. We first claim that

$$d_{\mathrm{TV}}\bigl(\mathrm{Ge}(p),\mathrm{Ge}(p-\varepsilon)\bigr) \le \frac{\varepsilon}{p} < \frac{\varepsilon}{p-\varepsilon}. \tag{6.1}$$

The second inequality of (6.1) is obvious. To see the first inequality, we construct two infinite sequences of independent random variables. The first sequence consists of $\mathrm{Be}(p)$ random variables, and the second sequence consists of $\mathrm{Be}(p-\varepsilon)$ random variables maximally coupled component-wise to the first so that the terms in the first sequence are no smaller than the corresponding terms in the second. For each of these sequences, the index of the first Bernoulli random variable, which is 1, follows a $\mathrm{Ge}(p)$ and $\mathrm{Ge}(p-\varepsilon)$ distribution, respectively. Since the index of the first occurrence of a 1 in the first sequence is less than or equal to that in the second sequence, the probability that these two random variables are not equal is the probability that a coordinate in the second sequence is 0, given the same coordinate is 1 in the first sequence, which is ε/p.

Using (6.1) and Lemma 6.4 we easily obtain

$$d_{\mathrm{TV}}\bigl(\mathrm{Ge}(1/\mathbb{E}W_{n,i}),\mathrm{Ge}\bigl(\sqrt{i/n}\bigr)\bigr) \le \frac{C}{i},$$

and applying Lemma 6.5 we find

$$d_{\mathrm{TV}}\bigl(\mathrm{Ge}(\mathbb{E}[W_{n,I}|I]^{-1}),\mathrm{Ge}\bigl(\sqrt{I/n}\bigr)\bigr) \le \frac{C\log(n)}{n},$$

which is Item (2) above. Now, coupling U to I by writing $U=I/n-V$, where V is uniform on $(0,1/n)$ and independent of I, and using first (6.1) and then Lemma 6.5 leads to

$$d_{\mathrm{TV}}\bigl(\mathrm{Ge}\bigl(\sqrt{U}\bigr),\mathrm{Ge}\bigl(\sqrt{I/n}\bigr)\bigr) \le \frac{C}{n}\sum_{i=1}^{n}\frac{\sqrt{i/n}-\sqrt{(i-1)/n}}{\sqrt{i/n}} \le \frac{C\log(n)}{n},$$

which is Item (3) above. Finally, applying Lemma 6.5 to Theorem 6.3 yields the claim related to Item (1) above so that Theorem 6.1 is proved. $\square$

The remainder of this section is devoted to the proof of Theorem 6.3; recall $W_{n,i}$ is the total degree of vertex i in the preferential attachment graph on n vertices. Since we want to apply our geometric approximation framework using the equilibrium distribution, we will use Proposition 2.3 and so we first construct a variable having the size-bias distribution of $W_{n,i} - 1$. To facilitate this construction we need some auxiliary variables.

For $j \geq i$, let $X_{j,i}$ be the indicator variable of the event that vertex j has an outgoing edge connected to vertex i in G_j so that we can denote $W_{j,i} = 1 + \sum_{k=i}^{j} X_{k,i}$. In this notation, for $1 \leq i < j \leq n$,

$$\mathbb{P}(X_{j,i} = 1 | G_{j-1}) = \frac{W_{j-1,i}}{2j-1},$$

and for $1 \leq i \leq n$,

$$\mathbb{P}(X_{i,i} = 1 | G_{i-1}) = \frac{1}{2i-1}.$$

The following well-known result will allow us to use this decomposition to size-bias $W_{n,i} - 1$; see, for example, Proposition 2.2 of [7] and the discussion thereafter.

Proposition 6.6. *Let $X_1, \ldots, X_n$ be zero-one random variables with $\mathbb{P}(X_i = 1) = p_i$. For each $i = 1, \ldots, n$, let $(X_j^{(i)})_{j \neq i}$ have the distribution of $(X_j)_{j \neq i}$ conditional on $X_i = 1$. If $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$ and K is chosen independently of the variables above with $\mathbb{P}(K = k) = p_k/\mu$, then $X^s = \sum_{j \neq K} X_j^{(K)} + 1$ has the size-bias distribution of X.*

Roughly, Proposition 6.6 implies that in order to size-bias $W_{n,i} - 1$, we choose an indicator $X_{K,i}$ where, for $k = i, \ldots, n$, $\mathbb{P}(K = k)$ is proportional to $\mathbb{P}(X_{n,k} = 1)$ (and zero for other values), then attach vertex K to vertex i, and sample the remaining edges conditionally on this event. Note that given $K = k$, in the graphs G_l, $1 \leq l < i$ and $k < l \leq n$, this conditioning does not change the original rule for generating the preferential attachment graph, given G_{l-1}. The following lemma implies the remarkable fact that in order to generate the graphs G_l for $i \leq l < k$ conditional on $X_{k,i} = 1$ and G_{l-1}, we attach edges following the same rule as preferential attachment, but include the edge from vertex k to vertex i in the degree count.

Lemma 6.7. *Retaining the notation and definitions above, for $i \leq j < k$, we have*

$$\mathbb{P}(X_{j,i} = 1 | X_{k,i} = 1, G_{j-1}) = \frac{1 + W_{j-1,i}}{2j},$$

where we define $W_{i-1,i} = 1$.

Proof. By Bayes's rule, we have

$$\mathbb{P}(X_{j,i} = 1 | X_{k,i} = 1, G_{j-1}) = \frac{\mathbb{P}(X_{j,i} = 1 | G_{j-1}) \mathbb{P}(X_{k,i} = 1 | X_{j,i} = 1, G_{j-1})}{\mathbb{P}(X_{k,i} = 1 | G_{j-1})}, \tag{6.2}$$

and we will calculate the three probabilities appearing in (6.2). First, for $i \le j$, we have

$$\mathbb{P}(X_{j,i} = 1 | G_{j-1}) = \frac{W_{j-1,i}}{2j-1},$$

which implies

$$\mathbb{P}(X_{k,i} = 1 | G_{j-1}) = \frac{\mathbb{E}[W_{k-1,i} | G_{j-1}]}{2k-1}$$

and

$$\mathbb{P}(X_{k,i} = 1 | X_{j,i} = 1, G_{j-1}) = \frac{\mathbb{E}[W_{k-1,i} | X_{j,i} = 1, G_{j-1}]}{2k-1}.$$

Now, to compute the conditional expectations appearing above, note first that

$$\mathbb{E}(W_{k,i} | G_{k-1}) = W_{k-1,i} + \frac{W_{k-1,i}}{2k-1} = \left(\frac{2k}{2k-1}\right) W_{k-1,i},$$

and thus

$$\mathbb{E}(W_{k,i} | G_{k-2}) = \left(\frac{2(k-1)}{2(k-1)-1}\right)\left(\frac{2k}{2k-1}\right) W_{k-2,i}.$$

Iterating, we find that, for $i, s < k$,

$$\mathbb{E}(W_{k,i} | G_{k-s}) = \prod_{t=1}^{s} \left(\frac{2(k-t+1)}{2(k-t+1)-1}\right) W_{k-s,i}. \tag{6.3}$$

By setting $j-1 = k-s$ and then replacing $k-1$ by k in (6.3), we obtain

$$\mathbb{E}(W_{k-1,i} | G_{j-1}) = \prod_{t=1}^{k-j} \left(\frac{2(k-t)}{2(k-t)-1}\right) W_{j-1,i},$$

which also implies

$$\mathbb{E}(W_{k-1,i} | X_{j,i} = 1, G_{j-1}) = \prod_{t=1}^{k-j-1} \left(\frac{2(k-t)}{2(k-t)-1}\right)(1 + W_{j-1,i}).$$

Substituting these expressions appropriately into (6.2) proves the lemma. □

The previous lemma suggests the following (embellished) construction of $(W_{n,i} | X_k = 1)$ for any $1 \le i \le k \le n$. Here and below we will denote quantities related to this construction with a superscript k. First we generate G_{i-1}^k, a graph with $i-1$ vertices, according to the usual preferential attachment model. At this point, if $i \ne k$, vertex i and k are added to the graph, along with a vertex labeled i' with an edge to it emanating from vertex k. Given G_{i-1}^k and these additional vertices and edges, we generate G_i^k by connecting vertex i to a vertex j randomly chosen from the vertices $1, \dots, i, i'$ proportional to their degree, where i has degree one (from the out-edge), and i' has degree one (from the in-edge emanating from vertex k). If $i = k$, we attach i to i' and denote the resulting graph by G_i^i. For $i < j < k$, we generate the graphs G_j^k recursively from G_{j-1}^k by connecting vertex j to a vertex l, randomly chosen from the vertices $1, \dots, j, i'$ proportional to their degree, where j has degree one (from the out-edge). Note that none of the vertices $1, \dots, k-1$ can connect to vertex k. We now define $G_k^k = G_{k-1}^k$, and for $j = k+1, \dots, n$, we generate G_j from G_{j-1}, according to preferential attachment among the vertices $1, \dots, j, i'$.

Lemma 6.8. *Let* $1 \le i \le k \le n$, *and retain the notation and definitions above.*

(1) $\mathscr{L}(W_{n,i}^k + W_{n,i'}^k) = \mathscr{L}(W_{n,i} | X_k = 1)$.
(2) *For fixed* i, *if* K *is a random variable such that*

$$\mathbb{P}(K = k) = \frac{\mathbb{E}X_{k,i}}{\mathbb{E}W_{n,i} - 1}, \qquad k \ge i,$$

then $W_{n,i}^K + W_{n,i'}^K - 1$ *has the size-bias distribution of* $W_{n,i} - 1$.
(3) *Conditional on the event* $\{W_{n,i}^k + W_{n,i'}^k = m+1\}$, *the variable* $W_{n,i}^k$ *is uniformly distributed on the integers* $1, 2, \dots, m$.
(4) $W_{n,i}^K - 1$ *has the discrete equilibrium distribution of* $W_{n,i} - 1$.

Proof. Items (1) and (2) follow from Proposition 6.6 and Lemma 6.7. Viewing $(W_{n,i}^k, W_{n,i'}^k)$ as the number of balls of two colors in a Polya urn model started with one ball of each color, Item (3) follows from induction on m and Item (4) follows from Proposition 2.3. □

Proof of Theorem 6.3. We will apply Theorem 2.2 to $\mathscr{L}(W_{n,i} - 1)$, so that we must find a coupling of a variable with this distribution to that of a variable having its discrete equilibrium distribution. For each fixed $k = i, \dots, n$ we will construct $(X_{j,i}^k, \widetilde{X}_{j,i}^k)_{j \ge i}$ so $(X_{j,i}^k)_{j \ge i}$ and $(\widetilde{X}_{j,i}^k)_{j \ge i}$ are distributed as the indicators of the events vertex j connects to vertex i in G_n^k and G_n, respectively. We will use the notation

$$W_{j,i}^k = \sum_{m=i}^{j} X_{j,i}^k \quad \text{and} \quad \widetilde{W}_{j,i}^k = \sum_{m=i}^{j} \widetilde{X}_{j,i}^k,$$

which will be distributed as the degree of vertex i in the appropriate graphs.

The constructions for $k = i$ and $k > i$ differ, so assume here that $k > i$. Let $U_{j,i}^k$ be independent uniform (0, 1) random variables, and first define

$$X_{i,i}^k = \mathrm{I}[U_{i,i}^k < 1/2i] \quad \text{and} \quad \widetilde{X}_{i,i}^k = \mathrm{I}[U_{i,i}^k < 1/(2i-1)].$$

Now, for $i < j < k$, and assuming that $(W_{j-1,i}^k, \widetilde{W}_{j-1,i}^k)$ is given, we define

$$X_{j,i}^k = \mathrm{I}\left[U_{j,i}^k < \frac{W_{j-1,i}^k}{2j}\right] \quad \text{and} \quad \widetilde{X}_{j,i}^k = \mathrm{I}\left[U_{j,i}^k < \frac{\widetilde{W}_{j-1,i}^k}{2j-1}\right]. \tag{6.4}$$

For $j=k$ we set $X^k_{k,i}=0$ and $\widetilde{X}^k_{k,i}$ as in (6.4) with $j=k$, and, for $j>k$, we define

$$X^k_{j,i}=\mathrm{I}\left[U^k_{j,i}<\frac{W^k_{j-1,i}}{2j-1}\right] \quad \text{and} \quad \widetilde{X}^k_{j,i}=\mathrm{I}\left[U^k_{j,i}<\frac{\widetilde{W}^k_{j-1,i}}{2j-1}\right].$$

Thus we have recursively defined the variables $(X^k_{j,i}, \widetilde{X}^k_{j,i})$ and it is clear they are distributed as claimed with $(W^k_{j,i}, \widetilde{W}^k_{j,i})$ distributed as the required degree counts. Note also that $\widetilde{W}^k_{j,i}\geq W^k_{j,i}$ and $\widetilde{X}^k_{j,i}\geq X^k_{j,i}$. We also define the events

$$A^k_{j,i}:=\bigl\{\min\{i\leq l\leq n\colon\ X^k_{l,i}\neq \widetilde{X}^k_{l,i}\}=j\bigr\}.$$

Using that $W^k_{j-1,i}=\widetilde{W}^k_{j-1,i}$ under $A^k_{j,i}$ (which also implies $A^k_{j,i}=\varnothing$ for $j>k$) we have

$$\begin{aligned}
\mathbb{P}(\widetilde{W}^k_{n,i}\neq W^k_{n,i}) &= \mathbb{P}\left(\bigcup_{j=i}^n A^k_{j,i}\right)\\
&= \mathbb{E}\widetilde{X}^k_{k,i}+\sum_{j=i}^k \mathbb{P}\left(A^k_{j,i}\cap\left\{\frac{W^k_{j-1,i}}{2j}<U^k_{j,i}<\frac{\widetilde{W}^k_{j-1,i}}{2j-1}\right\}\right)\\
&\leq \mathbb{E}\widetilde{X}^k_{k,i}+\sum_{j=i}^n \mathbb{P}\left(\frac{\widetilde{W}^k_{j-1,i}}{2j}<U^k_{j,i}<\frac{\widetilde{W}^k_{j-1,i}}{2j-1}\right),
\end{aligned}$$

where we write $W^k_{i-1,i}:=\widetilde{W}^k_{i-1,i}:=1$. Finally, starting from (6.5) and using the computations in the proof of Lemma 6.7 and the estimates in Lemma 6.4, we find

$$\begin{aligned}
\mathbb{P}(\widetilde{W}^k_{n,i}\neq W^k_{n,i}) &\leq \frac{\mathbb{E}W_{k-1,i}}{2k-1}+\sum_{j=i}^n \mathbb{E}W_{j-1,i}\left(\frac{1}{2j-1}-\frac{1}{2j}\right)\\
&\leq C\left[\sqrt{\frac{k}{i}}\frac{1}{k}+\sqrt{\frac{k}{i^3}}\frac{1}{k}+\sum_{j\geq i}\left(\sqrt{\frac{j}{i}}\frac{1}{j^2}+\sqrt{\frac{j}{i^3}}\frac{1}{j^2}\right)\right]\leq C/i.
\end{aligned}$$

If $k=i$, it is clear from the construction preceding Lemma 6.8 that an easy coupling, similar to that above, will yield $\mathbb{P}(\widetilde{W}^i_{n,i}\neq W^i_{n,i})<C/i$. Since these bounds do not depend on k, we also have

$$\mathbb{P}(\widetilde{W}^K_{n,i}-1\neq W^K_{n,i}-1)\leq C/i, \tag{6.5}$$

and the result now follows from Lemma 6.8, (6.5) and (2.9) of Remark 2.5. □

Acknowledgements

Erol A. Peköz and Adrian Röllin were supported in part by NUS Grant R-155-000-098-133. The authors would like to thank the anonymous referees for their helpful comments.

References

[1] Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. *Applied Mathematical Sciences* **77**. New York: Springer. MR0969362

[2] Barbour, A.D. and Čekanavičius, V. (2002). Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30** 509–545. MR1905850

[3] Barbour, A.D. and Grübel, R. (1995). The first divisible sum. *J. Theoret. Probab.* **8** 39–47. MR1308669

[4] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. *Oxford Studies in Probability* **2**. New York: Oxford Univ. Press. MR1163825

[5] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. MR1824277

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

[7] Chen, L.H.Y., Goldstein, L. and Shao, Q.M. (2011). *Normal Approximation by Stein's Method*. *Probability and Its Applications* (*New York*). Heidelberg: Springer. MR2732624

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

[9] Daly, F. (2010). Stein's method for compound geometric approximation. *J. Appl. Probab.* **47** 146–156. MR2654764

[10] Ford, E. (2009). Barabási–Albert random graphs, scale-free distributions and bounds for approximation through Stein's method. Ph.D. thesis. Univ. Oxford.

[11] Goldstein, L. (2009). Personal communication and unpublished notes. In *Stein Workshop*, January 2009, Singapore.

[12] 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

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

[14] Lalley, S.P. and Zheng, X. (2011). Occupation statistics of critical branching random walks in two or higher dimensions. *Ann. Probab.* **39** 327–368. MR2778804

[15] 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

[16] Mattner, L. and Roos, B. (2007). A shorter proof of Kanter's Bessel function concentration bound. *Probab. Theory Related Fields* **139** 191–205. MR2322695

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

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

[19] Peköz, E., Röllin, A. and Ross, N. (2011). Degree asymptotics with rates for preferential attachment random graphs. Preprint. Available at arXiv:org/abs/1108.5236.

[20] Phillips, M.J. and Weinberg, G.V. (2000). Non-uniform bounds for geometric approximation. *Statist. Probab. Lett.* **49** 305–311. MR1794749

[21] Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11** 1115–1128. MR2189083

[22] Röllin, A. (2008). Symmetric and centered binomial approximation of sums of locally dependent random variables. *Electron. J. Probab.* **13** 756–776. MR2399295

[23] Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surveys* **8** 210–293.

[24] Ross, S. and Peköz, E. (2007). *A Second Course in Probability*. Boston, MA: www.ProbabilityBookstore.com.

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

Received June 2011