EXPONENTIAL APPROXIMATION FOR THE NEARLY CRITICAL GALTON-WATSON PROCESS AND OCCUPATION TIMES OF MARKOV CHAINS

Erol Peköz and Adrian Röllin

Boston University and National University of Singapore

Abstract

In this article we provide new applications for exponential approximation using the framework of Peköz and Röllin (in press), which is based on Stein's method. We give error bounds for the nearly critical Galton-Watson process conditioned on non-extinction, and for the occupation times of Markov chains; for the latter, in particular, we give a new exponential approximation rate for the number of revisits to the origin for general two dimensional random walk, also known as the Erdős-Taylor theorem.

1 INTRODUCTION

A new framework for estimating the error of the exponential approximation was recently developed in Peköz and Röllin (in press), where it was applied to geometric sums, Markov chain hitting times, and the critical Galton-Watson conditioned on non-extinction. In this article we provide some generalizations to the approach of Peköz and Röllin (in press) and apply them to study Markov chain occupation times and a result of Erdős and Taylor (1960) for the number of visits to the origin by the two dimensional random walk, as well as to get a rate for the result of Fahady, Quine, and Vere-Jones (1971) for the nearly critical Galton-Watson branching process conditioned on non-extinction.

The main result in Peköz and Röllin (in press) that we use is based on Stein’s method (see e.g. Ross and Peköz (2007) for an introduction) and can be thought of as formalizing the intuitive notion that a random variable $X$ has approximately an exponential distribution if $X$ and $X^e$ are close in

Version from August 25, 2010, 00:18
distribution, where $X^e$ has the equilibrium distribution with respect to $X$ characterized by
\[
P[X^e \leq x] = \frac{1}{\mathbb{E}X} \int_{0}^{x} \mathbb{P}[X > y]dy.
\]
The equilibrium distribution appears in renewal theory as the time until the next renewal starting from steady-state. A renewal process with exponential inter-renewal times has the exponential distribution for its equilibrium distribution, and so the above intuition is not surprising. Peköz and Röllin \textit{in press} give bounds on the accuracy of the exponential approximation in terms of how closely $X$ and $X^e$ can be coupled together on the same probability space; one version of the result we will use below can be written as
\[
\sup_{x \geq 0} |\mathbb{P}[X > x] - e^{-x/\mathbb{E}X}| \leq 2.46\sqrt{\mathbb{E}|X - X^e|}.
\]

Some heuristics for Stein’s method can be understood using size-biased random variables. For a continuous random variable $X$ with probability density function $f(x)$, the size-biased random variable $X^s$ has density $xf(x)/\mathbb{E}X$. The size of the renewal interval containing a randomly chosen point as well as the children in the family of a randomly chosen child are examples of size-biased random variables; see Brown (2006) and Arratia and Goldstein (2010) for surveys and applications of size biasing.

Stein’s method for the exponential distribution, as well as for some other nonnegative distributions, can be viewed in terms of size-biasing. For the Poisson approximation to some random variable $X$, the Stein-Chen method (see Barbour, Holst, and Janson (1992)) gives a bound on the error in terms of how closely $X$ and $X^s - 1$ can be coupled together on the same probability space; these both have exactly the same distribution when $X$ has a Poisson distribution. For approximation by a binomial distribution (see Peköz, Röllin, Čekanavičius, and Shwartz (2009)), we can obtain a bound in terms of how closely $X^s - 1$ and $n - (n - X)^s$ can be coupled; both of these have exactly the same distribution if $X$ is binomial with parameters $n$ and $p$. For the exponential distribution, we can obtain a bound on the error in terms of how closely $X$ and $UX^s$ can be coupled, where $U$ is an independent uniform (0,1) random variable independent of all else; $X^e$ has the same distribution as $UX^s$. This last approach is the one we use below for the nearly critical Galton Watson process conditioned on non-extinction.

The organization of this article is as follows. In Section 2 we give the notation, background and preliminaries. In Section 3 we consider the setting
of a nearly critical Galton Watson branching process conditioned on non-
extinction. In Section 4 we study general dependent sums, occupation times
for Markov chains and the the number of times the origin is revisited for the
two-dimensional general random walk.

2 PRELIMINARIES

We first define the probability distance metrics we use below. For two prob-
obability distributions \( F \) and \( G \) define the Kolmogorov distance metric as

\[
d_K(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|.
\]

If both distributions have finite expectation, define the Wasserstein distance
metric

\[
d_W(F, G) = \int_{\mathbb{R}} |F(x) - G(x)| \, dx.
\]

We can relate the two metrics using

\[d_K(P, \text{Exp}(1)) \leq 1.74 \sqrt{d_W(P, \text{Exp}(1))};\]

see e.g. Gibbs and Shi (2002).

Central to the approach in Peköz and Röllin (in press) is the equilibrium
distribution from renewal theory, and we next give the definition we use.

**Definition 2.1.** Let \( X \) be a non-negative 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-continuous \( f \)

\[
\mathbb{E} f(X) - f(0) = \mathbb{E} X \mathbb{E} f'(X^e).
\]

It is straightforward that this implies

\[
\mathbb{P}[X^e \leq x] = \frac{1}{\mathbb{E} X} \int_{0}^{x} \mathbb{P}[X > y] \, dy.
\]

Indeed for nonnegative \( X \) having finite first moment, define the distribution
function

\[
F^e(x) = \frac{1}{\mathbb{E} X} \int_{0}^{x} \mathbb{P}[X > y] \, dy
\]
on \( x \geq 0 \) and \( F^e(x) = 0 \) for \( x < 0 \). Then

\[
\mathbb{E} f(X) - f(0) = \mathbb{E} \int_{0}^{X} f'(s) \, ds
\]

\[
= \mathbb{E} \int_{0}^{\infty} f'(s) \mathbb{P}[X > s] \, ds = \int_{0}^{\infty} f'(s) \mathbb{P}[X > s] \, ds
\]
so that \( F^c \) is the distribution function of \( X^e \) and our definition via (2.1) is consistent with that from renewal theory.

The size biased distribution will also be used below. We define it as follows.

**Definition 2.2.** Let \( X \) be a non-negative random variable with finite mean. We say that a random variable \( X^s \) has the size-biased distribution w.r.t. \( X \) if for all bounded \( f \)

\[
\mathbb{E}\{X f(X)\} = \mathbb{E} X \mathbb{E} f(X^s).
\]  

(2.2)

We next present the key result from [Peköz and Röllin (in press)](https://example.com) that we will use in the applications that follow.

**Theorem 2.1** ([Peköz and Röllin (in press)](https://example.com), Theorem 2.1). Let \( W \) be a non-negative random variable with \( \mathbb{E} W = 1 \) and let \( W^e \) have the equilibrium distribution w.r.t. \( W \). Then, for any \( \beta > 0 \),

\[
d_K(\mathcal{L}(W), \text{Exp}(1)) \leq 12\beta + 2\mathbb{P}[|W^e - W| > \beta],
\]

and, if in addition \( W \) has finite second moment,

\[
d_W(\mathcal{L}(W), \text{Exp}(1)) \leq 2\mathbb{E}|W^e - W|
\]

(2.3)

3 THE NEARLY CRITICAL GALTON-WATSON BRANCHING PROCESS

It was shown by [Yaglom (1947)](https://example.com) that the limiting distribution of a critical Galton-Watson branching process conditioned on non-extinction is exponential. A corresponding rate of convergence was first proved by Peköz and Röllin (in press). In the super- and sub-critical cases, the limiting distributions are very difficult to calculate and are only known explicitly in very special cases; see e.g. [Bingham (1988)](https://example.com), [Fahady et al. (1971)](https://example.com), however, were able to show that the limiting distribution of a nearly critical branching process conditioned on non-extinction converges to the exponential distribution as \( m \to 1 \) over a general class of offspring distributions. The following theorem gives explicit error bounds for the exponential approximation for any finite \( n \) and any \( m \neq 1 \). The bounds obtained confirm the exponential limit as \( m \to 1 \) for the limiting distribution under the assumptions of a finite third moment for the offspring distribution and a non-zero probability of having 2 or more offspring.
Theorem 3.1. Consider a Galton-Watson branching process starting from a single particle at time zero, and let $Z_n$ be the size of the $n$th generation. Assume $P[Z_1 \geq 2] > 0$, $m = \mathbb{E}Z_1 \neq 1$ and $\mathbb{E}|Z_1|^3 \leq \infty$. Let $\alpha = P[Z_1 = 1]/P[Z_1 \geq 2]$, 

$$C = (2 + \alpha)(2 + \alpha + \text{Var}Z_1 + \mathbb{E}|Z_1|^3),$$

and $\lambda = 1/\mathbb{E}(Z_n|Z_n > 0) = P[Z_n > 0]/m^n$. Then 

$$d_W(\mathcal{L}(\lambda Z_n|Z_n > 0), \text{Exp}(1)) \leq C \eta(n, m),$$

where 

$$\eta(m, n) = \frac{1 - m}{2(1 - m^n)} + \frac{(1 - m)^2}{m(1 - m^n)} \sum_{j=1}^{n} \frac{m^{2j}}{1 - m^j}.\tag{3.1}$$

In particular, for any $n \geq 1$ and $m > 1$, 

$$\eta(m, n) \leq 6(m - 1) + \frac{6.5 + \log(n)}{n},\tag{3.2}$$

and, for any $n \geq 2$ and $\frac{1}{2} \leq m < 1$, 

$$\eta(m, n) \leq (4.5 - \log(1 - m))(1 - m) + \frac{4.5 + \log(n)}{n}.\tag{3.3}$$

Note that as a consequence of (3.2) and (3.3) we obtain that the limiting distribution of a nearly critical branching process conditioned on non-extinction converges to the exponential as $m \to 1$ under the very mild condition that $C$ does not grow too fast. This is in particular true if the third moment of the offspring distribution remains bounded from above and $P[Z_1 \geq 2]$ remains bounded away from zero as $m \to 1$.

Proof of Theorem 3.1. With some modifications, we follow the line of argument from Peköz and Röllin (in press), which is based on the size-biased branching tree of Lyons, Pemantle, and Peres (1995).

We assume that the particles in the tree are labeled and ordered. That is, if $w$ and $v$ are two particles in the same generation, then all offspring of $w$ are to the left of the offspring of $v$, whenever $w$ is to the left of $v$. We start in generation 0 with one particle $v_0$ and let it have a size-biased number of offspring. Then we pick one of the offspring of $v_0$ uniformly at random and label it $v_1$. For each of the siblings of $v_1$ we continue with an independent Galton-Watson branching process with the original offspring distribution. For $v_1$ we proceed as we did for $v_0$, i.e., we give it a size-biased number of offspring, pick one uniformly at random, label it $v_2$, and so on.
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 \( v_n \). 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). 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 \( v_j \). We have the relations \( L_n = \sum_{j=1}^n L_{n,j} \) and \( R_n = 1 + \sum_{j=1}^n R_{n,j} \).

Next let \( R'_{n,j} \) be independent random variables such that

\[
\mathcal{L}(R'_{n,j}) = \mathcal{L}(R_{n,j} | L_{n,j} = 0).
\]

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

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

Define also \( R^*_n = 1 + \sum_{j=1}^n R^*_n \). Below are a few facts that we will subsequently use to give the proof of the theorem. In what follows, let \( \sigma^2 = \text{Var} Z_1 \) and \( \gamma = \mathbb{E} |Z_1|^3 \).

(i) The size-biased distribution of \( \mathcal{L}(X) \) is the same as that of \( \mathcal{L}(X | X > 0) \);

(ii) \( S_n \) has the size-biased distribution of \( \mathcal{L}(Z_n) \);

(iii) \( v_n \) is uniformly distributed among the particles of generation \( n \);

(iv) \( \mathcal{L}(R^*_n) = \mathcal{L}(Z_n | Z_n > 0) \); 

(v) \( \mathbb{E} \{ R'_{n,j} I_{A^c_{n,j}} \} \leq mn^{-j} \gamma \mathbb{P}[A^c_{n,j}] \);

(vi) \( \mathbb{E} \{ R_{n,j} I_{A^c_{n,j}} \} \leq mn^{-j} \gamma \mathbb{P}[A^c_{n,j}] \); 

(vii) \( \mathbb{P}[A^c_{n,j}] \leq m^{-1} \sigma^2 \mathbb{P}[Z_{n-j} > 0] \).

(viii) \( \mathbb{P}[Z_n > 0] \leq (1 + \alpha) \frac{m^n(1 - m)}{1 - mn} \).

For (i)-(iv) see Peköz and Röllin (in press). Using independence,

\[
\mathbb{E} \{ R'_{n,j} I_{A^c_{n,j}} \} = \mathbb{E} R'_{n,j} \mathbb{P}[A^c_{n,j}] \leq \mathbb{E} S_{n,j} \mathbb{P}[A^c_{n,j}] \leq mn^{-j} \sigma^2 \mathbb{P}[A^c_{n,j}],
\]

which proves (v). If \( X_j \) denotes the number of siblings of \( v_j \), having the size-biased distribution of \( \mathcal{L}(Z_1) \) minus 1, we have

\[
\mathbb{E} \{ R_{n,j} I_{A^c_{n,j}} \} \leq mn^{-j} \mathbb{E} \{ X_j I_{A^c_{n,j}} \} \leq \sum_k m^{n-j} k \mathbb{P}[X_j = k; A^c_{n,j}]
\leq \sum_k m^{n-j} k \mathbb{P}[X_j = k] \mathbb{P}[A^c_{n,j} | X_j = k]
\leq \sum_k m^{n-j} k^2 \mathbb{P}[X_j = k] \mathbb{P}[A^c_{n,j}]
\leq mn^{-j} \mathbb{E} X^2 \mathbb{P}[A^c_{n,j}] \leq mn^{-j} \gamma \mathbb{P}[A^c_{n,j}],
\]

6
hence (vi). Now,
\[ \mathbb{P}[A_{n,j}^c] = \mathbb{E}\{\mathbb{P}[A_{n,j}^c | X_j]\} \leq \mathbb{E}\{X_j \mathbb{P}[Z_{n-j} > 0]\} = m^{-1} \sigma^2 \mathbb{P}[Z_{n-j} > 0], \]
which proves (vii). Finally, using the Corollary on page 356 of Fujimagari (1980), we have
\[ \mathbb{P}[Z_n > 0] \leq (2 + \alpha) \frac{1 - m}{m^{n-1}}, \]
which is (viii) (note that the result cited is for bounded offspring distribution, but easily extends to the unbounded case).

Due to (iv) we can set \( W = \lambda R_n^* \). Due to (i) and (ii), \( S_n \) has the size-biased distribution with respect to \( R_n^* \). 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 \( \mathcal{L}(US_n); \) hence we can set \( W = \lambda (R_n - U) \). Therefore, we can apply (2.3) and use (v)–(vii) to obtain
\[ \mathbb{E}|W - W^e| \leq \frac{\lambda}{2} + \lambda \mathbb{E}|R_n^* - R_n| \]
\[ \leq \frac{\lambda}{2} + \lambda \sum_{j=1}^{n} \mathbb{E}\{R_{n,j}^* I_{A_{n,j}^c} + R_{n,j} I_{A_{n,j}^c}\} \]
\[ \leq \frac{(2 + \alpha)(1 - m)}{2(1 - m^n)} + \frac{C(1 - m)^2}{m(1 - m^n)} \sum_{j=1}^{n} \frac{m^{2j}}{1 - m^j}, \]
which proves (3.1).

Let now \( m > 1 \). Note that \( m^{2j}/(m^j - 1) \) is an increasing function in \( j \), except possibly at the beginning where it may be decreasing. Hence
\[ \sum_{j=1}^{n} \frac{m^{2j}}{m^j - 1} \leq \frac{m^2}{m - 1} + \frac{2m^{2n}}{m^n - 1} + \int_{1}^{n} \frac{m^{2x}}{m^x - 1} dx \]
\[ = \frac{m^2}{m - 1} + \frac{2m^{2n}}{m^n - 1} + \frac{m^n - m + \log\left(\frac{m^n - 1}{m - 1}\right)}{\log(m)}, \]
which implies that
\[ \eta(m,n) \leq \frac{m - 1}{2(m^n - 1)} + \frac{m(m - 1)}{m^n - 1} + \frac{2m^{2n}(m - 1)^2}{m(m^n - 1)^2} \]
\[ + \frac{(m - 1)^2(m^n - m)}{m(m^n - 1)\log(m)} + \frac{(m - 1)^2\log\left(\frac{m^n - 1}{m - 1}\right)}{m(m^n - 1)\log(m)} \]
\[ =: r_1 + r_2 + r_3 + r_4 + r_5. \]
Recall that
\[
\frac{m^n - 1}{m - 1} = \sum_{k=0}^{n-1} m^k \geq n.
\] (3.4)

This implies that
\[
r_1 \leq \frac{1}{2n}, \quad r_2 = \frac{m - 1}{n} + \frac{1}{n} \leq m - 1 + \frac{1}{n},
\]
and, using that \((a + b)^2 \leq 2a^2 + 2b^2,\)
\[
r_3 = \frac{2}{m} \left( \frac{m^n(m-1)}{m^n-1} \right)^2 = \frac{2}{m} \left( m - 1 + \frac{m-1}{m^n-1} \right)^2
\]
\[
\leq \frac{4(m-1)^2}{m} + \frac{4}{mn^2} \leq 4(m-1) + \frac{4}{n},
\]
Furthermore, recalling that \((m-1) \leq m \log(m),\)
\[
r_4 \leq \frac{(m-1)(m^n - m)}{m^n - 1} \leq m - 1.
\]
Finally,
\[
r_5 = \frac{(m-1)^2 \log\left( \frac{m^n-1}{m-1} \right)}{m(m^n-1) \log(m)} \leq \frac{m-1}{m^n-1} \log\left( \frac{m^n-1}{m-1} \right) \leq \frac{1}{n} + \frac{\log(n)}{n}.
\]
The last estimate is due to the fact that \(\log(x)/x\) is clearly bounded by \((1 + \log(x))/x\) for \(x > 1\), and the latter is a decreasing function, and then by applying (3.3). Putting the estimates for \(r_1\) through \(r_5\) together proves (3.2).

Let now \(0 < m < 1\). As \(m^{2j}/(1-m^j)\) is a decreasing function in \(j\) (to see this, note that \(x^{2j}/(1-x)\) is increasing on \(0 < x < 1\)) we have
\[
\sum_{j=1}^{n} \frac{m^{2j}}{1-m^j} \leq \frac{m^2}{1-m} + \frac{m^{2n}}{1-m^n} + \int_{1}^{n} \frac{m^{2x}}{1-m^x}dx
\]
\[
= \frac{m^2}{1-m} + \frac{m^{2n}}{1-m^n} + \frac{m-m^n + \log\left( \frac{1-m}{1-m^n} \right)}{\log(m)},
\]
which implies that
\[
\eta(m, n) \leq \frac{1-m}{2(1-m^n)} + \frac{m(1-m)}{1-m^n} + \frac{2m^{2n}(1-m)^2}{m(1-m^n)^2} + \frac{(1-m)^2(m-m^n)}{m(1-m^n) \log(m)} + \frac{(1-m)^2 \log\left( \frac{1-m}{1-m^n} \right)}{m(1-m^n) \log(m)}
\]
\[
= r_1 + r_2 + r_3 + r_4 + r_5.
\]

Note that \( \frac{1-m^n}{1-m} = \sum_{k=0}^{n-1} m^k \geq m^n n \), which yields

\[
\frac{1-m}{1-m^n} \leq \frac{1}{m^n n} \iff \frac{m^n}{1-m^n} \leq \frac{1}{(1-m)n} \iff \frac{1-m}{1-m^n} \leq 1 + \frac{1}{(1-m)n} \iff \frac{1-m}{1-m^n} \leq 1 - m + \frac{1}{n}.
\]  

(3.5)

Hence

\[
 r_1 + r_2 \leq 1.5(1 - m) + \frac{1.5}{n}
\]

and

\[
r_3 = m^{2n-1} \left( \frac{1-m}{1-m^n} \right)^2 \leq 2(1-m)^2 + \frac{2}{n^2} \leq 2(1-m) + \frac{2}{n}.
\]

It is easy to see that \( r_4 \leq 0 \).

Using again that \( m - 1 \leq m\log(m) \),

\[
r_5 = \frac{(m-1)^2 \log\left(\frac{1-m}{1-m^n}\right)}{m(1-m^n) \log(m)} \leq \frac{\log\left(\frac{1-m^n}{1-m}\right)}{\frac{1-m}{1-m^n}}.
\]

Under the restriction \( \frac{1}{2} \leq m < 1 \) and \( n \geq 2 \), we can now apply Lemma 3.2 below for \( a = \frac{1-m^n}{1-m}, b = \frac{1}{1-m} \) and \( c = n \) due to (3.5) and we obtain

\[
r_5 \leq \frac{1+\log\left(\frac{1}{1-m}\right)}{1-m} + \frac{1+\log(n)}{n}.
\]

Putting all the estimates together proves (3.3).

\[\square\]

**Lemma 3.2.** Let \( a, b \) and \( c \) be real numbers bigger than 1 such that

\[
\frac{1}{a} \leq \frac{1}{b} + \frac{1}{c} \leq 1.
\]  

(3.6)

Then

\[
\frac{\log(a)}{a} \leq \frac{1+\log(b)}{b} + \frac{1+\log(c)}{c}.
\]

**Proof.** It is clear from the monotonicity of the logarithm function that for \( x, y > 0 \) we have

\[
x \log(x) + y \log(y) \leq (x+y) \log(x+y).
\]

Hence,

\[
(x+y)(1-\log(x+y)) \leq x(1-\log(x)) + y(1-\log(y)).
\]
Rewriting this inequality for \( x = 1/b \) and \( y = 1/c \), we have
\[
\frac{1 + \log \left( \frac{bc}{b+c} \right)}{bc/b+c} \leq \frac{1 + \log(b)}{b} + \frac{1 + \log(c)}{c}.
\]
Noting that \( \frac{1+\log(a)}{a} \) is a decreasing function for \( a \geq 1 \) and noting that \( a \geq \frac{bc}{b+c} \geq 1 \) from (3.6),
\[
\frac{\log(a)}{a} \leq \frac{1 + \log(a)}{a} \leq \frac{1 + \log(b)}{b} + \frac{1 + \log(c)}{c},
\]
which proves the claim. \( \square \)

4 VISITS TO THE ORIGIN FOR A TWO DIMENSIONAL SIMPLE RANDOM WALK

Exponential approximation results for sums of nonnegative random variables \( X_1, X_2, \ldots, X_n \) satisfying the condition \( \text{Var}(\mathbb{E}(X_i|X_1, \ldots, X_{i-1})) = 0 \) for all \( i \) were given in Peköz and Röllin (in press, Theorem 3.1), but not for more general dependent sums. Here we give an extension to sums of arbitrarily dependent nonnegative random variables, apply it to occupation times for Markov chains and then illustrate it by getting a new exponential approximation rate for the number of times a general two-dimensional random walk revisits the origin.

**Theorem 4.1.** Let \( W = \lambda \sum_{i=1}^{n} X_i \) where \( X_1, X_2, \ldots, X_n \) are (possibly dependent) nonnegative random variables and let \( \lambda = 1/\mathbb{E} \sum_{i=1}^{n} X_i \). Suppose \( W_i(x) \) is independent of all else and
\[
\mathcal{L}(W_i(x)) = \mathcal{L} \left( \lambda \sum_{m=1}^{i-1} X_m \bigg| X_i = x \right).
\]
Let \( I \) be independent of all else with \( \mathbb{P}[I = i] = \lambda \mathbb{E} X_i \), let \( U \) be a uniform random variable on \((0, 1)\), and let \( X_i^{\ast} \) be independent of all else with the size-biased distribution with respect to \( X_i \). Then
\[
W^e = W_I(X_i^{\ast}) + \lambda UX_i^{\ast}
\]
has the equilibrium distribution with respect to \( W \).
Proof. Let \( S_m = \lambda \sum_{i=1}^m X_i \). By first conditioning on \( I \) and \( U \), and using (2.2) and \( \mathcal{L}(S_i) = \mathcal{L}(W_i(X_i)) \) in the third line we get

\[
\mathbb{E} f'(W_\infty) = \lambda \sum_{i=1}^n \mathbb{E} X_i \int_0^1 \mathbb{E} \left\{ f'(W_i(X_i^s) + \lambda u X_i^s) \right\} du = 0.
\]

Remark 4.1. The argument goes through in the same way when instead we define

\[
\mathcal{L}'(W_i(x)) = \mathcal{L}' \left( \lambda \sum_{m=i+1}^n X_m \mid X_i = x \right).
\]

We next apply the above result to Markov chain occupation times. Our next result gives a bound on the error of the exponential approximation for the number of times a Markov chain revisits its starting state. More general asymptotic results of this type, but without explicit bounds on the error, go back to Darling and Kac (1957).

Corollary 4.2. Consider a Markov chain started at time zero in a state 0 and let \( X_i \) be the indicator for the event that the Markov chain is in state 0 at time \( i \). Let \( W_m = \lambda \sum_{i=1}^m X_i \) and \( \lambda = 1/\mathbb{E}[\sum_{i=1}^n X_i] \). Then writing \( W \equiv W_n \) we have

\[
d_W(\mathcal{L}(W), \text{Exp}(1)) \leq 2\lambda + 2\lambda^2 \sum_{i=1}^n \sum_{j=n-i+1}^n \mathbb{E} X_i \mathbb{E} X_j.
\]

Proof. Using the notation of the previous Theorem 4.1 and Remark 4.1 in this setting the strong Markov property gives

\[
\mathcal{L}(W_i) = \mathcal{L} \left( \lambda \sum_{m=i+1}^n X_m \mid X_i = 1 \right) = \mathcal{L} \left( \lambda \sum_{m=1}^{n-i} X_m \right)
\]

and so we can let

\[
W^e = \lambda \sum_{i=1}^{n-L} X_i + \lambda U,
\]
where $P[I = i] = \lambda E X_i$, and conditioning on $I$ gives

$$E|W - W^e| \leq \lambda + \lambda^2 \sum_{i=1}^{n} \sum_{j=n-i+1}^{n} EX_i EX_j$$

and then (2.3) gives the result. \hfill \Box

We next consider a general random walk on the two dimensional integer lattice started at the origin. As a consequence of Lawler and Limic (2010, p. 24) we have the following lemma.

**Lemma 4.3.** Let $Z_n$ be an irreducible and aperiodic random walk on $\mathbb{Z}^2$ with mean zero and finite third moment. Then there are positive constants $c_1$ and $c_2$ such that

$$\frac{c_1}{n} \leq P[Z_n = 0] \leq \frac{c_2}{n}$$

for sufficiently large $n$.

We are now able to give a bound on the error of the exponential approximation for the number of times the random walk revisits the origin. This type of result, for simple random walk, goes back to Erdős and Taylor (1960).

**Corollary 4.4.** Let $Z_n$ be an irreducible and aperiodic random walk on $\mathbb{Z}^2$ with mean zero and finite third moment. Let $R$ be the number of return visits to the origin by time $n$, and let $W = \lambda R$, where $\lambda = 1/\mathbb{E} R$. Then, there is constant $C$ independent of $n$ such that

$$d_W(\mathcal{L}(W), \mathcal{L}(\text{Exp}(1))) \leq \frac{C}{\log n}$$

for all $n$.

**Proof.** Let $X_n = I_{\{Z_n=0\}}$ be the indicator for the event that the random walk revisits the origin at time $n$. Lemma 4.3 gives $\lambda \leq C/\log n$ and thus the result follows from Corollary 4.2 and, where $C$ may be different (but independent of $n$) in each instance used,

$$\lambda^2 \sum_{i=1}^{n} \sum_{j=n-i+1}^{n} EX_i EX_j \leq \frac{C}{(\log n)^2} \sum_{i=1}^{n} \frac{i}{(n-i)} \leq \frac{C}{\log n}.$$  

\hfill \Box
Remark 4.2. The result for the two-dimensional simple random walk

\[ \sup_{a<x<b} |\mathbb{P}[W > x] - e^{-x}| \leq \frac{C \log \log n}{\log n} \]

for fixed \( a \) and \( b \) follows from Erdős and Taylor (1960, Eq. (3.10)), so the above corollary can be viewed as a complement and extension. Using the method of moments, Gärtner and Sun (2009, Theorem 1.1) give an argument for the analogous exponential limit theorem for general random walks, but without a rate of convergence.

ACKNOWLEDGMENTS

The authors are grateful to Andrew Barbour and Rongfeng Sun for stimulating discussions and for the suggestions (respectively) to study the nearly critical branching process and the Erdős-Taylor theorem. Both authors were partially supported by NUS research grant R-155-000-098-133.

REFERENCES

R. Arratia and L. Goldstein (2010). Size bias, sampling, the waiting time paradox, and infinite divisibility: when is the increment independent? Preprint. Available at http://arxiv.org/abs/1007.3910.

A. D. Barbour, L. Holst, and S. Janson (1992). Poisson approximation, volume 2 of Oxford Studies in Probability. The Clarendon Press Oxford University Press, New York. ISBN 0-19-852235-5. Oxford Science Publications.

N. H. Bingham (1988). On the limit of a supercritical branching process. J. Appl. Probab. 25, 215–228.

M. Brown (2006). Exploiting the waiting time paradox: applications of the size-biasing transformation. Probability in the Engineering and Informational Sciences 20, 195–230.

D. A. Darling and M. Kac (1957). On occupation times for markoff processes. Transactions of the American Mathematical Society 84, 444–458.

P. Erdős and S. J. Taylor (1960). Some problems concerning the structure of random walk paths. Acta Math. Acad. Sci. Hungar. 11, 137–162. (unbound insert).

K. S. Fahady, M. P. Quine, and D. Vere-Jones (1971). Heavy traffic approximations for the galton-watson process. Advances in Applied Probability 3, 282–300.

T. Fujim agar i (1980). On the extinction time distribution of a branching process in varying environments. Adv. Appl. Probab. 12, 350–366.

J. Gärtner and R. Sun (2009). A quenched limit theorem for the local time of random walks on \( \mathbb{Z}^2 \). Stochastic Process. Appl. 119, 1198–1215. ISSN 0304-4149.
A. L. Gibbs and F. E. Su (2002). On choosing and bounding probability metrics. *International Statistical Review / Revue Internationale de Statistique* **70**, 419–435.

G. F. Lawler and V. Limic (2010). *Random walk: a modern introduction*. Cambridge.

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

E. Peköz and A. Röllin (in press). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*

E. A. Peköz, A. Röllin, V. Čekanavičius, and M. Shwartz (2009). A three-parameter binomial approximation. *J. Appl. Probab.* **46**, 1073–1085.

S. Ross and E. Peköz (2007). *A second course in probability*. www.ProbabilityBookstore.com, Boston, MA.

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