Exponential approximation for the nearly critical Galton-Watson process and occupation times of Markov chains

Erol Peköz[*]   Adrian Röllin[†]

Abstract

In this article we provide new applications for exponential approximation using the framework of Peköz and Röllin (2011), 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.

[*]School of Management, Boston University, 595 Commonwealth Avenue, Boston, Massachusetts 02215; partially supported by NUS research grant R-155-000-098-133

[†]Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546; partially supported by NUS research grant R-155-000-098-133

A new framework for estimating the error of the exponential approximation was recently developed in Peköz and Röllin (2011), 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 (2011) 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 (2011) 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 distribution, where X^e has the equilibrium distribution with respect to X characterized by

$$P[X^e \leqslant x] = \frac{1}{EX} \int_0^x P[X > y] dy. \tag{1.1}$$

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 (2011) 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

$$\sup_{x \geqslant 0} \left| P[X > x] - e^{-x/EX} \right| \leqslant 2.46 \sqrt{E|X - X^e|}.$$

Some heuristics for Stein's method can be understood using size-biased random variables. For a nonnegative continuous random variable X with probability density function $f(x)$, the size-biased random variable X^s has density $xf(x)/EX$. The size of the renewal interval containing a randomly chosen point as well as the number of 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 metrics we use below. For two probability distributions F and G define the Kolmogorov metric as

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

If both distributions have finite expectation, define the Wasserstein metric

$$d_W(F, G) = \int_R |F(x) - G(x)| dx.$$

We can relate the two metrics using

$$d_K(P, \operatorname{Exp}(1)) \leqslant 1.74 \sqrt{d_W(P, \operatorname{Exp}(1))};$$

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

Central to the approach in Peköz and Röllin (2011) 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

$$Ef(X) - f(0) = EX \, Ef'(X^e). \tag{2.1}$$

It is straightforward that this implies (1.1). Indeed for nonnegative X having finite first moment, define the distribution function

$$F^e(x) = \frac{1}{EX} \int_0^x P[X > y] dy$$

on $x \geqslant 0$ and $F^e(x) = 0$ for $x < 0$. Then

$$Ef(X) - f(0) = E \int_0^X f'(s) ds$$

$$= E \int_0^\infty f'(s) I[X > s] ds = \int_0^\infty f'(s) P[X > s] ds$$

so that F^e 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

$$E\{Xf(X)\} = EX \, Ef(X^s). \tag{2.2}$$

It may be helpful in what follows to note that this definition using $f(x) = x^n$ immediately gives $E(X^s)^n = EX^{n+1}/EX$. We next present the key result from Peköz and Röllin (2011) that we will use in the applications that follow.

Theorem 2.1 (Peköz and Röllin (2011), Theorem 2.1). *Let W be a non-negative random variable with $EW = 1$ and let W^e have the equilibrium distribution w.r.t. W. Then, for any $\beta > 0$,*

$$d_K\big(\mathscr{L}(W), \mathrm{Exp}(1)\big) \leqslant 12\beta + 2P[|W^e - W| > \beta],$$

and, if in addition W has finite second moment,

$$d_W\big(\mathscr{L}(W), \mathrm{Exp}(1)\big) \leqslant 2E|W^e - W| \tag{2.3}$$

3   THE NEARLY CRITICAL GALTON-WATSON BRANCHING PROCESS

Consider the Galton-Watson branching process starting from a single particle in generation zero, where each particle has an independent and identically distributed number of children according to some distribution having mean m; let Z_n be the size of the nth generation. For the critical case where $m = 1$ and when $EZ_1^2 < \infty$ and $P(Z_1 = 0) > 0$, it was shown by Yaglom (1947) that the conditional distribution of Z_n/n given $Z_n > 0$ converges as $n \to \infty$ to an exponential distribution. A corresponding rate of convergence was first proved with the additional condition $EZ_1^3 < \infty$ by Peköz and Röllin (2011). In the super- and sub-critical cases, respectively when $m > 1$ and $m < 1$, the limiting distributions are very difficult to calculate and are only known explicitly in very special cases; see e.g. Bingham (1988). Fahady, Quine, and Vere-Jones (1971), 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 general classes of offspring distributions. The following theorem gives explicit error bounds for the exponential approximation for any finite n and any $m \neq 1$. To avoid trivial cases, we make the general assumption that

$$0 < P[Z_1 = 0] < 1.$$

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 nth generation. Assume $P[Z_1 \geqslant 2] > 0$, $m = EZ_1 \neq 1$ and $EZ_1^3 < \infty$. Let $\alpha = P[Z_1 = 1]/P[Z_1 \geqslant 2]$,*

$$C = (2 + \alpha)^2 \big(1 + \mathrm{Var}\, Z_1 + EZ_1^3\big)^2 \tag{3.1}$$

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

$$d_W\big(\mathscr{L}(\lambda Z_n | Z_n > 0), \mathrm{Exp}(1)\big) \leqslant C\eta(m, n), \tag{3.2}$$

where

$$\eta(m, n) = \frac{1 - m}{1 - m^n} + \frac{(1 - m)^2}{m(1 - m^n)} \sum_{j=1}^{n-1} \frac{m^{2j}}{1 - m^j}. \tag{3.3}$$

It seems difficult to directly deduce rates of convergence from (3.3). The following estimates are more useful (a proof is given in the Appendix).

Lemma 3.2. *For any $n \geqslant 1$ and $m > 1$,*

$$\eta(m,n) \leqslant 2(m-1) + \frac{3 + \log(n)}{n}, \tag{3.4}$$

and, for any $n \geqslant 2$ and $\frac{1}{2} \leqslant m < 1$,

$$\eta(m,n) \leqslant \left(4 - 2\log(1-m)\right)(1-m) + \frac{4 + 2\log(n)}{n}. \tag{3.5}$$

Fahady, Quine, and Vere-Jones (1971) considered classes $K(a,b)$, $0 < a < \infty$, $0 < b < \infty$, of offspring distributions such that for all $\mathscr{L}(Z_1) \in K(a,b)$,

$$(A) \ EZ_1^3 \leqslant a, \qquad (B) \ EZ_1(Z_1 - 1) \geqslant b,$$

and showed that within each such class the limiting distribution of the conditioned Galton-Watson branching process converges to the exponential as $m \to 1$. While retaining (A), it is not too difficult to see that Condition (B) is equivalent to

$$(B') \ P[Z_1 \geqslant 2] \geqslant b'$$

for some $b' > 0$ (it is easy to see that (B') implies (B)—a proof of the reverse is given in the Appendix). Hence, it is clear that under these assumptions, the constant C in (3.1) will remain bounded as $m \to 1$ and hence Theorem 3.1 and Lemma 3.2 give explicit bounds under the conditions of Fahady, Quine, and Vere-Jones (1971). Thanks to our explicit bounds, we can furthermore weaken the assumptions on the offspring distributions in the sense that the third moment of Z_1 may grow and $P[Z_1 \geqslant 2] \to 0$ as long as $C = o\left(\frac{1}{m-1}\right)$ if $m \searrow 1$, respectively, $C = o\left(\frac{-1}{(1-m)\log(1-m)}\right)$ if $m \nearrow 1$.

Proof of Theorem 3.1. With some modifications, we follow the line of argument from Peköz and Röllin (2011), 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 (the other offspring from the same parent) 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

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

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

(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) $\mathcal{L}(R^*_n) = \mathcal{L}(Z_n|Z_n > 0)$; $\qquad$ (v) $E\{R'_{n,j}I_{A^c_{n,j}}\} \leqslant m^{n-j}\sigma^2 P[A^c_{n,j}]$;

(vi) $E\{R_{n,j}I_{A^c_{n,j}}\} \leqslant m^{n-j}\gamma P[A^c_{n,j}]$; $\qquad$ (vii) $P[A^c_{n,j}] \leqslant m^{-1}\sigma^2 P[Z_{n-j} > 0]$.

$(viii)$ $P[Z_n > 0] \leqslant (2+\alpha)\dfrac{m^n(1-m)}{1-m^n}$

For (i)-(iv) see Peköz and Röllin (2011). Using independence,

$$E\{R'_{n,j}I_{A^c_{n,j}}\} = ER'_{n,j}P[A^c_{n,j}] \leqslant ES_{n,j}P[A^c_{n,j}] \leqslant m^{n-j}\sigma^2 P[A^c_{n,j}],$$

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

$$
\begin{aligned}
E\{R_{n,j}I_{A^c_{n,j}}\} &\leqslant m^{n-j}E\{X_jI_{A^c_{n,j}}\} \leqslant m^{n-j}\sum_k kP[X_j = k, A^c_{n,j}] \\
&\leqslant m^{n-j}\sum_k kP[X_j = k]P[A^c_{n,j}|X_j = k] \\
&\leqslant m^{n-j}\sum_k k^2 P[X_j = k]P[A^c_{n,j}] \\
&\leqslant m^{n-j}EX_i^2 P[A^c_{n,j}] \leqslant m^{n-j}\gamma P[A^c_{n,j}],
\end{aligned}
$$

hence (vi). Now,

$$P[A^c_{n,j}] = E\{P[A^c_{n,j}|X_j]\} \leqslant E\{X_j P[Z_{n-j} > 0]\} = m^{-1}\sigma^2 P[Z_{n-j} > 0],$$

which proves (vii). Finally, using the Corollary on page 356 of Fujimagari (1980), we have

$$P[Z_n > 0] \leqslant (2+\alpha)\frac{1-m}{m^{-n}-1},$$

which is $(viii)$ (note that the result cited is for bounded offspring distribution, but easily extends to the unbounded case).

Set $W = \lambda R^*_n$, and note that, due to (iv), $\mathcal{L}(W) = \mathcal{L}(Z_n|Z_n > 0)$. Due to (i) and (ii), S_n has the size-biased distribution with respect to R^*_n. Let U be a uniform random variable on $[0,1]$, independent

of all else. Note that, if Y is a random variable, uniformly distributed on the integers $\{1, \ldots, n\}$, then $Y - U$ is continuous and uniformly distributed on $[0, n]$. Observing that, given S_n, R_n has uniform distribution on $\{1, \ldots, S_n\}$ because of (iii), we therefore deduce that $R_n - U$ has uniform distribution on $[0, S_n]$. Hence, $\mathscr{L}(R_n - U) = \mathscr{L}(US_n)$, which implies that we can set $W^e = \lambda(R_n - U)$. Applying (2.3) and using (v)–(vii), we obtain

$$
\begin{aligned}
E|R_n^* - R_n| &\leqslant \sum_{j=1}^{n} E\{R'_{n,j} I_{A^c_{n,j}} + R_{n,j} I_{A^c_{n,j}}\} \\
&\leqslant \sum_{j=1}^{n} m^{n-j} (\sigma^2 + \gamma) P[A^c_{n,j}] \\
&\leqslant \sigma^2 + \gamma + \sum_{j=1}^{n-1} \frac{m^{n-j}(\sigma^2 + \gamma)\sigma^2 P[Z_{n-j} > 0]}{m} \\
&\leqslant \sigma^2 + \gamma + \sum_{j=1}^{n-1} \frac{m^{n-j}(\sigma^2 + \gamma)\sigma^2 (2+\alpha) m^{n-j}(1-m)}{m(1 - m^{n-j})} \\
&\leqslant \sigma^2 + \gamma + \frac{1-m}{m}(2+\alpha)(\sigma^2 + \gamma)\sigma^2 \sum_{j=1}^{n-1} \frac{m^{2j}}{1 - m^j},
\end{aligned}
$$

and, using

$$
\lambda = \frac{P[Z_n > 0]}{m^n} \leqslant \frac{(2+\alpha)(1-m)}{1 - m^n},
$$

we obtain

$$
E|W - W^e| \leqslant \lambda/2 + \lambda E|R_n^* - R_n| \leqslant C\eta(m, n),
$$

which proves (3.2). $\qquad\square$

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 $\mathrm{Var}(E(X_i | X_1, \ldots X_{i-1})) = 0$ for all i were given in Peköz and Röllin (2011, Theorem 3.1), but not for more general dependent sums. Here we give a construction of the equilibrium distribution for sums of arbitrarily dependent nonnegative random variables having finite means, 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 irreducible aperiodic two-dimensional integer-valued 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 $\lambda = 1/E \sum_{i=1}^{n} X_i$. Suppose, for each i and each x, $W_i(x)$ is a random variable such that*

$$
\mathscr{L}(W_i(x)) = \mathscr{L}\left(\lambda \sum_{m=1}^{i-1} X_m \,\middle|\, X_i = x\right).
$$

For each i, let X_i^s be a random variable having the size-biased distribution of X_i. Let I be independent of all else with $P[I = i] = \lambda E X_i$ and let U be a uniform random variable on $(0, 1)$, independent of all else. Then $W_I(X_I^s) + \lambda U X_I^s$ has the equilibrium distribution with respect to W. In particular, if $X_i \in \{0, 1\}$ for all i, we have $X_i^s = 1$ and hence $W_I(1) + \lambda U$ has the equilibrium distribution with respect to W.

1387

Proof. Let $S_m = \lambda \sum_{i=1}^{m} X_i$. By first conditioning on I and U, and using (2.2) and $\mathscr{L}(S_i) = \mathscr{L}(W_i(X_i))$ for the third equality, we obtain

$$
\begin{aligned}
E f'\big(W_I(X_I^s) + \lambda U X_I^s\big) \\
&= \sum_{i=1}^{n} \lambda E X_i \int_0^1 E\{f'(W_i(X_i^s) + \lambda u X_i^s)\} du \\
&= \sum_{i=1}^{n} \lambda \int_0^1 E\{X_i(f'(W_i(X_i) + \lambda u X_i))\} du \\
&= \sum_{i=1}^{n} E\{f(W_i(X_i) + \lambda X_i) - f(W_i(X_i))\} \\
&= \sum_{i=1}^{n} E\{f(S_i) - f(S_{i-1})\} = Ef(W) - f(0). \qquad \square
\end{aligned}
$$

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

$$
\mathscr{L}(W_i(x)) = \mathscr{L}\left(\lambda \sum_{m=i+1}^{n} X_m \,\Big|\, X_i = x \right).
$$

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 $\lambda = 1/E[\sum_{i=1}^{n} X_i]$. Then writing $W \equiv W_n$ we have*

$$
d_{\mathrm{W}}\big(\mathscr{L}(W), \mathrm{Exp}(1)\big) \leqslant 2\lambda + 2\lambda^2 \sum_{i=1}^{n} \sum_{j=n-i+1}^{n} E X_i \, E X_j
$$

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

$$
\mathscr{L}(W_i) = \mathscr{L}\left(\lambda \sum_{m=i+1}^{n} X_m \,\Big|\, X_i = 1 \right) = \mathscr{L}\left(\lambda \sum_{m=1}^{n-i} X_m \right)
$$

and so we can let

$$
W^e = \lambda \sum_{i=1}^{n-I} X_i + \lambda U,
$$

where $P[I = i] = \lambda E X_i$, and conditioning on I gives

$$
E|W - W^e| \leqslant \lambda + \lambda^2 \sum_{i=1}^{n} \sum_{j=n-i+1}^{n} E X_i E X_j
$$

and then (2.3) gives the result. $\qquad \square$

We next consider a general aperiodic irreducible 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 Z^2 with mean zero and finite third moment. Then there are positive constants c_1 and c_2 such that*

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

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 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 $\lambda = 1/ER$. Then, there is constant C independent of n such that*

$$d_W(\mathscr{L}(W), \mathscr{L}(\mathrm{Exp}(1))) \leqslant \frac{C}{\log n}.$$

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 $\lambda \leqslant C/\log n$ and thus the result follows from Corollary 4.2 and, where C may be different (but independent of n) in each instance used,

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

□

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

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

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.

A PROOF OF LEMMA 3.2

We first need some simple estimates.

Lemma A.1. *Let a, b and c be real numbers, strictly greater than 1, such that*

$$\frac{1}{a} \leqslant \frac{1}{b} + \frac{1}{c} \leqslant 1. \tag{A.1}$$

Then

$$\frac{\log(a)}{a} \leqslant \frac{1+\log(b)}{b} + \frac{1+\log(c)}{c}.$$

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

$$x \log(x) + y \log(y) \leqslant (x + y) \log(x + y).$$

Hence,

$$(x + y)(1 - \log(x + y)) \leqslant x(1 - \log(x)) + y(1 - \log(y)).$$

Rewriting this inequality for $x = 1/b$ and $y = 1/c$, we have

$$\frac{1 + \log\left(\frac{bc}{b+c}\right)}{\frac{bc}{b+c}} \leqslant \frac{1 + \log(b)}{b} + \frac{1 + \log(c)}{c}.$$

Noting that $\frac{1+\log(a)}{a}$ is a decreasing function for $a \geqslant 1$ and noting that $a \geqslant \frac{bc}{b+c} \geqslant 1$ from (A.1),

$$\frac{\log(a)}{a} \leqslant \frac{1 + \log(a)}{a} \leqslant \frac{1 + \log\left(\frac{bc}{b+c}\right)}{\frac{bc}{b+c}},$$

which proves the claim. $\qquad\square$

Let f be a non-negative function on $[a, b]$ for two integers a and b. If f is either increasing, decreasing or has exactly one minimum, a simple geometric argument yields that

$$\sum_{j=a}^{b-1} f(j) \leqslant f(a) + \int_a^b f(x)dx \tag{A.2}$$

(this estimate is not optimal if the function is increasing, but we want to avoid further case distinctions).

Proof of Lemma 3.2 for $m > 1$. It is straightforward to see that $f(x) = m^{2x}/(m^x - 1)$, $x > 0$, has exactly one minimum at $x_0 = \log(2)/\log(m)$, hence $f(x)$ is decreasing on $0 < x \leqslant x_0$ and increasing on $x \geqslant x_0$. Using (A.2), we therefore have

$$\sum_{j=1}^{n-1} \frac{m^{2j}}{m^j - 1} \leqslant \frac{m^2}{m-1} + \frac{m^n - m + \log\left(\frac{m^n-1}{m-1}\right)}{\log(m)},$$

which implies that

$$\eta(m,n) \leqslant \frac{m-1}{m^n-1} + \frac{m(m-1)}{m^n-1} + \frac{(m-1)^2(m^n-m)}{m(m^n-1)\log(m)}$$
$$+ \frac{(m-1)^2\log\left(\frac{m^n-1}{m-1}\right)}{m(m^n-1)\log(m)} =: r_1 + r_2 + r_3 + r_4.$$

Recall that

$$\frac{m^n-1}{m-1} = \sum_{k=0}^{n-1} m^k \geqslant n. \tag{A.3}$$

This implies that

$$r_1 \leqslant \frac{1}{n}, \qquad r_2 \leqslant \frac{m}{n} = \frac{m-1}{n} + \frac{1}{n} \leqslant m - 1 + \frac{1}{n}.$$

Furthermore, recalling that $m - 1 \leqslant m\log(m)$,

$$r_3 \leqslant \frac{(m-1)(m^n-m)}{m^n-1} \leqslant m - 1.$$

Finally,

$$r_4 = \frac{(m-1)^2\log\left(\frac{m^n-1}{m-1}\right)}{m(m^n-1)\log(m)} \leqslant \frac{m-1}{m^n-1}\log\left(\frac{m^n-1}{m-1}\right) \leqslant \frac{1+\log(n)}{n}.$$

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 (A.3). Putting the estimates for r_1 through r_4 together proves (3.4). $\qquad\square$

Proof of Lemma 3.2 for $m < 1$. Note first that $\frac{y^2}{1-y}$ is increasing on $0 < x < 1$, hence $f(x) = m^{2x}/(1-m^x)$ is a decreasing function in x. Applying (A.2),

$$\sum_{j=1}^{n-1} \frac{m^{2j}}{1-m^j} \leqslant \frac{m^2}{1-m} + \frac{m - m^n + \log\left(\frac{1-m}{1-m^n}\right)}{\log(m)},$$

which implies that

$$\eta(m,n) \leqslant \frac{1-m}{1-m^n} + \frac{m(1-m)}{1-m^n} + \frac{(1-m)^2(m-m^n)}{m(1-m^n)\log(m)}$$
$$+ \frac{(1-m)^2\log\left(\frac{1-m}{1-m^n}\right)}{m(1-m^n)\log(m)} =: r_1 + r_2 + r_3 + r_4.$$

As $\frac{1-m^n}{1-m} = \sum_{k=0}^{n-1} m^k \geqslant m^n n$, we have

$$\frac{1-m}{m^n} + \frac{1-m}{1-m^n} \leqslant \frac{1-m}{m^n} + \frac{1}{m^n n} \quad \Rightarrow \quad \frac{1-m}{1-m^n} \leqslant 1 - m + \frac{1}{n}. \tag{A.4}$$

Hence

$$r_1 + r_2 \leqslant 2(1-m) + \frac{2}{n}.$$

1391

It is easy to see that $r_3 \leqslant 0$. Using now that $1 - m \leqslant -2m \log(m)$ for $\frac{1}{2} \leqslant m < 1$,

$$r_4 = \frac{(1-m)^2 \log\left(\frac{1-m}{1-m^n}\right)}{m(1-m^n)\log(m)} \leqslant \frac{2\log\left(\frac{1-m^n}{1-m}\right)}{\frac{1-m^n}{1-m}}.$$

Under the restriction $\frac{1}{2} \leqslant m < 1$ and $n \geqslant 2$, we can now apply Lemma A.1 below for $a = \frac{1-m^n}{1-m}$, $b = \frac{1}{1-m}$ and $c = n$ due to (A.4) and we obtain

$$r_4 \leqslant 2\left(\frac{1+\log\left(\frac{1}{1-m}\right)}{\frac{1}{1-m}} + \frac{1+\log(n)}{n}\right).$$

Putting all the estimates together proves (3.5). $\qquad\square$

B   PROOF THAT CONDITION (B) IMPLIES (B′)

We will show that $\neg(B')$ implies $\neg(B)$. More specific, under the Condition (A), we show that $P[X \geqslant 2] \to 0$ implies $EX(X-1) \to 0$.

To do this we need to find a vector of probabilities $p_0, p_1, \ldots p_n$ that maximizes

$$\sum_{k=0}^{n} k(k-1)p_k,$$

subject to the constraints

$$\sum_{k=0}^{n} k^3 p_k \leqslant a, \qquad \sum_{k=2}^{n} p_k \leqslant \varepsilon, \qquad p_k \geqslant 0, \ \ k = 0, 1, \ldots n, \qquad \sum_{k=0}^{n} p_k = 1.$$

This is a linear programming problem with $n+1$ variables and $n+4$ constraints. The constraints define a simplex and the fundamental theorem of linear programming tells us the maximum is achieved at a corner point of the simplex where there are $n+1$ binding constraints; this means at most three of the variables p_k can be non-zero at the maximum. As we assume $p_0 > 0$, we have therefore reduced the problem to just looking at three-point distributions, where one of the three points is at 0.

We consider first the case where neither points are at 1, where we are now trying to find x, y and p, q that maximizes

$$x(x-1)p + y(y-1)q$$

subject to the constraints

$$x^3 p + y^3 q \leqslant a, \qquad p + q \leqslant \varepsilon, \qquad p, q \geqslant 0, \qquad x, y \geqslant 2.$$

Since the first constraint gives $x \leqslant (a/p)^{1/3}$ and $y \leqslant (a/q)^{1/3}$ we have

$$x(x-1)p + y(y-1)q \leqslant a^{2/3}p^{1/3} + a^{2/3}q^{1/3} \leqslant 2a^{2/3}\varepsilon^{1/3} \to 0$$

as $\varepsilon \to 0$ and we get our intended result. The case of a point at $y = 1$ works out the same way.

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. MR1163825

N. H. Bingham (1988). On the limit of a supercritical branching process. *J. Appl. Probab.* **25**, 215–228. MR0974583

M. Brown (2006). Exploiting the waiting time paradox: applications of the size-biasing transformation. *Probability in the Engineering and Informational Sciences* **20**, 195–230. MR2261286

D. A. Darling and M. Kac (1957). On occupation times for markoff processes. *Transactions of the American Mathematical Society* **84**, 444–458. MR0084222

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). MR0121870

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. MR0288858

T. Fujimagari (1980). On the extinction time distribution of a branching process in varying environments. *Adv. Appl. Probab.* **12**, 350–366. MR0569432

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. MR2508570

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. MR2677157

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. MR1349164

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

E. A. Peköz, A. Röllin, V. Čekanavičius, and M. Shwartz (2009). A three-parameter binomial approximation. *J. Appl. Probab.* **46**, 1073–1085. MR2582707

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. MR0022045