Finite Branching Processes and AND/OR Tree Evaluation

Richard M. Karp *   Yanjun Zhang†

TR-93-043

March 1994

Abstract

This paper studies tail bounds on supercritical branching processes with finite distributions of offspring. Given a finite supercritical branching process $\{Z_n\}_0^\infty$, we derive upper bounds, decaying exponentially fast as c increases, on the right-tail probability $\Pr[Z_n > cE(Z_n)]$. We obtain a similar upper bound on the left-tail probability $\Pr[Z_n < \frac{E(Z_n)}{c}]$ under the assumption that each individual generates at least two offspring. As an application, we observe that the evaluation of an AND/OR tree by a canonical algorithm in certain probabilistic models can be viewed as a two-type supercritical finite branching process, and show that the execution time of this algorithm is likely to concentrate around its expectation.

*Supported in part by NSF/DARPA Grant CCR-9005448. Author's address: Computer Science Division, University of California, Berkeley, CA 94720 and International Computer Science Institute, Berkeley, CA 94704.

†Supported in part by NSF Grant CCR-9110839 for this research. Author's address: Department of Computer Science and Engineering, Southern Methodist University, Dallas, TX 75275. Email: zhang@seas.smu.edu.

Key Words: Branching process, large deviation of random variables, multitype branching process, AND/OR tree evaluation.

1 Introduction

A *branching process* $\{Z_n\}_0^\infty$ is a discrete Markov chain such that $Z_0 = 1$, Z_1 is a non-negative discrete random variable, and $Z_n = \sum_{i=1}^{Z_{n-1}} X_i$ where $X_i \sim Z_1$ are independent of each other and independent of Z_{n-1}. The probabilities $p_j = \Pr[Z_1 = j]$ for $j = 0, 1, 2, \ldots$, are the *offspring distribution* of the branching process. Random variable Z_n is the *nth generation* of the branching process. A branching process can be viewed as a process, starting with a single element, of generating offspring in successive generations such that each element of one generation independently generates j offspring of the next generation with probability p_j. The random variable Z_n is thus the number of offspring in the nth generation of this process. For convenience, we may also use Z_n to refer to the set of offspring in the nth generation. For an introduction to the theory of branching processes, see [5], [2].

The *branching factor* of a branching process is $\rho = E(Z_1) = \sum_{j=0}^{\infty} jp_j$, which is the expected number of offspring of a single element. The expected number of elements in the nth generation is $E(Z_n) = E(E(Z_n|Z_{n-1})) = E(E(\sum_{i=1}^{Z_{n-1}} X_i|Z_{n-1})) = E(Z_1)E(Z_{n-1}) = \rho E(Z_{n-1}) = \rho^n$ as $X_i \sim Z_1$ are independent of Z_{n-1}. A branching process is *supercritical* if $\rho = E(Z_1) > 1$. It can be shown that, unless $p_1 = 1$, a non-supercritical process will, with probability 1, vanish after some generations, i.e., $Z_n = 0$ for some n, whereas a supercritical process has a positive probability of never vanishing. A branching process is *finite* if $p_j = 0$ for $j \geq d$ where d is some fixed positive integer, i.e., each element generates at most d offspring for some fixed d.

In this paper, we study large deviations of the random variable Z_n associated with a finite supercritical branching process. We shall derive bounds on the probabilities that Z_n is substantially larger or smaller than its expectation ρ^n. In Section 2, we derive two upper bounds on the right-tail probability $\Pr[Z_n > c\rho^n]$ that decay exponentially fast as c increases. The first bound is for the special case $\rho > \sqrt{d}$, and is obtained by using Azuma's martingale inequality. The second bound is a general bound, and is obtained by a careful application of tail bounds on the sum of independent bounded variables. These bounds are stronger than a previous bound for general supercritical branching processes. In Section 4, we derive an exponential upper bound on the left-tail probability $\Pr[Z_n < \rho^n/c]$ under the assumption that each individual generates at least two offspring. In Section 3, we extend the special-case right-tail bound in Section 2 to multitype branching processes with a growth rate greater than $\sqrt{d}$. All these bounds are independent of n, because the tail probabilities $\Pr[Z_n > c\rho^n]$ and $\Pr[Z_n < \rho^n/c]$ in general do not diminish as n increases.

Our motivation for studying finite branching processes originated from the problem of evaluating AND/OR trees. The execution of a canonical algorithm for evaluating uniform AND/OR trees in certain probabilistic models can be viewed as a two-type finite branching process. The probability that the running time of this algorithm deviates from its expected value corresponds to the tail probabilities of the associated two-type branching process. As an application, we show that the running time of this algorithm for evaluating uniform AND/OR trees is likely to concentrate around its expectation.

2 Right-Tail Bounds

In this section we study the right tail of the nth generation Z_n of a finite supercritical branching process $\{Z_n\}$. The following theorem, due to K.B. Athreya, gives an exponential upper bound in c on the probability $\Pr[Z_n > c\rho^n]$ for supercritical branching processes.

Theorem 1 *Let $\{Z_n\}$ be a supercritical branching process with $\rho = E(Z_1) > 1$. Suppose that $E(e^{Z_1}) < \infty$. Then there exist constants $t_0 > 0$ and $C(t_0) > 0$ such that* (i) *$E(e^{t_0 Z_n/\rho^n}) < C(t_0)$ for all n and* (ii) *for any $c > 1$, $\Pr[Z_n > c\rho^n] < C(t_0)e^{-t_0 c}$.*

Proof. For (i), see [1, Theorem 4]. For (ii), by the Markov inequality and (i), for $t_0 > 0$, $\Pr[Z_n > c\rho^n] = \Pr[e^{t_0 Z_n/\rho^n} > e^{t_0 c}] \leq e^{-t_0 c}E(e^{t_0 Z_n/\rho^n}) < C(t_0)e^{-t_0 c}$. □

The upper bound in Theorem 1 holds for general branching processes that are not necessarily finite. For finite branching processes, we can derive stronger upper bounds on the right tail. We shall derive two such bounds. One is a special bound for the case $\rho > \sqrt{d}$. This bound will be extended to the multitype case in Section 3. The other is a general bound for any finite supercritical branching process.

2.1 A special bound

This section gives an upper bound on the right-tail probability $\Pr[Z_n \geq c\rho^n]$ for the special case $\rho > \sqrt{d}$. This bound uses the tools of martingales. A sequence of random variables $Y_0, Y_1, \ldots, Y_n$ is called a *martingale* if $E(Y_{i+1}|Y_1, Y_2, \ldots, Y_i) = Y_i$ for $i = 0, 1, \ldots, n-1$. An example of a martingale is $Y_i = \sum_{k=1}^{i} X_k$ for $i = 0, 1, \ldots, n$, where $X_1, X_2, \ldots, X_n$ are independent random variables with mean 0. Another example is the *Doob* martingale $Y_i = E(Z|X_1, X_2, \ldots, X_i)$ where Z is a random variable and $X_1, X_2, \ldots, X_n$ are a sequence of random variables [8, 4]. The following lemma is an important inequality on martingales due to K. Azuma [3] (for a proof, see [16, pp. 55]).

Lemma 1 (Azuma's Inequality) *Let $\{Y_i\}_1^n$ be a martingale such that $|Y_i - Y_{i-1}| \leq a_i$ for $i = 1, 2, \ldots, n$ where $\sum_1^n a_i > 0$. Then for all $\lambda > 0$,* (i) *$\Pr[Y_n - Y_0 \geq \lambda] \leq e^{-\frac{1}{2}\lambda^2/\sum_{i=1}^n a_i^2}$ and* (ii) *$\Pr[Y_0 - Y_n \geq \lambda] \leq e^{-\frac{1}{2}\lambda^2/\sum_{i=1}^n a_i^2}$.*

Theorem 2 *Let $\{Z_n\}_0^\infty$ be a finite branching process such that each element of the process has at most d offspring and $\rho = E(Z_1) > \sqrt{d}$. Then, for any $c > 1$,*

$$\Pr[Z_n > c\rho^n] \leq e^{-\alpha(c-1)^2},$$

where $\alpha > 0$ is a constant depending on d and ρ.

Proof. We associate with the branching process $\{Z_0, Z_1, \ldots, Z_n\}$ a Doob martingale defined as follows. For $0 \le m \le n-1$ and $1 \le i \le Z_m$, let $X_{m,i} \sim Z_1$ be the number of offspring generated by the ith element of Z_m in some order, and let $S_m = \{X_{m,1}, X_{m,2}, \ldots, X_{m,Z_m}\}$. Note $Z_m \le d^m$. For $0 \le m \le n-1$ and $1 \le i \le d^m$, let

$$Y_{m,i} = \begin{cases} E(Z_n \mid S_0, S_1, \ldots, S_{m-1}, X_{m,1}, X_{m,2}, \ldots, X_{m,i}), & \text{if } 1 \le i \le Z_m; \\ Y_{m,Z_m}, & \text{if } Z_m < i \le d^m. \end{cases} \tag{1}$$

Let $N = \sum_{m=0}^{n-1} d^m$. Then the sequence of $N+1$ random variables

$$Y_0 = E(Z_n), Y_{0,1}, Y_{1,1}, \ldots, Y_{1,d}, Y_{2,1}, \ldots, Y_{2,d^2} \ldots, Y_{n-1,1}, \ldots, Y_{n-1,d^{n-1}} = Y_N$$

is a Doob martingale with $Y_0 = E(Z_n) = \rho^n$ and $Y_N = Z_n$.

We shall bound the differences of this martingale. Define $D_{m,i}$ as follows for $0 \le m \le n-1$ and $1 \le i \le d^m$.

$$\begin{aligned} D_{0,1} &= |Y_{0,1} - Y_0|, \\ D_{m,1} &= |Y_{m,1} - Y_{m-1,d^{m-1}}| \text{ for } 1 \le m \le n-1, \\ D_{m,i} &= |Y_{m,i} - Y_{m,i-1}| \text{ for } 2 \le i \le d^m. \end{aligned}$$

Consider $D_{0,1} = |Y_{0,1} - Y_0| = |E(Z_n|X_{0,1}) - E(Z_n)|$. Random variable $X_{0,1} \le d$ is the number of offspring of the initial element, and each of these offspring is expected to contribute ρ^{n-1} descendants in Z_n. Hence, $D_{0,1} \le d\rho^{n-1}$. For each subsequent non-zero difference, the difference $D_{m,i}$ is due to the additional contribution of the ith element in Z_m with $X_{m,i} \le d$ offspring, each of which is expected to contribute ρ^{n-m-1} descendants in Z_n. So $D_{m,i} \le d\rho^{n-m-1}$. Hence, for $0 \le m \le n-1$ and $1 \le i \le d^m$,

$$D_{m,i} \le d\rho^{n-m-1} = a_{m,i} \tag{2}$$

and

$$\begin{aligned} A &= \sum_{m=0}^{n-1} \sum_{i=1}^{d^m} a_{m,i}^2 = \sum_{m=0}^{n-1} d^{m+2} \rho^{2(n-m-1)} \\ &= d\rho^{2n} \left(\frac{d}{\rho^2} + \left(\frac{d}{\rho^2}\right)^2 \cdots \left(\frac{d}{\rho^2}\right)^n \right) \le c_0 d \rho^{2n} \end{aligned} \tag{3}$$

since $\rho^2 > d$ and $\sum_{m=1}^{n} (d/\rho^2)^m \le c_0$ for some constant c_0 depending on d and ρ.

Let $\lambda = (c-1)\rho^n > 0$. Then $\Pr[Z_n > c\rho^n] = \Pr[Y_N - Y_0 > \lambda]$ as $Y_N = Z_n$ and $Y_0 = E(Z_n) = \rho^n$. By Lemma 1,

$$\Pr[Z_n > c\rho^n] \le e^{-\frac{1}{2}\lambda^2/A} \le e^{-\frac{1}{2}(c-1)^2\rho^{2n}/c_0 d\rho^{2n}} = e^{-\alpha(c-1)^2}$$

where $\alpha = \frac{1}{2c_0 d} > 0$. □

2.2 A General Bound

This section gives a right-tail bound that holds for any finite supercritical branching process. This bound is asymptotically comparable to the bound in Theorem 2 when $\rho > \sqrt{d}$. The technique used to derive this bound is a careful application of tail bounds on the sum of bounded random variables with identical means. This technique will be used again in Section 4 for an upper bound on the left-tail probability, and in Section 5 in connection with AND/OR tree evaluation.

Lemma 2 *Let* $X = \sum_{i=1}^{n} X_i$ *where the* X_i, $1 \leq i \leq n$, *are independent random variables over* $\{0, 1, \ldots, d\}$ *with the same mean* $\mu > 0$. *Then* (i) *for* $\beta > 0$, $\Pr[X \geq (1+\beta)\mu n] \leq e^{-\frac{1}{2}\beta^2\mu^2 n/d^2}$; (ii) *for* $0 \leq \beta < 1$, $\Pr[X \leq (1-\beta)\mu n] \leq e^{-\frac{1}{2}\beta^2\mu^2 n/d^2}$.

Proof. We use Azuma's martingale inequality to derive these bounds. Let $X_i' = X_i - \mu$ for $1 \leq i \leq n$. Then $Y_0 = 0$ and $Y_i = \sum_{k=1}^{i} X_i'$, $1 \leq i \leq n$, is a martingale with $Y_n = X - n\mu$. Since $0 \leq X_i \leq d$ and $0 < \mu \leq d$, $|Y_i - Y_{i-1}| = |X_i - \mu| \leq d = a_i$ for $1 \leq i \leq n$. Set $\lambda = \beta\mu n$. By Lemma 1(i), for $\beta > 0$, $\Pr[X \geq (1+\beta)\mu n] = \Pr[Y_n - Y_0 \geq \lambda] \leq e^{-\frac{1}{2}\lambda^2/nd^2} = e^{-\frac{1}{2}\beta^2\mu^2 n/d^2}$. Similarly, by Lemma 1(ii), for $0 \leq \beta < 1$, $\Pr[X \leq (1-\beta)\mu n] = \Pr[Y_0 - Y_n \geq \lambda] \leq e^{-\frac{1}{2}\lambda^2/nd^2} = e^{-\frac{1}{2}\beta^2\mu^2 n/d^2}$ □

Lemma 3 (i) $\prod_{i=2}^{\infty}\left(1 - \frac{1}{i^2}\right) = \frac{1}{2}$; (ii) *For any* $\epsilon > 0$, $\prod_{i=t_0}^{\infty}\left(1 + \frac{1}{i^2}\right) < 1 + \epsilon$ *for some integer* $t_0 > 0$.

Proof. For (i), $\prod_{i=2}^{\infty}\left(1 - \frac{1}{i^2}\right) = \frac{3 \cdot 1}{2^2} \times \frac{4 \cdot 2}{3^2} \times \frac{5 \cdot 3}{4^2}\frac{6 \cdot 4}{5^2} \times \cdots = \frac{1}{2}$ by cancellation of terms. For (ii), the infinite sum $\sum_{i=1}^{\infty}\frac{1}{i^2}$ converges, thus the infinite product $\prod_{i=1}^{\infty}\left(1 + \frac{1}{i^2}\right)$ converges and for any $\epsilon > 0$, there exists some integer $t_0 > 0$ such that $\prod_{i=t_0}^{\infty}\left(1 + \frac{1}{i^2}\right) < 1 + \epsilon$ (cf. [9, pp. 94-95]). □

Theorem 3 *Let* $\{Z_n\}_0^\infty$ *be a finite supercritical branching process with branching factor* $\rho = E(Z_1) > 1$ *such that each element of the process has at most* $d \geq 2$ *offspring. Assume that* $\rho < d$. *Then, for any* $c \geq 2$,

$$\Pr[Z_n > c\rho^n] \leq C_0(d) e^{-\frac{\alpha\rho^2}{d^5} c^{1+\frac{1}{\tau}}},$$

where $C_0(d)$ *is a constant depending on* d, $\alpha > 0$ *is a fixed constant, and* $\tau > 0$ *is any constant satisfying* $d^\tau \geq 2$ *and* $\rho^{1+\tau} \geq d$. *In particular, if* $\rho \geq \sqrt{d}$, *one may take* $\tau = 1$.

Proof. The idea behind the proof is that Z_n can be much larger than its expectation only if, at some generation where the population is already fairly large, the population increases by a factor significantly greater than ρ. We derive an upper bound on the probability that such an event occurs.

Let k_0 be a positive integer, let $\beta_0, \beta_1, \cdots, \beta_{n-k_0-1}$ be positive reals, and let $m_0, m_1, \cdots, m_{n-k_0}$ be positive reals such that $m_0 = d^{k_0}$, $m_i = (1+\beta_{i-1})\rho m_{i-1}$ and $m_{n-k_0} \leq c\rho^n$. Note that $Z_{k_0} \leq m_0$. Thus $Z_n > c\rho^n$ only if, for some i, $1 \leq i \leq n - k_0$, $Z_{k_0+i-1} \leq m_{i-1}$ and $Z_{k_0+i} > m_i$. For $i = 1, 2, \cdots, n-k_0$, let $b_i = \Pr[Z_{k_0+i} > m_i | Z_{k_0+i-1} = m_{i-1}]$. Since Z_{k_0+i} is the sum of Z_{k_0+i-1} independent bounded random variables, each distributed as Z_1 with mean ρ, we may apply Lemma 2 to obtain the inequality

$$b_i = \Pr[Z_{k_0+i} > m_i | Z_{k_0+i-1} = m_{i-1}] \leq e^{-\frac{1}{2}\beta_{i-1}^2\rho^2 m_{i-1}/d^2}.$$

Clearly, for any choice of Z_{k_0+i-1} less than or equal to m_{i-1}, $\Pr[Z_{k_0+i} > m_i | Z_{k_0+i-1}] \leq \Pr[Z_{k_0+i} > m_i | Z_{k_0+i-1} = m_{i-1}] \leq b_i$. Hence, by unconditioning, $\Pr[Z_{k_0+i-1} \leq m_{i-1}$ and $Z_{k_0+i} > m_i] \leq b_i$, and it follows that, assuming $m_{n-k_0} \leq c\rho^n$,

$$\Pr[Z_n > c\rho^n] \leq \sum_{i=0}^{n-k_0} b_i.$$

We now describe how the above parameters are chosen in order to yield the conclusion of the theorem. Let $k_0 = \max\left\{k \mid c \geq 2\,(d/\rho)^k\right\}$ where $k_0 \geq 0$ is well defined as $c \geq 2$ and $\rho < d$. Let $m_0 = d^{k_0} \geq 1$. By Lemma 3, $\prod_{i=t_0}^{\infty}\left(1+\frac{1}{i^2}\right) \leq 2$ for some integer $t_0 > 0$. Set $\beta_i = \frac{1}{(i+t_0)^2}$ for $0 \leq i \leq n-k_0$. Then $\prod_{i=0}^{j}(1+\beta_i) < \prod_{i=t_0}^{\infty}\left(1+\frac{1}{i^2}\right) \leq 2$, and

$$m_i = (1+\beta_{i-1})\rho m_{i-1} = m_0\rho^i \prod_{j=0}^{i}(1+\beta_j) < 2m_0\rho^i.$$

Then, noting $m_0 = d^{k_0}$,

$$m_{n-k_0} < 2m_0\rho^{n-k_0} = 2(d/\rho)^{k_0}\rho^n \leq c\rho^n,$$

as $2(d/\rho)^{k_0} \leq c$. As $c < 2(d/\rho)^{k_0+1}$, by the assumption that $\rho^{1+\tau} \geq d$,

$$c < 2(d/\rho)^{k_0+1} \leq 2(\rho^\tau)^{k_0+1} = 2(\rho^{k_0+1})^\tau,$$

and again as $c < 2(d/\rho)^{k_0+1}$,

$$c^{1+\tau} < 2(c\rho^{k_0+1})^\tau < 2(2d^{k_0+1})^\tau \leq (d^3 m_0)^\tau$$

where the last inequality is because $m_0 = d^{k_0}$, $d \geq 2$ and by assumption $d^\tau \geq 2$. Hence,

$$m_0 > \frac{c^{1+\frac{1}{\tau}}}{d^3}.$$

As $m_i = m_0\rho^i \prod_{j=0}^{i}(1+\beta_j) > m_0\rho^i$ for $1 \leq i \leq n-k_0$, and $\beta_i = \frac{1}{(i+t_0)^2}$,

$$b_i \leq e^{-\frac{1}{2}\beta_{i-1}^2\rho^2 m_{i-1}/d^2} \leq e^{-\frac{\rho^{i+1}m_0}{2d^2(i+t_0-1)^2}}.$$

Hence, noting $m_0 > c^{1+\frac{1}{\tau}}/d^3$,

$$\sum_{i=1}^{n-k_0} b_i \leq \sum_{i=1}^{n-k_0} e^{-\frac{\rho^{i+1}m_0}{2d^2(i+t_0-1)^2}} \leq e^{-\frac{\rho^2 m_0}{2d^2t_0^2}}\left(1+\sum_{i=2}^{n-k_0} e^{-f(i)}\right) \leq C_0(d)e^{-\frac{\rho^2}{2d^5t_0^2}c^{1+\frac{1}{\tau}}}$$

where the function $f(i) = \frac{\rho^{i+1}}{2d^2(i+t_0-1)^2} - \frac{\rho^2}{2d^2t_0^2}$ grows exponentially fast as i increases, and the summation $\sum_{i=2}^{n-k_0} e^{-f(i)}$ is bounded by a constant factor $C_0(d)$ depending on d. Hence, $\Pr[Z_n > c\rho^n] \leq \sum_{i=1}^{n-k_0} b_i \leq C_0(d)e^{-\frac{\alpha\rho^2}{d^5}c^{1+1/\tau}}$, where $\alpha = 1/2t_0^2$. $\square$

3 A Special Bound for the Multitype Case

We extend the special-case right-tail bound in Section 2.1 to finite branching processes with elements of different types.

A *multitype branching process* with k types is a discrete vector Markov chain whose state set is N^k where N denotes the nonnegative integers. The chain is specified by k *offspring distributions*, each of which is a probability distribution over N^k. The probability of the k-tuple $(r_1, r_2, \cdots, r_k)$ in the ith offspring distribution is denoted by $p^i(r_1, r_2, \ldots, r_k)$, and represents the probability that an element of type i has r_j offspring of type j, for $j = 1, 2, \cdots, k$. A multitype branching process is *finite* if, for all i, $p^i(r_1, r_2, \ldots, r_k) = 0$ when $\sum_{j=1}^k r_j > d$ where d is some fixed positive integer. Let e_i denote the k-dimensional unit vector with a 1 in the ith coordinate. The successive states of the branching process are $Z_0, Z_1, \cdots, Z_n, \cdots$, where $Z_0 = e_i$ and, in general, Z_n is a k-tuple $(r_{n,1}, r_{n,2}, \cdots, r_{n,k})$ where $r_{n,j}$ denotes the number of elements of type j in the nth generation. The *population* of the nth generation is the sum $\overline{Z}_n = \sum_{i=1}^k r_{n,i}$. The rule for generating Z_{n+1} from $Z_n = (r_{n,1}, r_{n,2}, \cdots, r_{n,k})$ is as follows: for each j, draw $r_{n,j}$ samples from the jth offspring distribution, and set Z_{n+1} equal to the sum of these samples. The *mean matrix* is the $k \times k$ matrix $M = (m_{i,j})$ where $m_{i,j}$ is the expected number of type j offspring of an element of type i.

A matrix M is *strictly positive* if all entries of M^n are positive for some integer $n > 0$. By the Perron-Frobenius theorem, a strictly positive matrix M has a *dominant* eigenvalue ρ which is a simple characteristic root greater in absolute value than any other characteristic root (for a proof, see [8]). A multitype branching process is *positive regular* if its mean matrix M is strictly positive. For a positive regular multitype branching process, we call the dominant eigenvalue ρ of its mean matrix M the *branching factor* of the process. A positive regular multitype branching process is *supercritical* if $\rho > 1$.

The following theorem characterizes the asymptotic growth of a supercritical multitype branching process (for a proof, see [5, pp. 37-38,44] or [2, pp. 185,192]).

Theorem 4 *Let $\{Z_n\}$ be a multitype branching process with k types that is positive regular and supercritical. Let $\rho > 1$ be its branching factor. Then, with probability 1, $\lim_{n\to\infty} \left(\frac{Z_n}{\rho^n}\right) = \nu W$, where $\nu = (\nu_1, \nu_2, \ldots, \nu_k)$ is certain fixed vector of constants (an eigenvector of ρ with respect to M), and W is a nonnegative random variable such that $E(W|Z_0 = e_i)$ is bounded for all $i = 1, \ldots, k$.*

Corollary 1 *Let $\{Z_n\}$ be a multitype branching process with k types that is positive regular and supercritical. Then, there are constants $\beta_0 > 0$ and $\beta_1 > 0$ such that for all $n \geq 0$ and $i = 1, 2, \ldots, k$,*

$$\beta_0 \rho^n \leq E(\overline{Z}_n | Z_0 = e_i) \leq \beta_1 \rho^n,$$

where $\overline{Z}_n$ is the population of the nth generation Z_n.

Proof. Immediate from Theorem 4. □

The following is the main theorem of this section. It extends the special bound in Theorem 2 to the multitype case.

Theorem 5 *Let $Z_0, Z_1, \ldots, Z_n$, be a finite positive regular multitype branching process such that each element has at most d offspring. Suppose that the branching factor $\rho > \sqrt{d}$. Then, for any $c > 1$, regardless of the initial element,*

$$\Pr[\overline{Z}_n > cE(\overline{Z}_n)] \leq e^{-\alpha(c-1)^2},$$

where $\alpha > 0$ is a constant depending on d and ρ.

Proof. We use the same martingale argument used in Theorem 2. For $0 \leq m \leq n-1$ and $1 \leq i \leq \overline{Z}_n$, let $X_{m,i} = (x^1_{m,i}, x^2_{m,i}, \cdots, x^k_{m,i})$ where $x^j_{m,i}$ is the number of offspring of type j of the ith element of Z_n in some order, and let $S_m = \{X_{m,1}, X_{m,2}, \ldots, X_{m,\overline{Z}_m}\}$. Define $Y_{m,i}$ according to (1) with substitutions of Z_n with $\overline{Z}_n$ and Z_m with $\overline{Z}_m$. For $0 \leq m \leq n-1$, let $T_m = \{Y_{m,1}, Y_{m,2}, \ldots, Y_{m,d^m}\}$. Let $N = \sum_{m=0}^{n-1} d^m$. The sequence of $N+1$ random variables $Y_0 = E(\overline{Z}_n), T_1, T_2, \ldots, T_{n-1}$ is a Doob martingale with $Y_N = Y_{n-1,d^{n-1}} = \overline{Z}_n$.

An element of type i in Z_m initiates a multitype branching process with an initial element of type i, and is expected to contribute $E(\overline{Z}_{n-m}|Z_0 = e_i)$ to Z_n. By Corollary 1, there is a universal constant $\beta_1 > 0$ such that for $0 \leq m \leq n-1$ and $0 \leq i \leq k$, $E(\overline{Z}_{n-m}|Z_0 = e_i) \leq \beta_1 \rho^{n-m}$. Thus the difference $D_{k,i}$ of (2) is bounded by $d\beta_1\rho^{n-k-1}$, and the summation of (3) converges to a value bounded by $c_0\beta_1 d\rho^{2n}$ where $c_0 > 0$ is a constant.

Let $\lambda = (c-1)E(\overline{Z}_n)$. By Corollary 1, $\lambda \geq (c-1)\beta_0\rho^n$. As $Y_0 = E(\overline{Z}_n)$ and $Y_N = \overline{Z}_n$, $\Pr[\overline{Z}_n > cE(\overline{Z}_n)] = \Pr[Y_N - Y_0 > \lambda] \leq e^{-\alpha(c-1)^2}$ where $\alpha = \frac{\beta_0^2}{2c_0\beta_1 d}$. $\Box$

4 A Left-Tail Bound

We return to single-type branching processes and study the left tail of Z_n. We shall derive an upper bound, exponential in c, on the left-tail probability $\Pr[Z_n < \rho^n/c]$ under the condition that each element generates at least two offspring.

The requirement that each element generates at least two offspring, i.e., $p_i = \Pr[Z_1 = i] = 0$ for $i = 0, 1$, is a necessary condition for the probability $\Pr[Z_n < \rho^n/c]$ to decrease exponentially in c. Consider the case that $p_0 + p_1 > 0$. If $p_0 = \Pr[Z_1 = 0] > 0$, the process dies out immediately with probability p_0, and $\Pr[Z_n < \rho^n/c] \geq p_0 > 0$. If $p_0 = 0$ but $p_1 > 0$, then $\Pr[Z_k = 1] = p_1^k$ for $k \geq 1$. In this case, choose $c = \rho^{k-1}$. Then $\Pr[Z_n \leq \rho^n/c] = \Pr[Z_n \leq \rho^{n-k+1}] \geq p_1^k \Pr[Z_n \leq \rho^{n-k+1}|Z_k = 1] = p_1^k \Pr[Z_{n-k} \leq \rho^{n-k+1}] = p_1^k \Pr[Z_{n-k} \leq \rho E(Z_{n-k})] > (1-1/\rho)p_1 c^{\log_\rho p_1}$ by the Markov inequality and $p_1^{k-1} = c^{\log_\rho p_1}$. The last bound does not diminish exponentially in c.

Given an integer $\lambda \geq 2$, a branching process $\{Z_n\}_0^\infty$ is called λ-*definite* if $\Pr[Z_1 = k] = 0$ for $k = 0, 1, \ldots, \lambda - 1$. We may assume that $\rho = E(Z_1) > \lambda$. A λ-definite branching process is supercritical with $Z_n \geq \lambda^n \geq 2^n$. A λ-definite branching process will never die, and grows by at least a factor of $\lambda \geq 2$ in each generation.

Theorem 6 *Let* $\{Z_n\}_0^\infty$ *be a finite* λ*-definite branching process in which each element has at most* d *offspring. Assume that* $\rho = E(Z_1) > \lambda$*. Then, for any* $c > 2$*,*

$$\Pr\left[Z_n < \frac{\rho^n}{c}\right] \le C_1(d) e^{-\frac{\rho^2}{16d\lambda^3} c^{\frac{1}{\tau}}},$$

where $C_1(d)$ *is a constant depending on* d*, and* $\tau > 0$ *is any constant satisfying* $\lambda^{2\tau} \ge 2$ *and* $\lambda^{1+\tau} \ge \rho$*. In particular, if* $\lambda \ge \sqrt{d}$*, one may take* $\lambda = \sqrt{d}$ *and* $\tau = 1$*.*

Proof. The proof is similar to the proof of Theorem 3. We will be concise. By λ-definiteness, $Z_k \ge \lambda^k$ for any k. Let $k_0 = \max\left\{k \mid c \ge 2\,(\rho/\lambda)^k\right\}$ where $k_0 \ge 0$ is well-defined as $c \ge 2$ and $\rho > \lambda$. Let $m_0 = \lambda^{k_0} \ge \frac{2\rho^{k_0}}{c}$. By λ-definiteness, $Z_{k_0} \ge m_0$. Let $m_i = (1-\beta'_{i-1})\rho m_{i-1}$ for $i = 1, 2, \ldots, n-k_0$, where parameters β'_{i-1} will be assigned later so that $m_{n-k_0} \ge \frac{\rho^n}{c}$. Thus $Z_n < \frac{\rho^n}{c}$ only if, for some i, $1 \le i \le n-k_0$, $Z_{k_0+i-1} \ge m_{i-1}$ and $Z_{k_0+i} \le m_i$. For $i = 1, 2, \cdots, n-k_0$, let $b_i = \Pr[Z_{k_0+i} < m_i | Z_{k_0+i-1} = m_{i-1}]$. Since Z_{k_0+i} is the sum of Z_{k_0+i-1} independent random variables, each distributed as Z_1, by Lemma 2,

$$b_i = \Pr[Z_{k_0+i} < m_i | Z_{k_0+i-1} = m_{i-1}] \le e^{-\frac{1}{2}\beta_{i-1}^2 \rho^2 m_{i-1}/d^2}.$$

Clearly, $\Pr[Z_{k_0+i} < m_i | Z_{k_0+i-1}] \le b_i$ for any $Z_{k_0+i-1} \ge m_{i-1}$. By unconditioning, $\Pr[Z_{k_0+i-1} \ge m_{i-1}$ and $Z_{k_0+i} < m_i] \le b_i$, and it follows that

$$\Pr[Z_n < \rho^n/c] \le b_1 + b_2 + \cdots + b_{n-k_0}.$$

For $0 \le i \le n-k_0$, set $\beta_i = \frac{1}{(i+2)^2}$. By Lemma 3, $\prod_{i=0}^{j}(1-\beta_i) > \prod_{i=2}^{\infty}\left(1-\frac{1}{i^2}\right) = \frac{1}{2}$. Thus

$$m_i = (1-\beta_{i-1})\rho m_{i-1} = m_0 \rho^i \prod_{j=0}^{i}(1-\beta_j) > \frac{m_0\rho^i}{2}$$

and in particular, noting $m_0 \ge \frac{2\rho^{k_0}}{c}$, $m_{n-k_0} \ge \frac{m_0\rho^{n-k_0}}{2} > \frac{\rho^n}{c}$. Thus, for $i = 1, 2, \ldots, n-k_0$,

$$b_i = e^{-\frac{1}{2}\beta'^2_{i-1}\rho^2 m_{i-1}/d^2} \le e^{-\frac{\rho^{i+1} m_0}{4d^2(i+1)^2}}.$$

As $c < 2(\rho/\lambda)^{k_0+1}$, and by the assumptions that $\lambda^{2\tau} \ge 2$ and $\lambda^{1+\tau} \ge \rho$,

$$c < 2(\rho/\lambda)^{k_0+1} \le \lambda^{2\tau}(\lambda^\tau)^{k_0+1} = (\lambda^{k_0+3})^\tau = (\lambda^3 m_0)^\tau,$$

which gives $m_0 > c^{1/\tau}/\lambda^3$. Hence,

$$\Pr\left[Z_n < \frac{\rho^n}{c}\right] \le \sum_{i=1}^{n-k_0} e^{-\frac{\rho^{i+1} m_0}{4d^2(i+1)^2}} \le C_1(d) e^{-\frac{\rho^2 m_0}{16d^2}} \le C_1(d) e^{-\frac{\rho^2}{16d^2\lambda^3} c^{\frac{1}{\tau}}}$$

where the summation converges to a limit proportional to the first term $e^{-\frac{\rho m_0}{16d^2}}$, with the constant factor $C_1(d)$ depending on d. □

5 AND/OR Tree Evaluation

An *AND/OR tree* is a rooted tree in which an internal node of even (odd) distance from the root is an AND-node (OR-node), and each leaf has a boolean value 0 or 1. The value of an AND-node (OR-node) is recursively defined as the value of logical-AND(OR) of the values of its children. The evaluation problem for an AND/OR tree is to determine the boolean value of the root by examining values of the leaves. The goal is to minimize the number of leaves examined. A canonical algorithm SOLVE for evaluating AND/OR trees is as follows: To evaluate a node v, evaluate v directly if v is a leaf; otherwise, evaluate the children of v recursively in left-to-right order until the value of v can be determined. A uniform tree of degree d and height n is a tree in which each internal node has d children and there are n edges on each root-leaf path. A basic fact is that SOLVE, or any other algorithm, must evaluate at least at least $d^{\lfloor n/2 \rfloor}$ leaves on any instance of a uniform tree of height n and degree d, as $d^{\lfloor n/2 \rfloor}$ leaf-values are required to certify the value of the root.

We shall study the behavior of SOLVE for evaluating uniform AND/OR trees in two probabilistic models. One is the i.i.d. model in which the value of each leaf is determined randomly by a coin toss with a fixed bias q, independent of the values of other leaves. The other is the randomized model in which the children are evaluated in a random order instead of left-to-right order. We shall observe that the execution of SOLVE in the i.i.d. model with a threshold bias and the execution of randomized SOLVE on certain instances can each be viewed as a two-type finite branching process, and show that the execution time of SOLVE is likely to concentrate around its expectation in these cases.

5.1 The i.i.d. Model

Let $T(n,d,q)$ denote a random uniform AND/OR-tree of degree d and height n in the i.i.d. model in which the probability that a leaf is 0 is q. Let $I(d,n,q)$ be the expected number of leaves evaluated by SOLVE to evaluate $T(n,d,q)$. The following theorem can be found in [12, pp. 262-263].

Theorem 7 *Let ξ_d be the unique positive root of $x^d + x - 1 = 0$. Then, for any $n > 0$ and $d > 1$,*

(i) *if $q = \xi_d$, $I(d,n,q) = \left[\frac{\xi_d}{1-\xi_d}\right]^n$, where $\rho = \frac{\xi_d}{1-\xi_d} > \sqrt{d}$;*

(ii) *if $q \neq \xi_d$, $\lim_{n\to\infty}[I(d,n,q)]^{1/n} = \sqrt{d}$.*

We shall focus on the case of the *threshold* bias ξ_d. We observe that the execution of SOLVE on random uniform tree with the threshold bias can be viewed as a *two-type* branching process. To see this, we represent an AND/OR tree in an equivalent form as a NOR-tree by replacing each AND-node and each OR-node with a NOR-node where a NOR-node is the negation of an OR-node. Let $\xi = \xi_d$ be the threshold bias and let $T(n,d)$ denote the NOR-tree $T(n,d,\xi)$. A node v is of *type-0* if the value of v is 0; otherwise, it is of *type-1*. In $T(n,d)$, with the NOR-tree representation, the probability being of type-0 is ξ for each node. This is because ξ satisfies $x^d = 1 - x$, which is precisely the condition that allows the probability being of type-0 to propagate from the children to their parent node in $T(n,d)$. Let v be a node evaluated by SOLVE. We call the children of v that are evaluated by SOLVE the *offspring* of v. For $i,j \in \{0,1\}$ and $0 \le k \le d$, let $p_{ij}(k)$ be the probability that a type-i node has k type-j offspring. We have $p_{10}(d) = p_{11}(0) = 1$, $p_{01}(1) = 1$, $p_{00}(d) = 0$ and $p_{00}(k) = \xi^{k-1}(1-\xi)$ for $0 \le k \le d-1$. Probabilities $p_{ij}(k)$ define the offspring distribution of a finite two-type branching process $\{Z_n\}$ associated with the execution of SOLVE where the initial element of $\{Z_n\}$ is of type-i if the root of the NOR-tree is evaluated to i, and the population $\overline{Z}_n$ corresponds to the number of the leaves evaluated by SOLVE. Clearly, $\{Z_n\}$ is finite, positive regular, and has a branching factor $\rho = \xi_d/(1-\xi_d) > \sqrt{d}$ by Theorem 7(i).

Lemma 4 *Let S_n^0 be the number of elements of type-0 in Z_n with an initial element of type-0. Then there are constants $c_1 \ge c_0 > 0$ such that for all $n \ge 0$,*

$$c_0\gamma^n \le E(S_n^0) \le c_1\gamma^n$$

where $\sqrt{d} \le \gamma = (\tau + \sqrt{\tau^2+4d})/2 \le d$ and $\tau > 0$ is the expected number of offspring of type-0 of a single element of type-0.

Proof. Let f_k and g_k be the expected number of elements of type-0 and type-1 in Z_k, respectively. Then $f_0 = 1$, $g_0 = 0$, and for $n > 0$, $f_n = \tau f_{n-1} + d g_{n-1}$ and $g_n = f_{n-1}$. Thus $f_n = \tau f_{n-1} + d f_{n-2}$. The characteristic equation of the last recurrence is $x^2 = \tau x + d$ with a unique positive solution $\gamma = (\tau + \sqrt{\tau^2+4d})/2 > \sqrt{d}$. Hence, $E(S_n^0) = f_n = \Theta(\gamma^n)$. As $f_n \le d^n$, $\gamma \le d$. □

The following is the main theorem of this section.

Theorem 8 *For $i \in \{0,1\}$, let $S_i(n,d)$ be the random variable that is the number of leaves evaluated by SOLVE to evaluate $T(n,d)$ given that the value of the root is i. Then, for $n > 1$, $d \ge 2$,*

i) $\Pr[S_i(n,d) > cE(S_i(n,d))] \le e^{-C_0(c-1)^2}$ *for $c \ge 2$, where $C_0 > 0$ is a constant depending on d;*

ii) $\Pr[S_i(n,d) < E(S_i(n,d))/c] \le C_1 e^{-C_2 c}$ *for $c \ge 4c_1 d/c_0$, where c_0, c_1 are the constants stated in Lemma 4, and C_1, C_2 are constants depending on d.*

Proof. Let $Z_0, Z_1, \ldots, Z_n$ be the two-type branching process associated with SOLVE. Then $\overline{Z}_n = S_i(n,d)$. The right-tail bound of (i) follows immediately from Theorem 5 in Section 3 and the fact that $\{Z_n\}$ is finite and positive regular with a branching factor greater than $\sqrt{d}$.

Our task is to prove the left-tail bound of (ii). We observe that $\overline{Z}_{2k} \ge d^k$ for any k. This is because a type-1 element generates d type-0 offspring and a type-0 element generates one type-1 offspring. Conceptually, one may think of this two-type branching process $\{Z_n\}$ as $\sqrt{d}$-definite. However, one cannot directly apply the left-tail bound of Theorem 6 in Section 4 to a two-type branching process. Nevertheless, we will show that the proof technique can be extended, with some efforts, to the specific two-type branching process $\{Z_n\}$ to yield the left-tail bound of (ii).

To prove (ii), we may focus on the number of elements of type-0. Let S_k^i be the number of elements of type-i in Z_k. Then $S_i(n,d) = \overline{Z}_n = S_n^0 + S_n^1 = S_n^0 + S_{n-1}^0$, as $S_n^1 = S_{n-1}^0$, and $\Pr[S_i(n,d) < E(S_i(n,d))/c] \le \Pr[S_n^0 < E(S_n^0)/c] + \Pr[S_{n-1}^0 < E(S_{n-1}^0)/c]$. Thus the bound in

(ii) holds if it holds for $\Pr[S_n^0 < E(S_n^0)/c]$. Moreover, we only need to prove (ii) for $S_0(n,d)$. As $S_1(n,d) = \sum_{i=1}^{d} X_i$ where $X_i \sim S_0(n-1,d)$, $S_1(n,d) < E(S_1(n,d))/c$ only if $X_i < E(X_i)/c$ for some i. It follows that $\Pr[S_1(n,d) < E(S_1(n,d))/c] \leq d \Pr[S_0(n-1,d) < E(S_0(n-1,d))/c]$.

Our goal is thus to prove the upper bound of (ii) for $\Pr[S_n^0 < E(S_n^0)/c]$ where S_n^0 is the number of elements of type-0 in Z_n when the initial element is of type-0. We shall follow the same line of proof as the proofs of Theorem 3 in Section 2 and Theorem 6 in Section 4. A critical transformation of the proof is the inequalities (11) and (12). These inequalities reduce the problem of bounding S_k^0, which depends on both S_{k-1}^0 and S_{k-1}^1, into a problem of bounding a sum involving only type-0 elements, which can be bounded as in the single-type case.

We first define the parameter k_0 that satisfies relations (4) and (5). Let Y be the random variable that is the number of type-0 offspring of a single type-0 element. Let $\tau = E(Y) > 0$ and let $\gamma = (\tau + \sqrt{\tau^2 + 4d})/2$. By Lemma 4, for any $n > 0$, $c_0\gamma^n \leq E(S_n^0) \leq c_1\gamma^n$ for some constants c_0 and c_1, and $\sqrt{d} < \gamma \leq d$. Let

$$k_0' = \max\left\{k \mid c \geq \frac{4c_1}{c_0}\left(\frac{\gamma}{\sqrt{d}}\right)^{2k}\right\},$$

which is well-defined as $\gamma > \sqrt{d}$. By the assumption, $c \geq 4c_1 d/c_0$. As $\gamma \leq d$, $(\gamma/\sqrt{d})^2 \leq d$. Thus $c \geq 4c_1(\gamma/\sqrt{d})^2/c_0$ and $k_0' \geq 1$. Define

$$k_0 = \begin{cases} 2k_0' & \text{if } S_{2k_0'}^0 \geq d^{k_0'}/2 \\ 2k_0' - 1 & \text{otherwise} \end{cases}$$

Note that $k_0 \geq 1$ is a random variable which is either $2k_0'$ or $2k_0' - 1$ depending on $Z_{2k_0'}$, namely, whether $S_{2k_0'}^0 \geq d^{k_0'}/2$.

As $\overline{Z}_{2k} \geq d^k$ for any k, $\overline{Z}_{2k_0'} = S_{2k_0'}^0 + S_{2k_0'}^1 \geq d^{k_0'}$. If $S_{2k_0'}^0 < d^{k_0'}/2$, then it must be that $S_{2k_0'}^1 = S_{2k_0'-1}^0 > d^{k_0'}/2$. Hence, by the definition of k_0,

$$S_{k_0}^0 \geq \frac{d^{k_0'}}{2} \geq \frac{d^{\frac{k_0}{2}}}{2}. \tag{4}$$

By the definitions of k_0',

$$\frac{4c_1}{c_0}\left(\frac{\gamma}{\sqrt{d}}\right)^{2k_0'} \leq c < \frac{4c_1}{c_0}\left(\frac{\gamma}{\sqrt{d}}\right)^{2(k_0'+1)}.$$

As $2k_0' - 1 \leq k_0 \leq 2k_0'$, $\gamma \geq \sqrt{d}$ and $(\gamma/\sqrt{d})^2 \leq d$,

$$\frac{4c_1}{c_0}\left(\frac{\gamma}{\sqrt{d}}\right)^{k_0} \leq c < \frac{4c_1}{c_0}\left(\frac{\gamma}{\sqrt{d}}\right)^{k_0+3} \leq \frac{4c_1}{c_0} d^{\frac{k_0+3}{2}}. \tag{5}$$

For $0 \leq i \leq n - k_0 - 1$, define

$$f_i = E(S_{k_0+i}^0 | Z_{k_0}),$$

which is the expected number of elements of type-0 in Z_{k_0+i} given Z_{k_0}. Note that the f_i are random variables depending on Z_{k_0}. By definition, $f_0 = E(S_{k_0}^0 | Z_{k_0}) = S_{k_0}^0$, $f_1 = E(S_{k_0+1}^0 | Z_{k_0}) =$

$\tau S^0_{k_0} + dS^1_{k_0} = \tau f_0 + dS^1_{k_0}$, and for $i \geq 2$, $f_i = \tau E(S^0_{k_0+i-1}|Z_{k_0}) + dE(S^1_{k_0+i-1}|Z_{k_0}) = \tau f_{i-1} + df_{i-2}$ as $S^1_{k_0+i-1} = S^0_{k_0+i-2}$. Hence,

$$
\begin{aligned}
f_0 &= S^0_{k_0} > 0 \\
f_1 &= \tau f_0 + dS^1_{k_0} \\
f_i &= \tau f_{i-1} + df_{i-2} \quad \text{for } i \geq 2
\end{aligned}
$$

We now define parameters $m_0, m_1, \ldots, m_{n-k_0}$. For $0 \leq i \leq n-k_0-1$, let $\beta_i = 1/(i+2)^2$. Define $m_0 = S^0_{k_0} = f_0$ and for $1 \leq i \leq n-k_0$,

$$m_i = (1-\beta_{i-1})m_{i-1}f_i/f_{i-1}.$$

Note that the m_i are random variables depending on Z_{k_0} and well-defined as $f_i > 0$. For $0 \leq i \leq n-k_0-1$,

$$m_{i+1} = m_0 \prod_{j=0}^{i}(1-\beta_j)\prod_{j=0}^{i}\rho_j = f_{i+1}\prod_{j=0}^{i}(1-\beta_j) > f_{i+1}/2, \qquad (6)$$

as $\prod_{i=0}^{n-k_0-1}(1-\beta_i) > 1/2$ by Lemma 3.

We show that $m_{n-k_0} \geq E(S^0_n)/c$. By (6), $m_{n-k_0} > f_{n-k_0}/2 = E(S^0_n|Z_{k_0})/2$. Let v be an element of type-0 in Z_{k_0}. The expected number of descendants of type-0 of v in S^0_n is $E(S^0_{n-k_0})$. Thus, $E(S^0_n|Z_{k_0}) \geq S^0_{k_0}E(S^0_{n-k_0}) \geq d^{k_0/2}E(S^0_{n-k_0})/2$ as $S^0_{k_0} \geq d^{k_0/2}$ by (4). By Lemma 4, $c_0\gamma^i \leq E(S^0_i) \leq c_1\gamma^i$ for $i \geq 0$. By (5), $c \geq \frac{4c_1}{c_0}(\gamma/\sqrt{d})^{k_0}$. Hence,

$$
\begin{aligned}
m_{n-k_0} &> E(S^0_n|Z_{k_0})/2 \\
&\geq d^{k_0/2}E(S^0_{n-k_0})/4 \\
&\geq d^{k_0/2}c_0\gamma^{n-k_0}/4 \\
&\geq c_0\gamma^n(\sqrt{d}/\gamma)^{k_0}/4 \\
&\geq c_1\gamma^n/c \\
&\geq E(S^0_n)/c.
\end{aligned}
$$

By the same argument,

$$f_i = E(S^0_{k_0+i}|Z_{k_0}) \geq E(S^0_i)S^0_{k_0} \geq \frac{c_0}{2}\gamma^i d^{\frac{k_0}{2}} \geq c_0' d^{\frac{i+k_0}{2}}$$

as $\gamma > \sqrt{d}$. Thus f_i grows exponentially in i.

Given $m_0 = S^0_{k_0}$ and $m_{n-k_0} \geq E(S^0_n)/c$, we can have $S^0_n < E(S^0_n)/c$ only if

i) either $S_{k_0+1} < m_1$ or

ii) for some i, $2 \leq i \leq n-k_0$, $S^0_{k_0+i-2} \geq m_{i-2}$ and $S^0_{k_0+i-1} \geq m_{i-1}$ but $S_{k_0+i} < m_i$.

Let $b_1 = \Pr[S^0_{k_0+1} < m_1]$ and for $i = 2, 3, \ldots, n-k_0$,

$$b_i = \Pr[S^0_{k_0+i} < m_i \,|\, S^0_{k_0+i-1} = m_{i-1} \text{ and } S^0_{k_0+i-2} = m_{i-2}].$$

Clearly, for $i \geq 2$, $\Pr[S^0_{k_0+i} < m_i \,|\, S^0_{k_0+i-1} \geq m_{i-1}$ and $S^0_{k_0+i-2} \geq m_{i-2}] \leq \Pr[S^0_{k_0+i} < m_i \,|\, S^0_{k_0+i-1} = m_{i-1}$ and $S^0_{k_0+i-2} = m_{i-2}] = b_i$. By unconditioning, $\Pr[S^0_{k_0+i} < m_i$ and $S^0_{k_0+i-1} \geq m_{i-1}$ and $S^0_{k_0+i-2} \geq m_{i-2}] \leq b_i$, and it follows that,

$$\Pr\left[S^0_n < \frac{E(S^0_n)}{c}\right] \leq b_1 + b_2 + \cdots + b_{n-k_0}.$$

Let $Y_j \sim Y$ be independent random variables, where Y is the number of type-0 offspring of a single type-0 element. As $S^0_{k_0} = m_0$,

$$S^0_{k_0+1} = \sum_{j=1}^{S^0_{k_0}} Y_j + dS^1_{k_0} = \sum_{j=1}^{m_0} Y_j + dS^1_{k_0}. \tag{7}$$

For $2 \leq i \leq n - k_0$, given $S^0_{k_0+i-1} = m_{i-1}$ and $S^0_{k_0+i-2} = m_{i-2}$, noting $S^1_{k_0+i-1} = S^0_{k_0+i-2}$,

$$S^0_{k_0+i} = \sum_{j=1}^{S_{k_0+i-1}} Y_j + dS^1_{k_0+i-1} = \sum_{j=1}^{m_{i-1}} Y_j + dm_{i-2}. \tag{8}$$

Given $m_0 = f_0$ and $f_1 = \tau f_0 + dS^1_{k_0}$,

$$m_1 = (1-\beta_0)m_0 f_1/f_0 = (1-\beta_0)\tau m_0 + (1-\beta_0)dS^1_{k_0} \tag{9}$$

and for $2 \leq i \leq n - k_0$, given $f_i = \tau f_{i-1} + df_{i-2}$,

$$\begin{aligned} m_i &= (1-\beta_{i-1})m_{i-1}f_i/f_{i-1} \\ &= (1-\beta_{i-1})\tau m_{i-1} + (1-\beta_{i-1})dm_{i-1}f_{i-2}/f_{i-1} \\ &= (1-\beta_{i-1})\tau m_{i-1} + (1-\beta_{i-1})(1-\beta_{i-2})dm_{i-2}. \end{aligned} \tag{10}$$

We now compare (7) and (9). As $dS^1_{k_0} > (1-\beta_0)dS^1_{k_0}$, we conclude that $S^0_{k_0+1} < m_1$ only if $\sum_{j=1}^{m_0} Y_j < (1-\beta_0)\tau m_0$. Thus

$$b_1 = \Pr[S^0_{k_0+1} < m_1] \leq \Pr[\sum_{j=1}^{m_0} Y_j < (1-\beta_0)\tau m_0]. \tag{11}$$

Similarly, we compare (8) and (10). As $dm_{i-2} > (1-\beta_{i-1})(1-\beta_{i-2})dm_{i-2}$, we conclude that $S^0_{k_0+i} < m_i$, given $S^0_{k_0+i-1} = m_{i-1}$ and $S^0_{k_0+i-2} = m_{i-2}$, only if $\sum_{j=1}^{m_{i-1}} Y_j < (1-\beta_{i-1})\tau m_{i-1}$. Thus, for $2 \leq i \leq n - k_0$,

$$\begin{aligned} b_i &= \Pr[S^0_{k_0+i} < m_i \,|\, S_{k_0+i-1} = m_{i-1} \text{ and } S_{k_0+i-2} = m_{i-2}] \\ &\leq \Pr[\sum_{j=1}^{m_{i-1}} Y_j < (1-\beta_{i-1})\tau m_{i-1}]. \end{aligned} \tag{12}$$

Note that Y_j are independent, each distributed as Y where $Y \leq d-1$ and $\tau = E(Y) > 0$. By Lemma 2, for $1 \leq i \leq n - k_0$,

$$\Pr[\sum_{j=1}^{m_{i-1}} Y_j < (1-\beta_{i-1})\tau m_{i-1}] \leq e^{-\frac{1}{2}\beta^2_{i-1}\tau^2 m_{i-1}/(d-1)^2}.$$

By (6), $m_i \geq f_i/2$. Hence, noting $b_i = 1/(i+2)^2$,

$$b_i \leq e^{-\frac{1}{2}\beta_{i-1}^2 \tau^2 m_{i-1}/(d-1)^2} \leq e^{-\frac{\tau^2 f_{i-1}}{4(i+1)^2(d-1)^2}}$$

and

$$\sum_{i=1}^{n-k_0} b_i \leq \sum_{i=1}^{n-k_0} e^{-\frac{\tau^2 f_{i-1}}{4(i+1)^2(d-1)^2}} \leq e^{-\frac{\tau^2 f_0}{16(d-1)^2}} \left(1 + \sum_{i=1}^{n-k_0} e^{-F(i)}\right) \leq C_1(d) e^{-\frac{\tau^2 f_0}{16(d-1)^2}}$$

where $F(i) = \frac{\tau^2 f_{i-1}}{4(i+1)^2(d-1)^2} - \frac{\tau^2 f_0}{16(d-1)^2}$ grows exponentially in i since f_i does, and $\sum_{i=1}^{n-k_0} e^{-F(i)}$ converges to a constant $C_1(d)$ depending on d.

By (5), $c < \frac{4c_1}{c_0} d^{\frac{k_0+3}{2}} \leq \frac{8c_1}{c_0} d^{3/2} f_0$, as $f_0 = S_{k_0}^0 \geq \frac{1}{2} d^{k_0/2}$. Let $\alpha = \frac{8c_1}{c_0} d^{3/2}$. Then $f_0 > c/\alpha$ and

$$\Pr\left[S_n^0 < \frac{E(S_n^0)}{c}\right] \leq C_1(d) e^{-\frac{\tau^2 f_0}{16(d-1)^2}} \leq e^{-\frac{\tau^2 c}{16\alpha d^2}} = e^{-C_2 c},$$

where $C_2 = \frac{\tau^2}{16\alpha d^2}$. □

We make two remarks about the proof of Theorem 8. First, the parameters k_0 and m_i are random variables conditioned on random variables $Z_{2k_0'}$ and Z_{k_0}. The given proof holds regardless of the randomness of k_0 and m_i. Secondly, the crucial inequalities (11) and (12) are derived by the fact that the last terms in (7)–(10) are non-random, thus can be compared. That these terms are non-random is due to the very fact that an element of type-1 deterministically generates d offspring of type-0.

5.2 Randomized Model

A randomized version of SOLVE is as follows: To evaluate a node v, evaluate v directly if v is a leaf; otherwise, do the following until the value of v can be determined: select an unevaluated child of v randomly and evaluate this child recursively. The complexity of a randomized algorithm is the maximum of the expected number of leaves evaluated by the algorithm over all instances. The randomized complexity is the minimum of the complexities of algorithms over all randomized algorithms. The randomized complexity of AND/OR tree evaluation was first studied by Saks and Wigderson in [14] in which they showed that randomized SOLVE described above is optimal for evaluating uniform AND/OR trees, and that the maximum of the expected number of leaves evaluated by randomed SOLVE on uniform AND/OR trees of height n and degree d is $\Theta\left((r_d)^n\right)$ where $r_d = (d-1+\sqrt{d^2+14d+1})/4 \sim d/2 + O(1)$.

We represent an AND/OR tree as a NOR-tree, and let $T(n,d)$ denote the set of uniform NOR-trees of height n and degree d. An instance of $T(n,d)$ is *uniformly-structured with parameter* k, where $1 \leq k \leq d$, if each internal type-0 node has k type-1 children and $d-k$ type-0 children. In other words, a uniformly-structured instance is an instance in which the internal nodes have the same number of children of the same value. For $i = 0, 1$, an i-instance of $T(n,d)$ is a NOR-tree in $T(n,d)$ whose root value is i. A *worst-case* (*best-case*) instance of $T(n,d)$ for randomized SOLVE is an i-instance of $T(n,d)$ that maximizes (minimizes) the expected number of leaves evaluated by randomized SOLVE on all i-instances. It is clear that the worst-case instances for randomized SOLVE are the uniformly-structured instances with $k = 1$ whereas the best-case instances are those with $k = d$.

The uniformly-structured instances of $T(n, d)$ represent those instances of $T(n, d)$ on which the execution of randomized SOLVE is a two-type branching process. For the best-case instances with $k = d$, the branching process is deterministic in which a type-0 element has no offspring of type-0. When $1 \leq k < d$, the branching process is a random process in which a type-0 element has j offspring of type-0, where $0 \leq j \leq d - k$, with probability $\frac{d-k}{d} \times \frac{d-k-1}{d-1} \times \cdots \times \frac{d-k-j+1}{d-j+1} \times \frac{k}{d-j}$; in particular, for the worst-case instances with $k = 1$, a type-0 element has j offspring of type-0 with probability $\frac{d-1}{d} \times \frac{d-2}{d-1} \times \cdots \times \frac{d-j}{d-j+1} \times \frac{1}{d-j} = \frac{1}{d}$. For the uniformly-structured instances with $1 \leq k < d$, the associated two-type branching process is finite, positive regular, and supercritical with a branching factor $\rho > \sqrt{d}$ where ρ depends on k.

Theorem 9 *For $i = 0, 1$, let $R_i(n, d)$ be the expected number of leaves evaluated by randomed SOLVE on an i-instance of an unform AND/OR tree of height n and degree d that is uniformly-structured with parameter k. Then, for $n \geq 1$, $d \geq 2$, and $1 \leq k \leq d$,*

i) $\Pr[R_i(n, d) > cE(R_i(n, d))] \leq e^{-C_0(c-1)^2}$ *for $c \geq 2$, where $C_0 > 0$ is a constant depending on d and k;*

ii) $\Pr[R_i(n, d) < E(R_i(n, d))/c] \leq C_1 e^{-C_2 c}$ *for $c \geq \alpha$, where α, C_1, C_2 are constants depending on d and k.*

Proof. When $k = d$, the process is deterministic that the bounds of (i) and (ii) hold trivially. When $1 \leq k < d$, the bounds of (i) and (ii) can be proved by a proof identical to that of Theorem 8. □

6 Conclusion

We have obtained strong upper bounds on the tail probabilities of supercritical branching processes with a finite offspring distribution, and shown their application to AND/OR tree evaluation. Another interesting application of finite branching processes is a tree-search problem studied in [7]. These results demonstrate the usefulness of finite branching processes.

In the randomized model of AND/OR tree evaluation, we have shown that the execution time of randomized SOLVE on the uniformly-structured instances of uniform AND/OR tree, in which the internal nodes have the same number of children of the same value, is unlikely to deviate significantly from its expected value. It would be desirable to show that the same conclusion holds for all instances of uniform AND/OR trees.

Acknowledgements

The authors thank Krishna B. Athreya for his encouragement of this work and his comments on a preliminary draft of this paper.

References

[1] K.B. Athreya: "Large deviation rates for branching processes - I, single type case," 1993, To appear in *Annals of Applied Probability.*

[2] K.B. Athreya and P. Ney: *Branching Processes,* Springer-Verlag, Berlin, 1972.

[3] K. Azuma: "Weighted sums of certain dependent variables," *Tohoku Math Journal*, pp. 357-367, Vol. 3, 1967.

[4] J.L. Doob: *Stochastic Process*, John Wiley & Sons, 1953.

[5] T. Harris: *The Theory of Branching Processes,* Springer-Verlag, Berlin, 1963.

[6] W. Hoeffding: "Probability inequalities for sums of bounded random variables," *Journal of American Statistical Association,* pp. 13-30, Vol. 58, 1963.

[7] R.M. Karp and J. Pearl: "Searching for an optimal path in a tree with random costs," *Artificial Intelligence*, pp. 99-117, Vol. 21, No. 1-2, 1983.

[8] Karlin and Taylor: *A first course in stochastic processes*, Second Edition, Academic Press, 1975.

[9] K. Knopp: *Infinite Sequences and Series*, Dover, 1956.

[10] D.E. Knuth and R.N. Moore: "An analysis of alpha-beta pruning," *Artificial Intelligence*, pp. 293-326, Vol. 6, 1975.

[11] F.T. Leighton and M. Newman and A. Ranade and E. Schwabe: "Dynamic tree embeddings in butterflies and hypercubes," *SIAM Journal on Computing*, pp. 639-654, Vol. 21, No. 4, 1992.

[12] J. Pearl: *Heuristics*, Addison-Wesley, 1984.

[13] J. Pearl: "The solution for the branching factor of the alpha-beta pruning algorithm and its optimality," *Communications of ACM*, pp. 559-564, Vol. 25, No. 8, 1982.

[14] M. Saks and A. Wigderson: "Probabilistic Boolean Decision Trees and the Complexity of Evaluating game trees," *Proceedings of the 27th Annual Symposium on Foundations of Computer Science*, pp. 29-38, 1986.

[15] P. Raghavan: "Probabilistic Construction of Deterministic Algorithms: Approximation Packing Integer Programs," *Proceedings of 27th Annual Symposium on Foundations of Computer Science*, pp. 10-18, 1986.

[16] J. Spencer: *Ten Lectures on the Probabilistic Method,* SIAM, Philadelphia, 1987.

[17] E. Shamir and J. Spencer: "Sharp concentration of the chromatic number on random graphs $G_{n,p}$," *Combinatorica*, pp. 121-130, Vol. 7, No. 1, 1987.

[18] M. Tarsi: "Optimal Search on Some Game Trees," *Journal of ACM*, pp. 389-396, Vol. 30, 1983.