

Subcritical Galton-Watson Trees

STEVEN FINCH

March 1, 2007

Fix a probability $0 < p < 1$. For any 2-vector $u = (v, w)$, write $u_L = v$ and $u_R = w$. A **Galton-Watson tree** is an ordered, strongly binary tree $T = \tau(p)_L$ that is defined recursively in terms of left and right subtrees of the root as follows:

$$\tau(p) = \left(\left\{ \begin{array}{ll} \tau(p) & \text{if } X \leq p, \\ \emptyset & \text{otherwise} \end{array} \right\}, \left\{ \begin{array}{ll} \tau(p) & \text{if } X \leq p, \\ \emptyset & \text{otherwise} \end{array} \right\} \right).$$

Each variable X is a new, independent Uniform $[0, 1]$ random number. For example, $T = \emptyset$ with probability $1 - p$, $T = (\emptyset, \emptyset)$ with probability $p(1 - p)^2$, and $T = ((\emptyset, \emptyset), \emptyset)$ with probability $p^2(1 - p)^3$.

The **number of vertices** N is equal to twice the number of left parentheses (parents) in the expression for T , plus one. Equivalently, N is twice the number of \emptyset s (leaves), minus one. It can be shown that N is finite with probability 1 if $p \leq 1/2$ and $1/p - 1$ if $p > 1/2$. We will focus on the **subcritical case** $p < 1/2$ for the remainder of this essay.

Let N_k denote the number of vertices at distance k from the root, that is, the size of the k^{th} generation. Clearly $N_0 = 1$ and $N < \infty$ if and only if $N_k = 0$ for all sufficiently large k . Define

the **height** H of T to be $\max_{N_k > 0} k$,

the **width** W of T to be $\max_{k \geq 0} N_k$.

We wish to evaluate the joint distribution of (N, H, W) as a function of p . Some partial results (mostly of a numerical nature) are all we can report now.

The sequence N_0, N_1, \dots, N_H is called the **profile** of T . Dual to this is the sequence of $(N + 1)/2$ leaf distances from the root, ordered from left to right, called the **contour** of T . It would be good someday to better understand joint profile and contour distributions as well.

⁰Copyright © 2007 by Steven R. Finch. All rights reserved.

0.1. Number of Vertices. The probability that T has at least 3 vertices is p . Let $m \geq 1$. From the conditional relation:

$$\begin{aligned} & \text{P}(T \text{ has } 2m + 1 \text{ vertices}) \\ &= \sum_{j=1}^{2m-1} \text{P}(T_L \text{ has } 2m - j \text{ vertices} \wedge T_R \text{ has } j \text{ vertices} \mid T \text{ has at least 3 vertices}) \cdot p, \end{aligned}$$

we deduce that

$$\text{P}(N = 2m + 1) = p \sum_{j=1}^{2m-1} \text{P}(N = 2m - j) \text{P}(N = j)$$

and hence

$$\text{P}(N = n) = \begin{cases} 0 & \text{if } n = 2m, \\ \frac{1}{m+1} \binom{2m}{m} p^m (1-p)^{m+1} & \text{if } n = 2m + 1. \end{cases}$$

Well-known asymptotics for the Catalan numbers

$$\frac{1}{m+1} \binom{2m}{m} \sim \frac{1}{\sqrt{\pi}} \frac{2^{2m}}{m^{3/2}}$$

give a sense of the rate at which $\text{P}(N = n) \rightarrow 0$ as $n \rightarrow \infty$, n odd. More precisely [1, 2],

$$\text{P}(N = n) \sim \left(\sqrt{\frac{2}{\pi}} n^{-3/2} + c n^{-5/2} + d n^{-7/2} + \dots \right) (2p)^m (2(1-p))^{m+1}.$$

We also have moments [3, 4, 5]

$$\text{E}(N) = \frac{1}{1-2p}, \quad \text{Var}(N) = \frac{4p(1-p)}{(1-2p)^3}.$$

0.2. Height. Let a_k denote the probability that $N_k = 0$, equivalently, the probability that $H < k$. The conditional distribution of N_k , given $N_1 = j$, is the same as the sum of j independent random copies of N_{k-1} [6]. Of course, $j = 0$ and $j = 2$ are the only possible values for N_1 ; thus we have

$$\text{P}(N_k = 0) = \underbrace{\text{P}(N_k = 0 \mid N_1 = 0)}_1 \overbrace{\text{P}(N_1 = 0)}^{1-p} + \underbrace{\text{P}(N_k = 0 \mid N_1 = 2)}_{\text{P}(N_{k-1}=0)^2} \overbrace{\text{P}(N_1 = 2)}^p$$

and hence [7]

$$a_0 = 0, \quad a_k = (1 - p) + p a_{k-1}^2 \quad \text{for } k \geq 1, \quad \lim_{k \rightarrow \infty} a_k = 1.$$

Let us prove that the convergence rate of $\{a_k\}$ is exponential, that is,

$$0 < \lim_{k \rightarrow \infty} \frac{1 - a_k}{(2p)^k} < 1.$$

First, note that $0 \leq a_k < 1$ for all k by induction ($a_k \geq 1 - p > 0$ is obvious; supposing $0 \leq a_{k-1} < 1$, we obtain $a_k < (1 - p) + p = 1$). Now, writing $b_k = 1 - a_k$, we have $b_0 = 1$, $0 < b_k \leq 1$ and

$$\begin{aligned} b_k &= p(1 - a_{k-1}^2) = p(1 - a_{k-1})(1 + a_{k-1}) \\ &= p b_{k-1}(2 - b_{k-1}) \\ &< 2p b_{k-1} < (2p)^2 b_{k-2} < (2p)^3 b_{k-3} \end{aligned}$$

thus $b_k < (2p)^k$ for all k . Observe that

$$\begin{aligned} b_k &= 2p b_{k-1} \left(1 - \frac{b_{k-1}}{2}\right) \\ &= (2p)^2 b_{k-2} \left(1 - \frac{b_{k-2}}{2}\right) \left(1 - \frac{b_{k-1}}{2}\right) \\ &= (2p)^3 b_{k-3} \left(1 - \frac{b_{k-3}}{2}\right) \left(1 - \frac{b_{k-2}}{2}\right) \left(1 - \frac{b_{k-1}}{2}\right) \\ &= (2p)^k \prod_{j=0}^{k-1} \left(1 - \frac{b_j}{2}\right) \end{aligned}$$

hence

$$\begin{aligned} C &= \lim_{k \rightarrow \infty} \frac{1 - a_k}{(2p)^k} = \lim_{k \rightarrow \infty} \frac{b_k}{(2p)^k} = \prod_{j=0}^{\infty} \left(1 - \frac{b_j}{2}\right) \\ &= \prod_{j=0}^{\infty} \left(1 - \frac{1 - a_j}{2}\right) = \prod_{j=0}^{\infty} \frac{1 + a_j}{2} \end{aligned}$$

exists and is nonzero since

$$\sum_{j=0}^{\infty} \frac{b_j}{2} < \frac{1}{2} \sum_{j=0}^{\infty} (2p)^j$$

converges. This completes the proof. The expression for C as an infinite product turns out to be useful for high precision estimates of C , given p (see Table 1).

The algorithm for $\{b_k\}$:

$$b_0 = 1, \quad b_k = p b_{k-1}(2 - b_{k-1}) \quad \text{for } k \geq 1, \quad \lim_{k \rightarrow \infty} b_k = 0$$

is helpful from a numerical perspective. While formulas in a_ℓ are easily converted into formulas in b_ℓ and vice versa:

$$\begin{aligned} P(H = k) &= a_{k+1} - a_k = b_k - b_{k+1}, \\ E(e^{tH}) &= \sum_{k=0}^{\infty} e^{tk} (a_{k+1} - a_k) = \sum_{k=0}^{\infty} e^{tk} (b_k - b_{k+1}) \end{aligned}$$

the difference $a_{k+1} - a_k$ is harder to calculate than $b_k - b_{k+1}$. (Reason: the subtraction of nearly equal quantities, each approaching 1, leads to a loss of floating point precision.)

Since the series for the moment generating function is telescoping:

$$E(e^{tH}) = b_0 + \sum_{k=1}^{\infty} (e^{tk} - e^{t(k-1)}) b_k$$

we obtain

$$E(H) = \sum_{k=1}^{\infty} b_k, \quad E(H^2) = \sum_{k=1}^{\infty} (2k - 1) b_k$$

upon differentiation. No closed-form expressions for the quantities in Table 1 are known.

Table 1 *Height-Related Parameters*

p	C	$E(H)$	$\text{Var}(H)$
0.2	0.4238945378...	0.3179675669...	0.6053027749...
0.25	0.3929068527...	0.4610125877...	1.0724312517...
0.3	0.3539671772...	0.6568327963...	1.9336638291...
0.35	0.3039572818...	0.9422336526...	3.7158517879...
0.4	0.2376466589...	1.4045313857...	8.2383270278...

We mention an interesting result for the **critical case** $p = 1/2$. The recurrence

$$a_0 = 0, \quad a_k = \frac{1}{2} (1 + a_{k-1}^2) \quad \text{for } k \geq 1, \quad \lim_{k \rightarrow \infty} a_k = 1$$

satisfies [8, 9, 10]

$$1 - a_k \sim \frac{2}{k + \ln(k) + 1.76799378...}$$

It is clear, therefore, that $E(H) = \infty$. The relevance of [8, 9] to Galton-Watson trees seems not to have been noticed before.

0.3. Height via Markov. The sequence N_0, N_1, N_2, \dots is a time-homogeneous Markov chain with transition probability matrix Q , where

$$q_{i,j} = \mathbb{P}(N_1 = j \mid N_0 = i) = \begin{cases} \binom{i}{j/2} p^{j/2} (1-p)^{i-j/2} & \text{if } 2 \leq j \leq 2i \text{ is even,} \\ 0 & \text{otherwise} \end{cases}$$

is the $(i, j)^{\text{th}}$ element of Q and $i \geq 1, j \geq 1$. Observe that

$$\mathbb{P}(H = 0 \mid N_0 = i) = (1-p)^i$$

and

$$\mathbb{P}(H = k \mid N_0 = i) = \sum_{j \geq 1} \underbrace{\mathbb{P}(H = k \mid N_1 = j)}_{\mathbb{P}(H = k-1 \mid N_0 = j)} \overbrace{\mathbb{P}(N_1 = j \mid N_0 = i)}^{q_{i,j}}$$

for $k \geq 1$. We will use these formulas to derive an alternative matrix expression for $\mathbb{E}(H)$, as outlined in [11]. Let $\mu_i = \mathbb{E}(H \mid N_0 = i)$ and $\nu_i = 1 - (1-p)^i$. From

$$\begin{aligned} \mu_i &= \sum_{k \geq 0} k \mathbb{P}(H = k \mid N_0 = i) \\ &= \sum_{k \geq 1} k \sum_{j \geq 1} q_{i,j} \mathbb{P}(H = k-1 \mid N_0 = j) \\ &= \sum_{j \geq 1} q_{i,j} \sum_{k \geq 1} k \mathbb{P}(H = k-1 \mid N_0 = j) \\ &= \sum_{j \geq 1} q_{i,j} \left(1 + \sum_{k \geq 0} k \mathbb{P}(H = k \mid N_0 = j) \right) \\ &= \nu_i + \sum_{j \geq 1} q_{i,j} \mu_j, \end{aligned}$$

it follows that $(I - Q)\mu = \nu$ and thus $\mu = (I - Q)^{-1}\nu$. Only the first component of μ is desired since $\mathbb{E}(H) = \mathbb{E}(H \mid N_0 = 1)$. Of course, we must restrict $i \leq \ell, j \leq \ell$ when evaluating μ_1 , where ℓ is large. As $\ell \rightarrow \infty$, indeed $\mu_1 \rightarrow \mathbb{E}(H)$ numerically as found in the previous section.

0.4. Width. Clearly $\mathbb{P}(W = 0) = 0$ since $N_0 = 1$ and $\mathbb{P}(W = 1) = \mathbb{P}(N_1 = 0) = 1 - p$. An elementary expression $f(p)$ for $\mathbb{P}(W = 2)$ arises from

$$\begin{aligned} f(p) &= p(1-p)^2 + 2p^2(1-p)^3 + 4p^3(1-p)^4 + 8p^4(1-p)^5 + \dots \\ &= p(1-p)^2 (1 + 2p(1-p) (1 + 2p(1-p) (1 + 2p(1-p) \dots))) \\ &= p(1-p)^2 (1 + 2f(p)/(1-p)); \end{aligned}$$

hence

$$(1 - 2p(1 - p))f(p) = p(1 - p)^2;$$

hence

$$P(W = 2) = \frac{p(1 - p)^2}{1 - 2p(1 - p)} = \frac{p(1 - p)^2}{2p^2 - 2p + 1}.$$

An analogous argument leading to $P(W = 4)$ does not seem to work. We turn therefore to the alternative approach.

0.5. Width via Markov. Define the matrix Q exactly as before with $i \geq 1$, $j \geq 1$. Observe that [12, 13]

$$P(W \leq 0 \mid N_0 = i) = 0$$

and

$$P(W \leq m \mid N_0 = i) = (1 - p)^i + \sum_{j=1}^m \underbrace{P(W \leq m \mid N_1 = j)}_{P(W \leq m \mid N_0 = j)} \overbrace{P(N_1 = j \mid N_0 = i)}^{q_{i,j}}$$

for $m \geq 1$. Let $\alpha_i(m) = P(W > m \mid N_0 = i)$, $\beta_0 = 1$, $\beta_m = \alpha_1(m)$ and

$$\gamma_i(m) = \begin{cases} \sum_{j=m+1}^{2i} q_{i,j} & \text{if } m + 1 \leq 2i, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\begin{aligned} \alpha_i(m) &= 1 - (1 - p)^i - \sum_{j=1}^m q_{i,j}(1 - \alpha_j(m)) \\ &= \sum_{j=1}^{2i} q_{i,j} - \sum_{j=1}^m q_{i,j} + \sum_{j=1}^m q_{i,j}\alpha_j(m) \\ &= \gamma_i(m) + \sum_{j=1}^m q_{i,j}\alpha_j(m) \end{aligned}$$

and thus $\alpha(m) = (I - Q)^{-1}\gamma(m)$. Only the first component of $\alpha(m)$ is desired since $P(W > m) = P(W > m \mid N_0 = 1)$. A theorem in [14, 15] leads to a conjecture that

$$D = \lim_{\substack{m \rightarrow \infty \\ m \text{ even}}} m \left(\frac{1}{p} - 1 \right)^m \beta_m = \lim_{\substack{m \rightarrow \infty \\ m \text{ odd}}} m \left(\frac{1}{p} - 1 \right)^{m-1} \beta_m$$

exists and is nonzero. We have

$$P(W = m) = \beta_{m-1} - \beta_m,$$

$$E(e^{tW}) = \sum_{m=1}^{\infty} e^{tm}(\beta_{m-1} - \beta_m) = e^t\beta_0 + \sum_{m=1}^{\infty} (e^{t(m+1)} - e^{tm})\beta_m$$

and hence

$$P(W = 4) = \frac{p^3(1-p)^4(2p^2 - 2p - 1)}{(2p^2 - 2p + 1)(8p^6 - 24p^5 + 30p^4 - 20p^3 + 4p^2 + 2p - 1)},$$

$$E(W) = \sum_{m=0}^{\infty} \beta_m, \quad E(W^2) = \sum_{m=0}^{\infty} (2m + 1)\beta_m.$$

No closed-form expressions for the quantities in Table 2 are known.

Table 2 *Width-Related Parameters*

p	$D(p)$	$E(W)$	$\text{Var}(W)$
0.2	0.8...	1.2243696655...	0.2507547512...
0.25	1.2...	1.3038399841...	0.3903119417...
0.3	1.7...	1.4072057242...	0.6311389283...
0.35	2.6...	1.5526227137...	1.1020414724...
0.4	4.3...	1.7823528114...	2.2389987484...

For the critical case, it can be proved [16, 17] that $E(W) = \infty$ and, in fact, $E(\max_{0 \leq k \leq \ell} N_k) \sim \ln(\ell)$ as $\ell \rightarrow \infty$.

0.6. Cross-Correlation. Lacking any better methods to study association, we generated 10^6 Galton-Watson trees for each probability $p = 0.2, \dots, 0.4$. The cross-correlation coefficients between N , H and W are each large, but we observe that roughly

$$0.95 \approx \rho(N, H) > \rho(N, W) > \rho(H, W) \approx 0.85.$$

No clear pattern in these, as functions of p , are yet evident. Clearly this is an area for further research [18].

0.7. Acknowledgements. I am grateful to Robert Israel and Anthony Quas for the simple proof that $C(p)$ exists, and to Frank Purcell for his algorithm that plots a binary tree (given as a nested list).

REFERENCES

- [1] R. Otter, The multiplicative process, *Annals of Math. Statistics* 20 (1949) 206–224; MR0030716 (11,41a).
- [2] N. Minami, On the number of vertices with a given degree in a Galton-Watson tree, *Adv. in Appl. Probab.* 37 (2005) 229–264; MR2135161 (2006a:60158).
- [3] D. Hawkins and S. Ulam, Theory of multiplicative processes, Los Alamos classified report LA-171, 1944; in *Analogies between Analogies*, ed. A. R. Bednarek and F. Ulam, Univ. of California Press, 1990, pp. 1–15; available online at <http://content.cdlib.org/escholarship/>; MR1123485 (92i:01033).
- [4] I. J. Good, The number of individuals in a cascade process, *Proc. Cambridge Philos. Soc.* 45 (1949) 360–363; MR0029489 (10,614a).
- [5] A. G. Pakes, Some limit theorems for the total progeny of a branching process, *Adv. Appl. Probab.* 3 (1971) 176–192; MR0283892 (44 #1122).
- [6] W. Feller, *An Introduction to Probability Theory and Its Applications*, v. 1, 3rd ed., Wiley, 1968, pp. 293–302; MR 37 #3604.
- [7] S. R. Finch, Otter’s tree enumeration constants, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 295–316.
- [8] L. Moser, On a problem of Cayley, *Scripta Math.* 22 (1956) 289–292.
- [9] I. Guttman, On a problem of L. Moser, *Canad. Math. Bull.* 3 (1960) 35–39; MR0129027 (23 #B2064).
- [10] S. R. Finch, Optimal stopping constants, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 361–363.
- [11] A. T. Bharucha-Reid, *Elements of the Theory of Markov Processes and their Applications*, McGraw-Hill, 1960, pp. 26–27; MR0112177 (22 #3032).
- [12] J. Bishir, Maximum population size in a branching process, *Biometrics* 18 (1962) 394–403; MR0145637 (26 #3166).
- [13] S. R. Adke, The maximum population size in the first N generations of a branching process, *Biometrics* 20 (1964) 649–651; MR0181013 (31 #5243).
- [14] T. Lindvall, On the maximum of a branching process, *Scand. J. Statist.* 3 (1976) 209–214; MR0426188 (54 #14134).

- [15] O. Nerman, On the maximal generation size of a non-critical Galton-Watson process, *Scand. J. Statist.* 4 (1977) 131–135; MR0474531 (57 #14169).
- [16] K. B. Athreya, On the maximum sequence in a critical branching process, *Annals of Probab.* 16 (1988) 502–507; MR0929060 (89f:60098).
- [17] E. M. Bondarenko and V. A. Topchii, Estimates for the expectation of the maximum of a critical Galton-Watson process on a finite interval (in Russian), *Sibirsk. Mat. Zh.* 42 (2001) 249–257; Engl. transl. in *Siberian Math. J.* 42 (2001) 209–216; MR1833155 (2002b:60145).
- [18] S. R. Finch, Shapes of binary trees, unpublished note (2004).