On the complexity of branch-and-bound search for random trees

Luc Devroye
Carlos Zamora-Cura
School of Computer Science
McGill University
Montreal, Canada H3A 2K6
{luc, czamora}@cs.mcgill.ca.

Abstract. Let T_n be a b-ary tree of height n, which has independent, nonnegative, identically distributed random variables associated with each of its edges, a model previously considered by Karp, Pearl, McDiarmid and Provan. The value of a node is the sum of all the edge values on its path to the root. Consider the problem of finding the minimum leaf value of T_n. Assume that the edge random variable X is nondegenerate, has $\mathbf{E}\left\{X^\theta\right\} < \infty$ for some $\theta > 2$, and satisfies $bP\{X = c\} < 1$ where c is the leftmost point of the support of X. We analyze the performance of the standard branch-and-bound algorithm for this problem and prove that the number of nodes visited is in probability $(\beta + o(1))^n$, where $\beta \in (1, b)$ is a constant depending only on the distribution of the edge random variables. Explicit expressions for β are derived. We also show that any search algorithm must visit $(\beta + o(1))^n$ nodes with probability tending to one, so branch-and-bound is asymptotically optimal where first-order asymptotics are concerned.

Keywords and phrases. Search algorithms, probabilistic analysis, tree search, branch-and-bound, expected complexity, lower bounds.

Research of the authors was sponsored by NSERC grant A3456. The second author received a DGAPA-UNAM Scholarship.

§1. Introduction

Optimization of rooted trees is a routine problem in computer science, operations research, and the study of efficient algorithms for finding the best leaf in a rooted tree is the subject of many projects and papers in the artificial intelligence community. If leaves have values associated with them and a minimum must be found, one may do exhaustive search, branch-and-bound search (which is a depth-first search with on-line pruning of useless subtrees), backtracking, backtracking with bounded lookahead, and variations of these methods (Reingold, Nievergelt and Deo (1977), Kumar (1992)). To compare various methods, toy models have been proposed, of which the model of Karp and Pearl (1983) is perhaps the most interesting. Karp and Pearl consider a complete binary tree with n levels of edges and associate with each edge an independent Bernoulli (p) random variable. Each node v has a value V_v equal to the sum of the edge values on the path to the root. Each value V_v is available upon request, but each request costs one time unit. The objective is to find the leaf of minimal value. Karp and Pearl noted that if $2p > 1$, any algorithm must necessarily take exponential time in n, while for $2p = 1$ and $2p < 1$, ordinary uniform cost breadth-first search takes on the average $\Theta(n^2)$ and $\Theta(n)$ time. In this algorithm, one first visits all nodes of value 0, then all nodes of value 1, and so forth.

In 1990, McDiarmid and Provan and McDiarmid (1990) generalized the work of Karp of Pearl to b-ary trees and more general nonnegative edge distributions. If X is a typical edge random variable, and $p = \mathbf{P}\{X = 0\}$, where 0 is the leftmost point of the support of X, they show that if $bp < 1$, any exact search algorithm must take exponential time. It is this situation that interests us. We assume throughout that X is nondegenerate, that is, X has no atom of mass one. Also, a nonnegative nondegenerate random variable X is said to be *regular* if $bp < 1$, where $p = \mathbf{P}\{X = c\}$ and c is the leftmost point of the support of X. As the tree T_n has b^n leaves, it is important to ask what fraction of the nodes is revealed before the minimal leaf is found. In branch-and-bound search, we visit the nodes as in depth-first search, and visit v if and only if its parent's value is less than the minimal leaf value seen thus far (if any have been visited; otherwise, the node is visited unconditionally). The algorithm is of course guaranteed to find the overall minimum. If N is the number of nodes visited by the branch-and-bound algorithm (the number of values V_v revealed), we will show the following:

THEOREM 1. *Let X be a regular random variable.*

A. *If N is the number of nodes visited by branch-and-bound search, and $\mathbf{E}\left\{X^\theta\right\} < \infty$ for some $\theta > 2$, then there exists a number $\beta \in (1, b)$ such that*

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|N^{1/n} - \beta\right| > \epsilon\right\} = 0. \tag{1}$$

B. *The number N' of nodes visited by any algorithm that is guaranteed to find the optimum must be such that*

$$\lim_{n\to\infty} \mathbf{P}\left\{N'^{1/n} \le \beta - \epsilon\right\} = 0$$

for all $\epsilon > 0$.

Sometimes it is instructive to see how N compares with the size of T_n. The theorem above states that in probability,

$$N = |T_n|^{\rho+o(1)}$$

as $n \to \infty$, where $|T_n| = \sum_{i=0}^{n} b^i$ and $\rho = \log_b \beta$. The closer ρ is to zero, the more pruning is achieved. Interestingly, the pruning parameter ρ differs from distribution to distribution, and may take any value in $(0, 1)$. As $\rho > 0$, we see that an exponential explosion (in n) is unavoidable. The proofs are based upon results from branching random walks due to Biggins (1977). Especially part B of Theorem 1 is an embarrassingly straightforward corollary of Biggins' results. The contribution of this paper is twofold: first, we give an explicit form for β and ρ for all regular distributions. Secondly, Theorem 1 shows that up to first-order asymptotics, ordinary depth-first search is as good as any search strategy for all regular random variables. This is remarkable, as depth-first search can be implemented using only $O(n)$ storage. For example, the fewest nodes are visited if one explores nodes in order of their values until depth n is reached. This algorithm, however, requires storage growing exponentially quickly with n.

The same tree model was also considered by Zhang and Korf (1992), who analyzed other search strategies, such as iterative-deepening-A* and recursive best-first search. Branch-and-bound was also analyzed by Smith (1984) on a different random tree model. Wah and Yu (1985) considered branch and-bound with best-first search, and Stone and Sipala (1986) looked at backtracking. Of course, backtracking was analyzed on a host of other models, and we refer to Purdom (1983) and Brown and Purdom (1981) for just two examples. Pearl (1984) is the basic reference for the probabilistic analysis of various search strategies.

§2. Theory of Branching Random Walks.

We will use some results adapted from the theory of branching random walks in order to prove our results. For background information see, for example, Asmussen and Hering (1983), Athreya and Ney (1972), and Harris (1963). For any random variable X, we define $m(\theta) = \mathbf{E}\left\{e^{-\theta X}\right\}$, $\theta \geq 0$.

Definition. *Let $X \geq 0$ be a nondegenerate random variable. For any $a \in \mathbb{R}$, the μ-function is defined by*

$$\mu(a) = b \inf_{\theta \geq 0} \{e^{\theta a}\, m(\theta)\} = b \inf_{\theta \geq 0} \mathbf{E}\left\{e^{\theta(a-X)}\right\}.$$

For each $t > 0$, and $n > 0$, we define

$$Z^{(n)}(t) = \sharp\{w : w \text{ is a leaf in } T_n, \text{ and } V_w \leq t\},$$

where $\sharp A$ denotes the cardinality of set A.

The following results are from Kingman (1975) and Biggins (1977).

Theorem 2. *If $\mu(a) < 1$, then with probability one,*

$$Z^{(n)}(na) = 0 \qquad \textit{for all but finitely many } n.$$

If $a \in \operatorname{int}\{a : \mu(a) > 1\}$, then

$$(Z^{(n)}(na))^{1/n} \longrightarrow \mu(a) \qquad \textit{almost surely.}$$

Theorem 3. *Let T be the infinite b-ary random tree having edge values distributed as $X \geq 0$, where X is nondegenerate. Let $B_n = \min\{V_v : v \text{ is a leaf of } T_n\}$. Then,*

$$\lim_{n\to\infty} \frac{B_n}{n} = \alpha \stackrel{\text{def}}{=} \inf\{a : \mu(a) > 1\},$$

almost surely.

In order to prove Theorem 1, we will relate the time complexity of the backtrack search procedure and the distribution of the edge random variables by means of the μ-function. We now will prove some rather obvious properties of the μ-function.

THEOREM 4. *Let $X \geq 0$ be a nondegenerate random variable. Then its μ-function satisfies the following properties:*

(1) *μ is an increasing function on $[0, \infty)$.*
(2) *μ is continuous on $\operatorname{int}\{a : \mu(a) > 0\}$.*
(3) *$\log \mu$ is concave.*
(4) *$\sup_{a \in \mathbb{R}} \mu(a) \leq b$.*
(5) *If $\mathbf{E}\{X\} < \infty$, then $\mu(a) \equiv b$, on $\operatorname{int}\{t : \mu(t) > 0\}$, for $a \geq \mathbf{E}\{X\}$.*
(6) *$\lim_{a \uparrow \infty} \mu(a) = b$.*
(7) *If $X \geq c > 0$, then $\mu(a) = 0$ for $a < c$.*
(8) *Let $s = \sup\{t : \mathbf{P}\{X < t\} = 0\}$, and define $p = \mathbf{P}\{X = s\}$. Then μ is continuous on (s, ∞), $\mu(s) = bp$, and $\mu(a) = 0$ for $a < s$.*

PROOF. For proofs of (1)-(3), we refer to Biggins (1977) or Kingman (1975).
Proof of (4). Clearly $\inf_{\theta \geq 0} \mathbf{E}\left\{e^{\theta(a-X)}\right\} \leq 1$ by evaluation at $\theta = 0$.
Proof of (5). For $a \geq \mathbf{E}\{X\}$, we have

$$\begin{aligned}
\mu(a) &= b \inf_{\theta \geq 0} \mathbf{E}\left\{e^{\theta(a-X)}\right\} \\
&\geq b \inf_{\theta \geq 0} e^{\theta(a-\mathbf{E}\{X\})} \qquad \text{(by Jensen's inequality)} \\
&\geq b \qquad \text{(as the infimum is attained at } \theta = 0\text{)}.
\end{aligned}$$

Proof of (6). If $\mathbf{E}\{X\} < \infty$, this follows from (4) and (5). If $\mathbf{E}\{X\} = \infty$, then let q_{1-p} be the $(1-p)$-th quantile of X, i.e. $q_{1-p} = \sup\{t : \mathbf{P}\{X \leq t\} \leq 1-p\}$. Clearly, as $p \downarrow 0$, $q_{1-p} \uparrow \infty$, since $\mathbf{E}\{X\} = \infty$. Now, for $a \geq q_{1-p}$,

$$\begin{aligned}
\mu(a) &\geq b \left(\inf_{\theta \geq 0} e^{\theta(a-q_{1-p})}\right)(1-p) \\
&\geq b(1-p).
\end{aligned}$$

We conclude that proof by letting $p \downarrow 0$.
Proof of (7). For $a < c$,

$$\mu(a) = b \inf_{\theta \geq 0} \mathbf{E}\left\{e^{\theta(a-X)}\right\} \leq b \inf_{\theta \geq 0} e^{\theta(a-c)} \leq b \liminf_{\theta \to \infty} e^{\theta(a-c)} = 0.$$

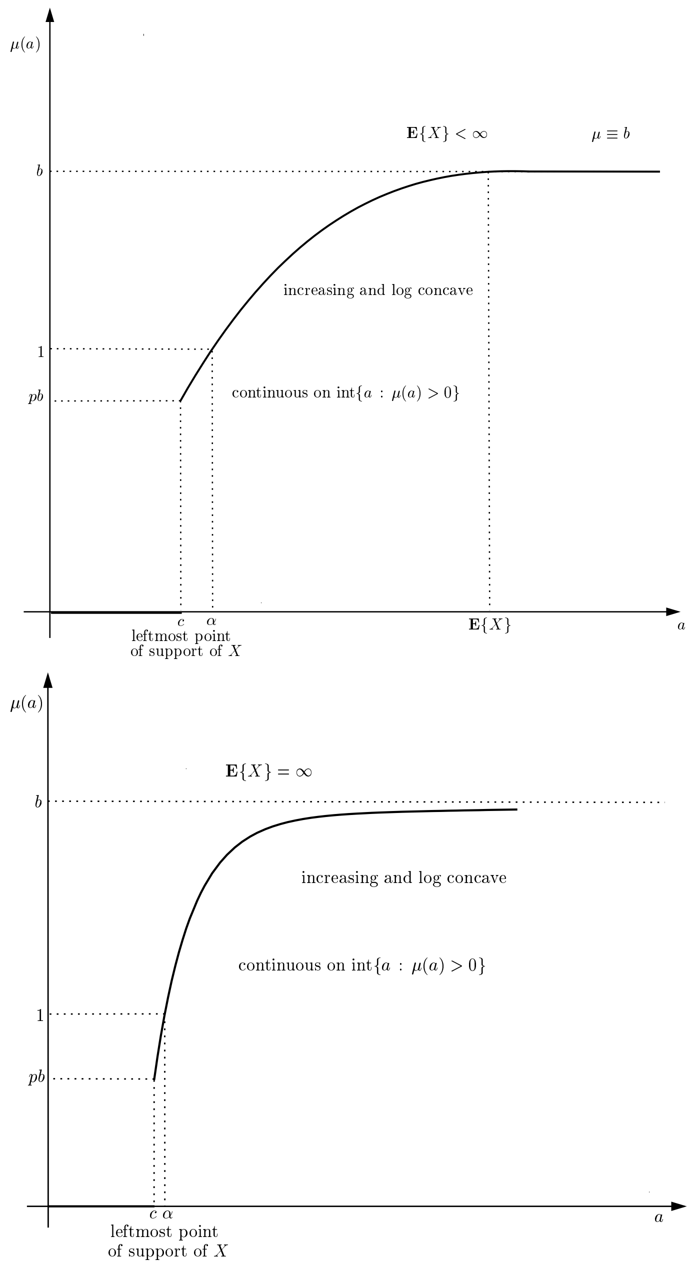

FIGURE 1. General form of the μ function.

Proof of (8). We only need to show that $\mu(s) = bp$, as the other statements follow from (2) and (7). Define $p + \delta = \mathbf{P}\{s \le X < s + \varepsilon\}$. Then for all $\varepsilon > 0$ small enough

$$\begin{aligned}\frac{\mu(s)}{b} &= \inf_{\theta\ge 0} \mathbf{E}\left\{e^{\theta(s-X)}\right\} \\ &\le \inf_{\theta\ge 0}\left\{e^{\theta(s-s)}(p+\delta) + (1-p-\delta)e^{\theta(s-s-\varepsilon)}\right\}\end{aligned}$$

$$= \inf_{\theta \geq 0} \left\{ p + \delta + (1-p-\delta) e^{-\theta\varepsilon} \right\}.$$

But the last expression goes to p as $\varepsilon \to 0$, by taking $\theta = 1/\varepsilon^2$ (notice that $\delta \to 0$ as $\varepsilon \to 0$). Thus $\mu(s) \leq bp$. Also, observe that

$$\frac{\mu(s)}{b} \geq \inf_{\theta \geq 0} p e^{\theta(s-s)} = p.$$

$\square$

We conclude that μ must always follow the pattern as described in Figure 1. Additional properties that may be needed are described below.

For regular distributions we define

$$\begin{aligned}
\alpha &= \inf\{a : \mu(a) > 1\}, \\
\beta &= \sup_{x \in (0,1)} \mu^x \left(\frac{\alpha}{x} \right), \\
\gamma &= \inf \left\{ y : \mu^y \left(\frac{\alpha}{y} \right) = \beta \right\}.
\end{aligned}$$

Clearly α is well-defined, finite and positive. Also (as we will see below), the solution γ of $\mu^\gamma(\alpha/\gamma) = \beta$ is unique, and $0 < \beta < b$, strictly.

We now prove two additional properties of the μ-function for regular distributions.

Lemma 1. *Let $X \geq 0$ be regular. Then:*

(9) *For all $\varepsilon > 0$, there is $\xi > 0$ such that*

$$\sup_{x \in (0,1)} \left\{ \mu \left(\frac{\alpha + \xi}{x} \right) - \mu \left(\frac{\alpha}{x} \right) \right\} \leq \varepsilon.$$

(10) *For all $\eta > 0$, there is $\xi > 0$ such that for all $0 \leq \nu \leq \xi$,*

$$\sup_{x \in (0,1)} \mu^x \left(\frac{\alpha + \nu}{x} \right) \leq \sup_{x \in (0,1)} \mu^x \left(\frac{\alpha}{x} \right) + \eta.$$

Proof. Proof of (9). As μ is continuous, bounded, and log-concave on $[\alpha, \infty)$, it is uniformly continuous there and thus for $\varepsilon > 0$ there is $\delta_\varepsilon > 0$ such that for all $\xi > 0$, small enough,

$$\left| \frac{\alpha + \xi}{x} - \frac{\alpha}{x} \right| = \frac{\xi}{x} \leq \delta_\varepsilon$$

implies that $\mu\left(\frac{\alpha+\xi}{x}\right)-\mu\left(\frac{\alpha}{x}\right)\le\varepsilon$. If $\xi/x>\delta_\varepsilon$, then $\alpha/x\ge(\alpha/\xi)\delta_\varepsilon$, and by choice of $\xi=\delta_\varepsilon^2$, we see that $\alpha/x\ge\alpha/\delta_\varepsilon$, so that

$$\sup_{\frac{\xi}{x}>\delta_\varepsilon}\left\{\mu\left(\frac{\alpha+\xi}{x}\right)-\mu\left(\frac{\alpha}{x}\right)\right\}\le b-\mu\left(\frac{\alpha\delta_\varepsilon}{\xi}\right)=b-\mu\left(\frac{\alpha}{\delta_\varepsilon}\right),$$

which is small enough by the choice of δ_ε.

Proof of (10). By the triangle inequality, we need only to show that

$$\sup_{x\in(0,1)}\left\{\mu^x\left(\frac{\alpha+\xi}{x}\right)-\mu^x\left(\frac{\alpha}{x}\right)\right\}\le\eta,$$

for $\xi>0$ small enough. By $(a+b)^p-a^p\le pba^{p-1}$, for all $a,b>0$, $p\in[0,1]$, we have

$$\begin{aligned}\sup_{x\in(0,1)}\left\{\mu^x\left(\frac{\alpha+\xi}{x}\right)-\mu^x\left(\frac{\alpha}{x}\right)\right\}&\le\sup_{x\in(0,1)}\left\{x\left(\mu\left(\frac{\alpha+\xi}{x}\right)-\mu\left(\frac{\alpha}{x}\right)\right)\mu^{x-1}\left(\frac{\alpha}{x}\right)\right\}\\&\le\sup_{x\in(0,1)}\left\{\mu\left(\frac{\alpha+\xi}{x}\right)-\mu\left(\frac{\alpha}{x}\right)\right\}\sup_{x\in(0,1)}\left\{x\,\mu^{x-1}\left(\frac{\alpha}{x}\right)\right\}\\&\le\sup_{x\in(0,1)}\left\{\mu\left(\frac{\alpha+\xi}{x}\right)-\mu\left(\frac{\alpha}{x}\right)\right\},\end{aligned}$$

as $1\le\mu\left(\alpha/x\right)$, for $x\in(0,1)$. □

Lemma 2. *Let $s=\sup\{t:\mathbf{P}\{X<t\}=0\}$. Assume that $s<\alpha$. Then*

$$\beta=\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha}{x}\right)<b.$$

Proof. By continuity of μ, there is $\varepsilon>0$, such that $\mu\left(\alpha/(1-\varepsilon)\right)\le\sqrt{b}$. Then, because $s<\alpha$,

$$\begin{aligned}\beta&=\max\left\{\sup_{0<x\le1-\varepsilon}\mu^x\left(\frac{\alpha}{x}\right),\sup_{1-\varepsilon<x<1}\mu^x\left(\frac{\alpha}{x}\right)\right\}\\&\le\max\left\{b^{1-\varepsilon},\mu\left(\frac{\alpha}{1-\varepsilon}\right)\right\}\\&\le\max\left\{b^{1-\varepsilon},\sqrt{b}\right\},\end{aligned}$$

the result follows. □

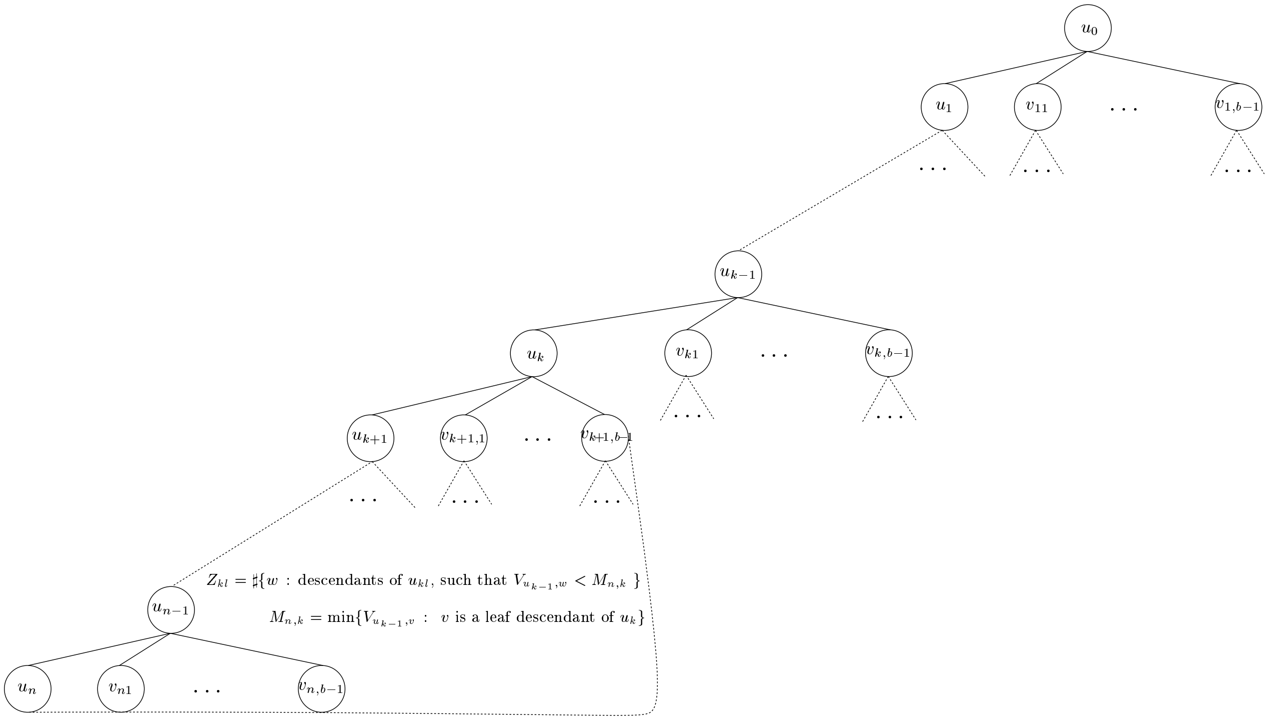

FIGURE 2. Notation for trees.

Notice that as backtrack search visits about $b^{n\rho}$ nodes, the previous result tell us that indeed, backtrack search visits only a fraction of the nodes in the tree, as $\rho = \log_b \beta < 1$. We finish this section by making the following interesting observation:

$$\begin{aligned}
\mathbf{E}\left\{Z^{(n)}(na)\right\} &= b^n\,\mathbf{P}\left\{X_1 + \cdots + X_n \le na\right\} \\
&\le \inf_{\theta \ge 0} b^n\,\mathbf{E}\left\{e^{\theta na - \theta X_1 - \cdots - \theta X_n}\right\} \text{ (for } \theta > 0\text{, by Markov's inequality)} \\
&= \inf_{\theta \ge 0} \left(b\,\mathbf{E}\left\{e^{\theta a - \theta X}\right\}\right)^n \text{ (by independence)} \\
&= (\mu(a))^n \text{ (by choice of } \theta\text{)}.
\end{aligned}$$

§3. Notation and preliminary results.

Let us first set the notation. Let $u_0, \ldots, u_n$ be the nodes in the left roof of T_n, let $v_{k,1}, \ldots, v_{k,b-1}$ be the siblings of u_k, and thus the children of u_{k-1} (see Figure 2). Let $V_{u,v}$ denote the sum of all edge values on the path from u to v. Notice that $V_{u,v} = V_v - V_u$, if v is a descendent of u. We will be particularly interested in $V_{u_{k-1},w}$, for descendants w of u_{k-1}. Let us define the following random variables that will appear in our analysis:

$$M_{n,k} = \min\left\{V_{u_{k-1},v} \;:\; v \text{ is a leaf descendant of } u_k\text{, where } v \text{ and } u_k \text{ are nodes in } T_n\right\},$$

$$Z_{kl} = \sharp\left\{w \;:\; \text{descendants of } v_{kl} \text{ (including } v_{kl} \text{ itself), such that } V_{u_{k-1},w} < M_{n,k}\right\}$$

We now present some preliminary results in order to prove part A of Theorem 1.

Proposition 1. *Assume* $\mathbf{E}\left\{X^\theta\right\} < \infty$ *for some* $\theta > 2$. *Consider* T_n, *and define* $N_k = \max\{V_w : w$ *is a descendant of* u_1 *at distance* k *from* $u_1\}$, *where* $k = \lfloor d \log_b(n-1) \rfloor$, *for* $0 < d < \theta - 2$ *fixed. Then, for every* $\xi > 0$ *there exists* $\zeta > 0$, $n_0 > 0$, *such that*

$$\mathbf{P}\left\{N_k > (n-1)\xi\right\} \le \frac{\zeta}{(n-1)^2}, \text{ for all } n \ge n_0.$$

Proof. Clearly, if $N_k > (n-1)\xi$, there must be at least one node decsendant of u_1 at depth $k+1$ having value greater than $(n-1)\xi$. Therefore $\mathbf{P}\left\{N_k > (n-1)\xi\right\} \le b^k\,\mathbf{P}\left\{\sum_{i=1}^{k+1} X_i > (n-1)\xi\right\}$, where $X_i \stackrel{\mathcal{L}}{=} X$, for $i = 1, \ldots, k$, and the X_i's are independent. Then by Markov's and Jensen's inequalities, for all n large enough,

$$\begin{aligned}
b^k\,\mathbf{P}\left\{\sum_{i=1}^{k+1} X_i > (n-1)\xi\right\} &\le \frac{b^k\,\mathbf{E}\left\{\left(\sum_{i=1}^{k+1} X_i\right)^\theta\right\}}{((n-1)\xi)^\theta} \\
&\le \frac{b^k (k+1)^\theta\,\mathbf{E}\left\{X^\theta\right\}}{((n-1)\xi)^\theta} \\
&\le \frac{b^{\lfloor d\log_b(n-1)\rfloor}\left(\lfloor d \log_b(n-1)\rfloor + 1\right)^\theta\,\mathbf{E}\left\{X^\theta\right\}}{((n-1)\xi)^\theta} \\
&\le \frac{(2d)^\theta\,\mathbf{E}\left\{X^\theta\right\}}{\xi^\theta}\frac{(\log_b(n-1))^\theta}{(n-1)^{\theta-d}} \\
&\le \frac{(2d)^\theta\,\mathbf{E}\left\{X^\theta\right\}}{\xi^\theta}\frac{1}{(n-1)^2}.
\end{aligned}$$

$\square$

Proposition 2. *Let* T_n, N_k, k, θ *and* d *as in the previous proposition. Then, for all* $\xi > 0$, *there is* $\varphi \in (0,1)$, *such that*

$$\mathbf{P}\left\{M_{n,1} > (n-1)(\alpha+\xi), N_k \le (n-1)\xi/2\right\} \le \varphi^{(n-1)^d},$$

for all n *large enough.*

Proof. Denote by $v_1, \ldots, v_{b^k}$ all nodes in T_n at depth k. Define B^i_{n-k} to be the minimum of all $V_{v_i,w}$ such that w is a leaf descendant of v_i. Then, by independence of the B^i_{n-k}'s,

$$
\begin{aligned}
&\mathbf{P}\left\{M_{n,1} > (n-1)(\alpha+\xi),\, N_k \le (n-1)\xi/2\right\} \\
&\quad \le \mathbf{P}\left\{B^1_{n-k} + (n-1)\xi/2 > (n-1)(\alpha+\xi), \ldots, B^{b^k}_{n-k} + (n-1)\xi/2 > (n-1)(\alpha+\xi)\right\} \\
&\quad = \left(\mathbf{P}\left\{B_{n-k} > (n-1)(\alpha+\xi/2)\right\}\right)^{b^k} \\
&\quad \le \left(\mathbf{P}\left\{B_{n-k} > (n-k)(\alpha+\xi/2)\right\}\right)^{b^k},
\end{aligned}
$$

where B_{n-k} is the random variable defined in Theorem 3. Substituting k by $\lfloor d\log_b(n-1)\rfloor$ and using Theorem 3 we get that

$$\mathbf{P}\left\{B_{n-\lfloor d\log_b(n-1)\rfloor} > (n-\lfloor d\log_b(n-1)\rfloor)(\alpha+\xi/2)\right\} \longrightarrow 0, \quad \text{as } n\to\infty.$$

Thus we can find $\varphi \in (0,1)$ such that

$$\mathbf{P}\left\{B_{n-\lfloor d\log_b(n-1)\rfloor} > (n-\lfloor d\log_b(n-1)\rfloor)(\alpha+\xi/2)\right\} \le \varphi,$$

for all n large enough. Hence,

$$
\begin{aligned}
&\mathbf{P}\left\{M_{n,1} > (n-1)(\alpha+\xi),\, N_k \le (n-1)\xi/2\right\} \\
&\qquad\qquad \le \left(\mathbf{P}\left\{B_{n-\lfloor d\log_b(n-1)\rfloor} > (n-\lfloor d\log_b(n-1)\rfloor)(\alpha+\xi/2)\right\}\right)^{(n-1)^d} \\
&\qquad\qquad \le \varphi^{(n-1)^d},
\end{aligned}
$$

for all n large enough. □

The following corollary is immediate from the two previous propositions.

Corollary 1. *Assume $\mathbf{E}\left\{X^\theta\right\} < \infty$ for some $\theta > 2$. Pick $d \in (0, \theta-2)$. For all $\xi > 0$, there are $\zeta > 0$, $\varphi \in (0,1)$, and n_0 such that*

$$\mathbf{P}\left\{M_{n,1} > (n-1)(\alpha+\xi)\right\} \le \frac{\zeta}{(n-1)^2} + \varphi^{(n-1)^d}, \text{ for all } n \ge n_0.$$

§4. Proof of main theorem

We first prove one half of part A of Theorem 1. The second half of part A follows from part B, and will be proved in Theorem 6 below.

THEOREM 5. *Let X be a regular random variable with $\mathbf{E}\left\{X^\theta\right\} < \infty$ for some $\theta > 2$. Then for every $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\left\{N^{1/n} > \beta + \varepsilon\right\} = 0.$$

PROOF. Let T_n be the random b-ary tree as defined in Section 1. It is clear that the number descendants of v_{kl} visited by the algorithm is smaller than or equal to $bZ_{kl} + 1$, because if the value of a node is less than the minimal leaf value seen thus far, all its b children will be visited. It follows that

$$N \le b\left(\sum_{k=1}^{n}\sum_{l=1}^{b-1} Z_{kl} + n\right) + 1.$$

Thus,

$$\begin{aligned}
\mathbf{P}\left\{N^{1/n} > \beta + \varepsilon\right\} &= \mathbf{P}\left\{N > (\beta+\varepsilon)^n\right\} \\
&\le \mathbf{P}\left\{b\sum_{k=1}^{n}\sum_{l=1}^{b-1} Z_{kl} + bn + 1 > (\beta+\varepsilon)^n\right\} \\
&\le \sum_{k=1}^{n}\sum_{l=1}^{b-1} \mathbf{P}\left\{b\, Z_{kl} > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\},
\end{aligned}$$

if $(\beta+\varepsilon)^n/((b-1)n+1) \ge bn+1$, which is true for all n large enough. Now, notice that $Z_{kl} \le 1 + b + \cdots + b^{n-k} \le b^{n-k+1}$, so that for $k \ge \lceil (1-\delta)n \rceil$,

$$bZ_{kl} \le b^{n-k+1} \le b^{\delta n+1} \le \frac{(\beta+\varepsilon)^n}{(b-1)n+1},$$

where δ is taken

$$0 < \delta \le \log_b(\beta+\varepsilon) - \frac{\log_b\left((b-1)n+1\right)+1}{n}.$$

Observe that the previous can be done for all n large enough. The previous observation implies that

$$\sum_{k=1}^{n}\sum_{l=1}^{b-1} \mathbf{P}\left\{b\, Z_{kl} > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\} \le \sum_{k=1}^{\lfloor(1-\delta)n\rfloor}\sum_{l=1}^{b-1} \mathbf{P}\left\{b\, Z_{kl} > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\} \qquad (*)$$

for all n large enough. Next, for all $\xi > 0$,

$$\begin{aligned}\mathbf{P}\left\{b\,Z_{kl} > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\}&\\ \leq \mathbf{P}\left\{M_{n,k} > (n-k)(\alpha+\xi)\right\} &+ \mathbf{P}\left\{b\,Z_{kl}\,(\alpha+\xi) > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\},\end{aligned}$$

where the random variable $Z_{kl}\,(\alpha+\xi)$ represents the number of descendants w of v_{kl}, such that $V_{u_{k-1},w}$ is not larger than $(\alpha+\xi)(n-k)$. Corollary 1 implies that there are $\zeta > 0$, $\varphi \in (0,1)$, and n_0, such that $\mathbf{P}\left\{M_{n,k} > (n-k)(\alpha+\xi)\right\}$ is smaller or equal to $\zeta/(n-k)^2 + \varphi^{(n-k)^d}$, for $n \geq n_0$, uniformly over $1 \leq k \leq (1-\delta)n$. Thus, for $n \geq n_0$,

$$\begin{aligned}\sum_{k=1}^{\lfloor(1-\delta)n\rfloor}\sum_{l=1}^{b-1}\mathbf{P}\left\{M_{n,k} > (n-k)(\alpha+\xi)\right\} &\leq \sum_{k=n-\lfloor(1-\delta)n\rfloor}^{n-1} b\left(\frac{\zeta}{k^2}+\varphi^{k^d}\right)\\ &\leq \frac{bn\zeta}{(n-\lfloor(1-\delta)n\rfloor)^2} + b\varphi^{(n-\lfloor(1-\delta)n\rfloor)^d}\lfloor(1-\delta)n\rfloor\\ &\leq \frac{b\zeta}{\delta^2 n} + b(1-\delta)\varphi^{(\delta n)^d}n\\ &\longrightarrow 0 \qquad \text{as } n\to\infty.\end{aligned}$$

Now, for the other part of the sum in $(*)$, note that

$$\begin{aligned}\mathbf{P}\left\{b\,Z_{kl}\,(\alpha+\xi) > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\} &\leq \frac{\mathbf{E}\left\{b\,Z_{kl}\,(\alpha+\xi)\right\}}{(\beta+\varepsilon)^n}\,((b-1)n+1)\\ &= \frac{\mathbf{E}\left\{b\sum_{j=1}^{n-k} Z_{kl}^j\,(\alpha+\xi)\right\}((b-1)n+1)}{(\beta+\varepsilon)^n},\end{aligned}$$

where $Z_{kl}^j\,(\alpha+\xi)$ is the number of descendants w of v_{kl} such that $V_{u_{k-1},w}$ is smaller than or equal to $(\alpha+\xi)(n-k)$, and the distance from w to v_{kl} is j. Now, for $\theta > 0$,

$$\begin{aligned}\mathbf{E}\left\{Z_{kl}^j\,(\alpha+\xi)\right\} &\leq b^j\,\mathbf{P}\left\{X_1+\cdots X_j + X_{j+1} \leq (n-k)(\alpha+\xi)\right\}\\ &\leq b^j\,\mathbf{E}\left\{e^{\theta((n-k)(\alpha+\xi)-X_1-\cdots-X_{j+1})}\right\} \text{ (by Markov's inequality)}\\ &= b^j e^{\theta(n-k)(\alpha+\xi)}\prod_{i=1}^{j+1}\mathbf{E}\left\{e^{-X_i\theta}\right\} \text{ (by the independence of the } X_i\text{'s)}\\ &= b^j e^{\theta(n-k)(\alpha+\xi)}\left(\mathbf{E}\left\{e^{-X\theta}\right\}\right)^{j+1} \text{ (because the } X_i\text{'s are identically distributed)}\\ &\leq \left(b\,\mathbf{E}\left\{e^{\theta\left(\frac{(n-k)(\alpha+\xi)}{j+1}-X\right)}\right\}\right)^{j+1}\\ &= \left(\mu\left(\frac{(n-k)(\alpha+\xi)}{j+1}\right)\right)^{j+1},\end{aligned}$$

where for the last equality to hold we took $\theta = \theta^*$ such that

$$b\,\mathbf{E}\left\{e^{\theta^*\left(\frac{(n-k)(\alpha+\xi)}{j+1}-X\right)}\right\} = \mu\left(\frac{(n-k)(\alpha+\xi)}{j+1}\right).$$

Thus, since $\mu(a)$ is an increasing function of $a \geq 0$,

$$\begin{aligned}
\mathbf{P}\left\{b\,Z_{kl}(\alpha+\xi) > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\} &\leq \frac{b((b-1)n+1)}{(\beta+\varepsilon)^n}\sum_{j=1}^{n-k}\mathbf{E}\left\{Z_{kl}^j(\alpha+\xi)\right\} \\
&\leq \frac{b((b-1)n+1)}{(\beta+\varepsilon)^n}\sum_{j=1}^{n-k}\left(\mu\left(\frac{(n-k)(\alpha+\xi)}{j+1}\right)\right)^{j+1} \\
&= \frac{b((b-1)n+1)}{(\beta+\varepsilon)^n}\sum_{j=1}^{n-k}\left(\mu^{\frac{j+1}{n+1}}\left(\frac{(n+1)(\alpha+\xi)}{j+1}\right)\right)^{n+1} \\
&\leq \frac{b((b-1)n+1)}{(\beta+\varepsilon)^n}\sum_{j=1}^{n-k}\left(\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi}{x}\right)\right)^{n+1} \\
&\leq \frac{nb((b-1)n+1)}{(\beta+\varepsilon)^n}\left(\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi}{x}\right)\right)^{n+1}.
\end{aligned}$$

We now sum all the terms and get

$$\begin{aligned}
&\sum_{k=1}^{\lfloor(1-\delta)n\rfloor}\sum_{l=1}^{b-1}\mathbf{P}\left\{b\,Z_{kl}(\alpha+\xi) > \frac{(\beta+\varepsilon)^n}{(b-1)n+1}\right\} \\
&\qquad\leq \frac{n^2b(b-1)((b-1)n+1)}{(\beta+\varepsilon)^n}\left(\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi}{x}\right)\right)^{n+1} \\
&\qquad\leq \frac{b^3n^3}{(\beta+\varepsilon)^n}\left[\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi}{x}\right)\right]^{n+1}.
\end{aligned}$$

Notice that if ξ were equal to 0, then the quantity in the square brackets would be β^{n+1}. Lemma 1 implies that for $\varepsilon > 0$, there exists $\xi^* > 0$ such that

$$\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi^*}{x}\right) \leq \sup_{x\in(0,1)}\mu^x\left(\frac{\alpha}{x}\right) + \frac{\varepsilon}{2} = \beta + \frac{\varepsilon}{2}.$$

Therefore,

$$\begin{aligned}
\frac{b^3n^3}{(\beta+\varepsilon)^n}\left(\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi^*}{x}\right)\right)^{n+1} &\leq b^3n^3\left(\frac{\sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi^*}{x}\right)}{\beta+\varepsilon}\right)^n \sup_{x\in(0,1)}\mu^x\left(\frac{\alpha+\xi^*}{x}\right) \\
&\leq b^3n^3\left(\frac{\beta+\varepsilon/2}{\beta+\varepsilon}\right)^n(\beta+\varepsilon/2) \\
&\underset{n\to\infty}{\longrightarrow} 0. \qquad \square
\end{aligned}$$

To finish the proof of Theorem 1, we show the universal lower bound of part B. Note that any algorithm must visit all nodes v in T_n with value strictly less than B_n, the minimal leaf value in T_n. Thus, it suffices to prove the following.

THEOREM 6. *Let $X \geq 0$ be a regular random variable. Let N' be the number of nodes v in T_n with value $V_v < B_n$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{N'^{1/n} < \beta - \epsilon\right\} = 0.$$

PROOF. We use the notation from Theorem 2, and note that

$$N' \geq \sum_{j=0}^{n} Z^{(j)}(B_n - 1).$$

Define α, β and γ as in section 2, and set $k = \lfloor \gamma n \rfloor$. Thus, for any $\xi \in (0, \alpha)$,

$$N' \geq Z^{(k)}\left((\alpha - \xi)n\right) I_{[B_n - 1 > (\alpha-\xi)n]}.$$

By Theorem 3, with probability one, $B_n - 1 > (\alpha - \xi)n$ for all n large enough. Thus, we are done if we can show that given $\epsilon > 0$, we can find $\xi > 0$ such that

$$\lim_{n\to\infty} \mathbf{P}\left\{\left(Z^{(k)}\left((\alpha - \xi)n\right)\right)^{1/n} < \beta - \epsilon\right\} = 0.$$

To this effect, observe that

$$\begin{aligned}
\left(Z^{(k)}\left((\alpha - \xi)n\right)\right)^{1/n} &= \left[\left(Z^{(k)}\left(\frac{(\alpha-\xi)n}{k}k\right)\right)^{k/n}\right]^{1/k} \\
&\geq \left[\left(Z^{(k)}\left(\frac{\alpha-\xi}{\gamma}k\right)\right)^{\gamma-\frac{1}{n}}\right]^{1/k} \\
&\to \left[\mu\left(\frac{\alpha-\xi}{\gamma}\right)\right]^{\gamma}
\end{aligned}$$

by Theorem 2. By the continuity of μ, the lower bound tends to β as $\xi \downarrow 0$, since $\beta = \mu^{\gamma}(\alpha/\gamma)$. $\square$

b	ρ	b	ρ
2	.334648	3	.257101
4	.220361	5	.198027
6	.182672	7	.171302
8	.162452	9	.155311
10	.149393	11	.144383
12	.140071	13	.136306
14	.132983	15	.130019
16	.127354	17	.124941
18	.122741	19	.120725
20	.118868	21	.117149

TABLE 1. Values of b vs. ρ for the exponential distribution.

§5. Some examples.

In this section we will present some examples of μ-functions and β-values for some well-known distributions.

EXAMPLE. Exponential Distribution. If X is exponentially distributed, $m(\theta) = 1/(\theta+1)$, and

$$\mathbf{E}\left\{e^{\theta(a-X)}\right\} = \frac{e^{\theta a}}{\theta+1},$$

for $\theta \geq 0$. Also,

$$\mu(a) = b \inf_{\theta\geq 0}\left\{\frac{e^{\theta a}}{\theta+1}\right\} = \begin{cases} bae^{1-a}, & a \leq 1; \\ b \qquad , & \text{otherwise,} \end{cases}$$

because $\log\left(\frac{be^{\theta a}}{\theta}\right)$ is minimum for $\theta = \frac{1}{a} - 1$. The value α is defined to be the unique solution of

$$\alpha b e^{1-\alpha} = 1. \tag{2}$$

Observe that as $b \to \infty$, $\alpha \sim 1/2b$. Note that

$$\begin{aligned} \sup_{x\in(0,1)} \mu^x\left(\frac{\alpha}{x}\right) &= \sup_{x\in(\alpha,1)}\left\{\frac{\alpha}{x}be^{1-\frac{\alpha}{x}}\right\}^x \vee \sup_{x\in(0,\alpha)} b^x \\ &= \sup_{x\in(\alpha,1)}\left\{\frac{e^{\alpha x}}{x^x}e^{-\alpha}\right\} \vee b^\alpha, \end{aligned}$$

b	ρ	b	ρ
2	.590941	3	.455860
4	.372014	5	.311999
6	.265677	7	.228184
8	.196821	9	.169940
10	.146465	11	.125653
12	.106973	13	.090028
14	.074516	15	.061973
16	.046876	17	.034384
18	.022566	19	.011245

TABLE 2. Values of b vs. ρ for the Bernoulli (.05) distribution.

because $\alpha b = e^{\alpha-1}$. The value of x which maximizes the first quantity is $e^{\alpha-1}$ (which is in $(\alpha, 1)$). By manipulating (2) we get

$$\begin{aligned}\beta &= \left\{ e^{-\alpha} \frac{e^{\alpha e^{\alpha-1}}}{e^{\alpha e^{\alpha-1}} e^{-e^{\alpha-1}}} \right\} \vee b^{\alpha} \\ &= \max\left\{ e^{e^{\alpha-1}-\alpha}, b^{\alpha} \right\} \\ &= e^{\alpha(b-1)},\end{aligned}$$

because $e^{\alpha-1} = \alpha b$. Note that $e^{\alpha(b-1)} < e$ for any value of b, and $\beta \to \sqrt{e}$ as $b \to \infty$. Also, $\rho = \log_b \beta \sim 1/(2\log b)$, as $b \to \infty$.

EXAMPLE. Bernoulli distribution with parameter $p \in (0,1)$. Let X be a Bernoulli(p) distributed random variable, then

$$\mathbf{E}\left\{ e^{-\theta X} \right\} = p + (1-p)e^{-\theta},$$

and

$$\begin{aligned}\mu(a) &= b \inf_{\theta \geq 0} \mathbf{E}\left\{ e^{\theta(a-X)} \right\} \\ &= b \inf_{\theta \geq 0} e^{\theta a}(p + (1-p)e^{\theta}) \\ &= \begin{cases} b\left(\frac{p}{1-a}\right)^{1-a}\left(\frac{1-p}{a}\right)^{a}, & 0 < a \leq 1-p, \\ b &, \quad a \geq 1-p. \end{cases}\end{aligned}$$

b	ρ	b	ρ
2	.522782	3	.445452
4	.405182	5	.379173
6	.360477	7	.346137
8	.334648	9	.325150
10	.317110	11	.310176
12	.304107	13	.298731
14	.293919	15	.289575
16	.285625	17	.282010
18	.278682	19	.275604
20	.272745	21	.270078

TABLE 3. Values of b vs. ρ for the Gamma (3) distribution.

So as to define α, we must assume that $bp < 1$. Thus α is defined as the solution of the equation

$$b\left(\frac{p}{1-\alpha}\right)^{1-\alpha}\left(\frac{1-p}{\alpha}\right)^{\alpha} = 1.$$

EXAMPLE. Gamma distribution with parameter r. We first compute

$$\begin{aligned}
\mu(a) &= b \inf_{\theta \geq 0} \int_0^\infty \frac{x^{r-1}e^{-x}e^{\theta(a-x)}}{\Gamma(r)} dx \\
&= b \inf_{\theta\geq 0} \frac{e^{\theta a}}{(1+\theta)^r} \\
&= b\frac{e^{a\left(\frac{r}{a}-1\right)}}{\left(\frac{r}{a}\right)^r} \vee b \\
&= \begin{cases} be^{r-a}\left(\frac{a}{r}\right)^r, & \text{if } 0 \leq a \leq r; \\ b & , \quad a \geq r. \end{cases}
\end{aligned}$$

because $\theta a - r\log(1+\theta)$ is minimal when $a = \frac{r}{1+\theta}$. Thus α is defined as the solution to the equation:

$$be^{r-\alpha}\left(\frac{\alpha}{r}\right)^r = 1.$$

Finally,

$$\beta = \sup_{x\in(0,1)} \mu^x\left(\frac{\alpha}{x}\right) = \sup_{x\in(0,1)} b^x e^{rx-\alpha}\left(\frac{\alpha}{rx}\right)^{rx}.$$

Note that $\mu^x(\alpha/x)$ is maximal at $x = b^{1/r}/r$ and therefore

$$\beta = \frac{b^{b^{1/r}\alpha/r}e^{\alpha(b^{1/r}-1)}}{b^{b^{1/r}\alpha/r}} = e^{\alpha(b^{1/r}-1)}.$$

§6. References.

Asmussen, S., and Hering, H. *Branching Processes.* Birkhäuser, Boston, 1983.

Athreya, K. B., and Ney, P. E. *Branching Processes.* Springer-Verlag, Berlin, 1972.

Biggins, J. D. Chernoff's theorem in the branching random walk. *Journal of Applied Probability 14* (1977), 630–636.

Brown, C. A., and Purdom, P. W. An average time analysis of backtracking. *SIAM Journal of Computing 10* (1981), 583–593.

Hammersley, J. M. Postulates for subadditive processes. *Annals of Probability 2* (1974), 652–680.

Harris, T. E. *The Theory of Branching Processes.* Springer-Verlag, Berlin, 1963.

Karp, R. M., and Pearl, J. Searching for an optimal path in a tree with random costs. *Artificial Intelligence 21* (1983), 99–117.

Kingman, J. F. C. The first-birth problem for an age-dependant branching process. *Annals of Probability 3*, 5 (1975), 341–345.

Kumar, V. Search, branch and bound, in: *Encyclopedia of Artificial Intelligence (2nd edition).* Ed. S. C. Shapiro, pp. 1468–1472. Wiley-Interscience, 1992.

McDiarmid, C. J. H. Probabilistic analysis of tree search, in: *Disorder in Physical Systems.* Ed. D. J. A. Welsh and G. R. Grimmett, pp. 249–260. Oxford Science Publications, 1990.

McDiarmid, C. J. H., and Provan, G. M. A. An expected-cost analysis of backtracking and non-backtracking algorithms. *IJCAI-91: Proceedings of the Twelfth International Conference on Artificial Intelligence* (1991), 172–177. Morgan Kaufmann Publishing, San Mateo, CA,.

Pearl, J. *Heuristics: Intelligent Search Strategies for Computer Problem Solving.* Addison-Wesley, Reading, Mass, 1984.

Purdom, P. W. Search rearrangement backtracking and polynomial average time. *Artificial Intelligence 21* (1983), 117–133.

Reingold, E. M., Nievergelt, J., and Deo, N. *Combinatorial Algorithms: Theory and Practice*. Prentice Hall, Englewood Cliffs, N.J., 1977.

Smith, D. R. Random trees and the analysis of branch and bound procedures. *Journal of the ACM, 31* (1984), 163–188.

Stone, H. S., and Sipala, P. The average complexity of depth-first search with backtracking and cutoff. *IBM Journal of Research and Development 30* (1986), 242–258.

Wah, B. W., and Yu, C. F. Stochastic modeling of branch-and-bound algorithms with best-first search. *IEEE Transactions of Software Engineering SE-11* (1985), 922–934.

Zhang, W., and Korf, R. E. An average-case analysis of branch-and-bound with applications. *Proceedings of the 10th National Conference on AI—AAAI-92* (1992), 1–6. San Jose, CA.