Table of Contents

List of Contributors

Luc Devroye
School of Computer Science,
McGill University,
Montreal, H3A 2K6, Canada

Alan M. Frieze
Mathematical Sciences Department
Carnegie-Mellon University,
Pittsburgh, PA 15213, USA

Mark Jerrum
Department of Computer Science,
University of Edinburgh,
The King's Buildings,
Edinburgh EHY 3JZ,
United Kingdom

Colin McDiarmid
Statistics Department,
University of Oxford,
1 South Park Road,
Oxford OX1 3TG,
United Kingdom

Michael Molloy
Department of Computer Science,
University of Toronto,
Toronto, Canada

Rajeev Motwani
Department of Computer Science,
Standford University,
CA 94305, USA

Prabhakar Raghavan
IBM Almaden Research Center,
650 Harry Road,
San Jose, CA 95120, USA

Bruce Reed
Equipe Combinatoire,
Université Pierre et Marie Curie,
4,Place Jussieu 75252 Paris Cedex 5,
France

Dominic Welsh
Mathematical Institute,
University of Oxford,
24-29 St. Giles',
Oxford OX1 3LB,
United Kingdom

Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms

Luc Devroye

School of Computer Science, McGill University, Montreal, Canada.

Summary. We give a partial overview of some results from the rich theory of branching processes and illustrate their use in the probabilistic analysis of algorithms and data structures. The branching processes we discuss include the Galton-Watson process, the branching random walk, the Crump-Mode-Jagers process, and conditional branching processes. The applications include the analysis of the height of random binary search trees, random m-ary search trees, quadtrees, union-find trees, uniform random recursive trees and plane-oriented recursive trees. All these trees have heights that grow logarithmically in the size of the tree. A different behavior is observed for the combinatorial models of trees, where one considers the uniform distribution over all trees in a certain family of trees. In many cases, such trees are distributed like trees in a Galton-Watson process conditioned on the tree size. This fact allows us to review Cayley trees (random labeled free trees), random binary trees, random unary-binary trees, random oriented plane trees, and indeed many other species of uniform trees. We also review a combinatorial optimization problem first suggested by Karp and Pearl. The analysis there is particularly beautiful and shows the flexibility of even the simplest branching processes.

1. Branching Processes

1.1 Branching Processes

Around 1873, Galton and Watson came up with a model for explaining the disappearance of certain family names in England (see the historical survey by Kendall, 1966). Their model, now known as the Galton-Watson process , is extremely simple: in a population, we begin with one pater familias, or root. The root has Z_1 children, where Z_1 has a fixed distribution (the reproduction distribution): it is convenient to let Z denote a prototypical random variable with this distribution, and to set

$$p_i = \mathbf{Pr}(Z = i) \ , \ \ i \geq 0 \ .$$

Each child in turn reproduces independently according to the same distribution, and so forth. This leads to a random tree, the Galton-Watson tree, and a random process, the Galton-Watson process. Let Z_i denote the number of particles in the i-th generation, with $Z_0 = 1$. Only one of two possible situations can occur: either the population survives forever ($Z_i > 0$ for all i), or it becomes extinct after a finite time. To analyze the Galton-Watson process it is convenient to use the RGF (the reproduction generating function), or simply generating function

$$f(s) = \sum_{k=0}^{\infty} p_k s^k = \mathbf{E}(s^{Z_1}) \ , \ s \in [0,1] \ .$$

This is a function of s that contains exactly the same information as the vector $(p_0, p_1, \ldots)$. It is strictly convex (if $p_1 \neq 1$) and increases from p_0 at $s = 0$ to 1 at $s = 1$. Different RGF's define different Galton-Watson branching processes. Intuitively, it should be clear that a population explodes if the expected number of children per particle is greater than one, and that it is bound to shrink if it is less than one. An important parameter thus is the expected number of children (or Malthusian parameter)

$$m = \mathbf{E}(Z) = \mathbf{E}(Z_1) = \sum_{k=0}^{\infty} k p_k = f'(1) \ .$$

We will prove that this intuition is partly correct. In fact, whether a population explodes or becomes extinct depends solely on the value of m, and not on the individual probabilities of the RGF! Consider the RGF for Z_n, the size of the n-th generation:

$$f_n(s) \stackrel{\text{def}}{=} \mathbf{E}(s^{Z_n}) \ , \ 0 \le s \le 1 \ .$$

With this notation, we clearly have $f_1(s) \equiv f(s)$, and $f_0(s) = s$. Conditional expectations help us in relating f_n to f. To this end, let Z_{n-1} be the number of particles in generation $n-1$. These have offspring of sizes $Y_n(1)$, ..., $Y_n(Z_{n-1})$, and these form an independently identically distributed (i.i.d.) sequence distributed as Z_1 (i.e., all the $Y_n(j)$ have the same distibution as Z_1 and the choices of the $Y_n(j)$ are made independently). Therefore,

$$\begin{aligned} f_n(s) &= \mathbf{E}\left(\mathbf{E}\left(s^{Z_n} | Z_{n-1}\right)\right) \\ &= \mathbf{E}\left(\mathbf{E}\left(s^{Y_n(1)+\cdots+Y_n(Z_{n-1})} | Z_{n-1}\right)\right) \\ &= \mathbf{E}\left(\prod_{j=1}^{Z_{n-1}} \mathbf{E}\left(s^{Y_n(j)} | Z_{n-1}\right)\right) \quad \text{(by independence)} \\ &= \mathbf{E}\left(\prod_{j=1}^{Z_{n-1}} \mathbf{E}\left(s^{Z_1}\right)\right) \quad \text{(identical distributions)} \\ &= \mathbf{E}\left((f(s))^{Z_{n-1}}\right) \\ &= f_{n-1}(f(s)) \\ &= \cdots \\ &= \overbrace{f(f(f(\cdots)))}^{n \text{ times}} \ . \end{aligned}$$

When $m < 1$, the graph of $f(s)$ lies above s and $f(s) = s$ only at $s = 1$. It is not difficult to see that $f_n(s) \to 1$ for any s. In particular, $f_n(0) = \mathbf{Pr}(Z_n = 0) \to 1$. When $m > 1$, there is a unique solution q of $f(s) = s$ that is less than one. See the figure below. It is easy to see that for any $s \in [0,1)$, $f_n(s) \to q$. In particular, $\mathbf{Pr}(Z_n = 0) \to q$.

We now show that q is the probability that the process becomes extinct. The point I am making here is subtle, but important, as the event "extinction" relates to the entire history of the process, not a particular n. Note the following:

$$\begin{aligned}\mathbf{Pr}(\text{extinction}) &= \mathbf{Pr}(Z_n = 0 \text{ for some } n) \\ &= \mathbf{Pr}(\cup_{i=1}^{\infty}[Z_i = 0]) \\ &= \lim_{n\to\infty} \mathbf{Pr}(\cup_{i=1}^{n}[Z_i = 0]) \\ &= \lim_{n\to\infty} \mathbf{Pr}(Z_n = 0) \\ &= q \ .\end{aligned}$$

Therefore, q is the extinction probability. We have thus shown the fundamental property of Galton-Watson processes:

Theorem 1.1. *In a Galton-Watson process, if $m > 1$, then*

$$q = \mathbf{Pr}(Z_n = 0 \text{ for some } n) = \lim_{n\to\infty} \mathbf{Pr}(Z_n = 0) < 1 \ .$$

When $m \leq 1$, the process becomes extinct with probability one, unless we have the degenerate case $p_1 = 1$, in which case every generation contains one particle.

Processes are called supercritical, critical and subcritical when $m > 1$, $m = 1$ and $m < 1$ respectively. We also introduce the hypercritical processes, which have $m = \infty$, and the exploding processes (which may be of any of the four types above) which have $\mathbf{E}(Z_1 \log Z_1) = \infty$. The last two terms are non-standard, but will be convenient to work with. It is worth noting that in all cases,

$$\mathbf{E}(Z_n) = (\mathbf{E}(Z_1))^n = m^n$$

(by induction and conditioning, as $\mathbf{E}(Z_n | Z_{n-1}) = mZ_{n-1}$). In the critical case, the expected size of the population remains constant, while the population becomes extinct with probability one.

1.2 Some Limit Results

Theorem 1.2. *Assume that $p_1 < 1$. In a Galton-Watson branching process, $\mathbf{Pr}(\lim_{n\to\infty} Z_n \in \{0,\infty\}) = 1$.*

Proof. Clearly,

$$\mathbf{Pr}(\lim_{n\to\infty} Z_n \notin \{0,\infty\}) \leq \sum_{k=1}^{\infty} \mathbf{Pr}(Z_n = k \text{ infinitely often})$$

and this is zero if every term is zero. Thus, it suffices to show that for every finite k,

$$\mathbf{Pr}(Z_n = k \text{ infinitely often}) = 0 \ .$$

We say that the population is in state k if $Z_n = k$. Let r_k be the probability that the population returns to state k given that we are in state k now, so that $1-r_k$ is the probability that we wander off forever ($Z_j \neq k$ for all $j > n$). If $p_0 = 0$, then

$$r_k \leq \mathbf{Pr}(Z_1 = k | Z_0 = k) = p_1^k < 1 \ .$$

If $p_0 > 0$, then

$$r_k \leq \mathbf{Pr}(Z_1 > 0 | Z_0 = k) = 1 - p_0^k < 1 \ .$$

Therefore, $r_k < 1$.

If X is the number of visits to state k, then

$$\mathbf{Pr}(X \geq n) \leq r_k^{n-1}$$

because we need to have at least $n-1$ transitions from state k to state k in the process driven by the transition probability r_k. Note that

$$\mathbf{E}(X) = \sum_{n=1}^{\infty} \mathbf{Pr}(X \geq n) \leq \sum_{n=0}^{\infty} r_k^n = \frac{1}{1-r_k} \ .$$

Take M arbitrary. Finally,

$$\begin{aligned} \mathbf{Pr}(Z_n = k \text{ infinitely often}) & \leq \mathbf{Pr}(X \geq M) \\ & \leq \tfrac{\mathbf{E}(X)}{M} \\ & \leq \tfrac{1}{M(1-r_k)} \ , \end{aligned}$$

which is as small as desired by our choice of M. We conclude that

$$\mathbf{Pr}(Z_n = k \text{ infinitely often}) = 0 \ .$$

□

Theorem 1.2, which is valid for any $m \in [0, \infty]$, shows that it is impossible to have oscillating populations, that is, populations in which the size drops below some finite level infinitely often when $m > 1$: in fact, with probability one, the limit of Z_n is zero or infinity. The remainder of this section is more advanced and rather technical. It can be skipped without harm (except for the definition of convergence in distribution and the statement of Fatou's Lemma, which can be returned to when and if required).

We can improve on Theorem 1.2 by using that Z_n behaves roughly speaking as m^n (recall that $\mathbf{E}(Z_n) = m^n$), and its behavior is best captured in Doob's limit law:

Theorem 1.3. *(Doob's limit law). Let m be finite. The random variables $W_n = Z_n / m^n$ form a martingale sequence with $\mathbf{E}(W_n) \equiv 1$, and $W_n \to W$ almost surely as $n \to \infty$, where W is a nonnegative random variable.*

For readers not familiar with martingales, we refer to the chapter on concentration inequalities by McDiarmid in the present volume.

The remainder of this section is more advanced and rather technical. It can be skipped without harm.

We use the symbol $\xrightarrow{\mathcal{L}}$ for convergence in distribution. For random variables $(X_n)_n$ and X, and a distribution function F, we say that $X_n \xrightarrow{\mathcal{L}} X$ or $X_n \xrightarrow{\mathcal{L}} F$ when for all $x \in \mathbb{R}$ at which $F(x) = \mathbf{Pr}(X \leq x)$ is continuous, $\mathbf{Pr}(X_n \leq x) \to F(x)$.

While we don't know the limit distribution of W_n in general, we know a lot about it: in case $m \leq 1$, $p_1 < 1$, we have $\mathbf{Pr}(W = 0) = 1$, an uninteresting case. If $m > 1$ and $\sigma^2 = \mathbf{Var}(Z) < \infty$, then $\mathbf{Pr}(W = 0) = q$, $\mathbf{E}(W) = 1$, $\mathbf{Var}\,(W) = \sigma^2/(m^2 - m)$, and $\mathbf{E}(W_n - W)^2 \to 0$. In fact, the second moment condition on Z is too strict, as the following result shows:

Theorem 1.4. *(Kesten-Stigum theorem, 1966). For a supercritical Galton-Watson process , the following properties are equivalent:*

A. $\lim_{n\to\infty} \mathbf{E}(|W_n - W|) = 0$;
B. $\mathbf{E}(Z \log(1 + Z)) < \infty$;
C. $\mathbf{E}(W) = 1$;
D. $\mathbf{Pr}(W = 0) = q$.

When $m > 1$, then the above results imply

$$\frac{\log Z_n}{n} \to \log m$$

almost surely on non-extinction. Note that in general, by Fatou's lemma (which in a special form states that for positive sequences of functions f_n with $\liminf_{n\to\infty} f_n = f$, $\liminf_{n\to\infty} \int f_n \geq \int f$), we have (as expected values are just integrals)

$$\mathbf{E}(W) \leq \liminf_{n\to\infty} \mathbf{E}(W_n) = 1$$

but we cannot conclude that $\mathbf{E}(W) = 1$. Indeed, when $m \leq 1$ and $p_1 < 1$, $W = 0$ almost surely, and when $m > 1$, there exist distributions for Z for which $W = 0$ almost surely as well! In the critical case, $Z_n \to 0$ almost surely, so finer results are needed.

We can avoid the extinction problem by studying the branching process conditional on survival at time n ($Z_n > 0$). Some results for the critical case are provided in the following theorem:

Theorem 1.5. *(Kesten, Ney and Spitzer, 1966). Assume that $m = 1$ and $\sigma^2 = \mathbf{Var}(Z) \leq \infty$. Let E be an exponentially distributed random variable (that is, a random variable with density e^{-x} on $[0, \infty)$). Then*

$$\lim_{n\to\infty} n\mathbf{Pr}(Z_n > 0) = \frac{2}{\sigma^2}\ .$$

Furthermore, if $\sigma^2 < \infty$, $Z'_n/n \xrightarrow{\mathcal{L}} \sigma^2 E/2$, *where* Z'_n *is distributed as* Z_n *given* $Z_n > 0$. *If* $\sigma^2 = \infty$, *then* $Z'_n/n \to \infty$ *in probability, and*
$\lim_{n\to\infty} n\mathbf{Pr}(Z_n > 0) = 0$.

Under the stronger condition $\mathbf{E}(Z^3) < \infty$, the theorem above is referred to as the Kolmogorov-Yaglom theorem after Kolmogorov (1938) and Yaglom (1947). The conditional random variable Z'_n is also useful to understand sub-critical branching processes . The main results in this respect are again provided by Yaglom (1947) and Heathcote, Seneta and Vere-Jones (1967) (see also Asmussen and Hering, 1983 and Lyons, 1997):

Theorem 1.6. *(Yaglom-Heathcote-Seneta-Vere-Jones theorem). If* $m < 1$, *then* $Z'_n \xrightarrow{\mathcal{L}} V$, *where* $\mathbf{Pr}(V < \infty) = 1$. *Furthermore,* $\mathbf{Pr}(Z_n > 0)/m^n$ *is nonincreasing (for any* m*). Finally, the following properties are equivalent:*

A. $\lim_{n\to\infty} \mathbf{Pr}(Z_n > 0)/m^n > 0$;
B. $\sup_n \mathbf{E}(Z'_n) = \sup_n \mathbf{E}(Z_n|Z_n > 0) < \infty$;
C. $\mathbf{E}(Z\log(Z+1)) < \infty$.

Proof. We will not give a complete proof here. However, it is worthwhile to note Lyons' proof of the equivalence of A and B. We know that for any m,

$$\mathbf{Pr}(Z_n > 0) = \frac{\mathbf{E}(Z_n}{)\mathbf{E}(Z_n|Z_n > 0)} = \frac{m^n}{\mathbf{E}(Z'_n)} .$$

Thus, $\mathbf{Pr}(Z_n > 0)/m^n \downarrow$ if $\mathbf{E}(Z'_n) \uparrow$. Thus, A is equivalent to B if we can prove that $\mathbf{E}(Z'_n) \uparrow$. Let Y_n be the size of the n-th generation in the subtree rooted at the leftmost child of the root with a descendant in the n-th generation, and let I_n be the index of this child (counted from left to right). Then, as $Z_n \geq Y_n$, for any $k \geq 1$,

$$\begin{aligned}
\mathbf{Pr}(Z_n \geq k|Z_n > 0) &\geq \mathbf{Pr}(Y_n \geq k|Z_n > 0) \\
&= \textstyle\sum_j \mathbf{Pr}(Y_n \geq k, I_n = j|Z_n > 0) \\
&= \textstyle\sum_j \mathbf{Pr}(Y_n \geq k|I_n = j, Z_n > 0)\mathbf{Pr}(I_n = j|Z_n > 0) \\
&= \textstyle\sum_j \mathbf{Pr}(Z_{n-1} \geq k|Z_{n-1} > 0)\mathbf{Pr}(I_n = j|Z_n > 0) \\
&= \mathbf{Pr}(Z_{n-1} \geq k|Z_{n-1} > 0) .
\end{aligned}$$

□

1.3 Bibliographic Remarks

For an account of the theory of branching processes, see Athreya and Ney (1972), Grimmett and Stirzaker (1992), Harris (1963), Jagers (1975), or Asmussen and Hering (1983). Kendall (1966) gives an enjoyable historical overview. Neveu (1986) provides a rigorous background for studying random trees in general and Galton-Watson trees in particular. A modern proof of the Kesten-Stigum, Kolmogorov-Yaglom and Heathcote-Seneta-Vere-Jones theorems based on Galton-Watson processes with immigration and/or trees with distinguished paths may be found in Lyons, Pemantle and Peres (1993, 1995). In these papers, size-biased trees are introduced that scale probabilities of events in the n-th generation by Z_n/m^n, which turns out be equivalent to looking at $\lim_{n\to\infty} \mathbf{Pr}(.|Z_n > 0)$. The idea of size-biasing is also due to Hawkes (1981) and Joffe and Waugh (1982).

For critical processes, Weiner (1984) showed that there exist positive constants $a \le b$ such that $\mathbf{E}(\max_{1\le i\le n} Z_i) \in [a \log n, b \log n]$ and $\mathbf{Var}(\max_{1\le i\le n} Z_i) \in [an, bn]$.

For a supercritical process, Heyde (1970) shows that if Z has a finite variance σ^2, and $Z_n/m^n \to W$ almost surely, then $(W - Z_n/m^n)m^{n/2}$ converges in distribution a random variable Y. Thus, Z_n/m^n is rather concentrated around W. Conditional on $Z_n > 0$,

$$\frac{m^n(W - W_n)\sqrt{m^2 - m}}{\sqrt{Z_n}\sigma} \xrightarrow{\mathcal{L}} \mathcal{N} ,$$

where $\mathcal{N}$ denotes the normal distribution (Heyde, 1971). A Berry-Esseen type inequality to quantify this convergence is given by Heyde and Brown (1971). Again on the non-extinction set $W > 0$, we have almost surely

$$\limsup_{n\to\infty} \frac{m^n W - Z_n}{\sqrt{2\sigma^2(m^2 - m)^{-1} Z_n \log n}} = 1$$

and a similar statement for the limit infimum with 1 replaced by -1 on the right-hand side.

The tail behavior of W was investigated by Bingham (1988), who showed faster than exponential drop-offs. For finite n, super-exponential tail inequalities for $\mathbf{Pr}(Z_n > c\mathbf{E}(Z_n))$ and $\mathbf{Pr}(Z_n < \mathbf{E}(Z_n)/c)$ for large c were derived by Karp and Zhang (1995). See also Biggins and Bingham (1993) about the description of W.

Darling (1970) describes the behavior when Z has very large tails, so that, in fact, $\log(Z_n + 1)/b^n$ tends to a limit law for some $b > 1$. Here, Z_n increases as a doubly exponentially quickly. This sort of transformation is necessary, because, as shown by Seneta (1969), if $m = \infty$, then no constants c_n can exist such that Z_n/c_n converges in distribution to a non-degenerate random variable.

2. Search Trees

2.1 Height of the Random Binary Search Tree

A binary search tree for distinct real numbers $x_1, \ldots, x_n$ is a binary tree in which x_1 is the root, whose left subtree is a binary search tree for

$\{x_2,\ldots,x_n\} \cap (-\infty, x_1)\}$ and whose right subtree is a binary search tree for $\{x_2,\ldots,x_n\} \cap (x_1,\infty)\}$ (thus the structire of the search tree depends heavily on the order in which the real are presented). If the left subtree has k points (nodes), then the rank of the root in the total ordering of the x_i's is $k+1$. We can grow the tree incrementally: if x_{n+1} is to be added (inserted), we start at the root and recursively find the subtree to which x_{n+1} must belong by comparing x_{n+1} to the current root and choosing the left or right subtree as appropriate. Eventually, we locate an empty subtree, which is then formally replaced by a one-node subtree having x_{n+1} as its root. The insertion time is equal to the distance in the tree (path length) between the root (x_1) and the inserted node (x_{n+1}), this distance is referred to as the *depth* of x_{n+1}. The *height* of a binary search tree is the maximal depth of a node, and it measures the worst-case insertion time, an important quantity if we are to maintain a binary search tree when new data arrive.

By a random binary search tree, we mean a binary search tree on a set of random variables $\{x_1,\ldots,x_n\}$ which is obtained by taking a permutation of $\{1,\ldots,n\}$ with each permutation equally probable. It is easy to see that the structure of the tree we obtain will be the same if we pick the x_i independently, all from the same distribution f such that the probability that we choose the same number twice in n trials under f is zero, e.g., if the x_i are uniformily chosen elements of $[0,1]$. The depth D_n of the last node to be inserted satisfies $\mathbf{E}(D_n) \sim 2\log n$ (Lynch, 1965; Knuth, 1973), (further $(D_n - 2\log n)/\sqrt{2\log n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ (Mahmoud and Pittel, 1984; Devroye, 1988)). For the height H_n, the maximal path distance between any node and the root, Robson (1979) showed that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{Pr}(H_n \geq (\gamma+\epsilon)\log n) = 0 \ ,$$

where $\gamma = 4.31107\ldots$ is the unique solution greater than 2 of the equation $c\log(2e/c) = 1$. To actually show that $H_n/\log n \to \gamma$ in probability (we recall that $X_n \to c$ in probability means that for any positive ϵ: $\lim_{n\to\infty} \mathbf{Pr}(|X_n - c| > \epsilon) = 0$.), branching processes were the first successful methodology (Devroye, 1986, 1987). Drmota (1997) was the first to prove this result by generating function analysis. The theorem below will be considerably generalized further on in the chapter.

Theorem 2.1. *(Devroye, 1986, 1987). In a random binary search tree on n nodes, $H_n/\log n \to \gamma = 4.31107\ldots$ in probability.*

Proof. We briefly show here that the height can be studied with the aid of Galton-Watson branching processes. To make the connection, we introduce a new representation of a binary search tree. Call the (random) binary search tree T. Augment the tree T by associating with each node the size of the subtree rooted at that node, and call the augmented tree T'. The root of T' has value n. Since the rank of the root element of T is equally likely to be $1,\ldots,n$, the number N of nodes in the left subtree of the root of T is uniformly distributed on $\{0, 1, \ldots, n-1\}$. A moments thought shows we can choose U by setting $N = \lfloor nU \rfloor$, where U is uniformly distributed on $[0, 1]$. Also, the size of the right subtree of the root of T is $n-1-N$, which is distributed as $\lfloor n(1-U) \rfloor$. All subsequent splits can be represented similarly by introducing independent uniform $[0, 1]$ random variables. This is a typical embedding argument: we have identified a new fictitious collection of random variables $U_1, U_2, \ldots$, and we can derive all the values of nodes in T' from it. This in turn determines (the shape of) T. More precisely, he rule is simply this: in an infinite binary tree, give the root the value n. Also, associate with each node an independent copy of U. If a node has value V, and its assigned copy of U is U' (say), then the value of the two children of the node are $\lfloor VU' \rfloor$ and $\lfloor V(1-U') \rfloor$ respectively. Thus, the value of any node at distance k from the root of T' is distributed as

$$\lfloor \cdots \lfloor \lfloor nU_1 \rfloor U_2 \rfloor \cdots U_k \rfloor \ ,$$

where $U_1, \ldots, U_k$ are i.i.d. uniform $[0, 1]$. We have just described a second way of generating a random tree with exactly the same distribution as a random binary search tree. This second method of generating the trees is much more amenable to analys.

The above representation has a myriad of applications. One of them involves the study of the height. Let H_n be the height of T when $|T| = n$. Then $H_n \geq k$ if and only if one of the 2^k values V_i of nodes at distance k from the root of T' is at least equal to one; which we write as

$$[H_n \geq k] \equiv [\max_{1 \leq i \leq 2^k} V_i \geq 1] \ .$$

This is a beautiful duality indeed. Some care must be exercised when manipulating it though, as the V_i's are very dependent—just consider the values V_i and V_j for nodes that are near one another in the tree. To steer around this, we will derive separate upper and lower bounds for H_n.

In doing so, we need to be able to analyze the distribution of the V_i which boils down to analyzing the distribution of the product of l uniform $[0, 1]$ random varables for various l. To do so, we pass to the logarithm. It turns out the logarithms we are interested in studying are drawn from a very well studied class of distributions, the Gamma distributions. To be precise, the product of k $[0, 1]$ uniforms is e^{-x} where x is a gamma k random variable (an aside: a $[0, 1]$ uniform is a e^{-r} where r is a random *** exponential and a gamma k random variable is simply the sum of k exponentials, see (Grimmett and Stirzaker, 1992); for formalists, a positive variable G is gamma k distributed if it has density $x^{k-1}e^{-x}/\Gamma(a)$ on $(0, \infty)$.

The upper bound. By the dual relationship shown above, we see that

$$\begin{aligned}
\mathbf{Pr}(H_n \geq k) &= \mathbf{Pr}\left(\cup_{i=1}^{2^k}[V_i \geq 1]\right) \\
&\leq 2^k \mathbf{Pr}\left(V_1 \geq 1\right) \\
&\quad \text{(by the union bound (Bonferroni's inequality)} \\
&\quad \text{and symmetry)} \\
&\leq 2^k \mathbf{Pr}\left(n \prod_{i=1}^{k} U_i \geq 1\right) \\
&\quad (U_1, \ldots, U_k \text{ are i.i.d. uniform } [0,1]) \\
&\quad (\text{omit the } \lfloor . \rfloor \text{ in the definition of } V_1) \\
&= 2^k \mathbf{Pr}\left(n e^{-G_k} \geq 1\right) \\
&\quad (G_k \text{ is a gamma } (k) \text{ random variable}) \\
&= 2^k \mathbf{Pr}\left(G_k \leq \log n\right) \ .
\end{aligned}$$

The point here is to find the smallest k such that the upper bound tends to zero. Recall that a G_k random variable has mean k. Thus, if $k = \log n$, the upper bound is $\Theta(2^k)$, which is obviously useless. In fact, k will have to be much larger than $\log n$ in order that the effect of the 2^k term be canceled. Let us try the next best thing: $k \sim c \log n$ for some $c > 1$. The whole enterprise now focuses on the probability in the left tail of the gamma distribution. We provide the details as they explain the choice of c. Let G_k be a gamma (k) random variable. We have

$$1 \leq \frac{\mathbf{Pr}(G_k \leq y)}{\frac{y^k e^{-y}}{k!}} \leq \frac{1}{1 - \frac{y}{k+1}} \ ,$$

where the lower bound is valid for all $y > 0$, and the upper bound is applicable when $0 < y < k+1$. In particular,

$$\mathbf{Pr}(G_k \leq \log n) \leq \frac{(\log n)^k}{n\, k!} \times \frac{1}{1 - \frac{\log n}{k+1}}$$

valid for $\log n < k+1$. Thus, we have, taking $k = \lceil c \log n \rceil$, and using $k! \geq (k/e)^k$ (which follows from Stirling's formula),

$$\begin{aligned}
\mathbf{Pr}(H_n \geq k) &\leq \tfrac{(2 \log n)^k}{n\, k!} \times \tfrac{1+o(1)}{1-\frac{1}{c}} \\
&\leq n^{-1}(2e \log n / k)^k \times \tfrac{1+o(1)}{1-\frac{1}{c}} \\
&\leq \left(\tfrac{1}{e}\left(\tfrac{2e}{c}\right)^c\right)^{\log n} \times \tfrac{1+o(1)}{1-\frac{1}{c}} \\
&\to 0
\end{aligned}$$

if $(1/e)(2e/c)^c < 1$. Let $\gamma = 4.31107\ldots$ be the only solution greater than one of

$$\left(\frac{1}{e}\right)\left(\frac{2e}{c}\right)^c = 1 \ .$$

We conclude that $\lim_{n\to\infty} \mathbf{Pr}(H_n > c \log n) = 0$, for all $c > \gamma$. A more careful use of Stirling's inequality shows that $\lim_{n\to\infty} \mathbf{Pr}(H_n > \gamma \log n) = 0$.

The lower bound. We know now that H_n is very likely less than $\gamma \log n$. Pick $\epsilon > 0$. To show that it is more than $k = \lfloor (\gamma - \epsilon) \log n \rfloor$ with high probability, all we have to do is exhibit a path in the augmented tree with the property that at distance k from the root, the augmented value is at least one. Now, you will say, this is a piece of cake. Why don't we just follow the path dictated by the largest split, that is, when we are at a node with uniform split value U, we go left if $U > 1/2$ and right otherwise? It turns out that if we do so, the augmented value drops below 1 for k near $c \log n$, with $c \approx 3.25$ only. So, this is not a good way to prove the existence of a node far from the root. Instead, we will use branching processes to show that the height is greater than $c \log n$ with probability tending to one, when $c < \gamma$. Thus, we need to track down nodes with large values in the augmented tree. For now, we define $V = nU_1U_2 \ldots U_k$ for a node at distance k from the root, where the U_i's are the uniform $[0,1]$ random variables describing the splits on the path to the root. The purpose is to construct a surviving Galton-Watson process. The root of T becomes the pater familias of the branching process. Consider all descendants in T L levels away, and declare these nodes Galton-Watson children if the product of uniform splitting random variables encountered on the path from the root to the possible child is $\geq d^L$ for a given constant d. The number of Galton-Watson children per node is bounded between 0 and 2^L. Clearly, all nodes in the Galton-Watson process reproduce independently according to identical reproduction distributions. If T were infinite, the corresponding Galton-Watson process would survive with probability $1 - q > 0$ if the expected number of Galton-Watson children per node were greater than one. But this expected number is

$$
\begin{aligned}
2^L \mathbf{Pr}\left(U_1 \ldots U_L > d^L\right) &= 2^L \mathbf{Pr}\left(G_L < L \log(1/d)\right) \\
&\quad (G_L \text{ is a gamma } (L) \text{ random variable}) \\
&\geq \tfrac{(2L\,d \log(1/d))^L}{L!} \\
&\quad (\text{by an inequality for the tail of the} \\
&\quad \text{gamma distribution}) \\
&\sim \tfrac{(2ed \log(1/d))^L}{\sqrt{2\pi L}} \\
&\quad (\text{by Stirling's approximation, as } L \to \infty) \\
&> 1
\end{aligned}
$$

for L large enough, when $2ed \log(1/d) > 1$. We choose $d = e^{-1/c}$, and note

$$2e^{1-1/c}/c > 1 \ .$$

So, with probability $1 - q > 0$, there exists a node at distance kL from the root with value $V \geq nd^{kL} = ne^{-kL/c}$. If we take truncations into account to get the real augmented value of that node, it takes only a minute to verify by induction that it is at least equal to $V - kL$ as we can lose one unit at worst in every truncation. In conclusion,

$$\mathbf{Pr}\left(H_n \geq kL\right) \geq 1 - q$$

if $ne^{-kL/c} - kL \geq 1$. Take for example $kL = c' \log n - \theta L$ for $c' < c$, where $\theta \in [0, 1)$ is possibly dependent upon n. Then the last condition is verified as

$$ne^{-kL/c} - kL \geq n^{1-c'/c} - c' \log n > 1$$

for all n large enough. As c' is arbitrarily close to c, which in turn is arbitrarily close to γ, we have $\liminf_{n\to\infty} \mathbf{Pr}(H_n > (\gamma - \epsilon) \log n) \geq 1 - q$ for all $\epsilon > 0$ and some $q < 1$. But we are not finished yet! Indeed, what if $1 - q = 0.00001$? Clearly, we want the latter probability to be $1 - o(1)$. So, we take t such that tL is integer-valued. The 2^{tL} nodes at distance tL from the root of T are roots of subtrees each of height kL (in T; height k in the Galton-Watson tree): each of the subtrees leads to an independent run of a Galton-Watson process. If tL is large enough, the probability that at least one of these processes survives is close to one. Let $a \in (0, 1/2)$ be another constant, and let A be the event that the $2^{tL} - 1$ uniform $[0, 1]$ random variables associated with the top tL levels of nodes take values in $(a, 1 - a)$. We see that

$$\mathbf{Pr}(A^c) = 2a \times \left(2^{tL} - 1\right) < a2^{tL+1} ,$$

and this is as small as desired by our choice of a. If A is true, then the augmented values V associated with the nodes at distance tL from the root are all at least na^{tL}. Let B be the event that one of the 2^{tL} Galton-Watson processes defined with the aid of the parameters c and L, and rooted at one of the given 2^{tL} nodes survives. From the previous discussion, using independence,

$$\mathbf{Pr}(B^c) = q^{2^{tL}} ,$$

which is as close to zero as desired by choice of t. If A and B happen simultaneously, then there exists a node at distance $tL + kL$ from the root whose augmented value at least equal to

$$na^{tL}e^{-kL/c} - (t + k)L .$$

Take for example $kL = c' \log n - \theta L$ as above. Then the augmented value is at least equal to

$$a^{tL}n^{1-c'/c} - c' \log n - tL .$$

This is greater than one for n large enough. Therefore,

$$\lim_{n\to\infty} \mathbf{Pr}(H_n \geq c' \log n - L + tL) \geq \mathbf{Pr}(A \cap B) \geq 1 - \mathbf{Pr}(A^c) - \mathbf{Pr}(B^c) .$$

The lower bound is as close to one as desired by the choice of a and t. Also, c' is arbitrarily close to γ. Hence, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{Pr}(H_n \geq (\gamma - \epsilon) \log n) = 1 .$$

This concludes the proof of the result that $H_n / \log n \to \gamma$ in probability. $\square$

2.2 Quadtrees

We round off this chapter by showing the universality of the above methodology with the aid of quadtrees. The point quadtree in R^d (Finkel and Bentley, 1974; see Samet (1990) for a survey) generalizes the binary search tree. Each data point is a node in a tree having 2^d subtrees corresponding to the quadrants formed by considering this data point as the new origin. Insertion into point quadtrees is as for binary search trees.

We assume that a random quadtree is constructed on the basis of an i.i.d. sequence with a given distribution in the plane. If this distribution is uniform in the unit square, we call it a uniform random quadtree. In the latter case, the root is easily seen to induce splits into 4 sections of sizes approximately equal to n times the products of two independent uniform $[0, 1]$ random variables.

The height H_n of a random quadtree has a distribution which depends upon the distribution of the data points. For this reason, we look only at uniform random quadtrees. It is easy to show that

$$\mathbf{Pr}(H_n \geq k) \leq 2^{dk}\mathbf{Pr}(n\prod_{i=1}^{dk} U_i \geq 1) ,$$

where the U_i's are i.i.d. uniform $[0, 1]$ random variables. We deduce that $\mathbf{Pr}(H_n > (c/d)\log n) \to 0$ whenever $c > \gamma$. Furthermore,

$$\mathbf{Pr}(H_n \geq k) \geq \mathbf{Pr}(\max_{1\leq i\leq 2^{dk}} nV_i \geq 1 + k) ,$$

where V_i is a product of independent products of two uniform $[0, 1]$ random variables along the i-th path of length k down the quadtree (Devroye, 1977). We deduce that $\mathbf{Pr}(H_n < (c/d)\log n) \to 0$ whenever $c < \gamma$ by mimicking the proof of Theorem 2.1. We conclude that $H_n/\log n \to \gamma/d$ in probability. This result still requires appropriate generalization to non-uniform distributions.

2.3 Bibliographic Remarks

The use of branching processes in the study of binary search trees was advocated in Devroye (1986, 1987). A nice account of this approach can be found in Mahmoud (1992). One can also prove that $\mathbf{E}(H_n^p)/\log^p n \leq \gamma^p + o(1)$ for all $p > 0$ and find a positive number δ such that

$$\lim_{n\to\infty} \mathbf{Pr}(H_n > \gamma\log n - \delta\log\log n) = 0 .$$

By mimicking the proof of the latter fact, show that $F_n/\log n \to 0.3711\ldots$ in probability, where F_n is the fill level, i.e., the maximal depth at which the binary search tree truncated to that depth is complete—thus, level F_n has 2^{F_n} nodes. The constant $0.3711\ldots$ is the only solution < 1 of $(2e/c)^c(1/e) = 1$. See Devroye (1986, 1987).

3. Heuristic Search

3.1 Introduction

In this chapter we present two other beautiful applications of the theory of branching processes. Both involve heuristics for finding the optimal path in a tree with random costs. The tree model studied here was first proposed and analyzed by Karp and Pearl (1983), who decided to look at the simplest possible nontrivial model so as to make the greatest didactical impact.

Consider an infinite complete binary tree in which we associate with every edge e an $0-1$ random variable X_e which is 1 with probability p and 0 with probability $1-p$. The value of a node is the sum of the values of the edges on the path from the root to that node. The object is to find the best node at distance n from the root, that is, the node of minimal value. Interestingly, for $p < 1/2$, we can discover one of the optima in $O(n)$ expected time. This is largely due to the fact that there are many more zeroes than ones in the tree, allowing us to use simple yet fast search algorithms (see section 2). In section 3, we deal with the much more difficult case $p > 1/2$. Rather than trying to reach the optimum, Karp and Pearl propose looking for a near-optimum that would be reachable in $O(n)$ expected time. The heuristic proposed by them employs bounded lookahead and backtrack search.

3.2 Depth First Search

The infinite subtree rooted at a node v is called T_v. All the nodes in this subtree that can be reached via 0-valued edges form a subtree called Z_v. The heuristic we consider here simply performs a series of depth first searches of trees Z_v. We can also think of 1-valued edges as blocked pipes, and 0-valued edges as open pipes. When we pour water in the root, it trickles down and makes all the 0-valued nodes wet. If we reach level n in this manner, we stop. Otherwise, we open one blocked pipe and start all over from there. During the depth first search of a given Z_v, the nodes u with the property that edge (w, u) is 1-valued and $w \in Z_v$ are collected in a set B_v. Since the method consists of always going for the easiest bait, we will call it depth first search. Note that the above procedure first visits all nodes with value 0, then all nodes with value 1, and so forth. This guarantees that an optimum will be returned. The question we have to answer is how long the algorithm runs on the average.

In order to analyze this algorithm, we offer the following crucial result of Karp and Pearl (1983):

Theorem 3.1. *(The family tree traversal theorem). Consider a Galton-Watson branching process with reproduction probabilities $p_0, \ldots, p_M$ (where M is a deterministic bound on the number of children of a node). Consider the (possibly infinite) family tree T thus generated. Let D_n be the number of* *nodes encountered in the depth first search of T, stopped as soon as level n is reached. Then* $\mathbf{E}(D_n) = O(n)$.

Proof. We consider three cases. In **case 1**, we assume that m, the mean number of children per node, is ≤ 1. Let $Z_0, Z_1, \ldots$ denote the generation sizes in T. We bound D_n by the total size of T. We recall that

$$\mathbf{E}(Z_k) = m^k \leq 1 \ .$$

Therefore,

$$\mathbf{E}(D_n) \leq \sum_{k=0}^{n} \mathbf{E}(Z_k) = \sum_{k=0}^{n} m^k \leq n+1 \ .$$

In **case 2**, we assume that $m > 1$, yet T is finite. This corresponds to a process that becomes extinct. We introduce the notation $\mathbf{E}^*$ for the conditional expectation given that T is finite. We also introduce q, the probability of eventual extinction, and $f(s)$, the RGF (reproduction generating function). Once again, we bound

$$D_n \leq \sum_{k=0}^{\infty} Z_k \ .$$

Note first that for $k \geq 0$,

$$\mathbf{Pr}(Z_1 = k | T \text{ finite}) = \frac{\mathbf{Pr}(Z_1 = k)\mathbf{Pr}(T \text{ finite } | Z_1 = k)}{\mathbf{Pr}(T \text{ finite})} = \frac{p_k q^k}{q} = p_k q^{k-1} \ .$$

Note that

$$\mathbf{E}^*(Z_1) = \sum_{k=0}^{\infty} k p_k q^{k-1} = f'(q) \ .$$

Thus, the derivative of f at q tells us the expected number of children of the root of an extinct tree: note that this is less than one. But this formula should be universally valid for all generation sizes. Therefore,

$$\begin{aligned}
\mathbf{E}^*(Z_k) &= \left(\overbrace{f(f(\cdots(q))\cdots)}^{k \text{ times}} \right)' \\
&= f'\left(\overbrace{f(f(\cdots(q))\cdots)}^{k-1 \text{ times}} \right) \times f'\left(\overbrace{f(f(\cdots(q))\cdots)}^{k-2 \text{ times}} \right) \times \cdots \times f'(q) \\
&= (f'(q))^k \ .
\end{aligned}$$

Thus,

$$\mathbf{E}^*(D_n) \leq \sum_{k=0}^{\infty} (f'(q))^k = \frac{1}{1 - f'(q)} \ .$$

This concludes the proof of case 2. (Note that for supercritical Galton-Watson processes, the branching process given T finite is an unconditional branching process with RGF $f(sq)/q$.) Finally, in **case 3**, we assume that $m > 1$ and that T is infinite. Nodes in the search are designated as mortal or immortal according to whether their subtrees are finite or not. Note that the search at a given node at worst visits all the nodes in the subtrees with mortal nodes as roots. The expected size of each such subtree is not more than $1/(1 - f'(q))$ by case 2. When the search visits the first immortal child, it will never return to visit another child, as an infinite tree is bound to have at least one node at level n. As each node has not more than M mortal children, we have the following recurrence:

$$\mathbf{E}(D_n | T \text{ infinite}) \leq 1 + \mathbf{E}(D_{n-1} | T \text{ infinite}) + \frac{M}{1 - f'(q)} .$$

This recurrence leads trivially to

$$\mathbf{E}(D_n | T \text{ infinite}) \leq n + (n-1) \frac{M}{1 - f'(q)} .$$

Cases 2 and 3 may be combined easily, as

$$\begin{aligned} \mathbf{E}(D_n) &= \mathbf{Pr}(T \text{ finite})\ \mathbf{E}(D_n | T \text{ finite}) \\ &\quad + \mathbf{Pr}(T \text{ infinite})\ \mathbf{E}(D_n | T \text{ infinite}) \\ &\quad \leq \max\{\mathbf{E}(D_n | T \text{ finite}),\ \mathbf{E}(D_n | T \text{ infinite})\} \end{aligned}$$

This concludes the proof of the family tree traversal theorem. □

Next, we claim that the expected running time of iterated depth first search is $O(n)$ when $p < 1/2$. A depth first search trial is one iteration of this process: at a node, all the nodes in its subtree reachable via 0-valued edges are visited. We call this collection of nodes the expansion tree of the node. A node with an infinite expansion tree is called immortal. The other ones are mortal. Consider the branching process defined by zero edges only. The reproduction distribution has $p_2 = (1-p)^2$ (two zero edges), $p_1 = 2p(1-p)$, and $p_0 = p^2$. The expected number of children per node is

$$m = 2(1-p)^2 + 2p(1-p) = 2(1-p) > 1 .$$

Thus, the extinction probability for this branching process is $q < 1$. q is also the probability that a given node is mortal.

The running time is conveniently decomposed as follows: any trial started at any node takes expected time bounded by cn (Theorem 3.1). Thus, the total expected time before halting is not more than the expected number of trials times cn. The total number of trials in turn is not more than the total number of trials started at mortal nodes plus one. Therefore,

$$\mathbf{E}(\text{total time}) \le \frac{cn}{1-q} \ ,$$

since the probability of having an immortal node is $1-q$, and a search started at an immortal node surely reaches level n. This concludes the proof of the linear expected time claim.

Remark 3.1. The case $p = 1/2$. When $p = 1/2$, the given iterated depth-first-search procedure takes quadratic expected time.

We conclude this section with another analysis: what is the value C_n of the minimal node at distance n from the root? Clearly, C_n is a random variable sandwiched between 0 and n. When n grows, C_n increases as well (on a given tree). As all monotone sequences have a (possibly infinite) limit, we may call our limit C. Interestingly, when $p < 1/2$, C is finite with probability one! This means that we can find an infinite path in almost every tree with only a finite number of nonzero edges. We have the following:

A. For every k, $\mathbf{Pr}(C_n > k) \le \mathbf{Pr}(C > k)$. (Obvious, since $C_n \uparrow C$.)
B. $\lim_{n\to\infty} \mathbf{Pr}(C_n > k) = \mathbf{Pr}(C > k)$. (Thus, C really matters, as it describes the situation for all n large enough.)
C. For $p < 1/2$,

$$\mathbf{Pr}(C > k) \le (2p)^{2^{k+1}} \ , \ k = 0, 1, 2, \ldots \ .$$

Proof. Consider a branching process in which we keep only the 0-valued edges in the complete binary tree. As the number of children per node is binomially distributed with parameters 2 and $1-p$, the expected number of children is $2(1-p) > 1$. Let q be the extinction probability. Then

$$\mathbf{Pr}(C > k) < q^{2^k}$$

since $[C > k]$ implies that each of the 2^k subtrees rooted at the nodes at depth k must fail to have an infinite path of zero-cost branches (that is, each of the 2^k branching processes spawned at these nodes must become extinct). Since the RGF of this branching process is $f(s) = (p + (1-p)s)^2$, it is easy to see that $q < (2p)^2$. To prove this, we need only show that $f((2p)^2) < (2p)^2$, or that

$$p + (1-p)(2p)^2 < 2p \ ,$$

or that $4p(1-p) < 1$. But the last inequality is obviously true. □

3.3 Bounded Lookahead and Backtrack

In the case of a majority of 1-valued edges ($p > 1/2$), depth first search yields exponential expected time. In fact, it seems impossible to concoct any kind of polynomial expected time algorithm for locating the optimal value. We can do the next best thing, that is, we can try to find an almost optimal solution. To set the stage, we first define C_n, the optimum value of a solution found by an algorithm, and C_n^*, the value of the true optimum in the random tree. Clearly, $C_n^* \leq C_n$. For a given algorithm, two issues have to be dealt with:

A. What is the expected time $\mathbf{E}(T)$ taken by the algorithm?
B. How close is C_n to C_n^* (in some probabilistic sense)?

The bounded-lookahead-and-backtrack (or: BLAB) algorithm proposed by Karp and Pearl (1983) introduces three design parameters, d, α and L, where $d \geq 1$ is an integer, $\alpha \in (0,1)$ is a real number, and $L > 1$ is an integer. If v is a node in our tree and u is a descendant of v such that the path distance from v to u is L, then we say that u is an (α, L) son of v if the sum of the edge values on the linking path is $\leq \alpha L$. To make things more readable, we will simply say that u is a good child of v.

We now construct a fake branching process as follows: start with a given node and make it the root of the branching process. Declare all the good sons to be its offspring. So, this process jumps L levels at a time. (This is illustrated in the first figure of this section.) Repeat this definition for all the nodes thus obtained. The Malthusian parameter for this process is the expected number of good sons per node, or

$$m \stackrel{\text{def}}{=} 2^L \mathbf{Pr}(BIN(L,p) \leq \alpha L) .$$

The fake branching process is supposed to help us locate near-optimal nodes at level n. If it is to work for us, we surely would like the process to survive forever, thus leading to the condition $m > 1$. From the properties of the binomial distribution, we retain that if $\alpha < p$ is fixed, then, as $L \to \infty$,

$$m = 2^L \frac{\Theta(1)}{\sqrt{L}} \{R(\alpha,p)\}^L = 2^L \frac{\Theta(1)}{\sqrt{L}} \left(\left(\frac{p}{\alpha}\right)^\alpha \left(\frac{1-p}{1-\alpha}\right)^{1-\alpha} \right)^L ,$$

where the function $R(\alpha,p)$ increases monotonically from $1-p$ at $\alpha = 0$ to 1 at $\alpha = p$. Thus, it takes the value $1/2$ somewhere in the interval $(0,p)$, at a place we will call α^*. We have the freedom to choose α and L. So, we first pick $\alpha \in (\alpha^*, 1)$. Then we choose L so large that $m > 1$. This fixes the branching process. We let the probability of extinction be q. The BLAB algorithm proceeds as follows: we select d in some way (to be specified later), such that $n-d$ is a multiple of L. Repeat for each of the 2^d nodes at level d until successful the following process: traverse the "good sons" branching process in a depth-first-search manner until a node is found at level n or until the subtree is exhausted without ever reaching level n. If a node at level n is reached, then its value is guaranteed to be no more than $d+\alpha(n-d)$. But the probability of a given depth-first-search succeeding is at least $1-q$. Thus, the overall procedure returns a failure with probability less than q^{2^d}. In that case, if a node has to be returned, we might as well return the leftmost node in the tree, with value $\le n$. Putting this together, we see that

$$\begin{aligned} \mathbf{E}(C_n) &\le n\mathbf{Pr}(\text{search fails}) + d + \alpha(n-d) \\ &\le nq^{2^d} + d + \alpha(n-d)\ . \end{aligned}$$

For fixed $\epsilon > 0$, this is less than $\alpha^*(1+\epsilon)n$ by choice of α (e.g., $\alpha \le \alpha^*(1+\epsilon/2)$ will do), L (as above) and d (large, but fixed). We also see that

$$\lim_{n\to\infty} \mathbf{Pr}(C_n > \alpha^*(1+\epsilon)n) = 0$$

for all $\epsilon > 0$ if we choose α and L as above and $d \to \infty$ while $d/n \to 0$ (example: $d \sim \log n$).

The second thing we need to prove is that $\mathbf{E}(C_n^*) \ge \alpha^* n$ or something close to that. Note the following:

$$\begin{aligned} \mathbf{Pr}(C_n^* < \alpha^* n) &\le \mathbf{Pr}(\exists \text{ at least one } (\alpha^*, n) \text{ good son of the root}) \\ &\le 2^n \mathbf{Pr}(BIN(n,p) \le \alpha^* n) \\ &= 2^n \frac{\Theta(1)}{\sqrt{n}} \left\{R(\alpha^*, p)\right\}^n \\ &= \frac{\Theta(1)}{\sqrt{n}}\ . \end{aligned}$$

Thus, $\mathbf{Pr}(C_n^* \ge \alpha^* n) \to 1$. Also,

$$\begin{aligned} \mathbf{E}(C_n^*) &\ge \mathbf{E}(C_n^*) I_{C_n^* \ge \alpha^* n} \\ &\ge \alpha^* n \mathbf{Pr}(C_n^* \ge \alpha^* n) \\ &\ge \alpha^* n\,(1 - \Theta(1)/\sqrt{n}) \\ &\ge \alpha^* n - \Theta(\sqrt{n})\ . \end{aligned}$$

For given $\epsilon > 0$, we can design an algorithm that guarantees the following:

$$\limsup_{n\to\infty} \frac{\mathbf{E}(C_n)}{\mathbf{E}(C_n^*)} < 1+\epsilon\ .$$

Or, if one wants it,

$$\lim_{n\to\infty} \mathbf{Pr}\left(\frac{C_n}{C_n^*} > 1+\epsilon\right) = 0\ .$$

(The last event implies either $C_n > \alpha^*(1+\epsilon)n$ or $C_n^* < \alpha^* n$, and the probabilities of both of these events tend to zero with n.)

We conclude this section with a proof of the linear expected time complexity: $\mathbf{E}(T) = O(n)$. When finding a good son of a node in the branching process, an effort not exceeding 2^L is spent. Then, by the family tree traversal lemma, each depth-first-search takes time not exceeding cn, where c is a constant depending upon the branching process parameters. The expected number of depth-first-searches until a node is encountered that is the root of a surviving branching process is not more than $1/(1-q)$. Thus, the total expected time does not exceed

$$\frac{cn}{1-q} = O(n) \ .$$

REMARK. McDiarmid and Provan (1991) pointed out that bounded lookahead without backtrack is also feasible. Assume that we find the optimal path from the root to a node at depth L. Make this node the new starting point and repeat. L is a large integer constant. For $p > 1/2$, and $\epsilon > 0$, one can show that there exists an L such that this algorithm runs in linear expected time, and that the best value found by the algorithm (C_n) satisfies the inequality

$$C_n \leq (1+\epsilon) C_n^*$$

with probability tending to one.

3.4 Bibliographic Remarks

The problem dealt with here was proposed and analyzed by Karp and Pearl (1983). An alternate short proof of Theorem 3.1 is given by McDiarmid (1990), where additional information about the problem may be found as well. The analysis of the optimal value C_n in the case $p < 1/2$ is due to McDiarmid and Provan (1991). Consider now depth first search in a complete b-ary tree in which the probability of a "one" edge is p, and $b(1-p) > 1$. The following inequality is due to McDiarmid and Provan (1991): if C_n is the optimal value of a node at distance n from the root, then

$$\mathbf{Pr}(C_n > k) < \left(\frac{bp}{b-1}\right)^{b^{k+1}} \ , \ k \geq 0 \ .$$

Karp and Zhang (1995) analyze random AND/OR trees, where internal nodes at even (odd) distances from the root are AND (OR) nodes and each node has a boolean value 0 or 1. The value of a node is the outcome of the logical operation of the node on its children's values. The evaluation problem is to determine the root's value by examining the leaf values (which are randomly and independently assigned), while keeping computation to a minimum. This is Pearl's minimax tree model (1984). Karp and Zhang propose and analyze various algorithms using tail bounds on generation sizes in Galton-Watson processes. For minimax trees, Devroye and Kamoun (1996) analyze the value of the root in a random minimax tree, in which the leaf values in the n-th generation are those of a branching random walk, and intermediate level values are obtained by alternating the operations minimum and maximum.

4. Branching Random Walk

4.1 Definition

In a branching random walk, we superimpose a random walk on each path from the root down in a Galton-Watson tree. More specifically, we associate with each individual u in a Galton-Watson tree a value V_u, the value of the root being zero. If u has N offspring (where N follows the model of the Galton-Watson process), then the values of the offspring relative to the value V_u of the parent u jointly have a given distribution. In the simplest model, for every child v of u, we have $V_v = V_u + X_v$, and all displacements X_v are independent (this will be called the independent branching random walk). However, in general, if the children have displacements $X_{v_1}, \ldots, X_{v_N}$, then the joint distribution of $(N, X_{v_1}, \ldots, X_{v_N})$ is quite arbitrary. What is important is that each parent produces children (and their values) in the same manner.

The analysis of branching random walks is greatly facilitated by the following function:

$$m(\theta) = \mathbf{E}\left(\sum_{i=1}^{N} e^{-\theta X_{v_i}}\right)$$

where $v_1, \ldots, v_N$ are the children of the root. We assume throughout that $m(\theta) < \infty$ for some θ. This function may be considered as the Laplace-Stieltjes transform of $F(t) = \mathbf{E}(Z_1(t))$, the expected number of individual in the first generation, with value less than or equal to t. In general, we introduce the notation $Z_n(t)$, the number of individuals in the n-th generation, with value $\leq t$. Note that $Z_n = Z_n(\infty)$, so that this definition generalizes that of the previous section. Let Z^n be the point process with atoms V_u for all u in the n-th generation. Then, following Kingman (1975), introduce

$$W_n(\theta) = \frac{1}{m(\theta)^n} \sum_{u \text{ in generation } n} e^{-\theta V_u} .$$

This is a martingale for $\mathcal{F}_n$, the σ-field generated by all events in the first n generations. There is an almost sure limit, $W(\theta)$ (as $W_n(\theta) \geq 0$), and by Fatou's lemma, $\mathbf{E}(W(\theta)) \leq 1$. The study of W_n and W reveals that there may be several modes of behavior, and this was studied by Biggins (1977) in more detail. In this chapter, we do not wish any distractions due to extinction of the underlying Galton-Watson process, and assume therefore that N, the number of children per parent, is a fixed positive integer: $N = b$. For more general theorems, we refer to the cited papers.

In the next section, for $N = b$, we survey the main results on the first birth in the n-th generation, or $B_n = \min\{V_u : u \text{ in } n\text{-th generation}\}$, and on $Z_n(t)$, the distribution of values in the n-th generation. A straightforward application in the study of the height of trees then concludes this chapter.

4.2 Main Properties

Let X be a random variable equal to the value V_u of a randomly picked child of the root. Since $N = b$, the earlier definition of $m(\theta)$ specializes to

$$m(\theta) \stackrel{\text{def}}{=} b\mathbf{E}\left(e^{-\theta X}\right) .$$

Then, if $X \geq 0$ is nondegenerate, we define the μ-function by

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

Theorem 4.1. *(Biggins, 1977). If $\mu(a) < 1$, then with probability one, $Z_n(na) = 0$ for all but finitely many n. If $a \in \operatorname{int}\{a : \mu(a) > 1\}$, then*

$$\lim_{n\to\infty} \left(Z_n(na)\right)^{1/n} = \mu(a)$$

almost surely.

This theorem shows that $\mu(a)^n$ is about equal to the number of individuals in the n-th generation with value $\leq na$. Its simple proof is not given here, but it follows the lines of the proof of Theorem 2.1. In fact, Theorem 4.1 is nothing but a refined large deviation theorem, as along any path from the root, the values form a standard random walk.

As a corollary of the above result, we have:

Theorem 4.2. *(Kingman, 1975; Hammersley, 1974; Biggins, 1977). Assume $m(\theta) < \infty$ for some $\theta > 0$. Let $B_n = \min\{V_u : u$ is in the n-th generation $\}$. Then,*

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

almost surely, and γ is finite.

Interestingly, B_n gows linearly with n, while the n-th generation size (b^n) grows exponentially with n. As the μ-function has an impact on both results, it is useful to have its properties at hand.

Lemma 4.3. *Let $X \geq 0$ be a nondegenerate random variable. Then its μ-function satisfies the following properties:*

(i) μ *is an increasing function on* $[0, \infty)$.
(ii) μ *is continuous on int*$\{a : \mu(a) > 0\}$.
(iii) $\log \mu$ *is concave on int*$\{a : \mu(a) > 0\}$.
(iv) $\sup_{a \in R} \mu(a) \leq b$.
(v) If $\mathbf{E}(X) < \infty$, *then* $\mu(a) \equiv b$ *for* $a \geq \mathbf{E}(X)$.
(vi) $\lim_{a \uparrow \infty} \mu(a) = b$.
(vii) If $X \geq c > 0$, *then* $\mu(a) = 0$ *for* $a < c$.

(viii) Let $s = \sup\{t : \mathbf{Pr}(X < t) = 0\}$*, and define* $p = \mathbf{Pr}(X = s)$*. Then* μ *is continuous on* (s, ∞)*,* $\mu(s) = bp$*, and* $\mu(a) = 0$ *for* $a < s$*.*
(ix) If $bp < 1$*, and* $\gamma = \inf\{a : \mu(a) \geq 1\}$*, then* $\mu(\gamma) = 1$*.*

If all displacements with respect to a parent are identical, then we speak of a Bellman-Harris branching random walk. McDiarmid (1995) calls this a common branching random walk. Of course, all theorems above also apply to this situation. It is of interest to pin down the asymptotic behavior of B_n beyond Theorem 4.2. Consider for example an infinite b-ary tree on which we superimpose a branching random walk, with all displacements Bernoulli $(1/b)$, that is, they are 1 with probability $1/b$ and 0 otherwise. The case $b = 2$ is easiest to picture, as all displacements are independent equiprobable bits. Joffe, LeCam and Neveu (1973) showed that $B_n/n \to 0$ almost surely, and this also follows from Theorem 4.2, which was published later. Bramson (1978) went much further and showed that there exists a random variable W such that

$$\lim_{n\to\infty} B_n - \frac{\lceil \log\log n - \log(W + o(1)) \rceil}{\log 2} = 0$$

almost surely, where the $o(1)$ term is stochastic. In the binary case, each individual in the n-th generation has a binomial $(n, 1/2)$ distribution. If these 2^n binomials had been independent, we would have had $\liminf_{n\to\infty} B_n = 0$ almost surely and $\limsup_{n\to\infty} B_n = 1$ almost surely. This follows from the fact that $\mathbf{Pr}(B_n = 0) \to 1 - 1/e$ as $n \to \infty$ and $\mathbf{Pr}(B_n \geq 2) \leq e^{-(n+1)}$. Thus, Bramson's result exposes a crucial property of branching random walks. Dekking and Host (1990) consider the general branching random walk with nonnegative integer-valued displacements. Thus, $B_n \uparrow$. Let $N(k)$ be the number of children of the root with displacement k. Let $N = \sum_{j=0}^{\infty} N(j)$ be the number of offspring of the root. Again, we assume $N = b$ with probability one, although the results of Dekking and Host treat the general case. Some of their results can be summarized as follows:

Theorem 4.4. *(Dekking and Host, 1990). If* γ *denotes the constant of Theorem 4.2, then* $\gamma = 0$ *if and only if* $\mathbf{E}(N(0)) \geq 1$*.*

Assume now $\mathbf{Pr}(N(0) = 1) < 1$. Then $\mathbf{Pr}(B_n \to \infty) \in \{0, 1\}$, and the zero case happens if and only if $\mathbf{E}(N(0)) > 1$. Also,

A. If $\mathbf{E}(N(0)) > 1$, then there exists a proper random variable W such that $B_n \to W$ almost surely.
B. If $\mathbf{E}(N(0)) = 1$, $\mathbf{E}(N^2) < \infty$, and $g = \inf\{i > 0 : \mathbf{E}(N(i)) > 0\}$, then $B_n \log 2/\log\log n \to g$ almost surely.

If $\mu = \mathbf{E}(N(1)) > 0$ and $\tau = (1/2)\mathbf{Var}(N(0))$, then for integer $k \geq 0$,

$$\mathbf{Pr}(B_n \leq k) \sim \frac{\mu}{\tau(\mu n)^{\frac{1}{2^k}}} \ as\ n \to \infty\ .$$

McDiarmid (1995) extends the results of Dekking and Host in some cases. Consider only nonnegative displacements, and recall that the branch factor is b. Then, if b_n is the median of B_n, McDiarmid establishes the existence of positive constants c, c' such that for all n,

$$\mathbf{Pr}(|B_n - b_n| > x) < ce^{-c'x}$$

for all $x \in [0, n]$. This implies that almost surely, for all n large enough, $B_n - b_n = O(\log n)$. Clearly, by Theorem 4.2, b_n should be near γn. The following result describes the closeness of B_n to γn. We give only the version for the case that the underlying Galton-Watson tree is the complete infinite b-ary tree.

Theorem 4.5. *(McDiarmid, 1995). Consider a common branching random walk in which every individual has b children, and all displacements are on $[a, \infty)$, where a is the leftmost point of the support of the displacement random variable X, and $b\mathbf{Pr}(X = a) < 1$. Let $\tau > 0$ be the (necessarily unique) solution of $e^{\tau\gamma} m(\tau) = 1$, and let m be finite in a neighborhood of τ. Then there are positive constants c, c', c'' such that*

$$\mathbf{Pr}(B_n \le \gamma n + c \log n - x) \le e^{-c'x} \ , \ x \ge 0,$$

and

$$\mathbf{Pr}(B_n \ge \gamma n + c'' \log n + x) \le e^{-c'x} \ , \ 0 \le x \le n.$$

McDiarmid's proof does not imply $c = c''$, but it strengthens earlier results, such as a result by Biggins (1977), who showed that under the stated conditions, $B_n - \gamma n \to \infty$ almost surely. Interestingly, his argument is based on the second moment method, and the idea of leading sequences. A sequence $(x_1, \ldots, x_n)$ is leading if for all $j = 1, \ldots, n-1$,

$$\sum_{i=1}^{j} x_i \ge \frac{j}{n} \sum_{i=1}^{n} x_i \ .$$

If $(X_1, \ldots, X_n)$ are exchangeable random variables, then indeed, $\mathbf{Pr}((X_1, \ldots, X_n) \text{ is leading }) \ge 1/n$. Given an individual u in the n-th generation, we denote by $Y_1, \ldots, Y_n$ be the displacements encountered on the path from the root to v. We call v leading if this displacement sequence is leading, that is, if $W_j \ge (j/n) W_n$, where $W_1, \ldots, W_n$ are the values of the ancestors of v in generations 1 through n. Clearly, $Z_n(t) \ge Z_n^*(t)$, where $Z_n^*(t)$ is the number of leading individuals in the n-th generation with value $\le t$. It should be clear that $Z_n^*(t)$ is about $Z_n(t)/n$ when $Z_n(t)$ is large, and not much is lost by considering $Z_n^*(t)$, or by considering the minimum value B_n^* among leading individuals, instead of just B_n. A careful application of the second moment method ($\mathbf{Pr}(X > 0) \ge (\mathbf{E}(X))^2 / \mathbf{E}(X^2)$ for any random variable X with finite mean $\mathbf{E}(X) \ge 0$) then yields Theorem 4.5.

4.3 Application to Analysis of Height of Trees

One may use Theorem 4.2 in the study of the height of a large class of random trees. These trees can be modeled indirectly by the size tree, a tree in which we associate with each node u the size of its subtree S_u. For the root, we have $S_u = n$, and for each leaf, $S_u = 1$. Often, these size trees are close to a split tree T in a manner to be made precise. A split tree T starts with a root u of value $V_u = 1$. It is an infinite b-ary tree, and the values of the children $v_1, \ldots, v_b$ are $V_u X_{v_1}, \ldots, V_u X_{v_b}$. Furthermore, $\sum_{i=1}^{b} X_{v_i} = 1$ and $X_{v_i} \geq 0$ for all i. In other words, considering the value as mass of a subtree, the mass 1 at the root is partitioned into smaller masses that again add up to one. This process continues forever, each particle splitting in the same manner. The distribution of values in the split tree is governed by the joint distribution of the b child values of the root. If we consider $V_u' = -\log V_u$, then the above model describes a branching random walk. Let $m()$ and $\mu()$ be defined as for that random walk, that is, if X is the value of a randomly picked child of the root (so, $0 \leq X \leq 1$), then

$$m(\theta) = b\mathbf{E}\left(e^{-\theta(-\log X)}\right) = b\mathbf{E}\left(X^{\theta}\right) .$$

Define

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

Finally, let $N_n(t)$ be the number of n-th generation individuals with value exceeding t in the split tree. The following is a corollary of Theorem 4.1:

Theorem 4.6. *If $\mu(a) < 1$, then with probability one, $N_n\left(e^{-na}\right) = 0$ for all but finitely many n. If $a \in \operatorname{int}\{a : \mu(a) > 1\}$, then $\lim_{n\to\infty}\left(N_n\left(e^{-na}\right)\right)^{1/n} = \mu(a)$ almost surely. Furthermore, if B_n is the maximal value of any individual in the n-th generation of the split tree, then*

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

almost surely.

The above results may be applied in the study of Kolmogorov's rock (see Athreya and Ney, 1972), which is subjected to many rounds of breaking, and each break results in two rocks with uniform size. If the initial rock has mass one, then Theorem 4.6 describes the maximal rock size among 2^n shattered rocks in the n-th generation. The random variables that govern the splitting are $(U, 1-U)$, where U is uniformly distributed on $[0,1]$. In this case, we have

$$m(\theta) = 2\mathbf{E}(U^{\theta}) = \frac{2}{\theta + 1} .$$

Also,

$$\mu(a) = \inf_{\theta \geq 0} \left\{ \frac{2e^{\theta a}}{\theta + 1} \right\} = 2ae^{1-a} .$$

From this, we determine γ as the solution of $2ae^{1-a} = 1$, and obtain $\gamma = 0.2319\ldots$. As a consequence, the size B_n of the largest rock is almost surely $e^{-n(\gamma+o(1))}$. For comparison, if we were to break the rocks evenly, then $B_n = 2^{-n} = e^{-n\times 0.6931\ldots}$, almost the third power of the maximal rock in the random model!

However, the way Tree splits are used is different. A search tree holding n nodes has mass n at the root, so we define our split tree in such a way that each node has n times the value of the corresponding node in the original split tree. These (typically non-integer) roughly represent the sizes of the subtrees. Nodes with value (after multiplication with n) less than 1 correspond to nothing and will be cut. In this manner, the size tree is finite. For example, in a random binary search tree, the sizes of the left and right subtrees of the root are distributed as $\lfloor nU \rfloor$ and $\lfloor n(1-U) \rfloor$ respectively, where U is uniform $[0,1]$. These sizes are jointly smaller than $(nU, n(1-U))$, and thus, by embedding, we can say that the values in the size tree are jointly (over the infinite tree!) smaller than the values in a split tree with multiplicative factor n and with root child values $(U, 1-U)$. Furthermore, the sizes of the left and right subtrees are jointly larger than $(nU-1, n(1-U)-1)$. If we repeat this sort of bounding for k generations, then it is easy to see that all values in the size tree at generation k are jointly larger than the values in the split tree just defined, minus k. The connection between size trees and split trees is thus established. In particular, what interests us most is that if H_n is the height of the binary search tree with n nodes, then

$$\begin{aligned}\mathbf{Pr}(H_n \geq k) &= \mathbf{Pr}(\text{maximum value in generation } k \text{ of size tree } \geq 1) \\ &\leq \mathbf{Pr}(nB_k \ \geq 1)\end{aligned}$$

where B_k is the maximum value of a k-th generation node in the original split tree (n is the multiplicative factor). Similarly,

$$\begin{aligned}\mathbf{Pr}(H_n < k) &= \mathbf{Pr}(\text{maximum value in generation } k \text{ of size tree } < 1) \\ &\leq \mathbf{Pr}(nB_k - k \ < 1).\end{aligned}$$

As $B_k = e^{-k(\gamma+o(1))}$ almost surely as $k \to \infty$, where γ is precisely as in the example of Kolmogorov's rock, it is easy to conclude from these inequalities the following (essentially Theorem 2.1): for $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{Pr}\left(\frac{H_n}{\log n} > \frac{1}{\gamma} + \epsilon\right) = 0$$

and

$$\lim_{n\to\infty} \mathbf{Pr}\left(\frac{H_n}{\log n} < \frac{1}{\gamma} - \epsilon\right) = 0 .$$

Thus, $H_n/\log n \to 1/\gamma = 4.31107\ldots$ in probability, where γ is defined in Theorem 4.6. For the random binary search tree, we thus have a second proof of Theorem 2.1.

The technique above consists in describing the sizes of the subtrees of a random tree by an embedding argument, and to relate these sizes to those of a split tree by suitable inequalities. This has been done in the literature for a number of random trees, and rather than dwelling on the details, we will review the known results. The remainder of this section is rather specialized and may be skipped upon first reading.

EXAMPLE 1: THE RANDOM b-ARY SEARCH TREE. Let n i.i.d. random variables with a common density be used to construct a random b-ary search tree, where each physical node holds up to $b-1$ elements. As soon as a node is full, new nodes reaching it on the path down from the root are sent down to one of the b child trees by a comparison of values of the $b-1$ (sorted) elements in the node. Here the tree size is measured in number of elements, not number of nodes. The first $b-1$ elements occupy the root. Without loss of generality, they are i.i.d. uniform $[0,1]$. Thus, as the other elements are independent, we see that the subtree sizes $(N_1, \ldots, N_b)$ are distributed as a multinomial random vector with count $n-b+1$ and probabilities given by $S_1, \ldots, S_b$, the spacings determined on $[0,1]$ by a uniform sample of size $b-1$. Now, the relationship between the size tree and the split tree is only slightly more intricate, but the split tree clearly should have multiplicative factor n and split random vectors $(S_1, \ldots, S_b)$ (see Devroye, 1990, for the details). In particular, the S_i's are beta $(1, b-1)$ distributed (Pyke, 1965), and we can thus easily compute

$$m(\theta) = b\mathbf{E}(X^\theta) = b\mathbf{E}(S_1^\theta) = b\int_0^1 x^\theta (b-1)(1-x)^{b-2} dx = \frac{\Gamma(b+1)\Gamma(\theta+1)}{\Gamma(b+\theta)} .$$

Unfortunately, the expression for μ is in general not simple. We have $H_n/n \to \xi$ in probability, where

$$\xi = \inf\left\{ c > 1/\sum_{j=2}^{b}(1/j) : t + c\log b! - c\sum_{j=1}^{b-1}\log(t+i) < 0 \right\}$$

and $t > 0$ is the unique solution of

$$\frac{1}{c} = \sum_{i=1}^{b-1} \frac{1}{t+i}$$

(Devroye, 1990). Particular values of ξ include $\xi = 4.31107\ldots$ $(b=2)$, $\xi = 2.4699\ldots$ $(b=3)$, $\xi = 0.9979\ldots$ $(b=9)$ and $\xi = 0.3615\ldots$ $(b=100)$. The depth of the last node, D_n, is in probability asymptotic to $\log n/\sum_{j=2}^{b}(1/j)$ (Mahmoud and Pittel, 1984). Devroye (1997) showed that if $\lambda = 1/\sum_{i=2}^{b} 1/i$ and $\sigma^2 = \sum_{i=2}^{b} 1/i^2$, then

$$\frac{D_n - \lambda\log n}{\sqrt{\sigma^2\lambda^3\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) ,$$

where $\mathcal{N}$ denotes a normal random variable. As an example, if $b = 3$,

$$\frac{D_n - (6/5)\log n}{\sqrt{(78/125)\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

Example 2: the random quadtree. The point quadtree in R^d (Finkel and Bentley, 1974; see Samet (1990b) for a survey) generalizes the binary search tree. Defined in the previous chapter, we only consider uniform data in $[0,1]^d$. Note that if the root is $X = (X_1, \ldots, X_d)$, then the probabilities (volumes) of the 2^d quadrants are given by the identically distributed (but dependent) random variables

$$\prod_{i=1}^{d} X_i^{b_i}(1 - X_i)^{1-b_i} \ ,$$

where $b_1, \ldots, b_d$ is a vector of d bits identifying one of the 2^d quadrants. Devroye (1987) establishes probability inequalities between the values in the size tree and the values in the split tree, which imply for first order results that it suffices to study the split tree. Then we note that

$$m(\theta) = 2^d \mathbf{E}\left(\prod_{i=1}^{d} X_i^{\theta}\right) = 2^d \prod_{i=1}^{d} \mathbf{E}\left(X_i^{\theta}\right) = \left(\frac{2}{\theta + 1}\right)^d \ ,$$

thus generalizing the binary search tree (obtained when $d = 1$). Thus,

$$\mu(a) = \inf_{\theta \geq 0} \left\{\frac{2e^{\frac{\theta a}{d}}}{\theta + 1}\right\}^d = \left(\frac{2a}{d} e^{1-a/d}\right)^d \ .$$

Therefore, by simple inspection, $\mu(d\gamma) = 1$, where γ is the parameter for the binary search tree . As a result, the height H_n of a random quadtree is in probability asymptotic to $(1/d\gamma)\log n$, where $1/\gamma = 4.31107\ldots$ is the constant in the height of the random binary search tree (Devroye, 1987). Let D_n be the depth of the last node. It is also known that

$$\frac{D_n}{\log n} \to \frac{2}{d} \quad \text{in probability} \ ,$$

a result first noted by Devroye and Laforest, 1990. See also Flajolet, Gonnet, Puech and Robson (1991). Furthermore,

$$\frac{D_n - (2/d)\log n}{\sqrt{(2/d^2)\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ ,$$

valid for any $d \geq 1$. This result was obtained via complex analysis by Flajolet and Lafforgue (1994) and by standard central limit theorems by Devroye (1997). Example 3: the random median-of-(2k+1) binary search tree. Bell (1965) and Walker and Wood (1976) introduced the following method for constructing a binary search tree. Take $2k+1$ points at random from the set of n points on which a total order is defined, where k is integer. The median of these points serves as the root of a binary tree. The remaining points are thrown back into the collection of points and are sent to the subtrees. Following Poblete and Munro (1985), we may look at this tree by considering internal nodes and external nodes, where internal nodes hold one data point and external nodes are bags of capacity $2k$. Insertion proceeds as usual. As soon as an external node overflows (i.e., when it would grow to size $2k+1$), its bag is split about the median, leaving two new external nodes (bags) of size k each, and an internal node holding the median. After the insertion process is completed, we may wish to expand the bags into balanced trees. Using the branching process method of proof (Devroye, 1986b, 1987, 1990; see also Mahmoud, 1992) the almost sure limit of $H_n / \log n$ for all k may be obtained (Devroye, 1993). For another possible proof method, see Pittel (1992). The depth D_n of the last node when the fringe heuristic is used has been studied by the theory of Markov processes or urn models in a series of papers, notably by Poblete and Munro (1985), Aldous, Flannery and Palacios (1988). See also Gonnet and Baeza-Yates (1991, p. 109). Poblete and Munro (1985) showed that

$$\frac{D_n}{\log n} \to \frac{1}{\sum_{i=k+2}^{2k+2} \frac{1}{i}} \stackrel{\text{def}}{=} \lambda(k)$$

in probability. It should be clear by now that the height of this tree may be studied via a split tree with split vector distributed as $(B, 1-B)$, where B is beta $(k+1, k+1)$. That is, B is distributed as the median of $2k+1$ i.i.d. uniform $[0,1]$ random variables. This representation is obtained by associating with each point in the data an independent uniform $[0,1]$ random variable. Equivalently, if the U_i are independent uniform $[0,1]$ random variables, then B is distributed as

$$\prod_{i=k+1}^{2k+1} U_i^{1/i} \ .$$

Note that in this case

$$m(\theta) = 2\mathbf{E}(B^\theta) = \frac{\Gamma(2k+2+\theta)\Gamma(k+1)}{\Gamma(2k+2)\Gamma(k+\theta+1)} \ .$$

The computation of μ is a little bit more tedious, but the result can be phrased indirectly:

Theorem 4.7. *(Devroye, 1993). A random binary search tree constructed with the aid of the fringe heuristic with parameter k has the following property: $\frac{H_n}{\log n} \to c(k)$ in probability where $c(k)$ is the unique solution greater than $\lambda(k)$ of the equation*

$$\psi(c) - c \sum_{i=k+1}^{2k+1} \log\left(1 + \frac{\psi(c)}{i}\right) + c\log 2 = 0 ,$$

and $\psi(c)$ is defined by the equation

$$\frac{1}{c} = \sum_{i=k+1}^{2k+1} \frac{1}{\psi + i} .$$

In particular, $\lambda(0) = 4.31107\ldots$ *(the ordinary binary search tree),* $\lambda(1) = 3.192570\ldots$*,* $\lambda(3) = 2.555539\ldots$*,* $\lambda(10) = 2.049289\ldots$ *and* $\lambda(100) = 1.623695\ldots$*.*

With

$$\sigma^2 = \sum_{j=k+2}^{2k+2} \frac{1}{j^2} ,$$

Devroye (1997) obtained a central limit theorem for D_n for all k:

$$\frac{D_n - \lambda \log n}{\sqrt{\sigma^2 \lambda^3 \log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) .$$

As an example, for $k = 1$, we obtain

$$\frac{D_n - (12/7)\log n}{\sqrt{(300/343)\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) .$$

Example 4: random simplex trees. Triangulating polygons and objects in the plane is an important problem in computational geometry. Arkin, Held, Mitchell and Skiena (1994) obtained a simple fast $O(n \log n)$ expected time algorithm for triangulating any collection of n planar points in general position. We look more specifically at their triangulation and its d-dimensional extension to simplices, and ask what the tree generated by this partitioning looks like if the points are uniformly distributed in the unit simplex. Given are n vectors $X_1, \ldots, X_n$ taking values in a fixed simplex S of $\mathbb{R}^d$. It is assumed that this is an i.i.d. sequence with a uniform distribution on S for the purposes of analysis. X_1 is associated with the root of a $d+1$-ary tree. It splits S into $d+1$ new simplices by connecting X_1 with the $d+1$ vertices of S. Associate with each of these simplices the subset of $X_2, \ldots, X_n$ consisting of those points that fall in the simplex. Each nonempty subset is sent to a child of the root, and the splitting is applied recursively to each child. As every split takes linear time in the number of points processed, it is clear that the expected time is proportional to $n\mathbf{E}(D_n)$, where D_n is the expected depth of a random node in the tree. The partition consists of $dn + 1$ simplices, each associated with an external node of the tree. There are precisely n nodes in the tree and each node contains one point. If $|S|$ denotes the size of a simplex S, then the following crucial property is valid.

Lemma 4.8. *(Devroye, 1997). If simplex S is split into $d+1$ simplices $S_1, \ldots, S_{d+1}$ by a point X distributed uniformly in S, then $(|S_1|, \ldots, |S_{d+1}|)$ is jointly distributed as $(|S|V_1, \ldots, |S|V_{d+1})$, where $V_1, \ldots, V_{d+1}$ are the spacings of $[0,1]$ induced by d i.i.d. uniform $[0,1]$ random variables.*

It is immediate that the random simplex tree is a split tree with split vector distributed as the spacings defined by d i.i.d. uniform $[0,1]$ random variables on $[0,1]$ and branch factor $d+1$. Therefore, H_n (and also D_n) behave precisely as for the random d+1-ary tree discussed earlier. Thus, if $\sigma^2 = \sum_{i=2}^{d+1} 1/i^2$,

$$\frac{D_n}{\log n} \to \lambda \stackrel{\text{def}}{=} \frac{1}{\sum_{i=2}^{d+1} \frac{1}{i}} \quad \text{in probability}$$

and

$$\frac{D_n - \lambda \log n}{\sqrt{\sigma^2 \lambda^3 \log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

As an example, if $d=2$, then and

$$\frac{D_n - (6/5)\log n}{\sqrt{(78/125)\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

We also know that $H_n / \log n \to c(d)$ in probability for a function c of d that may be computed via the recipe described in the example on b-ary search trees.

4.4 Refinements for Binary Search Trees

The results of the previous section permit fundamentally only first order asymptotic analysis of H_n. For the study of the depth of the last node D_n, or the depth of a typical node, branching processes are really not necessary, although they could be used. Devroye (1997) derives a general central limit theorem for D_n, illustrated in the previous examples, based on a split tree model as in the previous section. By allowing n balls to drop according to a certain process down an infinite b-ary tree in which nodes may hold zero, one, or more balls, the model is rich enough to encompass both search trees and tries or digital search trees. Recall that $\gamma = 4.31107\ldots$ the unique solution greater than 2 of $c\log(2e/c) = 1$. Theorem 2.1 implies that the height H_n of the random binary search tree satisfies $H_n / \log n \to \gamma$ in probability. In fact, convergence is in the almost sure sense as well, a fact first noted by Pittel (1984). Using elementary inequalities and essentially the bounds found in this survey, Devroye (1987) showed that $H_n - \gamma \log n = O(\sqrt{\log n \log\log n})$ in probability. Robson (1979) reported that H_n was much more concentrated than that, and conjectured even $\mathbf{Var}(H_n) = O(1)$. There have been three attempts to crack this conjecture.

Michael Drmota (1997) uses generating functions to prove that $\mathbf{E}(H_n) \sim \gamma \log n$, and his proof is the first one based on this approach. This method may have two benefits: first of all, it may provide detailed behavior on the exact behavior of $\mathbf{E}(H_n)$ (the lower order terms may be useful elsewhere), and the method may perhaps one day be extended to treat $\mathbf{Var}(H_n)$ in a similar manner.

Devroye and Reed (1995) provided the first analysis of the height that did not require any results from the theory of branching processes. Instead, they mark certain paths to leaves in the split tree that corresponds to the binary search tree, and apply the second moment method to compute bounds on probabilities. Interestingly, the marked leaves are sufficiently spread out to make this method work. This method was later generalized, via the notion of leading sequences, to common branching random walks, by McDiarmid (1995) (see Theorem 4.5). They were able to show that

$$\lim_{n\to\infty} \mathbf{Pr}\left(|H_n - \gamma \log n| > \frac{15\gamma}{\log 2} \log\log n\right) = 0 \ .$$

(Note that $15\gamma/\log 2 = 92.2933\ldots$.) Using a surprisingly elementary recursive argument, Robson (1997) showed that for any $\epsilon > 0$, infinitely often, we have

$$\mathbf{E}(|H_n - \mathbf{E}(H_n)|) < \frac{8\gamma}{\log 2} - 4 + \epsilon \ .$$

In fact, if

$$\sup_n \left(\mathbf{E}(H_{2n}) - \mathbf{E}(H_n)\right) < \infty \ ,$$

then his method allows one to conclude that

$$\sup_n \mathbf{E}(|H_n - \mathbf{E}(H_n)|) < \infty \ .$$

If we were to know $\mathbf{E}(H_n)$ down to $O(1)$ terms, we would be done, at least for first moment deviations.

Finally, we just learned from Jean Jabbour (1998) at the University of Versailles that he has a proof of Theorem 2.1 based solely on martingales. This may be yet another path along which to proceed.

4.5 Bibliographic Remarks

For general background information see, for example, Asmussen and Hering (1983), Athreya and Ney (1972), and Harris (1963). Lemma 4.3 takes elements from Kingman (1975), Biggins (1977), and Devroye and Zamora (1997). The minimal displacement B_n was compared by Durrett (1979) with that of the independent tree model, in which all n-th generation individuals have independent values of their common distribution. Bramson (1978) also worked out the finer behavior of B_n when the displacements are gaussian, or in general when particles describe Brownian motion and split at random times. Biggins (1990) derives a central limit theorem for $Z_n(.)$ when $\mathbf{E}(N)\log N < \infty$, where N is the number of offspring. Lemma 4.8 is implicit in many older references, such as Rubistein (1982), Smith (1984) or Devroye (1986a)

5. Crump-Mode-Jagers Process

5.1 Introduction

The Crump-Mode-Jagers (or CMP) branching (Crump and Mode, 1968) starts with a single ancestor born at time $t = 0$. $Z_1(t)$, the number of children born to the ancestor before time t is an arbitrary counting process. The children of the ancestor, from their births, behave independently of one another and of their parent, producing children at random according to random processes with the same joint distribution as $Z_1(.)$. Their children produce children in the same way, and so on. We speak of a Poisson CMP branching process if the between-birth intervals are exponentially distributed with parameters $\lambda_0, \lambda_1, \ldots$ respectively. Thus, births occur at intervals distributed as $E_0/\lambda_0, E_1/\lambda_1, \ldots$, where the E_i's are independent and exponentially distributed random variables. Note that if $\lambda_i = 0$, for some i, then the number of offspring of an individual can never exceed i.

If we link each individual with its parent, then we obtain a tree, and the notion of a generation becomes meaningful again. Several random variables are of interest here:

A. t_n, the time at which the tree has exactly n nodes.
B. B_n, the time of the first birth in the n-th generation.
C. H_n, the height of the tree at time t_n.
D. Z_k, the number of individuals in generation k.
E. $Z(t)$, the number of individuals at time t.
F. $H(t)$, the height of the tree at time t.

The reason CMP processes are important to us is because of the following connection with random trees that can be grown in an incremental manner. The random trees are grown one edge at a time, starting from the root. If the degrees of the current nodes are denoted by D_i, then node i is selected with probability proportional to λ_{D_i}. This node becomes the parent of a new node. Observe that the order of the births in the Poisson CMP process follows exactly that of the incremental random trees just described. Also, both are probabilistically equivalent if we are only interested in studying depths and heights of nodes. The last remark is rooted in the observation that if we have a number of birth processes with rates λ_i, then process i gives the next

birth with probability proportional to λ_i. The model described above and the continuous time embedding idea are due to Pittel (1984).
EXAMPLES.

A. The uniform random recursive tree (URRT) has $\lambda_i \equiv 1$ for all i. It is grown by choosing a parent with equal probability from among all possible parents.
B. The random m-ary pyramid with $m \geq 2$ has $\lambda_i = 1$ for $i < m$ and $\lambda_i = 0$ for $i \geq m$. Here we choose a parent uniformly at random from among those parents with less than m children. See Mahmoud (1994).
C. In the random binary search tree, we have $\lambda_0 = 2$, $\lambda_1 = 1$ and $\lambda_2 = 0$. To see quickly why this incremental tree model corresponds to the standard random binary search tree, consider a random binary search tree constructed on the basis of an i.i.d. sequence of uniform $[0, 1]$ random variables $U_1, U_2, \ldots$. Given that the tree has $n - 1$ nodes, the n-th node has a rank that is uniformly distributed on $\{1, 2, \ldots, n\}$. That is, it falls in one of the n intervals on $[0, 1]$ defined by the first $n-1$ uniform random variables. But each such interval corresponds uniquely to a potential new node (these are called external nodes), and there are two external nodes for a node with no children, and one for a node with one child.
D. The linear recursive tree has $\lambda_i = 1 + bi$ for some positive constant b. To visualize this, consider $b = 1$. To grow a tree, we pick a parent with probability proportional to one plus the number of children. For $b = 1$, this is called a plane-oriented recursive tree by Mahmoud (1993) and Mahmoud, Smythe and Szymański (1993) (see also Szymański, 1987, and Bergeron, Flajolet and Salvy, 1992). The last name is selected because of the following planar visualization: draw the tree in the plane, and place a new edge uniformly at random as any possible child of any possible rank. In this manner, a plane-oriented tree is defined.

There are three recent papers that provide an analysis of the height of these random trees using Crump-Mode processes, Pittel (1994) for the URRT and linear recursive tree, Mahmoud (1994) for random pyramids, and Biggins and Grey (1996) in the more general setting followed in this chapter. The height H_n can be analyzed using the Biggins-Hammersley-Kingman theorem (Theorem 4.2). We conclude by working out the details for the various tree models mentioned above.

5.2 The Main Result

The relationship between the CMP process and the branching random walk is clear, if we let the displacements in the branching random walk be the inter-birth times. As the branch factor may be unbounded (as for the URRT case), we need to follow a general set-up. For simplicity, to ensure survival, we assume throughout that $Z_1(\infty) \geq 1$. For a general branching walk process, we define the Laplace transform of the mean reproduction measure,

$$m(\theta) = \mathbf{E}\left(\sum_i e^{-\theta Y_i}\right)$$

where the Y_i's are the realizations of $Z_1(.)$, and the sum ranges over all children of the root.

Example. For a Poisson CMP process, we have
$Y_1 = E_0/\lambda_0$, $Y_2 = Y_1 + E_1/\lambda_1$, and so forth, so that

$$\begin{aligned} m(\theta) &= \textstyle\sum_{i=0}^{\infty} \mathbf{E}\left(e^{-\theta(E_0/\lambda_0+\cdots+E_i/\lambda_i)}\right) \\ &= \textstyle\sum_{i=0}^{\infty} \prod_{j=0}^{i} \mathbf{E}\left(e^{-\theta E_j/\lambda_j}\right) \\ &= \textstyle\sum_{i=0}^{\infty} \prod_{j=0}^{i} \frac{1}{1+\frac{\theta}{\lambda_j}} \ . \end{aligned}$$

Assuming that $m(\theta) < \infty$ for some $\theta > 0$, we note that as $\theta \to \infty$, $m(\theta) \to 0$. Observe that a sufficient condition for this is that $\lambda_j = O(j)$ as $j \to \infty$ in the Poisson CMP case). Define

$$\mu(a) = \inf\left\{e^{\theta a} m(\theta) : \theta \geq 0\right\} \ ,$$

and observe that $\log \mu(a)$ is concave (the infimum of a family of lines is concave) and $\mu(a)$ is continuous on the interior of $\{a : \mu(a) > 0\}$.

Define $Z_k(t)$, the number of individuals in generation k with value at most t. Biggins (1977) uses classical large deviation results by Bahadur and Rao (1960) and Chernoff (1952) to prove the following:

Theorem 5.1. *(Biggins, 1977; Hammersley, 1974; Kingman, 1975).*
If $m(\theta) < \infty$ *for some* $\theta > 0$, *then* $(\mathbf{E}(Z_n(na)))^{\frac{1}{n}} \to \mu(a)$ *as* $n \to \infty$. *Furthermore, if* $\mu(a) < 1$, *then with probability one,* $Z_n(a)(na) = 0$ *for all but finitely many* n. *If* $a \in \mathrm{int}\{a : \mu(a) > 1\}$, *then* $\lim_{n\to\infty} (Z_n(na))^{1/n} = \mu(a)$ *almost surely. Finally,*

$$\lim_{n\to\infty} \frac{B_n}{n} = \gamma \stackrel{\mathrm{def}}{=} \sup\{a : \mu(a) < 1\}$$

almost surely, and γ *is finite.*

We must relate B_n to H_n. Observe that at the moment t_n, the family tree is of size n and of height H_n and that $B(H_n)$ and $B(H_n+1)$ are the first moments when the height becomes equal to H_n and H_n+1 respectively. Therefore,

$$B(H_n) \leq t_n \leq B(H_n+1) \ .$$

Since $t_n \to \infty$ almost surely, we have $H_n \to \infty$ almost surely as well. Thus, $B(H_n)/H_n \to \gamma$ almost surely, and $t_n/H_n \to \gamma$ almost surely. Therefore it suffices to study t_n. This can be done on a case by case basis, as is routinely done in the literature. However, there is a universal theorem:

Theorem 5.2. *(Nerman, 1981; Biggins, 1995) If $m(\theta) < \infty$ for some $\theta > 0$, and $Z(t)$ denotes the number of births up to time t, and*

$$\alpha \stackrel{\text{def}}{=} \inf\{\theta : m(\theta) < 1\}$$

(which is positive and finite, as $m(0+) \geq 1$ and $m(\theta) \to 0$ as $\theta \to \infty$), then

$$\frac{\log Z(t)}{t} \to \alpha$$

almost surely as $t \to \infty$. Equivalently,

$$\frac{t_n}{\log n} \to \frac{1}{\alpha}$$

almost surely as $n \to \infty$.

From this, we have:

Theorem 5.3. *(Biggins and Grey, 1996) Under the conditions of Theorem 5.2,*

$$\frac{H_n}{\log n} \to \frac{1}{\alpha\gamma}$$

almost surely as $n \to \infty$.

5.3 Application to Various Tree Models

In a few special cases, we have very refined information about t_n. This occurs principally when we can describe the spacings between consecutive births quite accurately. Consider first a branching process with one child per node, and the inter-birth times are exponential of unit parameter, then t_n is the sum of n independent standard exponential random variables, so that $t_n/n \to 1$ almost surely. Also, $H_n = n-1$, $m(\theta) = 1/(1+\theta)$ and

$$\mu(a) = \inf\left\{\frac{e^{\theta a}}{1+\theta} : \theta \geq 0\right\} .$$

The minimum occurs at $\theta = \max(1/a - 1, 0)$, so that

$$\mu(a) = \begin{cases} ae^{1-a} & (0 < a < 1) \ ; \\ 1 & (a \geq 1). \end{cases}$$

Since $\mu(1) = 1$, we have $\gamma = 1$. This was just a (stupid) roundabout way of checking what we already knew, that $H_n/n \to 1$ almost surely (as $H_n = n-1$).

In the second example, let Y_1, Y_2, the children of the root, be born at independent standard exponential times. In this case,

$$m(t) = \frac{2}{1+\theta}.$$

Clearly,

$$\mu(a) = \inf\left\{\frac{2e^{\theta a}}{1+\theta} : \theta \geq 0\right\} .$$

The minimum occurs at $\theta = \max(1/a - 1, 0)$, so that

$$\mu(a) = \begin{cases} 2ae^{1-a} & (0 < a < 1) ; \\ 2 & (a \geq 1). \end{cases}$$

Thus, γ is the solution less than one of $2ae^{1-a} = 1$. To study t_n, note that we have inter-birth times that are distributed as $E_2/2, E_3/3, \ldots, E_n/n$, where the E_i's are independent exponential random variables. From this, it is easy to show that

$$\frac{t_n}{\log n} \to 1$$

almost surely. Therefore, $H_n/\log n \to 1/\gamma$ almost surely. This may be cast in the Poisson CMP model, as the first birth to the ancestor occurs at a time distributed as $E_1/2$, and the second at a time distributed as $E_1/2+E_2$, where the E_i's are exponential random variables. Thus, $\lambda_0 = 2$, $\lambda_1 = 1$, and $\lambda_i = 0$ for $i \geq 2$. This, of course, yields the same results.

In a third example, let the root have children whose times of birth are distributed like a Poisson point process of unit rate. Thus,

$$m(\theta) = \sum_{j=1}^{\infty} \left(\frac{1}{1+\theta}\right)^j = \frac{1}{\theta} .$$

Therefore,

$$\mu(a) = \inf\left\{\frac{e^{\theta a}}{\theta} : \theta \geq 0\right\} .$$

The minimum occurs at $\theta = 1/a$, so that

$$\mu(a) = ea .$$

Thus, $\gamma = 1/e$. The study of t_n is equally simple, as t_n is distributed as $E_1/1 + E_2/2 + \cdots + E_{n-1}/(n-1)$. To see this, note that if k elements are alive, the time until the next birth is distributed as E_k/k, as the minimum of k independent exponentiual random variables. Thus, as before, $t_n/\log n \to 1$ almost surely. It is easily seen that $H_n/\log n \to 1/\gamma = e$ almost surely. This result for the uniform random recursive tree was first obtained in Devroye (1987).

Our fourth example involves the plane-oriented recursive tree. In this case, if a node u has degree $d(u)$, then its probability of making a child is proportional to $1 + d(u)$. This is like saying that the children of the root are born with inter-birth times distributed like E_1, $E_2/2$, $E_3/3$, and so forth. A simple computation shows that

$$m(\theta) = \sum_{j=1}^{\infty} \prod_{i=1}^{j} \left(\frac{i}{i+\theta} \right)$$

The computation of γ is a bit more complicated (see Pittel (1994) or Mahmoud (1994)). However, the inter-birth times are easy to deal with. Indeed, the sum of the intensities of the birth process is $\sum_u (1 + d(u)) = 2|u| - 1$, where $|u|$ denotes the number of nodes. Therefore, the inter-birth times for the tree are distributed like $E_1/1$, $E_3/3$, ... Hence, it is not hard to show that $t_n/\log n \to 1/2$ almost surely, so that $H_n/\log n \to 1/(2\gamma)$ almost surely.

In the random m-ary pyramid, we have $m(\theta) = (1 - (1+\theta)^{-m})/\theta$. One can easily see that for $m = 2$, $\alpha = (\sqrt{5} - 1)/2$ (Theorem 5.3), but γ requires numerical computation. See Mahmoud (1994).

Finally, for the linear recursive tree, Pittel (1994) and Biggins and Grey (1995) show that $m(\theta) = \frac{1}{\theta - 1/b}$ for $\theta > b$, so $\alpha = 1 + b$, $\mu(a) = ae^{1+ba}$, and γ is the unique root of $ae^{1+ba} = 1$. Thus, $H_n/\log n \to 1/(\gamma(b+1))$ almost surely as $n \to \infty$.

In a Bellman-Harris set-up, the whole litter is born simultaneously at time T. If there are b children per parent, then we have $m(\theta) = b\mathbf{E}(e^{-\theta T})$. When T is exponential and $b = 2$, this is the celebrated **Yule process**. Clearly, $m(\theta) = 2/(1+\theta)$, exactly as for the binary search tree discussed earlier. Thus, the height behaves in a manner similar to that of the binary search tree, even though the CMP processes are very different indeed. When T is not necessarily exponential, and the litter size follows a general distribution, we obtain the Bellman-Harris branching process, which is the subject of the next section.

5.4 The Bellman-Harris Branching Process

In 1952, Bellman and Harris described a generalization of the Galton-Watson branching process by embedding it in continuous time. The (so-called age-dependent branching) process is described by two parameters, a discrete distribution $\{p_i, i \geq 0\}$ for the number of children, as in a standard Galton-Watson process, and a distribution of a strictly positive random variable T, the time between birth and reproduction. With each edge in the Galton-Watson tree, we associate an independent copy of T. The process is started with a single root at time 0. The elements are still grouped in generations. The root element produces a litter of size determined by $\{p_i\}$ after a time T_1 distributed as T. Each individual in the litter reproduces in the same manner and independently.

This model can also be used for describing the growth of the random binary search tree. We take the point of view that we let the random binary search tree grow by at each iteration picking an external node uniformly and at random. This node becomes an internal node, gets removed from the pool of external nodes, and produces two new external nodes, its potential children. At any moment, there are n internal nodes if and only if there are $n+1$ external nodes. If T is standard exponential, then given that there are k external nodes at time t, by the memoryless property of the exponential distribution, we in fact pick as our next node any external node with equal probability. Thus, the order in which the nodes are chosen is identical to that for growing the random binary search tree. In notation of the previous section, the tree obtained at the time t when there are exactly $n+1$ external nodes is a random binary search tree on n internal nodes. Recall that the process in which T is exponential and the number of offspring is always two is the Yule process, or binary fission (Athreya and Ney, 1972, p. 109). For different distributions of T, we obtain different kinds of random binary trees. We will not explore the Yule process construction of random binary search trees any further, except for the mention of the following theorem below, valid when T is standard exponential.

Theorem 5.4. *Assume that $\{p_i\}$ has finite second moment and that T is standard exponential. Let $Z(t)$ be the number of particles alive at time t in a Bellman-Harris process. Then $Z(t)e^{-t}$ tends almost surely to a random variable W,*

$$\frac{Z(t) - e^t W}{\sqrt{Z(t)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \mathbf{Var}(W)$. Finally, conditioned on W, $U(t) \stackrel{\text{def}}{=} Z(\log(1 + t/W))$ is a unit rate Poisson process in t. That is, for any $0 < t_1 < \cdots < t_k < \infty$, and integers $n_i \geq 0$, $2 \leq i \leq k$, and Borel subset $B \subseteq [0, \infty)$,

$$\begin{aligned} \mathbf{Pr}(U(t_2) - U(t_1) &= n_2, \ldots, U(t_k) - U(t_{k-1}) = n_k,\ W \in B) \\ &= \mathbf{Pr}(W \in B) \textstyle\prod_{i=2}^k \mathbf{Pr}(P(t_i - t_{i-1}) = n_i) \end{aligned}$$

where $P(s)$ is a Poisson (s) random variable. Furthermore, $U(0) = Z(0) = 1$. For the Yule process, the random variable W has the standard exponential distribution.

The Poisson representation in the theorem above is due to Kendall (1966). If T is standard exponential, then in the Yule process, $Z(0) = 0$ and $Z(t)$ increases by one each time a particle gets replaced (as one dies but two are born). Two interesting properties of the exponential distribution are the following: if $E_1, E_2, \ldots$ are i.i.d. exponential random variables, then

A. For any n, $\min(E_1, \ldots, E_n) \stackrel{\mathcal{L}}{=} \frac{E_1}{n}$.
B. (The memoryless property.) For any $t > 0$, $E_1 - t$, given $E_1 > t$, is distributed as E_1.

Thus, the intervals between times of birth in a Yule process are distributed like E_1, $E_2/2$, $E_3/3$, ... Using these two properties repeatedly, we have

$$\begin{aligned}\mathbf{Pr}(Z(t) > k) &= \mathbf{Pr}(E_1 + E_2/2 + E_3/3 + \cdots + E_k/k \le t) \\ &= \mathbf{Pr}(\max(E_1, E_2, \ldots, E_k) \le t) \\ &= (1 - e^{-t})^k\end{aligned}$$

so that everything is known about the distribution of $Z(t)$. For example,

$$\mathbf{E}(Z(t)) = \sum_{k \ge 0} \mathbf{Pr}(Z(t) > k) = e^t \ .$$

In fact, at any t, $Z(t)$ has the geometric distribution with parameter e^{-t}.

5.5 Introduction

Of particular interest is the conditional Galton-Watson process, or conditional branching process, or simply CBP, in which we condition on $N = n$, where $N = \sum_{i=0}^{\infty} Z_i$ is the total size of the population, Z_i is the size of the population in generation i, and $Z_0 = 1$. These processes were studied by Kennedy (1975) and Kolchin (1978, 1985), who made key connections between them and so-called simply generated random trees, introduced by Meir and Moon (1978). These trees are uniformly picked in given collections such as, for example, all binary trees on n nodes.

Several examples will be given in the next section. In the other sections, we review some results for the distribution, size and height of the trees in this model.

Consider a multiset of trees, that is, a set in which repetitions are allowed. Let the *weight* $\Omega(t)$ of a tree t be the number of occurrences of t. Let $|t|$ denote the *size* of t, i.e., the number of nodes contained in t. Then

$$a_n = \sum_{t:|t|=n} \Omega(t)$$

is the number of trees in this multiset with n nodes. The *generating function* for $\{a_n\}$ is denoted by

$$y(z) = \sum_{n \ge 0} a_n z^n \ .$$

We define a *random tree* T_n of size n by

$$\mathbf{Pr}(T_n = t) = c\Omega(t) I_{|t|=n} = \frac{I_{|t|=n}}{a_n} \ ,$$

where c is a normalization constant. Thus, each of the a_n occurrences of elements in the multiset of trees of size n has the same probability. Therefore, it is appropriate to speak of a uniform model if we can somehow distinguish between all $\Omega(t)$ copies of t thrown into the multiset. This is illustrated in the next section.

A particularly interesting multiset of trees is the *simply generated family of trees* (Meir and Moon, 1978), which requires a descriptor

$$\phi(y) = \sum_{i=0}^{\infty} c_i y^i \ ,$$

where $c_0 > 0$, and the c_i's are nonnegative integers (usually, but not necessarily, uniformly bounded in i). The notation ϕ, y and c_i is by now standard, so we will adopt it as well. Consider ordered trees, that is, trees in which the order of the children matters. For each ordered tree t, let $D_i(t)$ be the number of nodes in t with i children (successors). Then define

$$\Omega(t) \stackrel{\text{def}}{=} \prod_{i \geq 0} c_i^{D_i(t)} \ .$$

The family of trees is *aperiodic* if $\gcd\{i > 0 : c_i > 0\} = 1$, and *periodic* otherwise. We define a *random simply generated tree* T_n of size n by

$$\mathbf{Pr}(T_n = t) = c\Omega(t) I_{|t|=n}$$

where c is a normalization constant. We note here that because we have ordered trees,

$$y(z) = z\phi(y(z)) \ .$$

A proof is given in Theorem 5.8.

Next, we define a Galton-Watson branching process with parameter $\theta > 0$ with offspring distribution

$$p_i = \frac{c_i \theta^i}{\phi(\theta)} \ , \ i \geq 0 \ .$$

Here we assume that $\phi(\theta) < \infty$. It is easy to verify that $(p_0, p_1, \ldots)$ is indeed a probability vector. Furthermore, the expected number of offspring, an increasing function of θ, is

$$\sum_{i \geq 0} i p_i = \sum_{i \geq 0} \frac{i c_i \theta^i}{\phi(\theta)} = \frac{\theta \phi'(\theta)}{\phi(\theta)} \ .$$

Let τ be the smallest positive root of $\phi(\tau) = \tau\phi'(\tau)$. Then for $\theta = \tau$, the branching process is critical, while for $0 < \theta < \tau$, it is subcritical. We now define CBP with parameter n as the above Galton-Watson process conditioned on the total population size n, and let T_n' denote a realization of CBP.

The crucial properties of the two random trees defined above are captured in Theorem 5.5, which states that the conditioned Galton-Watson tree T_n has the same distribution as the random simply generated tree!

Theorem 5.5. *(Kennedy, 1975) The distribution of T'_n is independent of $\theta \in (0, \tau]$. Furthermore, $T_n \stackrel{\mathcal{L}}{=} T'_n$, where $\stackrel{\mathcal{L}}{=}$ denotes equality in distribution.*

Proof. The first statement follows from the second one. Let t be an arbitrary fixed ordered tree with $|t| = n$. Let T^* be a family tree produced by the (unconditioned) Galton-Watson process. Then

$$\begin{aligned}
\mathbf{Pr}(T^* = t) &= \prod_{i\geq 0} (\mathbf{Pr}(Z_1 = i))^{D_i(t)} \\
&= \prod_{i\geq 0} \left(\frac{c_i\theta^i}{\phi(\theta)}\right)^{D_i(t)} \\
&= \prod_{i\geq 0} c_i^{D_i(t)} \times (\phi(\theta))^{-\sum_i D_i(t)} \times \theta^{\sum_i iD_i(t)} \\
&= \Omega(t) \times (\phi(\theta))^{-|t|} \times \theta^{n-1} \\
&= \Omega(t) \times (\phi(\theta))^{-n} \times \theta^{n-1} \ .
\end{aligned}$$

Also,

$$\begin{aligned}
\mathbf{Pr}(|T^*| = n) &= \sum_{t:|t|=n} \mathbf{Pr}(T^* = t) \\
&= \sum_{t:|t|=n} \Omega(t)(\phi(\theta))^{-n} \times \theta^{n-1} \\
&= a_n(\phi(\theta))^{-n} \times \theta^{n-1} \ ,
\end{aligned}$$

where a_n is the number of trees in the multiset of size n. Therefore, with $|t| = n$,

$$\mathbf{Pr}(T^* = t||T^*| = n) = \frac{\mathbf{Pr}(T^* = t)}{\mathbf{Pr}(|T^*| = n)} = \frac{\Omega(t)}{a_n} \ .$$

But this is proportional to $\Omega(t)$, so that T_n is indeed distributed as T^* conditional on $|T^*| = n$, that is, as T'_n. □

Trees are used in symbolic computations to represent formulas, with internal nodes representing operators or functions, and leaves operands. These are also called expression trees in the literature on parsing and the evaluation of expressions in higher level languages. In the analysis of such objects, it is natural to assume that all objects are equally likely. For example, in ordinary trigonometric expressions on three operands, x, y and z, there are internal nodes with two children (+ and -), internal nodes with one child (sin, cos, tan, cot), and leaves with zero children (x, y and z). The nodes are thus labeled, with a different number of labels according to the type of tree. In the formalism of the previous section, we have $c_0 = 3$, $c_1 = 4$ and $c_2 = 2$. As $y(z) = z\phi(y(z))$, we may get exact or asymptotically accurate expressions by analytic methods: see Vitter and Flajolet (1990) for a survey of such methods, based on Lagrange inversions and singularity analysis. For expected values of various additive parameters, this is indeed a natural route to follow.

5.6 Examples of Trees in the Uniform Random Tree Model

(1,1). Several choices of descriptors lead to various types of trees. Consider first the choice $(1,1)$. The weight of a tree t is one for every tree consisting of just leaves and one-child nodes. Thus, the multiset will contain one of each of these trees, which in fact are just linked chains. The CBP has probability vector

$$\left(\frac{1}{1+\theta}, \frac{\theta}{1+\theta}\right) .$$

But clearly, conditioned on the size of the tree being n, we see that it does not matter which θ we picked. The tree has height exactly $n-1$. One can easily verify that the same result would have been obtained if we had selected the descriptor (a,b) for any $a,b>0$. Therefore, interesting trees only occur when $c_i > 0$ for some $i > 1$.

(1,0,1). The next simplest choice is $(1,0,1)$. Here we place in our multiset trees with only leaves and two-child nodes. Such trees must have an odd cardinality. If $|t| = 2k+1$, there are necessarily $k+1$ leaves and k two-child nodes. The weight of each tree of size $n = 2k+1$ is thus identical and equal to 1 (as all nonzero c_i's are one). Hence, each tree in the multiset is different, and all possible trees of the type described above are present. The family is the family of full binary trees. Again, all such trees occur equally often in the multiset.

(1,0,m). If we take $(1,0,m)$, then the weight of each tree of size $n = 2k+1$ is m^k, and within this class, all trees occur equally often in the multiset. Therefore, there is no difference between random simply generated trees for $(1,0,m)$ for any $m > 0$.

(1,2,1). The next member on the ladder of complexity is $(1,2,1)$. Here we have trees with nodes having up to two children, and the weight of a tree with n nodes of which there are l leaves is given by $2^{n-(2l-1)}$, as the number of nodes with two children is $l-1$. Interestingly, not all trees with n nodes have equal representation. We can however *force* a distinction on them by additional ways of distinguishing between trees. For example, for each node with one child, we may make the child a left child or a right child of its parent. For a tree with $n-(2l-1)$ such nodes, there are $2^{n-(2l-1)}$ possible combinations of left/right distinctions. Let us attach exactly one of these combinations to each of the $2^{n-(2l-1)}$ trees with n nodes and l leaves in our multiset. Then, each tree in the multiset is distinct, and is in fact an ordinary binary tree. And all binary trees on n nodes are indeed in the multiset. An equivalent multiset (for our purposes) would have been obtained with the choice $(1,2m,m^2)$ for any $m > 0$. We will also refer to these trees as Catalan trees.

(1,m,1). If we pick $(1,m,1)$, then it is necessary to create a designation for each single child, and we could associate a label between 1 and m with each such lone child. This assures a bijection between all such "labeled" trees with up to two children per node and the trees in the multiset. With

$m = 1$, labeling is superfluous, and one obtains the so-called unary-binary trees, which are the ordered trees with up to two children per node.

$(\mathbf{1}, \mathbf{m}, \mathbf{m^2})$. If we pick $(1, m, m^2)$, then we color each child in one of m colors, and note that with all possible colorings, all trees in the multiset occur only once, and that there is a bijection. The family is that of trees with up to two children per node, and all nodes except the root are colored in one of m colors. In the CBP, we may set $\theta = 1/m$ to obtain the reproduction distribution $(1/3, 1/3, 1/3)$. Thus, the shape properties of all these trees are identical, regardless of the choice of m.

Binomial. Position trees of branch factor b are trees in which each node has up to b children, and each child is given a position, and only one child can occupy each position. With $b = 2$, this yields the binary trees. For general b, it is not hard to see that the descriptor must be binomial of the form $\left(1, \binom{b}{1}, \binom{b}{2}, \ldots, \binom{b}{b-1}, \binom{b}{b}\right)$. Ternary trees are obtained by using the descriptor $(1, 3, 3, 1)$, for example.

(1,1,1,...) or geometric. All ordered trees without restrictions on the number of children are obtained by the infinite descriptor $(1, 1, 1, \ldots)$. These are also called unlabeled rooted ordered trees or unlabeled planted plane trees, or unlabeled rooted plane trees, or just planted plane trees. For the CBP, we must take $\theta < 1$, so that $\phi(\theta) = 1/(1-\theta)$, and the basic reproduction distribution is given by $(1/(1-\theta), \theta/(1-\theta), \ldots, \theta^i/(1-\theta), \ldots)$, that is, a geometrically decreasing probability vector. From Theorem 5.5, we note that *any* $\theta \in (0, 1)$ yields the same random tree in the conditioned branching process model. We might thus as well take $\theta = 1/2$. It takes just a moment to verify that all unlabeled rooted plane trees with non-root nodes colored in one of m colors are obtained from $(1, m, m^2, m^3, \ldots)$. For the CBP, we require therefore $\theta < 1/m$. But then the CBP is exactly as in the case $m = 1$ (geometric), and thus this choice of descriptor is equivalent to $(1, 1, 1, \ldots)$ if we want to study shape properties of the trees, unrelated to color choices.

(1,0,0,...,1). If the only nonzero coefficient are the 0-th and the t-th, with $t > 0$, we obtain the so-called t-ary trees of Flajolet and Odlyzko (1982).

(1,1,2,3,4,5,...). A node with k children gets a label between 1 and k, which may indicate which of its children (in the ordered tree) is "best". We will call these trees favorite son trees.

If we remove structure in the order, by removing the order of the children altogether, or by replacing the total order by a circular order or a partial ordering, we in fact allow c_i's to take values less than one. This will not be pursued here. See, however, the section on Cayley trees, where a connection is made with Poisson-distributed CBP's.

5.7 Catalan Trees and Dyck Paths

There are specially pretty derivations of the equivalence between a CBP and a uniform random Catalan tree. We first consider a nonnegative random walk in which all steps are $+1$ or -1, we start at $X_0 = 0$, and have $X_{2n} = 0$. If we replace $+1$ and -1 by a and b respectively, then the sequence of $2n$ symbols thus obtained is a Dyck word. The walk is also called a Dyck path. If a_n is the number of different Dyck paths of length $2n$, by conditioning on the place $2p$ of the first return to the origin, we have

$$a_n = \sum_{p=0}^{n-1} a_p a_{n-1-p}$$

and $a_1 = 1$, $a_0 = 1$. It is well-known that

$$a_n = \frac{1}{n+1}\binom{2n}{n} ,$$

the n-th Catalan number. There is a bijection between a Dyck path of length $2n$ and a binary tree on n nodes. Draw the binary tree in the standard manner. Write an a to the left of every node, and a b underneath each node. Then start at the root and walk around the tree by following edges just like a boat would follow the shoreline, and note the sequence of a's and b's. The order of visit is called preorder. The sequence forms a Dyck word as the number of a's at any point must exceed the number of b's. This bijection is useful for many purposes but for the study of parameters as the height of the random binary tree, some extra work is needed. We just note that the rooted binary trees were correctly counted as far back as Cayley (1858).

Another bijection may be considered, but now with rooted ordered trees with $n+1$ nodes (and thus n edges), by placing next to each edge an a to the left and a b to the right, and forming a Dyck word by the walk of the former bijection. This walk will be referred to as a Harris walk. The correspondence with a CBP can be seen as follows. Let $X_1, X_2, \ldots$ be i.i.d. random variables taking the values $+1$ and -1 with equal probability. Let $S_n = \sum_{i=1}^n X_i$ be the partial sums. Consider only $X_1 = 1$. Define ρ as the time of the first return to zero: $\rho = \inf\{n : S_n = 0\}$. Let $\rho_1, \ldots, \rho_N$ be the times less than ρ with $S_n = 1$. We set $\rho_0 = 1$, and note that $\rho_N = \rho - 1$. Define $t_1 = \rho_1 - \rho_0$, $t_2 = \rho_2 - \rho_1$, and so forth. Note that

$$\mathbf{Pr}(N = k) = \frac{1}{2^{k+1}} ,$$

where $\mathbf{Pr}(.)$ denotes always conditional probability given $X_1 = 1$. This is best seen by noting that at each passage at one, the random walk has exactly 50% probability of returning to the origin. Thus, N is indeed geometrically distributed of parameter $1/2$. Furthermore, given $N = k \geq 1$, the excursions above one of lengths $t_1, \ldots, t_N$ are independent and have the same distribution as the original positive excursion $S_1, \ldots, S_\rho$. This is just a manifestation of the strong Markov property applied to the ordinary random walk. We now construct the corresponding ordered tree explicitly: take a root, and give it N children, and associate with the children the positive excursions of lengths

$t_1, \ldots, t_N$ respectively. Constructed in this manner, we note that the corresponding tree is nothing but a critical Galton-Watson tree with reproduction distribution $\mathbf{Pr}(Z = k) = 1/2^{k+1}, k \geq 0$. The bijection is formidable as it not only yields the desired connection, but it also is rather direct: for example, the maximum of an excursion corresponds to the height of the Galton-Watson tree, and the length of an excursion is twice the size of the Galton-Watson tree.

One may use the well-known bijection between rooted ordered trees on $n+1$ nodes and binary trees on n nodes: first copy all $n+1$ nodes from the ordered tree to the binary tree; then associate each parent-oldest child edge in the ordered tree with a parent-left child edge in the binary tree, and associate with each node-next sibling relationship in the ordered tree a parent-right child edge in the binary tree. Finally, remove the root and its left edge from the binary tree. This yields yet another (but slightly more indirect) bijection between Dyck paths and binary trees. The CBP relationship follows easily: if N is the number of children of the root in the ordered tree, then the binary tree's root (before removal) has a left child if $N > 0$. A node in the ordered tree regarded as a child in a family has a number Y of younger siblings that is again geometric $(1/2)$ by the memoryless property of the geometric distribution. Thus, it has a right child in the binary tree if $Y > 0$. To make a Galton-Watson process, place in the ordered tree a pair $(U, V) = (I_{N>0}, I_{Y>0})$, and observe that all these pairs in the tree are independent, and that U and V are also independent. Thus, the binary tree with a random number of nodes and after removal of the root is indeed a Galton-Watson tree with reproduction distribution $(p_0, p_1, p_2) = (1/2, 1/4, 1/2)$.

We should also mention that for symmetric random walks with zero mean having continuous distributions, Le Gall (1989) has proposed a beautiful tree construction that leads once again to a binary Galton-Watson tree with $(p_0, p_1, p_2) = (1/2, 1/4, 1/2)$.

5.8 Cayley Trees

The uniform random labeled tree $\mathcal{L}_n$ is the tree picked uniformly from the n^{n-2} trees on vertices $\{1, 2, \ldots, n\}$. The uniform random rooted labeled tree (or rooted nonplanar tree) $\mathcal{R}_n$ is the tree picked uniformly from the n^{n-1} trees on vertices $\{1, 2, \ldots, n\}$ in which one vertex is declared to be the root. Cayley (1889) studied $\mathcal{L}_n$ and Riordan (1960) counted various related species of trees , including $\mathcal{R}_n$. Rényi and Szekeres (1967) showed that the expected height H_n of $\mathcal{R}_n$ is $\sim \sqrt{2\pi n}$. They also showed that the limit distribution of $H_n/\sqrt{n}$ is the theta distribution (see further on). Rényi (1959) showed that the number of leaves is asymptotic to n/e, while Meir and Moon (1970) showed that the expected distance between two nodes taken at random is asymptotic to $\sqrt{\pi n/2}$.

Kolchin (1986), just like Meir and Moon (1978) and Moon (1970), studies $\mathcal{L}_n$ and $\mathcal{R}_n$ via generating functions, establishing a tight relationship with

CBP's. More probabilistic approaches may be found in Grimmett (1980) and Aldous (1988, 1991). The purpose of this section is to point out the key results in the latter papers.

Consider a Poisson (1) Galton-Watson tree $\mathcal{P}$. Make $\mathcal{P}$ a labeled tree by randomly labeling the vertices $1, \ldots, |\mathcal{P}|$. If t is a specific rooted labeled tree (having $|t|$ vertices), then

$$\mathbf{Pr}(\mathcal{P} = t) = \frac{e^{-|t|}}{|t|!} .$$

To see this, order all the sets of siblings in t by increasing labels, and let $N_1, \ldots, N_{|t|}$ be the number of children of all nodes, listed in preorder. Then,

$$\mathbf{Pr}(\mathcal{P} = t) = \prod_{i=1}^{|t|} \frac{1}{N_i!\, e} \frac{\prod_{i=1}^{|t|} N_i!}{|t|!}$$

where the first factor accounts for matching the geometrical layout of the tree (it uses the independence of the number of offspring, as well as the Poisson property), and the second factor is the probability of getting the random labels just right. Therefore, conditional on $|\mathcal{P}| = n$, we see that $\mathcal{P}$ is uniform on labeled trees of size n, and is thus distributed as $\mathcal{R}_n$. This property allows us to study the CBP with Poisson (1) offspring. The calculation above establishes the connection and may be made into a construction of $\mathcal{R}_n$. The theorems about CBP's then provide information on random Cayley trees.

There is a second construction due to Aldous (1988). It requires i.i.d. random variables $U_1, \ldots, U_n$ uniformly distributed on $\{1, \ldots, n\}$. First we make 1 the root. Then with i varying from 2 to n, we add edge $(i, \min(i-1, U_i))$. Then we remove the labels to obtain a random rooted (nonuniform) unlabeled tree. It can be made in a tree distributed as $\mathcal{R}_n$ by randomly assigning labels.

Grimmett (1980) proposes yet another related process, and Aldous (1991) builds on it to derive a tool for studying local properties of such trees. For each $k = 0, 1, 2, \ldots$, we create independent Poisson (1) Galton-Watson trees, regarded as trees with root r_k and other vertices unlabeled. Then we connect $r_0, r_1, r_2, \ldots$ as a path, make r_0 the root, and delete the labels. For fixed k, the vector of k i.i.d. copies of $\mathcal{P}$ is close in total variation distance to a random rooted unlabeled tree with a distinguished path of length $k-1$ attached to it. This connection will not be explored here.

Finally, we mention the Prüfer codes that are so useful in the generation and counting of all labeled trees (rooted or unrooted). The properties that may be deduced based on these codes are not directly linked to branching processes, and will thus not be studied here.

5.9 Fringe Subtrees

Following Aldous (1990), for a finite rooted ordered tree T, we call T^* the subtree rooted at a randomly and uniformly picked vertex from T. Aldous observed that in many (random or non-random) tree models, T^* tends in distribution to a certain random tree as $|T| \to \infty$. This has of course immediate consequences for the parameters of T^*. For example, we have the following:

Theorem 5.6. *(see Aldous, 1990) Let ξ be an offspring distribution of a Galton-Watson process, with $\mathbf{E}(\xi) = 1$, $\mathbf{Pr}(\xi = 1) < 1$, $\mathbf{E}(\xi^2) < \infty$ and ξ non-lattice. Let T be the Galton-Watson tree (note $|T| < \infty$ almost surely), and let T_n be T conditional on $|T| = n$. Let T_n^* be a tree rooted at a random vertex of T_n. Then for all trees t,*

$$\lim_{n \to \infty} \mathbf{Pr}(T_n^* = t) = \mathbf{Pr}(T = t) .$$

Discussion.

In this remarkable result, note that the limit distribution of a fringe tree of the CBP is the unconditional Galton-Watson tree! As a result, we may immediately deduce properties of local parameters from this. For example, the degree of a random vertex in a CBP tends in distribution to the degree of the root of T, that is, ξ. Also, $|T_n^*| \xrightarrow{\mathcal{L}} |T|$. Note also that the number of vertices in a CBP within distance k of a uniform random vertex tends in distribution to the number of vertices within distance k of the root of T, that is, $Z_0 + Z_1 + \cdots + Z_k$, where $Z_0, Z_1, \cdots$ are the population sizes in the tree T.

5.10 Size of a Galton-Watson Tree

Let T be a Galton-Watson tree that is either critical or subcritical. We know that if ξ is the offspring distribution and $\mathbf{Pr}(\xi = 1) < 1$, then $|T| < \infty$ almost surely. In fact, it is remarkable that the distribution of $|T|$ can be solely deduced from the distribution of ξ by a simple device discovered by Dwass (1969) and rediscovered by Kolchin (Kolchin, 1977, 1978, 1980; see 1986, p. 104).

Theorem 5.7. *For $n \geq 1$,*

$$\mathbf{Pr}(|T| = n) = \frac{\mathbf{Pr}(\xi_1 + \cdots + \xi_n = n - 1)}{n} ,$$

where $\xi_1, \xi_2, \cdots$ are i.i.d. and distributed as ξ. Let $T_1, T_2, \ldots$ be independent and distributed as T. Then, for $n \geq m \geq 0, n \geq 1$,

$$\mathbf{Pr}(|T_1| + \cdots + |T_m| = n) = \frac{m\mathbf{Pr}(\xi_1 + \cdots + \xi_n = n - m)}{n} .$$

Proof. It suffices to prove the more general statement. Clearly, if Z_1 is the number of offspring of the root of T_1, assuming $m \geq 1$, we have

$$\begin{aligned}\mathbf{Pr}(|T_1| + \cdots + |T_m| = n) &= \sum_{j=0}^{\infty} p_j \mathbf{Pr}(|T_1| + \cdots + |T_m| = n | Z_1 = j) \\ &= \sum_{j=0}^{n-m} p_j \mathbf{Pr}(|T_1| + \cdots + |T_{m+j-1}| = n-1) ,\end{aligned}$$

where $p_j = \mathbf{Pr}(\xi = j)$ and $Z_1 = \xi$ is the number of children of the root. We easily verify the Lemma for $m = 0$ and $m = 1, n = 1$ as $\mathbf{Pr}(|T| = 1) = \mathbf{Pr}(\xi_1 = 0)$. The remainder is by induction on n (for all $0 \leq m \leq n$), and we have

$$\begin{aligned}\mathbf{Pr}(|T_1| + \cdots + |T_m| = n) &= \sum_{j=0}^{n-m} p_j \mathbf{Pr}(|T_1| + |T_2| + \cdots + |T_{m+j-1}| = n-1) \\ &= \sum_{j=0}^{n-m} p_j \frac{m+j-1}{n-1} \mathbf{Pr}(\xi_1 + \xi_2 + \cdots + \xi_{n-1} = n-m-j) \\ &\quad \text{(by the induction hypothesis)} \\ &= \frac{m-1}{n-1} \mathbf{Pr}(\xi_1 + \xi_2 + \cdots + \xi_n = n-m) \\ &\quad + \frac{1}{n-1} \sum_{j=0}^{n-m} j p_j \mathbf{Pr}(\xi_1 + \xi_2 + \cdots + \xi_{n-1} = n-m-j) \\ &= \left(\frac{m-1}{n-1} + \frac{n-m}{n(n-1)}\right) \mathbf{Pr}(\xi_1 + \xi_2 + \cdots + \xi_n = n-m) \\ &\quad \text{(see below)} \\ &= \frac{m}{n} \mathbf{Pr}(\xi_1 + \xi_2 + \cdots + \xi_n = n-m) .\end{aligned}$$

We are done if we can explain the last step. But clearly,

$$\begin{aligned}\frac{n-m}{n} = \mathbf{E} \quad &(\xi_n | \xi_1 + \cdots + \xi_n = n-m) \\ &= \frac{\sum_{j=0}^{n-m} j \mathbf{Pr}(\xi_n = j) \mathbf{Pr}(\xi_1 + \cdots \xi_{n-1} = n-m-j)}{\mathbf{Pr}(\xi_1 + \cdots \xi_n = n-m)} .\end{aligned}$$

This concludes the proof of Theorem 5.7. □

Theorem 5.7 makes a crucial connection with sums of independent random variables, and for this, all is known. For example, following Kolchin (1986, p. 105), we note that if ξ has mean one (as in a critical branching process), variance σ^2 and maximal span d, when $n-1$ tends to infinity over multiples of d,

$$\mathbf{Pr}(|T| = n) \sim \frac{d}{\sqrt{2\pi} n^{3/2} \sigma} .$$

It is easily seen that $\mathbf{E}(|T|) = \infty$, a result that also follows by noting that $|T| = \sum_{i=0}^{\infty} Z_i$ and $\mathbf{E}(Z_i) = 1$ for all i.

Finally, the size of a Galton-Watson tree may also be determined by analytic methods. Let $y(s)$ be the generating function of $|T|$. Then we have

Theorem 5.8. *The generating function $y(s) = \mathbf{E}(s^{|T|})$ of $|T|$ satisfies*

$$y(s) = sf(y(s))$$

where f is the generating function of ξ in the Galton-Watson process.

Proof.

$$\begin{aligned} y(s) &= \mathbf{E}(s^{|T|}) \\ &= s\mathbf{E}\left(s^{|T_1|+\cdots+|T_\xi|}\right) \\ &= s\mathbf{E}\left(\left(\mathbf{E}\left(s^{|T_1|}\right)\right)^\xi\right) \\ &= s\mathbf{E}\left((y(s))^\xi\right) \\ &= sf(y(s)). \end{aligned}$$

□

The asymptotic form of y_n, the n-th coefficient of $y(s)$, and thus $y_n \equiv \mathbf{Pr}(|T| = n)$, may be obtained by singularity analysis (Meir and Moon, 1978; Pólya, 1937). For exact formulas, one may apply Lagrangian inversion and note that

$$y_n = \frac{1}{n} \times \text{coefficient of } u^{n-1}(f(u))^n \ .$$

See Vitter and Flajolet (1990) for more on this method, and for additional references.

5.11 Height of a Galton-Watson Tree

Let H_n be the height a Galton-Watson tree T conditional on $|T| = n$. By equivalence, we will refer to these trees by the names used in the combinatorial literature, based on the equiprobable equivalent trees thus obtained.

It is known that $\mathbf{E}(H_n) \sim \sqrt{\pi n}$ for the planted plane trees (Debruijn, Knuth and Rice, 1972), $\mathbf{E}(H_n) \sim \sqrt{2\pi n}$ for the rooted labelled trees (Cayley trees) (Rényi and Szekeres, 1967), $\mathbf{E}(H_n) \sim \sqrt{3\pi n}$ for the equiprobable unary-binary trees (Flajolet and Odlyzko, 1982), and $\mathbf{E}(H_n) \sim \sqrt{4\pi n}$ for the equiprobable binary trees (Flajolet and Odlyzko, 1982). For the last model, the expected depth of a random node is asymptotic to $\sqrt{\pi n}$ (Vitter and Flajolet, 1990). Rényi and Szekeres (1967) also computed a limit law for $H_n/\sqrt{n}$:

$$\lim_{n\to\infty} \mathbf{Pr}\left(\frac{H_n}{\sqrt{2n}} \le x\right) = \mathcal{H}(x) \ ,$$

where

$$\mathcal{H}(x) = \begin{cases} \frac{4\pi^{5/2}}{x^3} \sum_{j=1}^\infty j^2 e^{-\pi^2 j^2/x^2} \\ \sum_{j=-\infty}^\infty (1-2j^2x^2)e^{-j^2x^2} \ . \end{cases}$$

We will call $\mathcal{H}$ the theta distribution function. The theta distribution has first moment $\sqrt{\pi}$, variance $\pi(\pi-3)/3$ and general s-th moment $2\Gamma(1+s/2)(s-1)\zeta(s)$. Interestingly, the theta distribution describes the limit for all simply generated random trees. This result, due to Flajolet and Odlyzko (1982), who used analysis of singularities of generating functions in their proofs, may be formulated as follows. Let $c_0, c_1, \ldots$ define the simply generated family of ordered trees, and let

$$y(z) = z\phi(y(z)) \ ,$$

where $y(z) = \sum_{n\geq 1} y_n z^n$ and y_n is the total number of trees of size n, and $\phi(y) = \sum_{r\geq 0} c_r y^r$.

Theorem 5.9. *(Flajolet and Odlyzko, 1982). For simple families of trees corresponding to the equation $y = z\phi(y)$ and for $n = 1 \bmod d$ with $d = \gcd\{r : c_r \neq 0\}$, if we set*

$$\psi = \frac{2\phi'(\tau)^2}{\phi(\tau)\phi''(\tau)}$$

with τ the smallest positive root of the equation $\phi(\tau) - \tau\phi'(\tau) = 0$, we have

$$\frac{H_n}{\sqrt{\psi n}} \xrightarrow{\mathcal{L}} \mathcal{H}(.) \ .$$

Furthermore, all the moments of $H_n/\sqrt{\psi n}$ tend to those of $\mathcal{H}$. In particular,

$$\lim_{n\to\infty} \frac{\mathbf{E}(H_n)}{\sqrt{n}} = \sqrt{\psi\pi} \ .$$

The above result also applies to Cayley trees, even though their generating functions do not satisfy the required equality. However, if $y(z) = \sum_{n\geq 1} y_n z^n/n!$, then $y(z) = z\phi(y(z))$ with $\phi(y) = e^y$, which corresponds to the choices $c_r = 1/r!$. Combinatorists know that $ye^{-y} = z$ has a formal solution

$$y = \sum_{n=1}^{\infty} \frac{n^{n-2}}{(n-1)!} z^n \ ,$$

when $|z| \leq 1/e$ (Riordan, 1960). From this, we also obtain the number of unlabeled trees on n nodes.

By the connection of the previous section, we note that indeed, the limit law given above is applicable to random Cayley trees. In this case, we have

$$\psi = \frac{2\phi'(\tau)^2}{\phi(\tau)\phi''(\tau)} = 2$$

for any value of τ. Hence, $\mathbf{E}(H_n) \sim \sqrt{2\pi n}$, a result due to Rényi and Szekeres (1967).

5.12 Components in Random Graphs

We conclude with Karp's (1990) construction of a branching process for studying the components of random graphs. We place this material here, as it relates to sizes of extinct branching processes. Random graphs were introduced by Erdös and Rényi in 1960: we have an edge probability p, possibly depending upon n, and call $G_{n,p}$ the graph on n labeled vertices obtained by independently adding each of the $\binom{n}{2}$ possible edges with probability p.

Palmer (1985) gives a great account of the growth of $G_{n,p}$ as p increases. At least in the study of the behavior of $G_{n,p}$ for $p \leq 1/n$, thus for sparse graphs, branching processes come in handy. So we set $p = c/n$, $c \leq 1$. Around $p = 1/n$, $G_{n,p}$ undergoes a dramatic metamorphosis, as one giant component emerges which has size $\Theta(n)$ when $c > 1$. Karp's method is reconsidered in Alon, Spencer and Erdös (1992), where it is used to analyze the giant component in some detail (the case $c = 1$). We will fix $c < 1$ for simplicity.

Consider a fixed vertex u. We declare all other vertices alive, dead, or neutral. Originally, at discrete time $t = 0$, only u is alive, and all other nodes are neutral. Let Y_t be the number of live nodes at time t. We set $Y_0 = 1$. Each time unit, we take a live vertex w, and check all pairs (w, w') with w' neutral for membership in G. If (w, w') is indeed an edge, then we make w' live. after all such w' are awakened, w dies, and we declare Y_t the new number of live vertices. When there are no live vertices ($Y_t = 0$), the process terminates, and we equate $C(u)$, the component of u, as the collection of dead vertices. Clearly, we have

$$Y_t = Y_{t-1} + Z_t - 1 \ .$$

Each neutral w' has independent probability p of becoming live, and no pair (w, w') is ever examined twice, so that the conditional probability of the existence of edge (w, w') is always p. As $t-1$ vertices are dead and Y_{t-1} live, it is easy to see that

$$Z_t \stackrel{\mathcal{L}}{=} B(n - (t-1) - Y_{t-1}, p)$$

where $B(.,.)$ denotes the binomial distribution. Let T be the smallest t for which $Y_t = 0$, the time of extinction. Also, $T = |C(u)|$. We continue this definition recursively, and note that for all t,

$$Y_t \stackrel{\mathcal{L}}{=} B(n-1, 1-(1-p)^t) + 1 - t \ .$$

Proof. Define $N_t = n - t - Y_t$, the number of neutral vertices at time t. We will show that $N_t \stackrel{\mathcal{L}}{=} B(n-1, (1-p)^t)$. Clearly, $N_0 = n - 1$. We argue by induction, and note that

$$\begin{aligned} N_t &= n - t - Y_t \\ &= n - t - B(n-(t-1)-Y_{t-1}, p) - Y_{t-1} + 1 \\ &= N_{t-1} - B(N_{t-1}, p) \\ &= B(N_{t-1}, 1-p) \ . \end{aligned}$$

□

The property above is valid for all p. For $p = c/n$, when t and Y_{t-1} are small, the binomial law is close to a Poisson law with mean c. So, Z_t is close to $B(n, c/n)$, which is close to $P(c)$, a Poisson random variable with mean c. Thus, roughly speaking, the component grows at u like a branching process with offspring distributed as $P(c)$. For fixed c, let $Y_0^*, Y_1^*, \ldots, T^*, Z_1^*, Z_2^*, \ldots,$ refer to the $P(c)$ branching process, and let the unstarred random variables refer to the random graph process. More precisely, the branching process starts with one live individual, so that $Y_0^* = 1$, and at each time unit, one live individual is selected at random. It produces a $P(c)$ number of children, and then dies, so that

$$Y_t^* = Y_{t-1}^* + Z_t^* - 1$$

where $Z_1^*, Z_2^*, \ldots$ are i.i.d. $P(c)$ random variables. Let T^* be the least t for which $Y_t^* = 0$. If no such t exists, we say that $T^* = \infty$. From Theorem 1.1, if $\mathbf{E}(P(c)) = c < 1$, with probability one, the process dies out, so that $T^* < \infty$ almost surely.

Let $\mathcal{H}, \mathcal{H}^*$ denote the histories of the processes up to time t, that is, $\mathcal{H} = (Z_1, \ldots, Z_t)$ and $\mathcal{H}^* = (Z_1^*, \ldots, Z_t^*)$. Then

$$\mathbf{Pr}(\mathcal{H}^* = (z_1, \ldots, z_t)) = \prod_{i=1}^{t} \mathbf{Pr}(P(c) = z_i)$$

and

$$\mathbf{Pr}(\mathcal{H} = (z_1, \ldots, z_t)) = \prod_{i=1}^{t} \mathbf{Pr}(Z_i = z_i) \ ,$$

where Z_i is binomial $B(n - 1 - z_1 - \cdots - z_{i-1}, c/n)$. If $m \sim n$ and c and i are fixed, we have

$$\mathbf{Pr}(B(m, c/n) = i) \to \frac{e^{-c} c^i}{i!}$$

as $n \to \infty$. This may be used to show that

$$\lim_{n \to \infty} \mathbf{Pr}(\mathcal{H} = (z_1, \ldots, z_t)) = \mathbf{Pr}(\mathcal{H}^* = (z_1, \ldots, z_t)) \ .$$

Thus, for any fixed t, $\lim_{n\to\infty} \mathbf{Pr}(T = t) = \mathbf{Pr}(T^* = t)$. This may be used naively in two ways. First of all, T^* is the total size of a $P(c)$ Galton-Watson process. Therefore, as $n \to \infty$,

$$|C(u)| \xrightarrow{\mathcal{L}} T^* \ .$$

From Theorem 5.8, the generating function for $P(c)$ is $f(s) = e^{c(s-1)}$, while the generating function $y(s)$ for T^* is the solution of $y = f(sy)$, i.e., of

$$y = e^{c(sy-1)} \ .$$

This describes the asymptotic distribution of the size of $C(u)$ in its entirety.

Secondly, if we consider $C_n = \max_u |C(u)|$ over the nodes u of $G_{n,c/n}$, then we can easily prove the known result (see Palmer, 1985) that $\mathbf{Pr}(C_n > \beta \log n) = o(1)$ for some $\beta > 0$. To see this, observe that for any t, and for $h > 0$, by Chernoff's bounding method,

$$\begin{aligned}\mathbf{Pr}(T>t) &\le \mathbf{Pr}(Y_t>0)=\mathbf{Pr}(B(n-1,1-(1-p)^t)\ge t)\\ &\le \mathbf{Pr}(B(n,tc/n)\ge t)\le \mathbf{E}\left(e^{hB(n,tc/n)-ht}\right)\\ &= e^{-ht}\left(1+(e^h-1)tc/n\right)^n \le e^{-t(h-(e^h-1)c)}\\ &= e^{-t(\log(1/c)-(1-c))} \qquad (\text{take } h=\log(1/c))\\ &\stackrel{\text{def}}{=} e^{-\alpha t}\ .\end{aligned}$$

Thus,

$$\mathbf{Pr}(C_n>\beta\log n)\le ne^{-\alpha\beta\log n}=n^{1-\alpha\beta}\to 0$$

if we pick $\beta > 1/\alpha = 1/(\log(1/c)-(1-c))$.

We leave it as an interesting exercise to show that the $P(c)$ branching process of this section, with $c>1$, conditional on extinction, has the same distribution as the (unconditional) $P(c')$ branching process, where $c'=cq$, and q is the extinction probability of the $P(c)$ branching process, that is, $q=e^{c(q-1)}$. (Note that $ce^{-c}=c'e^{-c'}$.) This fact is used in Alon, Spencer and Erdös (1992) to show for example that the structure of $G_{n,c/n}$ with the giant component removed is fundamentally that of $G_{m,c'/m}$ (without any removals), where m, the number of vertices not in the giant component, satisfies $m \sim ny$.

5.13 Bibliographic Remarks

Meir and Moon (1978) studied the expected depth $\mathbf{E}(D_n)$ from root to nodes in simply generated random trees, and showed that $\mathbf{E}(D_n)/\sqrt{n}\to c$, where c is again a constant only depending upon the species of tree. The work of Flajolet and Odlyzko (1982) is continued by Gutjahr (1993), who derives asymptotics for expected values of various other tree parameters such as the number of nodes at level k and the total path length. Even tree models with trees of given size and height are considered there. The branching process approach was used by Kennedy (1975) (see also Kolchin, 1986) to obtain the limit law for $Z_{\lfloor\sqrt{n}t\rfloor}/(\sqrt{n}t)$ conditional on $N=n$ as $n\to\infty$, where Z_k is the size of the k-th generation. Thus, the bulk of the points is indeed at distance $\Theta(\sqrt{n})$ from the root. Finally, one might study the height of random binary trees, where each edge has an independent length drawn from a fixed distribution on the positive halfline. Height is then defined as the maximal sum of edge lengths of any path to the root. For the exponential distribution, Gupta, Mesa and Waymire (1990) showed that this height satisfies the same limit law as the standard height modulo a constant multiplicative factor. Their proof uses convergence of all moments.

References

1. Abramowitz M. (1970): and I. A. Stegun, Handbook of Mathematical Tables, Dover Publications, New York, NY

2. Aho A. V. (1983): J. E. Hopcroft, and J. D. Ullman, Data Structures and Algorithms, Addison-Wesley, Reading, MA.
3. Aldous D. (1988): The random walk construction of uniform spanning trees and uniform labelled trees SIAM Journal of Discrete Mathematics, **0**, 0–0.
4. Aldous D. (1991): The continuum random tree II: an overview, in: Proceedings Durham Symposium on Stochastic Analysis, ed. M. T. Barlow and N. H. Bingham, 23–70, Cambridge University Press, Cambrige, UK.
5. Aldous D. (1991): The continuum random tree I, Annals of Probability, **19**, 1–28.
6. Aldous D. (1983): The continuum random tree III, Annals of Probability, **21**, 248–289.
7. Aldous D. (1993): Probability distributions on cladograms, Technical Report, Institute of Mathematics and Applications, University of Minnesota.
8. Aldous D., Flannery B. and Palacios J.L. (1988): Two applications of urn processes: the fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains, Probability in the Engineering and Information Sciences, **2**, 293–307.
9. Alon N. (1992): J. H. Spencer, and P. Erdös, The Probabilistic Method, Wiley, New York.
10. Arkin E., Held M., Mitchell J. and Skiena S. (1994): Hamiltonian triangulations for fast rendering, in: Algorithms–ESA'94, ed. J. van Leeuwen, **855**, 36–47, Lecture Notes in Computer Science, Springer-Verlag.
11. Asmussen S. and Hering H. (1983): Branching processes, Birkhäuser Verlag, Basel.
12. Athreya K.B. and Ney P.E. (1972): Branching Processes, Springer Verlag, Berlin.
13. Avis D. and Gindy H.E. (1987): Triangulating point sets in space, Discrete Computational Geometry, **2**, 99–111.
14. Bahadur R.R. and Rao R.R. (1960): On deviation of the sample mean, Annals of Mathematical Statistics, **31**, 1015–1027.
15. Bell C.J. (1965): An investigation into the principles of the classification and analysis of data of an automatic digital computer, Doctoral Dissertation, Leeds University.
16. Bellman R. and Harris T.E. (1952): On age-dependent binary branching processes, Annals of Mathematics, **55**, 280–295.
17. Bergeron F., Flajolet P., and Salvy B. (1992): Varieties of increasing trees, in: CAAP 92, ed. J.-C. Raoult, **581**, 24–48, Lecture Notes in Computer Science, Springer-Verlag.
18. Biggins J.D. (1976): The first and last-birth problems for a multitype age-dependent branching process, Advances in Applied Probability, **8**, 446–459.
19. Biggins J.D. (1976): Asymptotic properties of the branching random walk, Ph.D. Thesis, University of Oxford.
20. Biggins J.D. (1977): Martingale convergence in the branching random walk, Journal of Applied Probability, **14**, 25–37.
21. Biggins J.D. (1977): Chernoff's theorem in the branching random walk, Journal of Applied Probability, **14**, 630–636.
22. Biggins J.D. (1978): The asymptotic shape of the branching random walk, Advances in Applied Probability, **10**, 62–84.
23. Biggins J.D. (1979): Growth rates in the branching random walk, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, **48**, 17–34.
24. Biggins J.D. (1990): The central limit theorem for the supercriticcal branching random walk, and related results, Stochastic Processes and theor Applications, **34**, 255–274.

25. Biggins J.D. (1995): The growth and spread of the general branching random walk, Annals of Applied Probability, **5**, 1008–1024.
26. Biggins J.D. (1996): How fast does a general branching random walk spread?, in: Classical and Modern Branching Processes, **84**, 19–40, IMA Volumes in Mathematics and its Applications, Springer-Verlag, New York.
27. Biggins J.D. and Bingham N.H. (1993): Large deviations in the supercritical branching process, Advances in Applied Probability, **25**, 757–772.
28. Biggins J.D. and Grey D.R. (1996): A note on the growth of random trees, Technical Report, School of Mathematics and Statistics, University of Sheffield, Sheffield, UK.
29. Bingham N. (1988): On the limit of a supercritical branching process, Journal of Applied Probability, **25** A, 215–228.
30. Bramson M. (1983): Convergence of solutions of the Kolmogorov equation to travelling waves, Mem. American Statistical Society, **285**, 1–190.
31. Bramson M.D. (1978): Maximal displacement of branching Brownian motion, Communications on Pure and Applied Mathematics, **21**, 531–581.
32. Bramson M.D. (1978): Minimal displacement of branching random walk, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, **45**, 89–108.
33. Bramson M., Durrett R. and Swindle G. (1989): Statistical mechanics of crabgrass, Annals of Probability, **17**, 444–481.
34. Brown C.A. and Purdom P.W. (1981): An average time analysis of backtracking, SIAM Journal of Computing, **10**, 583–593.
35. Bruijn N.G. de, Knuth D.E. and Rice S.O. (1972): The average height of planted plane trees, in: Graph Theory and Computing, ed. R.-C. Read, 15–22, Academic Press, New York.
36. Cayley A. (1858): On the analytical forms called trees, Philosophical Magazine, **28**, 374–378.
37. Cayley A. (1889): A theorem on trees, Quarterly Journal of Pure and Applied Mathematics, **23**, 376–378.
38. Chernoff H. (1952): A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Annals of Mathematical Statistics, **23**, 493–507.
39. Coffman E.G. and Eve J. (1970): File structures using hashing functions, Communications of the ACM, **13**, 427–436.
40. Crump K.S. and Mode C.J. (1968): A general age-dependent branching process, Journal of Mathematical Analysis and its Applications, **24**, 494–508.
41. Darling D.A. (1970): The Galton-Watson process with infinite mean, Journal of Applied Probability, **7**, 455–456.
42. Dekking F.M. and Host B. (1990): Limit distributions for minimal displacement of branching random walks, Probability Theory and Related Fields, **90**, 403–426.
43. Derrida B. and Spohn H. (1988): Polymers on disordered trees, spin glasses, and traveling waves, Journal of Statistical Physics, **51**, 817–841.
44. Devroye L. (1986): A note on the height of binary search trees, Journal of the ACM, **33**, 489–498.
45. Devroye L. (1986a): Non-Uniform Random Variate Generation, Springer-Verlag, New York.
46. Devroye L. (1987): Branching processes in the analysis of the heights of trees, Acta Informatica, **24**, 277–298.
47. Devroye L. (1988): Applications of the theory of records in the study of random trees, Acta Informatica, **26**, 123–130.
48. Devroye L. (1990): On the height of random m-ary search trees, Random Structures and Algorithms, **1**, 191–203.

49. Devroye L. (1993): On the expected height of fringe-balanced trees, Acta Informatica, **30**, 459–466.
50. Devroye L. (1997): Universal limit laws for depths in random trees, SIAM Journal on Computing, to appear.
51. Devroye L. and Kamoun O. (1996): Random minimax game trees, in: Random Discrete Structures, ed. D. Aldous and R. Pemantle, 55–80, John Wiley, New York.
52. Devroye L. and Laforest L. (1990): An analysis of random d-dimensional quadtrees, SIAM Journal on Computing, **19**, 821–832.
53. Devroye L. and Reed B. (1995): On the variance of the height of random binary search trees, SIAM Journal on Computing, **24**, 1157–1162.
54. Devroye L. and Zamora-Cura C. (1997): On the complexity of branch-and-bound search for random trees, Technical Report, McGill University.
55. Dharmadhikari S.W. and Jogdeo K. (1969): Bounds on moments of certain random variables, Annals of Mathematical Statistics, **40**, 1506–1508.
56. Drmota M. (1997): Analytic approach to the height of the binary search tree, Technical Report, University of Vienna.
57. Durrett R. (1979): Maxima of branching random walks versus independent random walks, Stochastic Processes and Their Applications, **9**, 117–135.
58. Durrett R. (1991): Probability: Theory and Examples, Wadsworth and Brooks, Pacific Grove, CA.
59. Dwass M. (1969): The total progeny in a branching process, Journal of Applied probability, **6**, 682–686.
60. Erdös P. and Rényi A. (1960): On the evolution of random graphs, Magyar Tud. Akad. Mat. Kut. Int. Közl, **5**, 17–61.
61. Feller W. (1971): An Introduction to Probability Theory and its Applications, Volume 2, John Wiley, New York.
62. Finkel R.A. and Bentley J.L. (1974): Quad trees: a data structure for retrieval on composite keys, Acta Informatica, **4**, 1–9.
63. Flajolet P., Gonnet G., Puech C. and Robson J.M. (1990): The analysis of multidimensional searching in quad-trees, in: Proceedings of the Second Annual ACM-SIAM Symposium on Discrete Algorithms, 100–109, ACM, New York, and SIAM, Philadelphia.
64. Flajolet P. and Lafforgue L. (1994): Search costs in quadtrees and singularity perturbation analysis, Discrete and Computational Geometry, **12**, 151–175.
65. Flajolet P. and Odlyzko A. (1982): The average height of binary trees and other simple trees, Journal of Computer and System Sciences, **25**, 171–213.
66. Flajolet P. and Odlyzko A. (1990): Singularity analysis of generating functions, SIAM Journal on Discrete Mathematics, **3**, 216–240.
67. Flajolet P. and Sedgewick R. (1986): Digital search trees revisited, SIAM Journal on Computing, **15**, 748–767.
68. Fredkin E.H. (1960): Trie memory, Communications of the ACM, **3**, 490–500.
69. Fuk D.K. and Nagaev S.V. (1961): Probability inequalities for sums of independent random variables, Theory of Probability and its Applications, **16**, 643–660.
70. Le Gall F.J. (1989): Marches aléatoires, mouvement Brownien et processus de branchement, in: Séminaire de Probabilités XXIII, ed. J. Azéma, P. A. Meyer and M. Yor, **1372**, 258–274, Lecture Notes in Mathematics, Springer-Verlag, Berlin.
71. Gonnet G.H. and Baeza-Yates R. (1991): Handbook of Algorithms and Data Structures, Addison-Wesley, Workingham, England.
72. Gradshteyn I.S. and Ryzhik I.M. (1980): Table of Integrals, Series and Products, Academic Press, New York.

73. Grimmett G.R. (1980): Random labelled trees and their branching networks, Journal of the Australian Mathematical Society series A, **30**, 299–237.
74. Grimmett G.R. and Stirzaker D.R. (1992): Probability and Random Processes, Oxford University Press.
75. Gupta, V.K., Mesa O.J. and Waymire E. (1990): Tree-dependent extreme values: the exponential case, Journal of Applied Probability, **27**, 124–133.
76. Gutjahr W. (1992): The variance of level numbers in certain families of trees, Random Structures and Algorithms, **3**, 361–374.
77. Gutjahr W. (1993): Expectation transfer between branching processes and random trees, Random Structures and Algorithms, **4**, 447–467.
78. Gutjahr W. and Pflug G.C. (1992): The asymptotic contour process of a binary tree is a Brownian excursion, Stochastic Processes and theoir Applications, **41**, 69–89.
79. Gutjahr W. and Pflug G.C. (1992): The limiting common distribution of two leaf heights in a random binary tree, Theoretical Informatics and Applications, **26**, 1–18.
80. Gutjahr W. and Pflug G.C. (1992): Average execution times of series-parallel networks, Technical Report, University of Vienna.
81. Gutjahr W. and Pflug G.C. (1992): The asymptotic distribution of leaf heights in binary trees, Graphs and Combinatorics, **8**, 243–251.
82. Hammersley J.M. (1974): Postulates for subadditive processes, Annals of Probability, **2**, 652–680.
83. Harris T.E. (1963): The Theory of Branching Processes, Springer Verlag, Berlin.
84. Hawkes J. (1981): Trees generated by a simple branching process, Journal of the London Mathematical Society, **24**, 373–384.
85. Heathcote C.R., Seneta E. and Vere-Jones D. (1967): A refinement of two theorems in the theory of branching processes, Theory of Probability and its Applications, **12**, 297–301.
86. Heyde C.C. (1970): A rate of convergence result for the super-critical Galton-Watson process, Journal of Applied Probability, **7**, 451–454.
87. Heyde C.C. (1971): Some central limit analogues for super-critical Galton-Watson processes, Journal of Applied Probability, **8**, 52–59.
88. Heyde C.C. (1971): Some almost sure converence theorems for branching processes, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, **20**, 189–192.
89. Heyde C.C. and Brown B.M. (1971): An invariance principle and some convergence rate results for branching processes, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, **20**, 271–278.
90. Jabbour J. (1998): Personal communication.
91. Jacquet P. and Régnier M. (1986): Trie partitioning process: limiting distributions, in: Lecture Notes in Computer Science, **214**, 196–210.
92. Jagers P. (1975): Branching Processes with Biological Applications, John Wiley, New York.
93. Jagers P. and Nerman O. (1984): The growth and composition of branching populations, Advances in Applied Probability, **16**, 221–259.
94. Janson S. (1983): Limit theorems for certain branching random walks on compact groups and homogeneous spaces, Annals of Probability, **11**, 909–930.
95. Joffe A. and Waugh W.A.O.N. (1982): Exact distributions of kin numbers in a Galton-Watson process, Journal of Applied Probability, **19**, 767–775.
96. Karp R.M. (1990): The transitive closure of a random digraph, Random Structures and Algorithms, **1**, 73–93.

97. Karp R.M. and Pearl J. (1983): Searching for an optimal path in a tree with random costs, Artificial Intelligence, **21**, 99–117.
98. Karp R.M. and Zhang Y. (1995): Bounded branching process and AND/OR tree evaluation, Random Structures and Algorithms, **7**, 97–116.
99. Kemp R. (1984): Fundamentals of the Average Case Analysis of Particular Algorithms, B.G.Teubner, Stuttgart.
100. Kendall D.G. (1966): Branching processes since 1873, Journal of the London Mathematical Society, **41**, 385–406.
101. Kennedy D.P. (1975): The Galton-Watson process conditioned on the total progeny, Journal of Applied Probability, **12**, 800–806.
102. Kesten H., Ney P. and Spitzer F. (1966): The Galton-Watson process with mean one and finite variance, Theory of Probability and its Applications, **11**, 513–540.
103. Kesten H. and Stigum B.P. (1966): A limit theorem for multidimensional Galton-Watson processes, Annals of Mathematical Statistics, **37**, 1211–1223.
104. Kingman J.F.C. (1973): Subadditive ergodic theory, Annals of Probability, **1**, 883–909.
105. Kingman J.F.C. (1975): The first-birth problem for an age-dependent branching process, Annals of Probability, **3**, 790–801.
106. Knuth D.E. (1973): The Art of Computer Programming, Vol. 3 : Sorting and Searching, Addison-Wesley, Reading, Mass.
107. Kolchin V.F. (1978): Moment of degeneration of a branching process and height of a random tree, Mathematical Notes of the Academy of Sciences of the USSR, **6**, 954–961.
108. Kolmogorov A.N. (1938): Zur Lösung einer biologischen Aufgabe, **2**, 1–6, Issledovatelskogo instituta matematiki i mechaniki pri Tomskom Gosudarstvennom universitete, Izvestiya nauchno.
109. Kumar V. (1992): Search, branch and bound, in: Encyclopedia of Artificial Intelligence (2nd edition), ed. S. C. Shapiro, 1468–1472, Wiley-Interscience.
110. Le Gall, F.J. (1989): Brownian excursion, trees and measure-valued branching processes, Technical Report, Université Pierre et Marie Curie, Paris.
111. Louchard G. (1987): Exact and asymptotic distributions in digital and binary search trees, Theoretical Informatics and Applications, **21**, 479–497.
112. Lynch W.C. (1965): More combinatorial problems on certain trees, Computer Journal, **7**, 299–302.
113. Lyons R. (1997): Probability and Trees, in press.
114. Lyons R., Pemantle R. and Peres Y. (1993): When does a branching process grow like its mean? Conceptual proofs of L log L criteria, Technical Report, Indiana University.
115. Lyons R., Pemantle R. and Peres Y. (1995): Conceptual proofs of L log L criteria for mean behavior of branching processes, Annals of Probability, **23**, 1125–1138.
116. Mahmoud H. (1993): Distances in plane-oriented recursive trees, Journal of Computers and Applied Mathematics, **41**, 237–245.
117. Mahmoud H.M. (1986): On the average internal path length of m-ary search trees, Acta Informatica, **23**, 111–117.
118. Mahmoud H.M. (1992): Evolution of Random Search Trees, John Wiley, New York.
119. Mahmoud H.M. (1994): A strong law for the height of random binary pyramids, Annals of Applied Probability, **4**, 923–932.
120. Mahmoud H.M. and Pittel B. (1988): On the joint distribution of the insertion path length and the number of comparisons in search trees, Discrete Applied Mathematics, **20**, 243–251.

121. Mahmoud H. and Pittel B. (1984): On the most probable shape of a search tree grown from a random permutation, SIAM Journal on Algebraic and Discrete Methods, **5**, 69–81.
122. Mahmoud H., Smythe R.T. and Szymański J. (1993): On the structure of random plane-oriented recursive trees and their branches, Random Structures and Algorithms, **4**, 151–176.
123. Marcinkiewicz J. and Zygmund A. (1937): Sur les fonctions indépendantes, Fundamentales de Mathématiques, **29**, 60–90.
124. McDiarmid C.J.H. (1990): Probabilistic analysis of tree search, in: Disorder in Physical Systems, G.R. Grimmett and D.J.A. Welsh, editors, 249–260, Oxford Science Publications.
125. McDiarmid C.J.H. (1995): Minimal positions in a branching random walk, Annals of Applied Probability, **5**, 128–139.
126. McDiarmid C.J.H. and Provan G.M.A. (1991): An expected-cost analysis of backtracking and non-backtracking algorithms, in: IJCAI-91: Proceedings of the Twelfth International Conference on Artificial Intelligence, 172–177, Morgan Kaufmann Publishing, San Mateo, CA.
127. Meir A. and Moon J.W. (1970): The distance between points in random trees, Journal of Combinatorial theory, **8**, 99–103.
128. Meir A. and Moon J.W. (1978): On the altitude of nodes in random trees, Canadian Journal of Mathematics, **30**, 997–1015.
129. Moon J.W. (1970): Counting labelled trees, Canadian Mathematical Congress.
130. Moon J.W. (1973): Random walks on random trees, Journal of the Australian Mathematical Society, **15**, 42–53.
131. Nagaev S.V. and Pinelis N.F. (1977): Some inequalities for sums of independent random variables, Theory of Probability and its Applications, **22**, 248–256.
132. Nerman O. (1981): On the convergence of the supercritical general (C-M-J) branching process, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, **57**, 365–395.
133. Neveu J. (1986): Arbres et processus de Galton-Watson, Annales de l'Institut Henri Poincaré, **22**, 199–207.
134. Neveu J. and Pitman J.W. (1989): The branching process in a Brownian excursion, in: Séminaire de Probabilités XXIII, ed. J. Azéma, P. A. Meyer and M. Yor, **1372**, 248–257, Lecture Notes in Mathematics, Springer-Verlag, Berlin.
135. Nievergelt J. and Hinrichs K.H. (1993): Algorithms and Data Structures with Applications to Graphics and Geometry, Prentice-Hall, Englewood Cliffs, NJ.
136. Nievergelt J., Hinterberger H. and Sevcik K.C. (1984): The grid file: an adaptable, symmetric multikey file structure, ACM Transactions on Database Systems, **9**, 38–71.
137. Okamoto M. (1958): Some inequalities relating to the partial sum of binomial probabilities, Annals of Mathematical Statistics, **10**, 29–35.
138. Pakes A.G. (1971): Some limit theorems for the total progeny of a branching process, Advances in Applied Probability, **3**, 176–192.
139. Palmer E.M. (1985): Graphical Evolution, John Wiley, New York.
140. Pearl J. (1984): Heuristics: Intelligent Search Strategies for Computer Problem Solving, Addison-Wesley, Reading, MA.
141. Petrov V.V. (1975): Sums of Independent Random Variables, Springer-Verlag, Berlin.
142. Pittel B. (1984): On growing random binary trees, Journal of Mathematical Analysis and its Applications, **103**, 461–480.

143. Pittel B. (1985): Asymptotical growth of a class of random trees, Annals of Probability, **13**, 414–427.
144. Pittel B. (1985): Paths in a random digital tree: limiting distributions, Advances in Applied Probability, **18**, 139–155.
145. Pittel B. (1994): Note on the heights of random recursive trees and random m-ary search trees, Random Structures and Algorithms, **5**, 337–347.
146. Poblete P.V. and Munro J.I. (1985): The analysis of a fringe heuristic for binary search trees, Journal of Algorithms, **6**, 336–350.
147. Prusinkiewicz P. and Lindenmayer A. (1990): The Algorithmic Beauty of Plants, Springer-Verlag, New York.
148. Purdom P.W. (1983): Search rearrangement backtracking and polynomial average time, Artificial Intelligence, **21**, 117–133.
149. Pyke R. (1965): Spacings, Journal of the Royal Statistical Society Series B, **7**, 395–445.
150. Reingold E.M., Nievergelt J. and Deo N. (1977): Combinatorial Algorithms: Theory and Practice, Prentice Hall, Englewood Cliffs, N.J.
151. Riordan J. (1960): The enumeration of trees by height and diameter, IBM Journal of research and develoment, **4**, 473–478.
152. Robson J.M. (1979): The height of binary search trees, The Australian Computer Journal, **11**, 151–153.
153. Robson J.M. (1982): The asymptotic behaviour of the height of binary search trees, Australian Computer Science Communications, p. 88.
154. Robson J.M. (1997): Bounds on the variation of binary search tree heights, Technical Report, Université Bordeaux II.
155. Rubistein R.Y. (1982): Generating random vectors uniformly distributed inside and on the surface of different regions, European Journal of Operations Research, **10**, 205–209.
156. Rényi A. (1959): Some remarks on the theory of trees, MTA Mat. Kut. Int. Közl, **4**, 73–85.
157. Rényi A. and Szekeres G. (1967): On the height of trees, Journal of the Australian Mathematical Society, **7**, 497–507.
158. Samet H. (1990): Applications of Spatial Data Structures, Addison-Wesley, Reading, MA.
159. Samet H. (1990): The Design and Analysis of Spatial Data Structures, Addison-Wesley, Reading, MA.
160. Sedgewick R. (1983): Mathematical analysis of combinatorial algorithms, in: Probability Theory and Computer Science, ed.G. Louchard and G. Latouche, 123–205, Academic Press, London.
161. Seneta E. (1969): Functional equations and the Galton-Watson process, Advances in Applied Probability, **1**, 1–42.
162. Sibuya M. (1979): Generalized hypergeometric, digamma and trigamma distributions, Annals of the Institute of Statistical Mathematics, **31**, 373–390.
163. Smith D.R. (1984): Random trees and the analysis of branch and bound procedures, Journal of the ACM, **31**, 163–188.
164. Smith R.L. (1984): Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions, Operations Research, **32**, 1296–1308.
165. Stepanov V.E. (1969): On the distribution of the number of vertices in strata of a random tree, Theory of Probability and its Applications, **14**, 65–78.
166. Stone H.S. and Sipala P. (1986): The average complexity of depth-first search with backtracking and cutoff, IBM Journal of Research and Development, **30**, 242–258.

167. Szpankowski W. (1988): Some results on V-ary asymmetric tries, Journal of Algorithms, **9**, 224–244.
168. Szymański J. (1988): On the nonuniform random recursive tree, Annals of Discrete Mathematics, **33**, 297–307.
169. Timofeev E.A. (1984): Random minimal trees, Theory of Probability and its Applications, **29**, 134–141.
170. Timofeev E.A. (1988): On finding the expected length of a random minimal tree, Theory of Probability and its Applications, **33**, 361–365.
171. Viennot X.V. (1990): Trees everywhere, in: CAAP 90, ed. A. Arnold, **431**, 18–41, Lecture Notes in Computer Science, Springer-Verlag, Berlin.
172. Vitter J.S. and Flajolet P. (1990): Average-case analysis of algorithms and data structures, in: Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity, ed. J. van Leeuwen, 431–524, MIT Press, Amsterdam.
173. Wah B.W. and Yu C.F. (1985): Stochastic modeling of branch-and-bound algorithms with best-first search, IEEE Transactions of Software Engineering, **SE-11**, 922–934.
174. Walker A. and Wood D. (1976): Locally balanced binary trees, Computer Journal, **19**, 322–325.
175. Weiner H. (1984): Moments on the maximum in a critical branching process, Journal of Applied Probability, **21**, 920–923.
176. Yaglom A.M. (1947): Certain limit theorems of the theory of branching processes, Dokl. acad. nauk SSSR, **56**, 795–798.
177. Zhang W. and Korf R.E. (1992): An average-case analysis of branch-and-bound with applications, in: Proceedings of the 10th National Conference on AI—AAAI-92, 1–6, San Jose, CA.