Search Trees and Branching Random Walks

Louigi Addario-Berry

March 8, 2010

Contents

Lecture 1

Random Binary Search Trees

1.1 Binary search trees

Binary search trees are a classic data structure, introduced in a primordial form by A.I. Dumey in 1952[1]. In this model, we imagine that we are given n items of data, indexed by distinct, real-valued *keys* $u_1, \ldots, u_n$. From this initial data, we then construct a rooted, labelled binary tree with n nodes, in the following recursive fashion. In what follows, we interpret a list of length zero as the empty set.

> BST$(u_1, \ldots, u_n)$
>
> 1. If the input is empty then output a single external node.
> 2. Otherwise, let the root have label u_1.
> 3. Let $i_1 < \ldots < i_k$ be the indices i with $u_i < u_1$, and let $j_1 < \ldots < j_{n-1-k}$ be the indices j with $u_j > u_1$.
> 4. Let the root have as left child the root of BST$(u_{i_1}, \ldots, u_{i_k})$ and as right child the root of BST$(u_{j_1}, \ldots, u_{j_k})$.

The binary search tree BST$(4, 1, 3, 7, 5, 9, 6, 2)$ is shown in Figure 1.1. (The dashed lines and rectangles will be explained momentarily.)

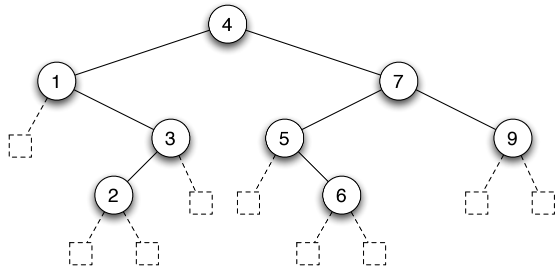


Figure 1.1: Bst example

[1]Knuth III, page 453

We will often write $\mathbf{u} = (u_1, \ldots, u_n)$, and write $\text{BST}(\mathbf{u})$ for $\text{BST}(u_1, \ldots, u_n)$. As will be discussed below, both practical and theoretical considerations motivate a sequential construction of binary search trees, which we describe below. It is for this construction that the dashed lines and rectangles of Figure 1.1 – which represent "edges to external nodes" and "external nodes", respectively – come in handy. An "external node" is a place where another node-key pair could appear in some binary search tree. For example, $\text{BST}(4, 1, 3, 7, 5, 9, 6, 2, 8)$ would have a node with key 8 in place of the leftmost of the two external nodes below 9. By "nodes", we usually mean "internal nodes," i.e., nodes with keys, but will be explicit when there is potential ambiguity.

Suppose we are given a binary search tree T with $n \geq 0$ internal nodes (we think of a binary search tree with no internal nodes as consisting of a single, external node). Given a new key u – which we assume is distinct from all keys already in T – we may use the following procedure to insert u into T, yielding a new binary search tree with $n + 1$ internal nodes.

> $\text{INSERT}(T, u)$
>
> 1. If T is empty then output a tree a single node of label u.
> 2. Otherwise, write u^* for the label of the root of T.
> 3. If $u < u^*$ then insert u into the left subtree of the root of T. In other words, replace the subtree T_ℓ of T by the output of $\text{INSERT}(T_\ell, u)$, then output the resulting tree.
> 4. If $u > u^*$ then insert u into the right subtree of the root of T.

Using this insertion procedure, we can provide a sequential description of the construction of $\text{BST}(u_1, \ldots, u_n)$.

> $\text{BST}(u_1, \ldots, u_n)$
>
> 1. Let $T_0 = \emptyset$.
> 2. If the input is empty then output T_0.
> 3. Otherwise, for $i = 1, \ldots, n$, let $T_i = \text{INSERT}(u_i, T_{i-1})$, and output T_n.

In the world of data structures, binary search trees are commonly built by using representing edges of the resulting tree as pointers from parent node to child node. Each internal node holds a key u and a pointer to the data with index u. The entire tree is then accessed via a single pointer to its root. For this to be a useful data structure, we must be able to retrieve data from the tree when told a key value. The following procedure achieves this.

> $\text{RETRIEVE}(T, u)$
>
> 1. If T is empty then the data is not in the tree; output $\emptyset$.
> 2. Otherwise, write u^* for the label of the root of T.
> 3. If $u = u^*$ then output the data with index u.
> 4. If $u < u^*$ then output the result of $\text{RETRIEVE}(T_\ell, u)$.
> 5. If $u > u^*$ then output the result of $\text{RETRIEVE}(T_r, u)$.

The the time taken when retrieving the data corresponding to key u is proportional to the number of recursive calls to RETRIEVE, which is precisely the *depth* of u in T, i.e., the number of edges on the path from the root to the node with label u. Thus, for a given binary search tree T, the worst-case retrieval time for any data item is proportional to the *height* of the tree T, defined as the greatest depth of any node in T. This observation is at the root of why researchers in analysis of algorithms are interested in the behaviour of heights of binary (and other) trees.

The permutation model

The behaviour of the above construction procedures depends only on the *order statistics* of $u_1, \ldots, u_n$, and not their precise values. In other words, let $\sigma = \sigma(\mathbf{u}) = (\sigma_1, \ldots, \sigma_n)$ be the vector where for each $i, j \in [n]$, $\sigma_i = j$ precisely if u_i has the j'th largest value among $u_1, \ldots, u_n$. Then σ is a permutation of $[n]$, and $\text{BST}(\mathbf{u})$ and $\text{BST}(\sigma)$ are identical. We may thus always view the input to BST as a permutation of $[n]$, and will do so when it is useful.

Aside: counting binary search trees. Let T be a binary search tree with n nodes labelled by $[n]$. Then the value of the key at the root is precisely the size of the left subtree of T, plus one. Applying this fact recursively, it follows that to determine the structure of T, it suffices to keep track of the size of the left subtree at each node. Now write bst_n for the number of distinct binary search trees on $[n]$. Then the following recursive equation holds.

$$bst_n = \sum_{i=1}^{n-1} (bst_i + bst_{n-1-i}).$$

Exercises

In all the below exercises, let T denote a binary tree with n vertices.

1.1.1 Let each node v of T be labelled with ℓ_v, the number of nodes in the subtree rooted at its left child. Now let v be any vertex in T. Explain how to determine the size n_v of the subtree rooted at v.

1.1.2 Find an explicit formula for t_n.

1.1.3 Show that T has height at least $\lfloor \log_2 n \rfloor$ and at most $n-1$, and that these bounds are tight.

1.1.4 Suppose that $n = 2^k - 1$. Show that the number N of permutations σ of $[n]$ for which T has height $k-1$ is

$$\frac{(2^k-2)!}{\prod_{i=1}^{k-1}(2^i)!} \prod_{i=1}^{k-1} \left(1 + \frac{1}{2^i - 1}\right).$$

1.1.5 Let T be a binary tree with n vertices. Find as simple a formula as you can for the number of permutations σ of $[n]$ satisfying $\text{BST}(\sigma) = T$.

1.2 Introducing randomness: average-case analysis

Another classic data structure is the *linked list*, in which elements are stored in a list of nodes, with each node holding a key and a pointer to the "next" node in the list. A linked list is then provided as a pointer to the first element of the list, and retrieving data requires traversing the list to find the associated key. In this case, the worst-case retrieval time is proportional to the length of the list, so to the number of data elements. As seen in Exercise 1.1.3, above, this worst-case time can also be attained in the case of binary search trees (if the keys happen to come in strictly increasing or strictly decreasing order). However, the tightness of the lower bound of Exercise 1.1.3 shows that it is possible for information retrieval to be much more efficient when using binary search trees than when using linked lists.

One way make the worst-case behaviour very unlikely is to assign random keys – say, for example, that independently as each piece of data is inserted, it is assigned a Uniform$[0,1]$ random variable as its key. After n insertions, then $\sigma(\mathbf{u})$ is a uniformly random permutation of $[n]$. When σ has this distribution, $\textsc{bst}(\sigma)$ is called a *random binary search tree*, or RBST.

The probability that a random binary search tree on n nodes has height $n-1$ (the worst possible) is $2/n!$. However, the probability that a tree of height $\lceil \log_2 n \rceil$ is attained is also very low (by considering equation in Exercise 1.1.4, it can be seen to be less than $1/((n+1)/2)!$ when $n+1$ is a power of two). So what behaviour should we expect in a random binary search tree? In the rest of the mini-course we will provide a detailed answer to this question.

Exercises

In the below exercises, let D_n denote the depth of the n'th key inserted into an RBST, and let H_n denote the height of an RBST.

1.2.1 Prove that for all $n \geq 2$, $\sum_{i=1}^{n-1} i^3 \leq n^3(n-1)/4$.

1.2.2 Show that for all n *and* $0 \leq k \leq n$,

$$\mathbf{P}\left\{D_n \geq k\right\} \leq \frac{n^2}{2^k}.$$

Hint: Use the permutation representation. Condition on the value of $\sigma(1)$, and split the probability according to whether $\sigma(n) > \sigma(1)$ or $\sigma(n) < \sigma(1)$. Then use induction and Exercise 1.2.1.

1.2.3 Prove using Exercise 1.2.2 that $\mathbf{E}[D_n] \leq 2\log_2(n) + O(1)$, *and that* $\mathbf{E}[H_n] \leq 3\log_2(n) + O(1)$. *(Remarkably, the latter bound is almost best possible; we will see that in fact* $\mathbf{E}[H_n] \sim c \log_2 n$ *with* c *a constant which is about* 2.99.*)*

1.3 Two couplings

The above construction of RBSTs based on a sequence of n uniform random variables has the pleasing consequence that (a slight extension of) it can be used to couple the constructions of RBSTs on n vertices for different values of n. To see this, we simply let $\{U_i\}_{i\in\mathbb{N}}$ be independent Uniform$[0,1]$ random variables and let $T_n^g = \textsc{bst}(U_1,\ldots,U_n)$ for each $n \in \mathbb{N}$. Then for each n, T_n^g is an RBST and for $n < n'$, T_n^g is a subtree of $T_{n'}^g$. (The superscript g stands for "growing", as we may view T_{n+1}^g can be grown from T_n^g by running $\textsc{insert}(T_n^g, U_n)$.

There is a second, often more useful coupling of RBSTs for different values of n, which is less elegant but more useful for the kind of analysis we will undertake. The basic observation underlying the coupling is that if σ is a uniformly random permutation of $[n]$, then the sizes of the left and right subtrees of $\textsc{bst}(\sigma)$, which are $\sigma(1)-1$ and $n-\sigma(1)$, respectively, are jointly distributed as $(\lfloor nU \rfloor, \lfloor n(1-U) \rfloor)$, where $U \sim$ Uniform$[0,1]$. This fact motivates the following definition.

Let T_∞ be the complete infinite binary tree. Label the root by $\emptyset$ and recursively define labels for all nodes in the tree by letting the left and right children of a node of label L have labels $L0$ and $L1$, respectively. Thus, nodes at depth d are labelled by the elements of $\{0,1\}^d$, binary strings of length d. Below we abuse notation and refer to nodes by their labels. Next, independently for each $L \in \bigcup_{d=0}^{\infty}\{0,1\}^n$, let $U_L \sim$ Uniform$[0,1]$ represent the "split at node L." Let $V_{L,L0} = U_L$ and $V_{L,L1} = 1 - U_L$. It will later be useful to think of the edges from L to $L0$ and to $L1$ as labelled by $V_{L,L0}$ and $V_{L,L1}$, respectively.

Now for fixed n, we may recover a tree distributed as an RBST on n nodes as follows. Let the root $\emptyset$ of T_∞ have mass $m_\emptyset^{(n)} = n$ and recursively define the mass of other nodes as follows. For a node L with mass $m_L^{(n)}$, let the mass $m_{L0}^{(n)}$ of $L0$ be $\lfloor m_L^{(n)} \cdot U_L \rfloor$ and let the mass $m_{L1}^{(n)}$ of $L1$ be $\lfloor m_L^{(n)} \cdot (1-U_L) \rfloor$. Then let T_n^s be the subtree of T_∞ consisting of all nodes L with $m_L^{(n)} \neq 0$.

We would like to use the second coupling, above, to prove upper and lower tail bounds on H_n. (For the remainder of the chapter, we write H_n for the height of T_n^s, so that the random variables $\{H_n\}_{n\in\mathbb{N}}$ are coupled and we can discuss almost sure convergence and convergence in probability.) The "floors" in the definition of T_n^s are a nuisance, and it is tempting to sacrifice correctness for elegance by rewriting the definition as follows. Let $m_\emptyset$ have mass 1 and for a node L with mass m_L, let $m_{L0} = m_L \cdot U_L$ and let $m_{L1} = m_L \cdot (1 - U_L)$. Then let T_n consist of all nodes L with $m_L > 1/n$. This is a useful bit of temptation to indulge in, because it turns out that the trees T_n and T_n^s are not too different. For $k \in \mathbb{N}$, write $T_{\infty,k}$ for the set of nodes of T_∞ at depth k.

Lemma 1. *For all $k \in \mathbb{N}$ and all nodes $L \in T_{\infty,k}$, $nm_L \geq m_L^{(n)} \geq nm_L - k$.*

Proof. The upper bounds are clear since for any positive real numbers $r_1, \dots, r_k$,

$$nr_1 \dots r_k \geq \lfloor \dots \lfloor \lfloor nr_1 \rfloor r_2 \rfloor \dots r_k \rfloor.$$

By symmetry, for each $k \in \mathbb{N}$ it suffices to prove the lower bound for the node with label $L_k = 0^k$, which we do by induction. For $k = 0$ we have $nm_{L_0} = n = m_{L_0}^{(n)}$. Next suppose the claim holds for all integers less than k. Then by induction

$$\begin{aligned} m_{L_k}^{(n)} &= \lfloor m_{L_{k-1}}^{(n)} U_{L_{k-1}} \rfloor \\ &\geq \lfloor n(m_{L_{k-1}} - (k-1)) U_{L_{k-1}} \rfloor \\ &\geq \lfloor nm_{L_{k-1}} U_{L_{k-1}} \rfloor - (k-1) \\ &\geq nm_{L_k} - k, \end{aligned}$$

proving the lower bound in the lemma. □

These two inequalities give us a strong enough link between T_n and T_n^s we will be able to analyze T_n and thereby obtain matching (at least in the first-order term) upper and lower tail bounds on H_n. Both tail bounds use the following, easy consequence of the lemma. Write m_k^+ for $\max_{L \in T_{\infty,k}} m_L$.

Corollary 2. *For all $n, k \in \mathbb{N}$,*

$$\mathbf{P}\left\{nm_k^+ \geq k+1\right\} \leq \mathbf{P}\left\{H_n \geq k\right\} \leq \mathbf{P}\left\{nm_k^+ \geq 1\right\}.$$

Proof. Since T_n^s is an RBST on n vertices, the probability that H_n is at least k is precisely

$$\mathbf{P}\left\{\max_{L \in T_{\infty,k}} m_L^{(n)} \geq 1\right\}.$$

By the bounds in Lemma 1 we have

$$nm_k^+ - k \leq \max_{L \in T_{\infty,k}} m_L^{(n)} \leq nm_k^+,$$

and the corollary follows. □

The lower bound in Lemma 1 also yields that, at least on a logarithmic scale, for lower bounds on H_n it suffices to study the trees T_n.

Corollary 3. *Fix any constant $a < 1$. Suppose that for all k sufficiently large, there is some node $L^{(k)} \in T_{\infty,k}$ with $m_L \geq a^k$. Then $\liminf_{n\to\infty} H_n / \ln n \geq \log_{a^{-1}}(e)$.*

Proof. By Lemma 1, it follows that for all k sufficiently large and for all n, we have $m_{L^{(k)}}^{(n)} \geq na^k - k$. In this case the height of T_n^s is at least h as long as $na^h - h \geq 1$.

Taking h near $\log_a(n^{-1}) = \log_{a^{-1}} n$ yields $na^h - h$ near $(-\log_{a^{-1}} n)$, just a bit too small. If instead we take $h = h(a,n) = \lfloor \log_{a^{-1}} n - (\log_{a^{-1}} \log_{a^{-1}}(n)) - 1 \rfloor$, we get

$$\begin{aligned} na^h - h &\geq na^{\log_{a^{-1}} n - (\log_{a^{-1}} \log_{a^{-1}}(n))} - (\log_{a^{-1}} n - (\log_{a^{-1}} \log_{a^{-1}}(n)) + 1 \\ &\geq a^{-(\log_{a^{-1}} \log_{a^{-1}}(n))} - \log_{a^{-1}} n + 1 \\ &= 1, \end{aligned}$$

the second inequality valid whenever $(\log_{a^{-1}} \log_{a^{-1}}(n)) \geq 0$, so in particular for all sufficiently large n. It follows that

$$\liminf_{n\to\infty} \frac{H_n}{\ln n} \geq \liminf_{n\to\infty} \frac{h(a,n)}{lnn} = \liminf_{n\to\infty} \frac{\log_{a^{-1}} n}{\ln n} = \log_{a^{-1}}(e).$$

□

1.4 A natural guess

Lemma 1 provide a strong connection between the heights of T_n and T_n^s (the latter being our true object of interest). One natural guess is that the height of T_n should be near the greatest value k such that

$$\mathbf{E}\left[|\{L \in T_{\infty,k} \ : \ m_L \geq n^{-1}\}|\right] > 1.$$

Since this is not a movie, we are not worried about giving away the ending: this guess ends up being correct (at least on a logarithmic scale). Knowing this, it is worth the effort to compute this value k. Fixing any node $L_k \in T_{\infty,k}$, by linearity of expectation and the symmetry of T_∞, we have

$$\mathbf{E}\left[|\{L \in T_{\infty,k} \ : \ m_L \geq n^{-1}\}|\right] = 2^k \mathbf{P}\left\{m_{L_k} \geq n^{-1}\right\}.$$

The random variable m_{L_k} is the product of k independent Uniform$[0,1]$ random variables, which makes precise calculations straightforward. It is helpful to transform m_{L_k} using the fact (whose verification is an easy calculation) that if $U \sim$ Uniform$[0,1]$ then $-\ln U \sim$ Exponential(1). Writing $S_{L_k} = -\ln m_{L_k}$, it follows that $S_{L_k} \sim$ Gamma(k), where Gamma(k) is the random variable with density

$$g_k(t) = \frac{t^{k-1}e^{-t}}{(k-1)!}, \qquad t > 0.$$

We thus have

$$\mathbf{E}\left[|\{L \in T_{\infty,k} \ : \ m_L \geq n^{-1}\}|\right] = 2^k \mathbf{P}\left\{S_{L_k} \leq \ln n\right\}. \tag{1.1}$$

The following lemma controls the latter probability for us, in the range of interest.

Lemma 4. *Fix $a \in (0,1)$ and let $(t_k)_{k\in\mathbb{N}}$ be a sequence of numbers such that $\lim_{k\to\infty} t_k/k = a$. Then*

$$\mathbf{P}\left\{\text{Gamma}(k) < t_k\right\} \sim \frac{1}{1-a}\frac{e^{-t_k}(t_k)^k}{k!}.$$

Proof. By repeated integration by parts, starting with $u = e^{-t}$, $dv = dt \cdot t^{k-1}/(k-1)!$, we have

$$\begin{aligned} \mathbf{P}\left\{\text{Gamma}(k) < t_k\right\} &= \int_0^{t_k} \frac{t^{k-1}e^{-t}}{(k-1)!} \\ &= e^{-t_k}\left(\frac{t_k^k}{k!} + \frac{t_k^{k+1}}{(k+1)!} + \frac{t_k^{k+2}}{(k+2)!} + \dots\right). \\ &= \frac{e^{-t_k}t_k^k}{k!}\left(1 + \sum_{i=1}^{\infty}\left(\frac{t_k}{k+1} \cdot \dots \cdot \frac{t_k}{k+i}\right)\right) \qquad (1.2) \\ &\sim \frac{1}{1-a}\frac{e^{-t_k}t_k^k}{k!}. \end{aligned}$$

□

Corollary 5. *If $t_k = ak$ then*

$$\frac{e^{-ak}(ak)^k}{k!} \le \mathbf{P}\left\{\text{Gamma}(k) < t_k\right\} \le \frac{1}{1-a}\frac{e^{-ak}(ak)^k}{k!}.$$

Also,

$$\lim_{k\to\infty} \frac{\ln(\mathbf{P}\left\{\text{Gamma}(k) < t_k\right\})}{k} = \ln(ae^{1-a}).$$

Proof. The lower bound follows by omitting the sum in (1.2). The upper bound also follows from (1.2) since for all $i \ge 1$, $ak/(k+i) < a$. The existence of the limit and the asymptotic statement follows by applying Stirling's formula $k! \sim \sqrt{2\pi k}(k/e)^k$, then taking logarithms. □

Remark. Taking logarithms and dividing by k in the last equality of Corollary 5 hides a factor $(2\pi k)^{-1/2}$ present in the first two inequalities, obtained from Stirling's formula applied to $k!$. This factor in fact appears in much greater generality, i.e., for sums of random variables with distributions other than exponential. It is essentially due to the central limit theorem, and will become very important later when we study lower order terms.

Due to (1.1) and the asymptotic in Corollary 5, we are interested in the value of c for which $\ln(ce^{1-c}) = -\ln 2$, or $2ce^{1-c} = 1$. Writing $f(c) = 2ce^{1-c}$, we have $f(1) = 2 > 1$, $f(0) = 0$, and $f'(c) = 2e^{1-c} - 2ce^{1-c} = 2e^{1-c}(1-2c)$, so f is increasing for $c < 1/2$ and decreasing for $c > 1/2$. It follows that there is a unique $c' \in (0,1)$ with $f(c') = 1$, and $c' < 1/2$. We now combine this fact and Corolary 5 with (1.1). Let $c^* = 1/c'$.

Lemma 6. *Fix $c > 0$ and let $k = k(c,n) = \lceil c \ln n \rceil$. Then*

$$\lim_{n\to\infty} \frac{\ln \mathbf{E}\left[|\{L \in T_{\infty,k} \;:\; m_L \ge n^{-1}\}|\right]}{k} = \ln(f(c^{-1})).$$

In particular, this limit exists and is positive, zero, or negative according as $c < c^*$, $c = c^*$, or $c > c^*$.

Proof. We prove only the case $c > c^*$ as the others are similar. In this case we have $\ln n \le c^{-1}k < c'k$, so by (1.1), we have

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\ln \mathbf{E}\left[|\{L \in T_{\infty,k} \;:\; m_L \ge n^{-1}\}|\right]}{k} &= \limsup_{n\to\infty} \frac{\ln\left(2^k \mathbf{P}\left\{S_{L_k} \le \ln n\right\}\right)}{k} \\
&\le \lim_{k\to\infty} \frac{\ln\left(2^k \mathbf{P}\left\{S_{L_k} \le c^{-1}k\right\}\right)}{k} \\
&= \lim_{k\to\infty} \frac{\ln\left(2^k \mathbf{P}\left\{\text{Gamma}(k) \le c^{-1}k\right\}\right)}{k}.
\end{aligned}$$

Letting $t_k = c^{-1}k$ in Corollary 5, we obtain

$$\limsup_{k\to\infty} \frac{\ln(2^k \mathbf{P}\left\{\text{Gamma}(k) < t_k\right\})}{k} = \ln(2c^{-1}e^{1-c^{-1}}) = \ln(f(c^{-1})).$$

The last term is negative as $c^{-1} < c'$, $f(c') = 1$ and f is increasing below c'. Since $\ln n \ge c^{-1}(k-1)$, we similarly have

$$\begin{aligned}
\liminf_{n\to\infty} \frac{\ln \mathbf{E}\left[|\{L \in T_{\infty,k} \;:\; m_L \ge n^{-1}\}|\right]}{k} &= \liminf_{n\to\infty} \frac{\ln\left(2^k \mathbf{P}\left\{S_{L_k} \le \ln n\right\}\right)}{k} \\
&\ge \lim_{k\to\infty} \frac{\ln\left(2^k \mathbf{P}\left\{S_{L_k} \le c^{-1}(k-1)\right\}\right)}{k} \\
&= \lim_{k\to\infty} \frac{\ln\left(2^k \mathbf{P}\left\{\text{Gamma}(k) \le c^{-1}k\right\}\right)}{k},
\end{aligned}$$

the last equality holding since, by Lemma 4, $\mathbf{P}\left\{\text{Gamma}(k) \le c^{-1}(k-1)\right\}$ and $\mathbf{P}\left\{\text{Gamma}(k) \le c^{-1}k\right\}$ are within a constant factor for k large. Thus the limit exists and equals $\ln(f(c^{-1})) < 0$. □

1.5 Upper tail bounds on H_n

We have in fact done all the hard work for our upper bounds in Section 1.4. Recall that $m_k^+ = \max\{m_L \ : \ L \in T_{\infty,k}\}$. By Corollary 2 and Markov's inequality, for any n and k we have

$$\begin{aligned}\mathbf{P}\{H_n \geq k\} &= \mathbf{P}\left\{m_k^+ \geq n^{-1}\right\} \\ &= \mathbf{P}\left\{|\{L \in T_{\infty,k} \ : \ m_L \geq n^{-1}\}| > 0\right\} \\ &\leq \mathbf{E}\left[|\{L \in T_{\infty,k} \ : \ m_L \geq n^{-1}\}|\right].\end{aligned}$$

Now fix any $c > c^*$ and let $k = k(c,n) = \lceil c \ln n \rceil$. It then follows from Lemma 6 that there is $\epsilon(c) < 0$ such that for all n sufficiently large,

$$\mathbf{P}\{H_n \geq k\} \leq e^{-k\epsilon(c)}.$$

As an immediate consequence we obtain the following theorem.

Proposition 7.

$$\limsup_{n\to\infty} \frac{H_n}{\ln n} \leq c^* \qquad \textit{in probability.}$$

Proof. Apply the preceding bound for any fixed $c > c^*$ to see that $\limsup_{n\to\infty} \frac{H_n}{\ln n} \leq c$ in probability; then let c decrease to c^*. □

The above proof exploited explicit formulae for Gamma-distributed random variables. It is instructive to also prove upper tail bounds for H_n using Chernoff's bounding technique, since this approach the one we will eventually generalize. We first use a union bound over nodes at depth k, which gives for any $t > 0$,

$$\mathbf{P}\left\{m_k^+ \geq t\right\} \leq \sum_{L \in T_{\infty,k}} \mathbf{P}\{m_L \geq t\} = 2^k \mathbf{P}\{m_{L_k} \geq t\}. \tag{1.3}$$

the last equality holding due to the symmetry of $T_{\infty,k}$. We then use the following identity, which may be checked using integration by parts and the fact that $-\ln m_{L_k}$ is Gamma(k) distributed.

Proposition 8. *For any* $\lambda > 0$,

$$\mathbf{E}\left[m_{L_k}^\lambda\right] = (\lambda+1)^{-k}.$$

Proof. Exercise. □

The key idea of Chernoff's bounding technique is to write $\mathbf{P}\{m_{L_k} \geq t\} = \mathbf{P}\left\{m_{L_k}^\lambda \geq t^\lambda\right\}$ (true for $\lambda \geq 0$), then apply Markov's inequality and optimize over $\lambda > 0$.[2] The remarkable fact is that very often, the resulting bound is best possible (up to sub-exponential terms).

Lemma 9. *For any* $t < 1$ *and integer* $k \geq \ln t^{-1}$,

$$\mathbf{P}\left\{m_k^+ \geq t\right\} \leq t\left(\frac{2e\ln(t^{-1})}{k}\right)^k.$$

Proof. From (1.3), for any $\lambda \geq 0$ we have

$$\mathbf{P}\left\{m_k^+ \geq t\right\} \leq 2^k \mathbf{P}\{m_{L_k} \geq t\} = 2^k \mathbf{P}\left\{m_{L_k}^\lambda \geq t^\lambda\right\}.$$

By Markov's inequality and Proposition 8, we then have

$$\mathbf{P}\left\{m_k^+ \geq t\right\} \leq 2^k t^{-\lambda} \mathbf{E}\left[m_{L_k}^\lambda\right] = 2^k t^{-\lambda}(\lambda+1)^{-k}.$$

[2] Traditionally, one would actually write $\mathbf{P}\{m_{L_k} \geq t\} = \mathbf{P}\left\{e^{\lambda m_{L_k}} \geq e^{\lambda t}\right\}$ then optimize over λ; however, the idea is essentially the same.

This bound is minimized by taking $\lambda + 1 = k/\ln(t^{-1})$ (exercise), which is valid since then $\lambda = k/\ln(t^{-1}) - 1 \geq 0$, and we get

$$\mathbf{P}\left\{m_k^+ \geq t\right\} \leq 2^k t^{1-k/\ln(t^{-1})} \left(\frac{k}{\ln(t^{-1})}\right)^{-k}.$$

Since $t^{1-k/\ln(t^{-1})} = te^k$, we thus have

$$\mathbf{P}\left\{m_k^+ \geq t\right\} \leq t \left(\frac{2e\ln(t^{-1})}{k}\right)^k,$$

as claimed. □

This lemma immediately gives upper tail bounds on the height.

Corollary 10. *For $c > 2$, $\mathbf{P}\{H_n \geq c\ln n\} \leq n^{c\ln(2e/c)-1}$.*

Proof. Let $k = \lceil c\ln n \rceil$. By Corollary 2 and Lemma 9 applied with $t = 1/n$,

$$\mathbf{P}\{H_n \geq c\ln n\} = \mathbf{P}\{H_n \geq k\} \leq \mathbf{P}\left\{m_k^+ \geq 1/n\right\} \leq \frac{1}{n}\left(\frac{2e\ln n}{k}\right)^k.$$

But

$$\frac{d}{dk}\left(\left(\frac{2e\ln n}{k}\right)^k\right) = \left(\frac{2e\ln n}{k}\right)^k \ln\left(\frac{2\ln n}{k}\right),$$

which is negative when $k \geq 2\ln n$. Thus,

$$\frac{1}{n}\left(\frac{2e\ln n}{k}\right)^k \leq \frac{1}{n}\left(\frac{2e\ln n}{c\ln n}\right)^{c\ln n} = n^{c\ln(2e/c)-1}.$$

□

The bound in Corollary 10 is negative whenever $c\ln(2e)/c < 1$. It is easily checked that the value c^* from above is the unique solution of $c\ln(2e/c) = 1$ with $c > 2$, and so this corollary provides a second proof of Proposition 7. In the next section, we use the lower bound on $\mathbf{P}\{H_n \geq k\}$ from Corollary 2, together with the bounds of Section 1.4 to prove a lower bound on $\liminf_{n\to\infty} \frac{H_n}{\ln n}$ that matches the bound of Proposition 7.

Exercises

1.5.1 Check that c^* is the unique solution of $c\ln(2e/c) = 1$ with $c > 2$.

1.5.2 Show that if $U \sim$ Uniform$[0,1]$ *then $(-\ln U) \sim$* Exponential(1).

1.5.3 Prove Proposition 8. (Hint: use $m_{L_k}^\lambda = e^{-\lambda S_k}$, then use that the density of S_k is $e^{-x}x^{k-1}/(k-1)!$ and integration by parts.)

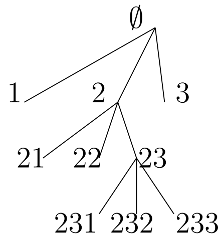
1.5.4 Prove the claim in the proof of Lemma 9, that $2^k t^{-\lambda}(\lambda+1)^{-k}$ is minimized when $\lambda + 1 = k/\ln(t^{-1})$, assuming that $k \geq \ln t^{-1}$.

1.5.5 Check that $c^*\ln 2 \approx 2.9882$, so that $\limsup_{n\to\infty} \frac{H_n}{\ln_2 n} \leq 2.989$ and Proposition 7 is indeed an improvement over Exercise 1.2.3.

1.6 Lower tail bounds on H_n

Before proving our lower tail bounds on H_n, we quickly introduce Galton-Watson trees and recall a fundamental fact about them, which we call the *fundamental theorem of branching processes.*

Reminder about branching processes. Fix a non-negative integer-valued random variable B. Then the following procedure generates a *Galton-Watson tree with offspring distribution B*.



⋆ Start from the root (call it θ), let $B_\theta \sim B$.
⋆ Give ϕ children $1, \ldots, B_\phi$.
⋆ Independently for each $i = 1, \ldots, B_\theta$, let $B_i \sim B$, and ⋆ give i children $i1, i2, \ldots, iB_i$.
⋆ Repeat forever or until done; call the resulting random tree $\mathcal{T}_B$.

Now let $F_B(z) = \mathbf{E}\left[z^B\right] = \sum_{k=0}^{\infty} \mathbf{P}\{B = k\} z^k$.

Theorem 11 (Fundamental theorem of branching processes)**.** *If* $\mathbf{P}\{B = 1\} < 1$ *then*

$$\mathbf{P}\{|\mathcal{T}_B| < \infty\} = \min\{x \geq 0 \ : \ F_B(x) = x\}.$$

In particular, $\mathbf{P}\{|\mathcal{T}_B| = \infty\} > 0$ *if and only if* $\mathbf{E}[B] > 1$ *(assuming* $\mathbf{P}\{B = 1\} < 1$*).*

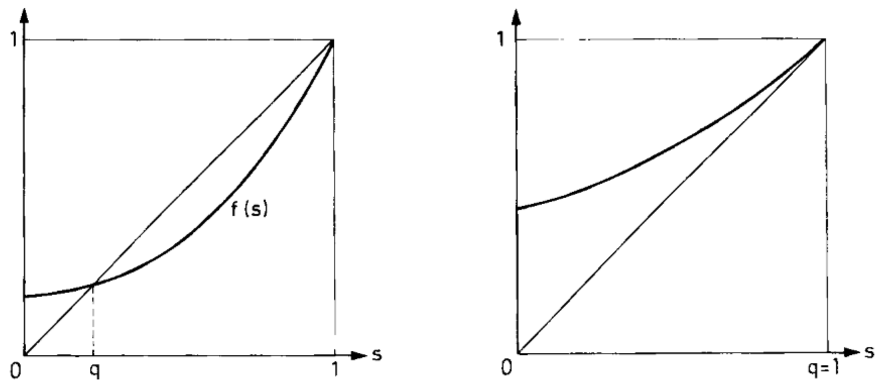


Figure 1.2: The two cases in the fundamental theorem of branching processes

The reason why the second part follows from the first is essentially captured by the pictures in Figure 1.2. In symbols, note that

$$F_B'(z) = (\sum_{n=0}^{\infty} \mathbf{P}\{B = n\} z^n)' = \sum_{n=1}^{\infty} n\mathbf{P}\{B = n\} z^{n-1},$$

so $F_B'(1) = \sum_{n=1}^{\infty} n\mathbf{P}\{B = n\} = \mathbf{E}[B]$. Also, $F_B(0) = \mathbf{P}\{B = 0\} \geq 0$, and $F_B''(z) \geq 0$.

Embedded branching processes. We may view the tree T_∞ as a (not particularly interesting) branching process in which B is deterministically equal to 2. However, there are more interesting branching processes contained within T_∞.

As a first example, for an edge e and a node L of T_∞, write $e \prec L$ if e is on the path from the root to L. Now fix any constant $c < 1/2$ and let $T_\infty(c)$ consist of all nodes L of T_∞ for which, for all edges e of T_∞ with $e \prec L$, we have $V_e \geq c$. Then let $B^{(c)}$ be a random variable with $\mathbf{P}\left\{B^{(c)} = 1\right\} = 2c$, $\mathbf{P}\left\{B^{(c)} = 2\right\} = 1 - 2c$

Lemma 12. *$T_\infty(c)$ is distributed as $\mathcal{T}_{B^{(c)}}$.*

Proof. For each node L of T_∞, let B_L be the number of edges e leaving the root and with $V_e \geq c$. Then $B_L \sim B^{(c)}$, and the random variables $\{B_L \ : \ L \text{ a node of } T_\infty\}$ are jointly independent.

Now let $T_{\infty,0}(c)$ be the set containing only the root of T_∞. For $k \geq 0$, inductively define $T_{\infty,k+1}(c)$ to be the set of nodes $L \in T_{\infty,k+1}$ for which first, the parent L' of L is in $T_{\infty,k}(c)$, and second, $V_{L',L} \geq c$. Then for each $k \geq 0$, conditional upon $T_{\infty,k}(c)$, each node $L' \in T_{\infty,k}(c)$ independently has $B_{L'}$ children in $T_{\infty,k+1}(c)$, and these are all the elements of $T_{\infty,k+1}(c)$. This shows that the tree with nodes $\bigcup_{k=0}^{\infty} T_{\infty,k}(c)$ is distributed as $\mathcal{T}_B$. But this tree is precisely $T_\infty(c)$, so the proof is complete. □

This lemma allows us to prove an initial lower bound – far from tight – on $\liminf_{n\to\infty} H_n/\ln n$.

Corollary 13. *For any $c < 1/2$, there is $\epsilon(c) > 0$ such that with probability at least $\epsilon(c)$,*

$$\liminf_{n\to\infty} \frac{H_n}{\ln n} \geq \log_{c^{-1}}(e).$$

Proof. For any $c < 1/2$, $\mathbf{E}\left[B^{(c)}\right] > 1$, and so by the lemma there is some $\epsilon(c) > 0$ such that $\mathbf{P}\{|T_\infty(c)| = \infty\} > \epsilon(c)$. If $|T_\infty(c)| = \infty$ then for all k, there is some node $L^{(k)} \in T_{\infty,k}$ with $m_L \geq c^k$. The corollary then follows by applying Corollary 3 with $a = c$.

□

The above corollary shows us that for any $x > \log_2(e) \approx 1.4427$, there is a positive probability that $\liminf_{n\to\infty} H_n/\ln n \geq x$. For lower bounds in probability, of course, this is not good enough: we need this probability to be one. However, it is possible to obtain a lower bound in probability from the *a priori* weaker bound of Corolary 13, with the help of an extremely useful and important lemma. The basic idea of the lemma is to consider the 2^k subtrees of T_∞ rooted at nodes of depth k independently. For $\liminf_{n\to\infty} H_n/\ln n$ to be small a corresponding quantity must *independently* be small in each of these subtrees, which is very unlikely if k is large.

Lemma 14 (Weak amplification lemma, binary version)**.** *If $\mathbf{P}\{\liminf_{n\to\infty}(H_n/\ln n) > c\} > 0$ then $\mathbf{P}\{\liminf_{n\to\infty}(H_n/\ln n) > c\} = 1$.*

Proof. Suppose that $\mathbf{P}\{\liminf_{n\to\infty}(H_n/\ln n) > c\} = \delta > 0$. For each node $L \in T_\infty$, let $T_{\infty,L}$ be the subtree of T_∞ rooted at L, together with its weights. Then for all L, $T_{\infty,L}$ is distributed as a copy of T_∞, and we define $T^s_{n,L}$ as the copy of T^s_n, rooted at L via the same coupling as above. Finally, write $H_{n,L}$ for the height of $T^s_{n,L}$.

Now suppose L has depth k. Then by the second inequality in Lemma 1, for all n, $m_L^{(n)} \geq nm_L - k$, so with probability 1 there is $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $m_L^{(n)} \geq \lceil nm_L/2 \rceil$. It follows that for all $n \geq n_0$, $T^s_{\lceil nm_L/2\rceil,L}$ is a subtree of T^s_n, so $H^s_n \geq H^s_{\lceil nm_L/2\rceil,L}$. Thus, with probability one,

$$\liminf_{n\to\infty} \frac{H_n}{\ln n} \geq \liminf_{n\to\infty} \frac{H^s_{\lceil nm_L/2\rceil,L}}{\ln n} = \liminf_{n\to\infty} \frac{H^s_{\lceil nm_L/2\rceil,L}}{\ln \lceil nm_L/2\rceil} = \liminf_{n\to\infty} \frac{H^s_{n,L}}{\ln n},$$

the second inequality holding since $\lceil nm_L/2\rceil$ is with probability one a constant greater than zero, and so with probability one $H^s_{\lceil nm_L/2\rceil,L} \to \infty$ as $n \to \infty$ and $\ln(nm_L/2) = \ln n - O(1)$.

It follows that with probability one, for any integer $k \geq 1$,

$$\liminf_{n\to\infty} \frac{H_n}{\ln n} \geq \max_{L\in T_{\infty,k}} \liminf_{n\to\infty} \frac{H^s_{n,L}}{\ln n}.$$

But if $L, L' \in T_{\infty,k}$ then $T^s_{n,L}$ and $T^s_{n,L'}$ are independent, so $H^s_{n,L}$ and $H^s_{n,L'}$ are also independent.

It follows that if $\mathbf{P}\left\{\liminf_{n\to\infty} \frac{H_n}{\ln n} \leq c\right\} = \delta < 1$, then for any integer $k \geq 1$,

$$\begin{aligned}\mathbf{P}\left\{\liminf_{n\to\infty} \frac{H_n}{\ln n} \leq c\right\} &\leq \mathbf{P}\left\{\max_{L\in T_{\infty,k}} \liminf_{n\to\infty} \frac{H^s_{n,L}}{\ln n} \leq c\right\} \\ &= \prod_{L\in T_{\infty,k}} \mathbf{P}\left\{\liminf_{n\to\infty} \frac{H^s_{n,L}}{\ln n} \leq c\right\} \\ &= \delta^k,\end{aligned}$$

$\mathbf{P}\left\{\liminf_{n\to\infty} \frac{H_n}{\ln n} \leq c\right\}$ must in fact equal zero. $\square$

It follows from Corollary 13 and Lemma 14 that $\liminf_{n\to\infty} H_n/\ln n \geq \log_2(e)$ with probability 1, which is still a far cry from our upper bound. However, the argument we used in Corollary 13 is nowhere near optimal. In particular, by restricting our attention to the tree $T_\infty(c)$ we are requiring that the uniform splits are between c and $1-c$ at every step, whereas for H_n to be large we only need a long path along which the splits are large on average. One way to try to capture this fact is to only look at the splitting behaviour at depths k which are a multiple of r for some integer $r > 1$, and only require that $T_{\infty,k}$ contain a node L_k with $m_{L_k} \geq c^k$ for these depths. The first thing to check is that a version of Corolary 3 still holds under this requirement.

Corollary 15. *Fix any constant $a < 1$ and any integer $r > 1$. Suppose that for all integers $d \geq 1$, letting $k = rd$ there is some node $L^{(k)} \in T_{\infty,k}$ with $m_L \geq a^k$. Then $\liminf_{n\to\infty} H_n/\ln n \geq \log_{a^{-1}}(e)$.*

Proof. Fix any constant $a' < a$. Then for all d sufficiently large, for all integers k' with $r(d-1) < k' \leq rd$, we have $(a')^{k'} < a^{rd}$. Thus, the ancestor $L^{(k')}$ at depth k' of $L^{(rd)}$, has

$$m_{L^{(k')}} \geq m_{L^{(rd)}} \geq a^{rd} > (a')^{k'}.$$

In other words, for all k' sufficiently large, there is $L^{(k')} \in T_{\infty,k'}$ such that $m_{L^{(k')}} \geq (a')^{k'}$. It then follows from Corollary 15 that $\liminf_{n\to\infty} H_n/\ln n \geq \log_{(a')^{-1}}(e)$. Since $a' < a$ was arbitrary the result follows. $\square$

With this idea and corollary in hand, it is now fairly easy to prove lower bounds that match our upper bounds. Fix any constant $c < c^*$.

Proposition 16. *With probability one, $\liminf_{n\to\infty} \frac{H_n}{n} \geq c^*$.*

Proof. Fix any constant $c < c^*$. Then by Lemma 6, writing $k = k(c,n) = \lceil c\ln n\rceil$, we have

$$\lim_{n\to\infty} \frac{\ln \mathbf{E}\left[|\{L \in T_{\infty,k} \ : \ m_L \geq n^{-1}\}|\right]}{k} = \ln(f(c^{-1})) > 0.$$

Thus, in particular, there is some fixed $n_0 > 3$ such that letting $k_0 = k(c, n_0)$, we have

$$\mathbf{E}\left[|\{L \in T_{\infty,k_0} \ : \ m_L \geq n_0^{-1}\}|\right] > 1.$$

Now let T^* be the Galton-Watson tree obtained by joining the root of T_∞ to all nodes of the set $\{L \in T_{\infty,k_0} \ : \ m_L \geq n_0^{-1}\}$, and recursively repeating in subtrees. Then T^* is supercritical and so with some probability $\epsilon > 0$ is infinite.

Now note that since $k_0 \geq c\ln n_0$, letting $a = e^{-\frac{1}{c}}$ we have $a^{k_0} \leq n_0^{-1}$. If T^* is infinite then by applying Corollary 15 with this choice of a and with $r = k_0$, we obtain that

$$\liminf_{n\to\infty} \frac{H_n}{\ln n} \geq \log_{a^{-1}}(e) = \log_{e^{1/c}} e = c.$$

Thus, the probability that $\liminf_{n\to\infty} \frac{H_n}{\ln n} \geq c$ is at least $\epsilon > 0$, so is one by Lemma 14. Since $c < c^*$ was arbitrary the lemma follows. $\square$

1.7 Devroye's theorem and a word on independence

With the preceding sections under our belt, we now have the following result.

Theorem 17 (Devroye's theorem, 1986)**.**

$$\frac{H_n}{n} \to c^* \qquad \textit{in probability, as } n \to \infty.$$

Proof. By Proposition 7 states that $\limsup_{n\to\infty} \frac{H_n}{\ln n} \leq c^*$ in probability. By Proposition 16 states that $\liminf_{n\to\infty} \frac{H_n}{\ln n} \geq c^*$ almost surely. The result follows. □

A few remarks on the proof are in order. The theorem was proved in Devroye [4], from whence the name. First, both proofs of the upper bound boiled down to bounding the probability that a *single* node L_k at depth k had m_{L_k} large. Such bounds are derived by considering exclusively the uniform random variables on the path from the root to L_k. In particular, the dependence between the uniforms on the edges from a node to its children played no role.

Similarly, the lower bound boiled down to an expected value calculation for the number of nodes L at a given level k with m_L not too small. This value is just 2^k times the probability that a given node L_k at level k has m_{L_k} not too small, and so again only the uniforms on the path from the root to L_k play a rôle.

It is possible to strengthen Devroye's theorem, replacing the convergence in probability with almost sure convergence. Indeed, when Devroye proved his theorem, Pittel [7] had already shown the existence of *some* constant c such that $\frac{H_n}{n} \to c$ almost surely. This knowledge immediately allows the aforementioned strengthening. Pittel's arguments were based on a strong law for subadditive processes, due to Kesten. In the next chapter, we will see a different argument for almost sure convergence.

Finally, write H_n^* for the greatest depth k at which there is some node L with $m_L \geq 1/n$. Then by combining Lemma 1 with Devroye's theorem, it is straightforward to see that $H_n^*/n \to c^*$ in probability also. (Later, we will see that for both H_n and H_n^*, the convergence is in fact almost sure.) The latter random variable, H_n^* is the one most naturally connected with branching random walks, and it is to this connection that we now turn.

Exercises

1.7.1 Prove that $H_n^*/n \to c^*$ in probability.

Lecture 2

The connection with branching random walks

2.1 Binary branching, Exponential(1) edge weights

Our calculations in Chapter 1 were simplified by the observation that writing $D_{L_k} = -\ln m_{L_k}$, we have that $D_{L_k} \stackrel{\mathrm{d}}{=} \text{Gamma}(k)$. This transformation is also at the heart of the link between search trees and branching random walks. For each edge e of T_∞, let $E_e = -\ln V_e$. Then $E_e \stackrel{\mathrm{d}}{=}$ Exponential(1), and $D_v = \sum_{e \prec v} E_e$, and $H_n^* = \max\{k \ : \ \exists L \in T_{\infty,k} \text{ with } D_L \le \ln n\}$. The tree T_∞ together with the edge weights $\mathcal{E} = \{E_e \ : \ e \in E(T_\infty)\}$ is (one example of one way to represent) a branching random walk. To justify the terminology, fix a node $L_k \in T_{\infty,k}$ and let its ancestors, starting from the root, be $L_0, L_1, \ldots, L_{k-1}$. Then, setting $D_{L_0} = 0$, the sequence $(D_{L_i})_{i=0}^k$ forms the first k steps of a random walk with step size $E \stackrel{\mathrm{d}}{=}$ Exponential(1). Using further terminology from random walks, we will refer to D_L as the *displacement* of L.

In the branching random walk setting, there is no particular reason to consider cutoffs only of the form $D_L \le \ln n$ for integer n: we may just as well replace $\ln n$ by any positive real. So, let $F_s == F_s(T_\infty, \mathcal{E}) \max\{k \ : \ \exists L \in T_{\infty,k} \text{ with } D_L \le s\}$. (The F stands for "final" – F_t is the final generation containing a node with displacement at most s.) So, for all $n \in \mathbb{N}$, we have $F_{\ln n} = H_n^*$. In particular, since F_s increases in s and H_n^* increases in n, we always have

$$H^*_{\lfloor e^s \rfloor} \le F_s \le H^*_{\lceil e^s \rceil} \le +H^*_{\lfloor e^s \rfloor} + 1.$$

Asymptotics for F_s then follow immediately from Devroye's theorem.

Lemma 18.

$$\frac{F_s}{s} \to c^* \qquad \textit{in probability, as } s \to \infty.$$

Proof. Fix $c_1 < c^*$. For $s > 0$ write $b = b(s) = \lfloor e^s \rfloor$, and $t = t(s) = \lceil e^s \rceil$. Then we have

$$\mathbf{P}\{F_s \le c_1 s\} \le \mathbf{P}\left\{H^*_{b(s)} \le c_1 s\right\}.$$

Since $c_1 s \le c_1 \ln b(s) + 1$, it follows that

$$\mathbf{P}\{F_s \le c_1 s\} \le \mathbf{P}\left\{H^*_{b(s)} \le c_1 \ln b(s) + 1\right\},$$

which tends to zero as $s \to \infty$. Thus $\liminf_{n\to\infty} F_s/s \ge c_1$ in probability. Similarly, if $c_2 > c^*$ then

$$\mathbf{P}\{F_s \ge c_2 s\} \le \mathbf{P}\left\{H^*_{t(s) \ge c_2 s.}\right\} \le \mathbf{P}\left\{H^*_{b(s)} \ge c_2 \ln t(s) - 1\right\},$$

which tends to zero ast $s \to \infty$. Thus $\limsup_{n\to\infty} F_s/s \geq c_2$ in probability. Since $c_1 < c^* < c_2$ were arbitrary, the lemma follows. □

There is a natural "dual" random variable to F_t: it is the *minimum position at depth* k,

$$M_k = M_k(T_\infty, \mathcal{E}) = \min\{D_L \ : \ L \in T_{\infty,k}\}.$$

We then have the easy equivalence

$$M_k \leq s \Leftrightarrow F_s \geq k. \tag{2.1}$$

Thus, knowledge of M_k for all $k \in \mathbb{N}$ determines F_s for all $s \in [0, \infty)$, and vice-versa. As a result, Lemma 18 easily yields asymptotics for M_k. Recall that c' is the constant $1/c^*$, defined in Chapter 1 to be the unique constant $0 < c < 1/2$ satisfying $2ce^{1-c} = 1$.

Proposition 19.

$$\frac{M_k}{k} \to c' \qquad \textit{in probability, as } k \to \infty.$$

Proof. Fix a constant $a > 0$ and let $c = 1/a$. By (2.1),

$$M_k \leq ak \Leftrightarrow F_{ak} \geq k \qquad\qquad \text{by (2.1)}$$

First suppose that $a < c'$, so that $c = a^{-1} > c^*$. Then $k = c(ak) > c^*(ak)$, so by Lemma 18, $\mathbf{P}\{F_{ak} \geq k\} = \mathbf{P}\{F_{ak} \geq c(ak)\}$ tends to zero as $k \to \infty$ and by the above equivalence, so must $\mathbf{P}\{M_k \leq ak\}$.

An identical argument shows that if $a > c'$ then $\mathbf{P}\{M_k \geq ak\} \to 0$ as $n \to \infty$. □

Proposition 19 is a kind of weak law of large numbers for M_k, and it is natural to ask if a strong law holds. The advantage of studying M_k over studying H_n or F_s is that due to the exponential growth of the levels, the probability that M_k deviates significantly from its mean decreases exponentially quickly in k, which makes easy to show almost sure convergence in place of convergence in probability. This result can then be funnelled through to yield almost sure convergence in Devroye's theorem.

In fact, we already proved one half of almost sure convergence: our lower bound on $\liminf H_n/\ln n$ was almost sure due to the weak amplification lemma (Lemma 14). Under the above duality, upper bounds become lower bounds and vice-versa, so to prove the other half of almost sure convergence we need to prove lower bounds on M_k.

Lemma 20. *With probability one,*

$$\liminf_{k\to\infty} \frac{M_k}{k} \geq c'.$$

Proof. Fix c with $0 < c < c'$. Then by Corollary 5,

$$\mathbf{P}\{D_{L_k} \leq ck\} = (ce^{1-c})^{(1+o(1))k},$$

and by a union bound,

$$\begin{aligned}\mathbf{P}\{M_k \leq ck\} &\leq 2^k \mathbf{P}\{D_{L_k} \leq ck\} \\ &= (2ce^{c-1})^{(1+o(1)k} \\ &= f(c)^{(1+o(1))k},\end{aligned}$$

where $f(c) = 2ce^{c-1}$ as in Chapter 1. When $c < c'$, $f(c) < 1$, so there is k_0 such that for $k \geq k_0$,

$$\mathbf{P}\{M_k \leq ck\} \leq f(c)^{k/2},$$

say. It follows by Borel-Cantelli that with probability one, $M_k > ck$ for all but finitely many k. Since $c < c'$ was arbitrary the result follows. □

Corollary 21. *With probability one,*

$$\limsup_{n\to\infty} \frac{H_n}{n} \le c^*.$$

The proof of the corollary is essentially a reprise in reverse of the argument for Proposition 19 from Devroye's theorem, and we leave it as an exercise. Of course, together with Proposition 16, this yields almost sure convergence in Devroye's theorem. In the next section, we will see how the same ideas apply to bound the minimum of very general branching random walks.

Exercises

2.1.1 Prove Corollary 21.

2.2 More general binary branching random walks

For now, we continue to restrict our attention to the complete binary tree T_∞. Instead of weighting the edges of T_∞ with exponentials, we may equally well consider weighting them with independent copies of some real random variable X, which we call the *step size*, and which may take both positive and negative values. [1] Let $\mathcal{X} = \{X_e : e \in E(T_\infty)\}$ and for $v \in V(T_\infty)$, let $S_v = S_v(T_\infty, \mathcal{X}) = \sum_{e\prec v} X_e$ as before. Then for positive integers k let $M_k = M_k(T_\infty, \mathcal{S}) = \min\{S_v : v \in T_{\infty,k}\}$. We seek to understand the behaviour of M_k for k large – in particular, we wish to understand when a law of large numbers holds for M_k.

Since $T_{\infty,k}$ has 2^k elements, our earlier heuristic suggests looking for M_k near the value c where $\mathbf{P}\{S_k \le ck\}$ is roughly 2^{-k}, at least at a logarithmic scale. In order for such c to exist, the step size must itself have exponential lower tails. In other words (see Exercise 2.2.1), there must be $\lambda > 0$ such that $\mathbf{E}\left[e^{-\lambda X}\right] < \infty$. In principle, it seems that the knowledge that such a λ exists, should be enough to prove a law of large numbers for M_k. Embarrassingly, no one has yet been able to do so under only this assumption, for technical reasons we shall meet shortly.

We now recall the workings of Chernoff's bound in more detail. We define the *logarithmic moment generating function* $\Lambda : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ by

$$\Lambda(t) = \Lambda_X(t) := \ln \mathbf{E}\left[e^{tX}\right],$$

We usually suppress the X in the subscript since it will be clear from context. Since $S_{L_k} = \sum_{i=1}^k X_{L_{i-1},L_i}$ is a sum of n independent copies of X, for any $c < \mathbf{E}[X]$ and $t < 0$, by using Markov's inequality and independence, we have

$$\mathbf{P}\{S_k \le ck\} = \mathbf{P}\left\{e^{tS_k} > e^{tck}\right\} \le \frac{\mathbf{E}\left[e^{tS_k}\right]}{e^{tck}} = \left(\mathbf{E}\left[e^{t(X-c)}\right]\right)^k = e^{-k(tc-\Lambda(t))},$$

by definition of $\Lambda(t)$. We choose the value of c that minimizes this upper bound:

$$\mathbf{P}\{S_k \le ck\} \le \exp\left(-k \sup_{t<0}\{tc - \Lambda(t)\}\right). \tag{2.2}$$

The optimal choice for t in (2.2) is then that for which $\Lambda'(t) = c$ —if such a t exists— as may be informally seen by differentiating $t \mapsto tc - \Lambda(t)$ with respect to t. Choosing t in this fashion and writing $\Lambda'(t)$ in place of c yields

$$\mathbf{P}\{S_k \le \Lambda'(t)k\} \le e^{-k(t\Lambda'(t)-\Lambda(t))}. \tag{2.3}$$

In many situations, this bound is only off by a factor of order $k^{-1/2}$. (This is the factor mentioned just after Corollary 5.) We can recover this factor, and see that it is tight, by an exponential

[1] As mentioned earlier, independence between sibling edges is not really important – but its presence or absence changes nothing so we assume complete independence to simplify notation.

change of measure applied to the summands of S_k. We will do so in a moment, but first we see how (2.3) yields lower bounds on M_k.

Let $\mathcal{D} = \mathcal{D}_X$ be the set of values t for which $\Lambda(t) < \infty$. Let $\mathcal{D}^o = \mathcal{D}^o_X$ be the interior of $\mathcal{D}$. Let also $g(t) = g_X(t) = t\Lambda'(t) - \Lambda(t)$. The function Λ is infinitely differentiable and convex on $\mathcal{D}^o$, and f is strictly convex on $\mathcal{D}^o$ (see [3], Lemma 2.2.5 and Exercise 2.2.24).

Suppose that there exists $t^* \in \mathcal{D}^o$, $t^* < 0$ for which $g(t^*) = \ln 2$. Such a t^*, if it exists, is necessarily unique by the convexity of f. Furthermore, for any $t < t^*$, $t \in \mathcal{D}^o$, we will have $g(t) > \ln 2$, again by convexity. Then from (2.3), we see that for such t,

$$\mathbf{P}\left\{S_k \le \Lambda'(t)\right\} \le e^{-kg(t)} \le c^{-k}$$

for some $c > 2$. It follows by a union bound and Borel-Cantelli, then optimizing over $t < t^*$ that

$$\liminf_k M_k/k \ge \Lambda'(t^*). \tag{2.4}$$

This lower bound on M_k precisely corresponds to the upper bound on H_n proved earlier. (Indeed, this is essentially the same proof as we gave in Section 1.5 to upper bound H_n.)

We now turn to the exponential change of measure. Let F be the distribution function of X. We remark that for $t \in \mathcal{D}^o$,

$$\Lambda'(t) = \frac{\mathbf{E}\left[Xe^{tX}\right]}{\mathbf{E}\left[e^{tX}\right]} \qquad \text{and} \qquad \Lambda''(t) = \frac{\mathbf{E}\left[X^2e^{tX}\right]}{\mathbf{E}\left[e^{tX}\right]} - \left(\frac{\mathbf{E}\left[Xe^{tX}\right]}{\mathbf{E}\left[e^{tX}\right]}\right)^2.$$

Consider the random variable Y_t, with distribution function G_t defined by

$$G_t(x) = \frac{1}{\mathbf{E}\left[e^{tX}\right]}\int_{-\infty}^{x} e^{ty}\,dF(y).$$

Then, we have

$$\begin{aligned} \mathbf{E}\left[Y_t\right] &= \int_{-\infty}^{\infty} y\,dG_t(y) = \frac{1}{\mathbf{E}\left[e^{tX}\right]}\int_{-\infty}^{\infty} xe^{tx}\,dF(x) = \Lambda'(t), \qquad (2.5)\\ \mathbf{E}\left[Y_t^2\right] &= \int_{-\infty}^{\infty} y^2\,dG_t(y) = \frac{1}{\mathbf{E}\left[e^{tX}\right]}\int_{-\infty}^{\infty} x^2e^{tx}\,dF(x) = \Lambda''(t) + \Lambda'(t)^2, \end{aligned}$$

so the random variable $Z_t = Y_t - \Lambda'(t)$ is such that $\mathbf{E}\left[Z_t\right] = 0$ and $\mathbf{Var}\left\{Z_t\right\} = \mathbf{Var}\left\{Y_t\right\} = \Lambda''(t)$.[2]

We now show that we may express the probability of events such as $\{X_1 + \cdots + X_k \le ck\}$ – which are exponentially unlikely when $c < \mathbf{E}\left[X\right]$ – in terms of the distribution of the sum $Z_1 + \cdots + Z_k$ of i.i.d. copies of Z_t in the central regime. This allows for the use of precise limit results related to the central limit theorem.

Let $S_k = X_1 + \ldots + X_k$ and fix any $a \in \mathbb{R}$ and $c < \mathbf{E}\left[X\right]$, and assume there exists $t \in \mathcal{D}^o$ with $t < 0$ such that $c = \Lambda'(t)$. Then using the same change of measure as in (2.5), we have

$$\begin{aligned} \mathbf{P}\left\{S_k \le ck + a\right\} &= \int_{\{x_1+\cdots+x_k \le ck+a\}} dF(x_1)\cdots dF(x_k) \\ &= e^{k\Lambda(t)}\int_{\{y_1+\ldots+y_k \le ck+a\}} e^{-t(y_1+\ldots+y_k)}\,dG_t(y_1)\cdots dG_t(y_k). \end{aligned}$$

Now let Z_t be a random variable having distribution function H_t satisfying $dH_t(z) = e^{t\Lambda'(t)}dG_t(z)$. Then this further change of measure yields

$$\begin{aligned} \mathbf{P}\left\{S_k \le \Lambda'(t)k + a\right\} &= e^{k\Lambda(t)}\int_{\{z_1+\ldots+z_k \le a\}} e^{-t(z_1+\ldots+z_k)}e^{-t\Lambda'(t)k}\,dH_t(z_1)\cdots dH_t(z_k) \\ &= e^{-kg(t)}\int_{\{z_1+\ldots+z_k \le a\}} e^{-t(z_1+\ldots+z_k)}\,dH_t(z_1)\cdots dH_t(z_k). \end{aligned}$$

[2]It may also easily be checked that $\mathbf{E}\left[|Y_t|^3\right] < \infty$, a fact we will use later.

Writing W_k for the distribution function of $Z_1+\ldots+Z_k$, k i.i.d. copies of Z_t, the preceding equation asserts that

$$\mathbf{P}\{S_k \le \Lambda'(t)k + a\} = e^{-kg(t)} \int_{-\infty}^{\infty} e^{-ts}\mathbf{1}_{[s\le a]}\, dW_k(s). \tag{2.6}$$

To make this result useful, we need one more ingredient, which is a local central limit theorem. We say X is a *lattice random variable* with period $1/d > 0$ if there is a constant z such that $dX - z$ is an integer random variable and d is the smallest positive real number for which this holds; in this case, we say that the set $\mathbb{L}_X = \{(n+z)/d : n \in \mathbb{Z}\}$ is *the lattice of* X. If X is not a lattice random variable then we say it is *non-lattice*. The following is a weakening of Theorem 1 from [8].

Theorem 22 ([8])**.** *Fix any $b > 0$. If $\mathbf{E}X = 0$ and $0 < \mathbf{E}\left[X^2\right] < \infty$ then for any $h > 0$, if X is non-lattice then for all x with $|x| \le b\sqrt{n}$,*

$$\mathbf{P}\{x \le S_n < x+h\} = (1+o(1))\frac{h\cdot e^{-x^2/(2n\mathbf{E}[X^2])}}{\sqrt{2\pi \mathbf{E}\left[X^2\right] n}},$$

and if X is lattice then for all $x \in \mathbb{L}_X$ with $|x| \le b\sqrt{n}$,

$$\mathbf{P}\{S_n = x\} = (1+o(1))\frac{e^{-x^2/(2n\mathbf{E}[X^2])}}{\sqrt{2\pi \mathbf{E}\left[X^2\right] n}}.$$

In both cases, $o(1) \to 0$ as $n \to \infty$ uniformly over all x in the allowed range.

Applying Theorem 22 in (2.6) (for simplicity assuming that X – and so Z – is non-lattice), we obtain the following result.

Proposition 23. *Assume X is non-lattice, and fix any $b > 0$ and $t < 0$, $t \in \mathcal{D}^o$. Then for positive integers k and uniformly over a with $|a| \le b\sqrt{k}$ we have*

$$\mathbf{P}\{S_k \le \Lambda'(t)k + a\} \asymp \frac{e^{-kg(t)+a|t|}}{\sqrt{k}}$$

Proof. By (2.6) and Theorem 22, we have

$$\begin{aligned}
\mathbf{P}\{S_k \le \Lambda'(t)k + a\} &= e^{-kg(t)} \int_{-\infty}^{\infty} e^{-ts}\mathbf{1}_{[s\le a]}\, dW_k(s) \\
&\asymp (e^{-kg(t)}) \sum_{i=1}^{\infty} \int_{(a-i,a-i+1]} e^{-ts}\, dW_k(s) \\
&\asymp \frac{e^{-kg(t)-at}}{\sqrt{k}} \\
&= \frac{e^{-kg(t)+a|t|}}{\sqrt{k}}.
\end{aligned}$$

∎

Remark. It is in fact possible to be more careful and derive not only the correct order of $\mathbf{P}\{S_k \le \Lambda'(t)k + a\}$, but its asymptotic value – this is the substance of the Bahadur–Rao theorem [1] (see also Dembo and Zeitouni [3], Theorem 3.7.4).

Now take any $t > t^*$. Then $g(t) < \ln 2$, so by Proposition 23 and linearity of expectation we have that

$$\mathbf{E}\left[|\{L \in T_{\infty,k} \;:\; S_L \le \Lambda'(t)k\}\right] = \Theta\left(\frac{e^{k(\ln 2 - g(t))}}{\sqrt{k}}\right),$$

which grows exponentially in k. In particular there is k_0 such that $\mathbf{E}\left[|\{L \in T_{\infty,k_0} \ : \ S_L \leq \Lambda'(t)k_0\}\right] > 1$, so we may find an embedded branching process and use an amplification argument, then optimize over $t > t^*$ to show that almost surely,

$$\limsup_{k\to\infty} \frac{M_k}{k} \leq \Lambda'(t^*).$$

Together with (2.4), we have thus showed that $\lim_{k\to\infty} \frac{M_k}{k} = \Lambda'(t^*)$ almost surely. In the next section, we explore the lower-order terms of M_k. Before doing so, we state without proof a theorem about how the above result generalizes to branching processes with branch factor B which is not identically 2.

Let B be a non-negative integer random variable with $1 < \mathbf{E}B < \infty$, let $\mathcal{T}$ be a Galton-Watson process with branching distribution B, and write $\mathcal{T}_k$ for the vertices in the k'th generation of $\mathcal{T}$. Fix a real random variable X and let $\{X_e \ : e \in E(\mathcal{T})\}$ be copies of X, independent except perhaps on sibling edges. Then for each $v \in V(\mathcal{T})$ let $S_v = \sum_{e \prec v} X_e$, and for positive integers k let $M_k = \min\{S_v \ : \ v \in \mathcal{T}_k$.

Theorem 24 (Hammersley–Kingman–Biggins Theorem [2, 5, 6])**.** *Suppose there is $t^* < 0$, $t^* \in \mathcal{D}^o$ such that $g(t^*) = \ln(\mathbf{E}[B])$. Then*

$$\lim_{k\to\infty} \frac{M_k}{k} = \Lambda'(t^*),$$

almost surely and in expectation.

The H–K–B theorem is in fact more general, applying in situations where B can be infinite and the displacements from parent to children are given by a point process on $\mathbb{R}$. However, we will not take the time to precisely formulate the most general version here. Biggins [2] also treats the case that $\mathcal{T}$ is a multitype branching random walk.

Exercises

2.2.1 Let X be a real random variable. Show that there exists $\epsilon > 0$ such that $\mathbf{P}\{X \leq c\} \leq e^{\epsilon c}$ for all sufficiently small $c < 0$, if and only if $\mathbf{E}\left[e^{-tX}\right] < \infty$ for some $t > 0$.

2.2.2 Check that if $X \stackrel{\mathrm{d}}{=} \mathrm{Exponential}(1)$, then $\mathbf{E}\left[e^{tX}\right] = \infty$ for all $t < 0$. Find the value t^* for which $g(t^*) = \ln 2$, and find $\Lambda'(t^*)$.

2.2.3 Show that the Hammersley–Kingman–Biggins Theorem does not hold if we weaken the independence assumption to only require independence of edges that do not share a common endpoint. (The difference between this and the previous independence assumption is that now, for a vertex v, the edge from v to its parents may not be independent of the edges from v to its children.)

Bibliography

[1] R.R. Bahadur and R. Ranga Rao. On deviations of the sample mean. *Annals of Mathematical Statistics*, 31:1015–1027, 1960.

[2] J. D. Biggins. The first and last-birth problems in a multitype age-dependent branching process. *Advances in Applied Probability*, 8:446–459, 1976.

[3] A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications.* Springer, second edition, 1998.

[4] L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33:489–498, 1986.

[5] J.M. Hammersley. Postulates for subadditive processes. *The Annals of Probability*, 2(4):652–680, 1974.

[6] J.F.C. Kingman. The first birth problem for an age-dependent branching process. *The Annals of Probability*, 3(5):790–801, 1975.

[7] B. Pittel. On growing random binary trees. *Journal of Mathematical Analysis and Applications*, 103:461–480, 1984.

[8] Charles J. Stone. On local and ratio limit theorems. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 217–224, 1965.