*Popular tradition may be used to defend what seems irrational, and you can also say
that sometimes it is not irrational, for it is likely that unlikely things should happen.*
— Aristotle, *Poetics* (c.335 BCE)

*But, on the other hand, Uncle Abner said that the person that had took a bull by the
tail once had learnt sixty or seventy times as much as a person that hadn't, and said
a person that started in to carry a cat home by the tail was gitting knowledge that
was always going to be useful to him, and warn't ever going to grow dim or doubtful.*
— Mark Twain, *Tom Sawyer Abroad* (1894)

*4   Tail Inequalities

The simple recursive structure of skip lists made it relatively easy to derive an upper bound
on the expected *worst-case* search time, by way of a stronger high-probability upper bound on
the worst-case search time. We can prove similar results for treaps, but because of the more
complex recursive structure, we need slightly more sophisticated probabilistic tools. These tools
are usually called **tail inequalities**; intuitively, they bound the probability that an integer random
variable takes a value far away from the mean, in the *tails* of the probability distribution.

4.1   Markov's Inequality

Perhaps the simplest tail inequality was named after the Russian mathematician Andrey Markov;
however, in strict accordance with Stigler's Law of Eponymy, it first appeared in the works of
Markov's probability teacher, Pafnuty Chebyshev.[1]

Markov's Inequality. *Let Z be a non-negative integer random variable. For any real number
$z > 0$, we have $\Pr[Z \geq z] \leq \mathrm{E}[Z]/z$.*

Proof: First, the definition of expectation implies

$$\Pr[Z \geq 1] \;=\; \sum_{z \geq 1} \Pr[Z = z] \;\leq\; \sum_{z \geq 0} z \cdot \Pr[Z = z] \;=\; \mathrm{E}[Z].$$

For any real $z > 0$, applying this argument to the integer random variable $\lfloor Z/z \rfloor$ gives us

$$\Pr[Z \geq z] \;=\; \Pr\left[\left\lfloor \frac{Z}{z} \right\rfloor \geq 1\right] \;\leq\; \mathrm{E}\left[\left\lfloor \frac{Z}{z} \right\rfloor\right] \;\leq\; \mathrm{E}\left[\frac{Z}{z}\right] \;=\; \frac{\mathrm{E}[Z]}{z}. \qquad \square$$

If that proof was a little opaque, the following visualization may be be helpful. The figure
on the next page shows the graph of $\Pr[Z \geq z]$ as a function of z. Because Z takes only integer
values, this graph consists entirely of horizontal and vertical line segments. By splitting the region
between this curve and the coordinate axes into horizontal rectangles, we see that the area of this
region is exactly $\sum_z \Pr[Z \geq z] = \sum_z z \cdot \Pr[Z = z] = \mathrm{E}[Z]$. On the other hand, for any particular
value of z, the rectangle with width z and height $\Pr[Z \geq z]$ (shaded darker in the figure) fits
entirely under the curve. We conclude that $z \cdot \Pr[Z \geq z] \leq \mathrm{E}[Z]$.

---

[1]The closely related tail bound traditionally called Chebyshev's inequality was actually discovered by the French
statistician Irénée-Jules Bienaymé, a friend and colleague of Chebyshev's. Just to be extra confusing, some sources
refer to what we're calling Markov's inequality as "Chebyshev's inequality" or "Bienaymé's inequality".

Proof of Markov's inequality.
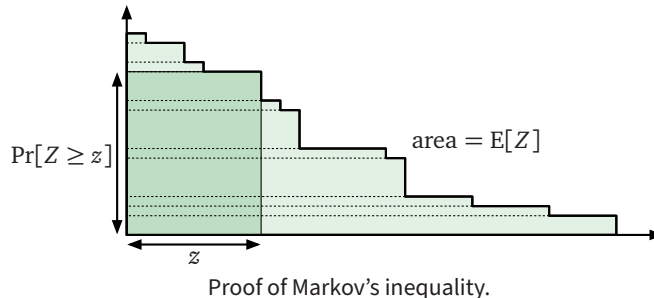
Markov's inequality implies the following bound on the probability that any random variable X is significantly larger than its expectation:

$$\Pr[X \geq (1+\delta)\operatorname{E}[X]]] \leq \frac{1}{1+\delta}.$$

Markov's inequality makes absolutely no assumptions about the underlying probability distribution; it holds for *arbitrary* integer random variables. Unfortunately, the bounds that Markov's inequality (directly) implies are generally too weak to be useful. (For example, Markov's inequality implies that with high probability, every node in an n-node treap has depth $O(n^2 \log n)$. Well, *duh!*) To get stronger bounds, we need to assume (and exploit) some additional structure in our random variables.

4.2   Independence

Two random variables are ***independent*** if knowing the value of one variable gives us no additional knowledge about the distribution of the other. More formally, X and Y are independent if and only if

$$\Pr[X = x \ \wedge \ Y = y] = \Pr[X = x] \cdot \Pr[Y = y],$$

or equivalently, if and only if

$$\Pr[X = x \mid Y = y] = \Pr[X = x]$$

for all possible values x and y. For example, two flips of the same (ideal) fair coin are independent, but the number of heads and the number of tails in a sequence of n coin flips are not independent (because they must add to n). Variables that are not independent are called ***dependent***. Independence has two important consequences.

- The expectation of the product of two independent random variables is equal to the product of their expectations $\operatorname{E}[X \cdot Y] = \operatorname{E}[X] \cdot \operatorname{E}[Y]$.

- If X and Y are independent, then for any function f, the random variables $f(X)$ and $f(Y)$ are also independent.

Neither of these properties hold for dependent random variables.

More generally, a collection of random variables $X_1, X_2, \ldots, X_n$ are said to be ***mutually independent*** (or ***fully independent***) if and only if

$$\Pr\left[\bigwedge_{i=1}^{n}(X_i = x_i)\right] = \prod_{i=1}^{n}\Pr[X_i = x_i]$$

or equivalently, if and only if

$$\Pr\left[X_n = x_n \mid \bigwedge_{i=1}^{n-1}(X_i = x_i)\right] = \Pr[X_n = x_n]$$

for all possible values $x_1, x_2, \ldots, x_n$. Mutual independence of a set of random variables implies that the expectation of their product is equal to the product of their expectations:

$$E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E[X_i].$$

Finally, if $X_1, X_2, \ldots, X_n$ are mutually independent, then for any function f, the random variables $f(X_1), f(X_2), \ldots, f(X_n)$ are also mutually independent.

In some contexts, especially in the analysis of hashing, we can only realistically assume a limited form of independence. A set of random variables is **pairwise independent** if every pair of variables in the set is independent. More generally, for any positive integer k, we say that a set of random variables is **k-wise independent** if every subset of size k is mutually independent. A set of variables is fully independent if and only if it is k-wise independent for *all* k.

Every k-wise independent set of random variables is also j-wise independent for all $j < k$, but the converse is not true. For example, let X, Y, and Z be independent random bits, and let $W = (X + Y + Z) \bmod 2$. The set of random variables $\{W, X, Y, Z\}$ is 3-wise independent but not 4-wise (or fully) independent.

4.3 Chebyshev's Inequality

Most of our analysis of randomized algorithms and data structures boils down to analyzing sums of indicator variables.[2] So far we have only been interested in the expected value of these sums, but we can prove stronger conditions under various assumptions about independence.

Consider a collection $X_1, X_2, \ldots, X_n$ of random bits, and for each index i, let $p_i = E[X_i] = \Pr[X_i = 1]$. Let X denote the sum $\sum_i X_i$ and let $\mu = E[X]$; linearity of expectation immediately implies that $\mu = \sum_i p_i$.

Chebyshev's Inequality. *If the indicator variables $X_1, X_2, \ldots, X_n$ are pairwise independent, then $\Pr[(X - \mu)^2 \geq z] < \mu/z$ for all $z > 0$.*

Proof: For each index i, let $Y_i := X_i - p_i$, and let $Y := \sum_i Y_i = X - \mu$.

$$E[Y^2] = E\left[\sum_{i,j} Y_i Y_j\right] \qquad\qquad \text{[definition of } Y]$$

$$= \sum_{i,j} E[Y_i Y_j] \qquad\qquad \text{[linearity]}$$

$$= \sum_{i} E[Y_i^2] + \sum_{i \neq j} E[Y_i Y_j]$$

$$= \sum_{i} E[Y_i^2] + \sum_{i \neq j} E[Y_i] \cdot E[Y_j] \qquad\qquad \text{[pairwise independence]}$$

---

[2]The focus on sums of indicator variables is a feature (or if you prefer, a handicap) of these lecture notes, not of randomized algorithm analysis more broadly!

$$= \sum_i \mathrm{E}[Y_i^2] + 0 \qquad\qquad\qquad [\mathrm{E}[Y_i] = 0, \text{ by linearity}]$$

$$= \sum_i \left((p_i(1-p_i)^2 + (1-p_i)(0-p_i)^2 \right) \qquad\qquad [\text{definition of E}]$$

$$= \sum_i p_i(1-p_i) \; < \; \sum_i p_i \; = \; \mu$$

Because the random variable Y^2 is non-negative, Markov's inequality now directly implies the bound $\Pr[Y^2 \geq z] \leq \mathrm{E}[Y^2]/z < \mu/z$, which completes the proof. ☐

The quantity $\mathrm{E}[Y^2] = \mathrm{E}[(X-\mu)^2]$ is commonly known as the **variance** of X.

Chebyshev's inequality immediately gives us significantly tighter bounds than Markov's inequality when the component indicator variables X_i are pairwise independent. The following bounds hold for all positive real numbers Δ and δ:

$$
\begin{array}{ll}
\Pr[X \geq \mu + \Delta] < \dfrac{\mu}{\Delta^2} & \Pr[X \geq (1+\delta)\mu] < \dfrac{1}{\delta^2 \mu} \\[3ex]
\Pr[X \leq \mu - \Delta] < \dfrac{\mu}{\Delta^2} & \Pr[X \leq (1-\delta)\mu] < \dfrac{1}{\delta^2 \mu}
\end{array}
$$

The inequalities on the left are called *additive* tail bounds; the inequalities on the right are called *multiplicative* tail bounds. The inequalities in the top row bounds the *upper tail* of X's probability distribution; the inequalities in the bottom row bound the *lower tail*.

An important consequence of Chebyshev's inequality is the so-called **Law of Large Numbers**, which states that if we repeat the same experiment many times, the statistical average of the outcomes tends toward the expected value of a single experiment with overwhelming probability. More formally, let $X_1, X_2, X_3, \ldots$ be independent random bits, with $\Pr[X_i = 1] = p$ for all i, and for any index n, let $\overline{X}_n = \sum_{i=1}^n X_i / n$ denote the mean of the first n bits. The **weak** law of large numbers states that

$$\lim_{n \to \infty} \Pr\left[|\overline{X}_n - p| > \varepsilon \right] = 0$$

for any real $\varepsilon > 0$; this law follows almost immediately from Chebyshev's inequality. The **strong** law of large numbers states that

$$\Pr\left[\lim_{n \to \infty} \overline{X}_n = p \right] = 1;$$

the proof of this law is considerably more involved.

4.4 Higher Moment Inequalities

Chebyshev's inequality and its corollaries are special cases of a more general family of so-called **moment inequalities**, which consider the distribution of higher even powers of the random variable X under stronger assumptions about the independence of the underlying indicator variables X_i. The kth **central moment** of X is the quantity $\mathrm{E}[(X-\mu)^k]$. Here I'll state a general result without proof; for each k, the derivation follows the same general outline as the proof of Chebyshev's inequality above, but with more, you know, math.

kth Moment Inequality. *For any fixed integer $k > 0$, if the variables $X_1, X_2, \ldots, X_n$ are $2k$-wise independent, then $\Pr[(X-\mu)^{2k} \geq z] = O(\mu^k/z)$ for all $z > 0$. The hidden constant in the $O(\cdot)$ notation depends on k.*

These inequalities immediately imply the following asymptotic tail bounds for all positive real numbers Δ and δ; again, the hidden $O(\cdot)$ constants depend on the fixed parameter k.

$$\Pr[X \geq \mu + \Delta] = O\left(\left(\frac{\mu}{\Delta^2}\right)^k\right) \qquad \Pr[X \geq (1+\delta)\mu] = O\left(\left(\frac{1}{\delta^2\mu}\right)^k\right)$$

$$\Pr[X \leq \mu - \Delta] = O\left(\left(\frac{\mu}{\Delta^2}\right)^k\right) \qquad \Pr[X \geq (1-\delta)\mu] = O\left(\left(\frac{1}{\delta^2\mu}\right)^k\right)$$

In short, more independence in the indicator variables X_i implies thinner tails in the distribution of X, and therefore higher concentration of the distribution of X near its mean.

4.5 Chernoff Bounds

For *fully* independent random variables, even tighter tail bounds were developed in the early 1950s by Herman Chernoff, who was then a mathematics professor at the University of Illinois. However, in strict accordance with Stigler's Law, the first published version of "Chernoff bounds", which appeared in a 1952 paper by Chernoff that gave the bounds their name, were actually due to his colleague Herman Rubin.[3]

Exponential Moment Inequality. *If the indicator variables $X_1, X_2, \ldots, X_n$ are fully independent, then $E[\alpha^X] \leq e^{(\alpha-1)\mu}$ for any $\alpha \geq 1$.*

Proof: The definition of expectation immediately implies that

$$E[\alpha^{X_i}] = p_i \alpha^1 + (1-p_i)\alpha^0 = (\alpha-1)p_i + 1.$$

Thus, by The World's Most Useful Inequality $1 + t \leq e^t$, we have

$$E[\alpha^{X_i}] \leq e^{(\alpha-1)p_i}.$$

Full independence of the X_i's now immediately implies

$$E[\alpha^X] = \prod_i E[\alpha^{X_i}] \leq \prod_i e^{(\alpha-1)p_i} = e^{(\alpha-1)\mu}$$

and we're done. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Chernoff Bound (upper tail). *If the indicator variables $X_1, X_2, \ldots, X_n$ are fully independent, then $\Pr[X \geq x] \leq e^{x-\mu}(\mu/x)^x$ for all $x \geq \mu$.*

Proof: Consider any fixed value $x \geq \mu$. The function $t \mapsto (x/\mu)^t$ is monotonically increasing, so

$$\Pr[X \geq x] = \Pr\left[\left(\frac{x}{\mu}\right)^X \geq \left(\frac{x}{\mu}\right)^x\right].$$

Markov's inequality implies that

$$\Pr\left[\left(\frac{x}{\mu}\right)^X \geq \left(\frac{x}{\mu}\right)^x\right] \leq \frac{E[(x/\mu)^X]}{(x/\mu)^x}.$$

---

[3]"Since that seemed to be a minor lemma in the ensuing paper I published (Chernoff, 1952), I neglected to give [Rubin] credit. I now consider it a serious error in judgment, especially because his result is stronger, for the upper bound, than the asymptotic result I had derived."

Finally, applying the exponential moment inequality at $\alpha = x/\mu$ gives us

$$E[(x/\mu)^X] \leq e^{((x/\mu)-1)\mu} = e^{x-\mu},$$

which completes the proof.   □

A nearly identical argument implies a similar bound on the lower tail:

Chernoff Bound (lower tail). *If the indicator variables $X_1, X_2, \ldots, X_n$ are fully independent, then $\Pr[X \leq x] \leq e^{x-\mu}(\mu/x)^x$ for all $x \leq \mu$.*

Proof: For any $x \leq \mu$, we immediately have $\Pr[X \leq x] = \Pr[(x/\mu)^X \geq (x/\mu)^x]$. (The direction of the inequality changes because $x/\mu \leq 1$.) The remainder of the proof is unchanged.   □

The particular value x/μ in these results may seem arbitrary, but in fact it's chosen vary carefully. The same arguments imply that

$$\Pr[X \geq x] \leq \frac{e^{(\alpha-1)\mu}}{\alpha^x} \text{ for } \textit{all } \alpha > 1 \qquad \text{and} \qquad \Pr[X \leq x] \leq \frac{e^{(\alpha-1)\mu}}{\alpha^x} \text{ for } \textit{all } \alpha < 1.$$

A bit of calculus implies that the right sides of these inequalities are minimized when $\alpha = x/\mu$.

Direct substitution now implies the following more traditional forms of Chernoff bounds, for any positive reals Δ and δ. Unlike the polynomial moment bounds we derived earlier, Chernoff bounds for the upper and lower tails of X are asymmetric.

$$\boxed{\begin{aligned} \Pr[X \geq \mu + \Delta] \leq e^{\Delta}\left(\frac{\mu}{\mu+\Delta}\right)^{\mu+\Delta} \qquad & \Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \\ \Pr[X \leq \mu - \Delta] \leq e^{-\Delta}\left(\frac{\mu}{\mu-\Delta}\right)^{\mu-\Delta} \qquad & \Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \end{aligned}}$$

It is sometimes more convenient to work with the following looser forms of the multiplicative Chernoff bounds, which hold for any $0 < \delta < 1$. Here we use the notation $\exp(t) := e^t$.

$$\boxed{\begin{aligned} \Pr[X \geq (1+\delta)\mu] &\leq \exp\left(-\delta^2\mu/3\right) \\ \Pr[X \leq (1-\delta)\mu] &\leq \exp\left(-\delta^2\mu/2\right) \end{aligned}}$$

These looser bounds can be derived from the more precise forms above using the Taylor-series approximations $1/\ln(1+\delta) \leq 1/2 + 1/\delta$ and $\ln(1-\delta) \geq -\delta + \delta^2/2$. I'll omit the straightforward but grungy details. (You're welcome.)

4.6   Coin Flips

To give an elementary example of applying these tail inequalities, consider an experiment where we independently flip N perfectly fair coins, and we want to bound the probability of getting at least αN heads, for some constant $1/2 < \alpha < 1$.

For each index i, let $X_i = 1$ if the ith flip comes up heads and $X_i = 0$ otherwise, so that $X = \sum_{i=1}^{N} X_i$ is the number of heads. Because the coins are fair, $\Pr[X_i = 1] = 1/2$ for all i. Linearity of expectation immediately implies $\mu = E[X] = N/2$.

- Markov's inequality implies the nearly trivial bound $\Pr[X \geq \alpha N] \leq 1/2\alpha$. In particular, we have $\Pr[X \geq 3N/4] \leq 2/3$ and $\Pr[X \geq 9N/10] \leq 5/9$.

- Chebyshev's inequality implies $\Pr[X \geq \alpha N] \leq 1/((2\alpha - 1)N) = O(1/(\alpha N))$. In particular, we have $\Pr[X \geq 3N/4] \leq 2/N$ and $\Pr[X \geq 9N/10] \leq 5/(4N)$.

- Higher moment inequalities implies smaller inverse-polynomial upper bounds. For example, the fourth-moment inequality implies $\Pr[X \geq \alpha N] \leq O(1/(\alpha N)^2)$.

- The simple version of Chernoff's inequality implies the inverse-exponential upper bound

$$\Pr[X \geq \alpha N] \leq \exp\left(-\frac{(2\alpha - 1)^2}{6} \cdot N\right).$$

In particular, we have $\Pr[X \geq 3N/4] < e^{-N/24}$ and $\Pr[X \geq 9N/10] < e^{-8N/75} < e^{-N/10}$.

- Symmetrically, Chernoff's inequality for the *lower* tail symmetrically bounds the probability of getting *few* heads as follows:

$$\Pr[X \leq \alpha N] \leq \exp\left(-\frac{(1 - 2\alpha)^2}{4} N\right).$$

In particular, we have $\Pr[X \leq N/4] < e^{-N/16}$ and $\Pr[X \leq N/10] < e^{-4N/25} \leq e^{-N/7}$.

- A more specialized argument involving binomial coefficients, which relies on all coins being fair, implies

$$\Pr[X \leq \alpha N] \leq \left(2\alpha^\alpha (1-\alpha)^{1-\alpha}\right)^{-N} \quad \text{and} \quad \Pr[X \geq \alpha N] \leq \left(2\alpha^\alpha (1-\alpha)^{1-\alpha}\right)^{-N}.$$

In particular, we have

$$\Pr[X \leq N/4] \leq \left(2\left(\frac{1}{4}\right)^{1/4}\left(\frac{3}{4}\right)^{3/4}\right)^{-N} \approx e^{-0.13081N} < e^{-N/8}$$

and

$$\Pr[X \leq N/10] \leq \left(2\left(\frac{1}{10}\right)^{1/10}\left(\frac{9}{10}\right)^{9/10}\right)^{-N} \approx e^{-0.36806N} < e^{-N/3}.$$

4.7 Back to Treaps, Take 1

Consider an arbitrary node k in an n-node treap. We can argue informally that $depth(k) = O(\log n)$ with high probability by following the "good/bad pivot" analysis of randomized quicksort as follows.

For each integer $0 \leq i \leq depth(k)$, let v_i denote the ancestor of node k at depth i in the treap, and let n_i denote the number of nodes in the subtree rooted at v_i. (Thinking in terms of randomized quicksort, n_i is the size of the subproblem containing k after i levels of recursion.) For any integer $i > 0$, we label level i **good** or **bad** as follows:

- If $i < depth(k)$, then level i is good if $n_{i-1}/4 \leq n_i \leq 3n_{i-1}/4$ and bad otherwise.

- If $i \geq depth(k)$, then level i is good or bad according to an independent fair coin flip.

Each level $i \geq 1$ is good with probability at least $1/2$, and the events $[i$ is good$]$ are mutually independent. There are at most $\log_{4/3} n$ good levels smaller than $depth(k)$, since otherwise $n_{depth(k)}$ would be less than 1.

Thus, the depth of node k is *at most* the number of fair coin flips required to get $\log_{4/3} n$ heads. Equivalently, for any constant c, the probability that $depth(k) \geq c \log_{4/3} n$ is at most the probability that we see at most $\log_{4/3} n$ heads in a sequence of $c \log_{4/3} n$ independent fair coin flips. Our earlier analysis of coin flips via Chernoff bounds, with $N = c \log_{4/3} n$ and $\alpha = 1/c$, implies

$$\Pr\left[depth(k) \geq c \log_{4/3} n\right] \leq \exp\left(-\frac{(1-2/c)^2}{4} c \log_{4/3} n\right) = n^{-(c-2)^2/(4c\ln(4/3))}.$$

For any $c > 2$, we obtain an inverse-polynomial upper bound, which we can make arbitrarily small by increasing the constant c. For example, setting $c = 20\ln(4/3) \approx 5.75364$ gives us $\Pr[depth(k) > 20\ln n] < 1/n^2$. We conclude that $depth(k) = O(\log n)$ with high probability.

Our more specialized argument implies that

$$\Pr\left[depth(k) \geq 10\log_{4/3} n\right] < e^{-10\log_{4/3} n/3} = n^{-10/(3\ln(4/3))} < \frac{1}{n^{11}}.$$

4.8   Back to Treaps: Take 2

In our analysis of randomized treaps, we wrote $i \uparrow k$ to indicate that the node with the ith smallest key ('node i') was a proper ancestor of the node with the kth smallest key ('node k'). We argued that

$$\Pr[i \uparrow k] = \frac{[i \neq k]}{|k-i|+1},$$

and from this we concluded that the expected depth of node k is

$$\mathrm{E}[depth(k)] = \sum_{i=1}^{n} \Pr[i \uparrow k] = H_k + H_{n-k} - 2 < 2\ln n.$$

To prove a worst-case expected bound on the depth of the tree, we need to argue that the *maximum* depth of any node is small. Chernoff bounds make this argument easy, once we establish that the relevant indicator variables are mutually independent.

Lemma 1. *For any index k, the $k-1$ random variables $[i \uparrow k]$ with $i < k$ are mutually independent, and the $n-k$ random variables $[i \uparrow k]$ with $i > k$ are mutually independent.*

Proof: Fix an arbitrary index k. We explicitly consider only the indicators $[i \uparrow k]$ with $i > k$; the proof for $i < k$ is nearly identical.

To simplify notation, let X_i denote the indicator variable $[i \uparrow k]$. We will prove by induction on n that the $n-k$ indicator variables $X_{k+1}, X_{k+2}, \ldots, X_n$ are mutually independent, starting with the vacuous base case $n = k$. Fix $n-k$ arbitrary indicator values $x_{k+1}, x_{k+2}, \ldots, x_n$. The definition of conditional probability immediately implies that

$$\Pr\left[\bigwedge_{i=k+1}^{n} (X_i = x_i)\right] = \Pr\left[\bigwedge_{i=k+1}^{n-1} (X_i = x_i) \wedge X_n = x_n\right]$$

$$= \Pr\left[\bigwedge_{i=k+1}^{n-1} (X_i = x_i) \,\middle|\, X_n = x_n\right] \cdot \Pr\left[X_n = x_n\right]$$

Now recall that $X_n = [n \uparrow k] = 1$ if and only if node n has the smallest priority of all nodes between k and n. The other $n-k-1$ indicator variables X_i depend only on the order of the priorities of nodes k through $n-1$. There are exactly $(n-k-1)!$ permutations of the $n-k$ priorities in which the nth priority is smallest, and each of these permutations is equally likely. Thus,

$$\Pr\left[\bigwedge_{i=k+1}^{n-1} (X_i = x_i) \,\middle|\, X_n = x_n\right] = \Pr\left[\bigwedge_{i=k+1}^{n-1} (X_i = x_i)\right]$$

The inductive hypothesis implies that the variables $X_{k+1}, \ldots, X_{n-1}$ are mutually independent, so

$$\Pr\left[\bigwedge_{i=k+1}^{n-1} (X_i = x_i)\right] = \prod_{i=k+1}^{n-1} \Pr[X_i = x_i].$$

We conclude that

$$\Pr\left[\bigwedge_{i=k+1}^{n} (X_i = x_i)\right] = \Pr[X_n = x_n] \cdot \prod_{i=k+1}^{n-1} \Pr[X_k = x_k] = \prod_{i=k+1}^{n} \Pr[X_i = x_i],$$

or in other words, that the variables $X_i = [i \uparrow k]$ are indeed mutually independent. $\square$

Now let me define two new sequences of random variables:

- $Y_2, Y_3, \ldots$ are fully independent indicator variables, where $\Pr[Y_j = 1] = 1/j$ for each j.

- $Z_2, Z_3, \ldots$ are prefix sums of the previous sequence: $Z_k := \sum_{j=2}^{k} Y_j$.

Lemma 2. *For all integers $2 \le k \le n$, we have $\Pr[Z_k > 4\ln n] < 1/n^2$.*

Proof: For every integer $k \ge 2$ and real number z, we have $\Pr[Z_k > z] < \Pr[Z_{k+1} > z]$. Thus, it suffices to consider the special case $k = n$.

Let $\mu = \mathrm{E}[Z_n] = H_n - 1 < \ln n$ and $\delta = 3$. Since Z_n is a sum of independent indicator variables, Chernoff's inequality implies

$$
\begin{aligned}
\Pr[Z_n > 4\ln n] &< \Pr[Z_n > 4\mu] \\
&= \Pr[Z_n \ge (1+\delta)\mu] \\
&\le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \\
&= \left(\frac{e^3}{4^4}\right)^\mu < \left(\frac{e^3}{4^4}\right)^{\ln n} \\
&= n^{3-4\ln 4} \approx n^{-2.54518} < \frac{1}{n^2}. \qquad \square
\end{aligned}
$$

Theorem 3. *An n-node treap has depth $O(\log n)$ with high probability.*

Proof: Specifically, I will bound the probability that the depth is at most $8\ln n$. The union bound $\Pr[A \vee B] \le \Pr[A] + \Pr[B]$ implies

$$\Pr\left[\max_k depth(k) > 8\ln n\right] = \Pr\left[\bigvee_{k=1}^{n} (depth(k) > 8\ln n)\right] \le \sum_{k=1}^{n} \Pr[depth(k) > 8\ln n].$$

The depth of any node k is a sum of n indicator variables $[i \uparrow k]$, as i ranges from 1 to n. Lemma I partitions these variables into two mutually independent subsets. Define

$$depth_<(k) := \sum_{i=1}^{k-1} [i \uparrow k] \quad \text{and} \quad depth_>(k) := \sum_{i=k+1}^{n} [i \uparrow k],$$

so that $depth(k) = depth_<(k) + depth_>(k)$. The union bound implies

$$\Pr[depth(k) > 8 \ln n] \leq \Pr[depth_<(k) > 4 \ln n] + \Pr[depth_>(k) > 4 \ln n].$$

For all indices $i \neq k$, the indicator variable $[i \uparrow k]$ has the same distribution as the variable $Y_{|i-k|+1}$ defined just before Lemma 2. It follows that $depth_<(k)$ has the same distribution as the random variable Z_k. Thus, for each k we have

$$\begin{aligned}
\Pr[depth_<(k) > 4 \ln n] &= \Pr\left[\sum_{i=1}^{k-1} Y_{k-i+1} > 4 \ln n \right] \\
&= \Pr\left[\sum_{j=2}^{k} Y_j > 4 \ln n \right] \\
&\leq \Pr[Z_k > 4 \ln n] \\
&< \Pr[Z_n > 4 \ln n] < 1/n^2.
\end{aligned}$$

Symmetrically, we have $\Pr[depth_>(k) > 4 \ln n] \leq \Pr[Z_{n-k-1} > 4 \ln n] \leq \Pr[Z_n > 4 \ln n] < 1/n^2$. We conclude that

$$\Pr\left[\max_k depth(k) > 8 \ln n \right] \leq 2n \Pr[Z_n > 4 \ln n] < 2/n,$$

which completes the proof.   □

By varying the constants in the previous argument, we can prove that for any constant c, the depth of the treap is greater than $c \ln n$ with probability at most $2/n^{c \ln c - c + 1}$.

We can now finally conclude that any search, insertion, deletion, or merge operation on an n-node treap requires $O(\log n)$ time with high probability. In particular, the expected *worst-case* time for each of these operations is $O(\log n)$.

Exercises

1. (a) Prove that for any index k such that $1 < k < n$, the $n-1$ indicator variables $[i \uparrow k]$ with $i \neq k$ are *not* mutually independent. *[Hint: Consider the case $n = 3$ and $k = 2$.]*

    (b) Prove that for any index k, the $k-1$ random variables $[k \uparrow i]$ with $i < k$ are *not* mutually independent. *[Hint: Consider the case $k = 4$.]*

2. (a) Consider an arbitrary node k in an n-node treap. Prove that the expected number of descendants of node k is *precisely* equal to the expected number of ancestors of node k.

    (b) Why doesn't the Chernoff-bound argument for depth imply that, with high probability, *every* node in a treap has $O(\log n)$ descendants? The conclusion is clearly bogus—Every treap has a node with n descendants!—but what's the hole in the argument?

3. The following algorithm finds the smallest element in an unsorted array:

> $\underline{\textsc{RandomMin}}(A[1..n])$:
> $min \leftarrow \infty$
>   for $i \leftarrow 1$ to n in random order
>   if $A[i] < min$
> $min \leftarrow min_I$ $(\star)$
>   return min

(a) Let X_i be the indicator variable that equals 1 if line $(\star)$ is executed in the ith iteration of the main loop. Prove that the variables X_i are mutually independent.

(b) Prove that line $(\star)$ is executed $O(\log n)$ time with high probability.

4. Let S be a set of n points in the plane. A point p in S is called *Pareto-optimal* if no other point in S is both above and to the right of p. Suppose each point in S is chosen independently and uniformly at random from the unit square $[0,1] \times [0,1]$. Prove that the number of Pareto-optimal points in S is $O(\log n)$ with high probability.

5. Recall from the previous lecture note that a **heater** is a sort of anti-treap, in which the priorities of the nodes are given, but their search keys are generated independently and uniformly from the unit interval $[0,1]$. Prove that an n-node heater has depth $O(\log n)$ with high probability.

6. Let $X_1, X_2, \ldots, X_n$ be independent indicator variables, each equal to 1 with probability $1/2$, and let $X = \sum_i X_i$. Clearly $\mathrm{E}[X] = n/2$.

(a) Let $Y = (X - \mathrm{E}[X])^2$. What is $\mathrm{E}[Y]$?

(b) Let $Z = (Y - \mathrm{E}[Y])^2$. What is $\mathrm{E}[Z]$?

(c) Let $\Omega = (Z - \mathrm{E}[Z])^2$. What is $\mathrm{E}[\Omega]$?