LECTURE 12 (October 3rd)

Tail Inequalities

RECAP Last time we showed randomized binary search trees/treaps satisfy

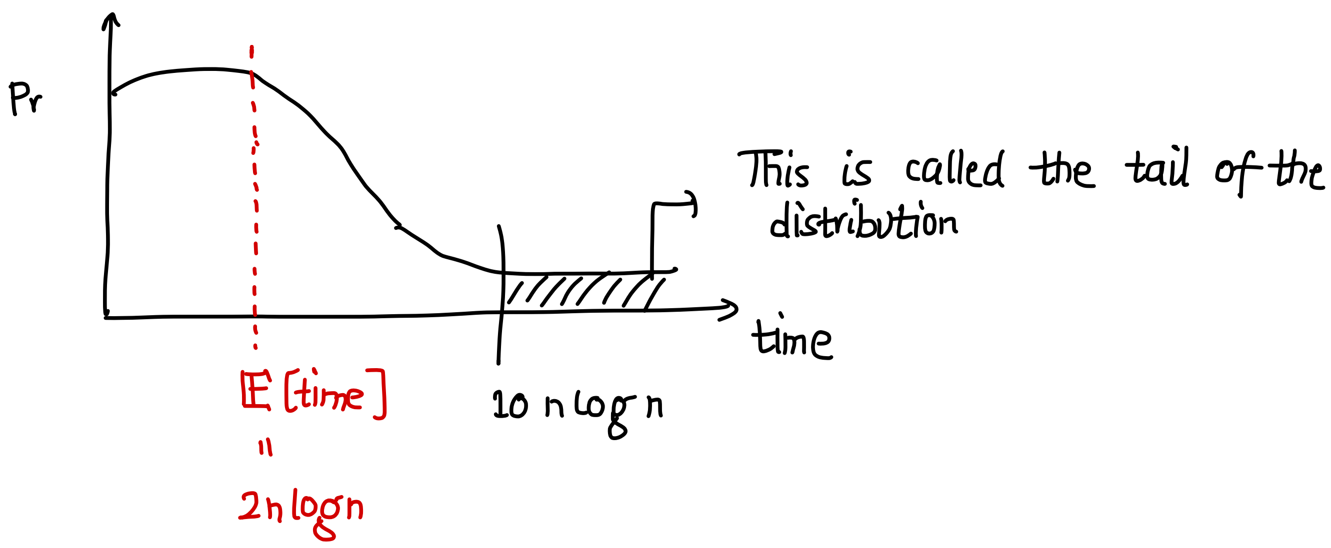
$$\mathbb{E}[\text{depth}(v)] = O(\log n) \quad \forall \text{ nodes } v$$

So, search & other operations take $O(\log n)$ expected time

This also implied that randomized quicksort runs in $O(n \log n)$ expected time

Today we are going to prove that these statements hold with high probability

Tail Inequalities Suppose the distribution of our runtime looks like



We want to find $\mathbb{P}[\text{run time} > 10 n \log n]$ for example

For example, if the tail decays like a Gaussian, then it decays super exponential & this probability would be small

But the tail behavior depends on the distribution which will not be Gaussian for our applications

What we are going to rely on is the fact that our random variables can be written as

$$X = X_1 + X_2 + \cdots$$

And what we want to bound, for instance

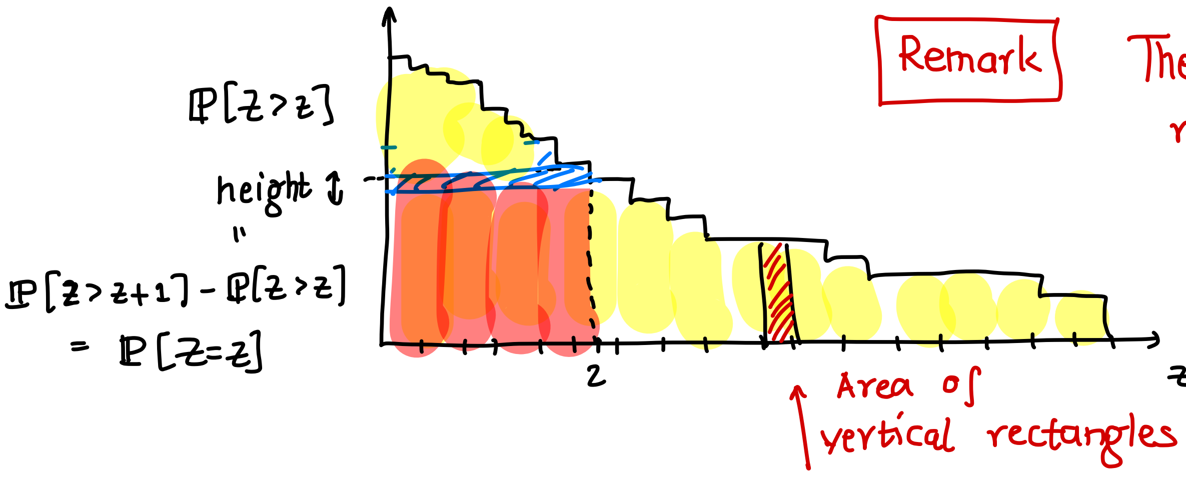
$$\mathbb{P}[X \geq \alpha \cdot \mathbb{E}[X]] \leq ?$$

Main message If the random variables X_i's are independent, we get a very sharp tail inequality

In general, the more independent X_i's are the better tail bound we obtain

Markov's Inequality If Z is a non-negative integer random variable, then

$$\mathbb{P}[Z \geq z] \leq \frac{\mathbb{E}[Z]}{z} = \frac{\mu}{z}$$ ← mean or expected value



$\boxed{\text{Remark}}$ The inequality holds for non-integer random variables as well

$$\text{Total Area} = \sum_z \mathbb{P}[Z \geq z]$$

$$= \sum_z \underbrace{z \cdot \mathbb{P}[Z=z]}_{\text{Area of horizontal rectangle}} = \mathbb{E}[Z]$$

We claim that Total Area $\geq \underbrace{z \cdot \mathbb{P}[Z \geq z]}_{= \text{Area of red shaded rectangle}}$

$$\sum_z \mathbb{P}[Z \geq z] \geq \sum_{i=0}^{z-1} \mathbb{P}[Z \geq i] \geq \sum_{i=?}^{z} \mathbb{P}[Z \geq z] \geq z \cdot \mathbb{P}[Z \geq z]$$

Therefore, we obtain $\mathbb{E}[Z] \geq z \cdot \mathbb{P}[Z \geq z]$ for any z

So, $\mathbb{P}[Z \geq \alpha \cdot \mathbb{E}[Z]] \leq \frac{\mathbb{E}[Z]}{\alpha \, \mathbb{E}[Z]} = \frac{1}{\alpha}$

Recall that for quicksort, $\mathbb{E}[\text{run time}] \leq 4n \cdot \log n$

So, $\mathbb{P}[\text{run time} > 8n \log n] \leq \frac{1}{2}$

o $\mathbb{P}[\text{run time} > n^3 \log n] \leq \frac{4n \log n}{n^3} \leq \frac{4}{n^2}$

So, we only get weak tail bounds, but that's understandable because we haven't made any assumptions on the random variable apart from non-negativity

To get stronger tail bounds, we need more assumptions, e.g. independence

Recall that X & Y are independent random variables if

$$\mathbb{P}[X=x \ \& \ Y=y] = \mathbb{P}[X=x]\cdot\mathbb{P}[Y=y]$$

or equivalently, $\mathbb{P}[X=x \mid Y=y] = \mathbb{P}[X=x]$

This also implies that $\mathbb{E}[XY] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$
& that $f(X)$ & $f(Y)$ are also independent for any function f

Similarly, $X_1, \dots X_n$ are fully (or mutually) independent if

$$\mathbb{P}[X_1=x_1, X_2=x_2, \dots, X_n=x_n] = \prod_{i=1}^{n} \mathbb{P}[X_i=x_i]$$

A weaker notion of independence that is sometimes useful for applications is k-wise independence

$X_1, \dots, X_n$ are k-wise independent if every subset of size k is fully independent

Example Suppose $X_1, X_2 \in \{0,1\}$ are independent random bits $\mathbb{P}[X_1=0]=\mathbb{P}[X_1=1]=1/2$
& same for X_2

Let $X_3 = X_1 \oplus X_2$

Then, (X_1, X_2, X_3) are 2-wise (also called pairwise) independent

Pairwise independence implies a stronger tail bound,

In particular, let $X_1 = \sum_{i=1}^{n} X_i$ where $X_i \in \{0,1\}$, $\mathbb{E}[X_i] = \mathbb{P}[X_i=1] = p_i$

& $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} \mu_i$

then, we have the following stronger tail bound

Chebyshev's Inequality If $X_1, \dots, X_n$ are pairwise independence, then

$$\mathbb{P}[(X-\mu)^2 \geq t] \leq \frac{\mathbb{E}[(X-\mu)^2]}{t} \leq \frac{\mu}{t}$$

This follows from Markov's inequality applied to the non-negative random variable $(X-\mu)^2$

$\mathbb{E}[(X-\mu)^2] \leq \mu$ follows from pairwise independence

Proof Let $Y_i = X_i - p_i$ & $Y = \sum_{i=1}^n Y_i = \sum_{i=1}^n X_i - \sum_{i=1}^n p_i = X - \mu$

Note that $\mathbb{E}[Y_i] = \mathbb{E}[Y] = 0$

$$\mathbb{E}[(X-\mu)^2] = \mathbb{E}[Y^2] = \sum_{i,j} \mathbb{E}[Y_i Y_j]$$

Second moment of Y

$$= \sum_{i=1}^n \mathbb{E}[Y_i^2] + \sum_{i \neq j} \underbrace{\mathbb{E}[Y_i Y_j]}_{= \mathbb{E}[Y_i]\cdot\mathbb{E}[Y_j]}$$

(by pairwise independence)

$$= \sum_{i=1}^n \underbrace{\mathbb{E}[Y_i^2]}_{} + \sum_{i \neq j} \underbrace{\mathbb{E}[Y_i]}_{=0} \cdot \underbrace{\mathbb{E}[Y_j]}_{=0}$$

$Y_i^2 = \begin{cases} (-p_i)^2 & \text{with prob. } 1-p_i \\ (1-p_i)^2 & \text{with prob } p_i \end{cases}$

$$= \sum_i p_i(1-p_i)^2 + (1-p_i)p_i^2$$

$$= \sum_i \left[p_i(1+p_i^2-2p_i) + p_i^2 - p_i^3 \right]$$

$$= \sum_i \left[p_i + \cancel{p_i^3} - 2p_i^2 + p_i^2 - \cancel{p_i^3} \right]$$

$$= \sum_i \left[p_i - p_i^2 \right] \leq \sum_i p_i = \mu \qquad \square$$

What does this inequality say about $\mathbb{P}[X \geq (1+\delta)\mu] \leq \mathbb{P}[(X-\mu)^2 \geq (\delta\mu)^2]$

$$\leq \frac{\mathbb{E}(X-\mu)^2}{\delta^2\mu^2} \leq \frac{1}{\delta^2\mu}$$

For quicksort, $\mu = 4n\log n$, let $\delta = \frac{1}{4}$, then we get that

$$\mathbb{P}[\text{runtime} \geq 5n\log n] \leq \frac{1}{n\log n}$$

Thus, if the random variables were pairwise independence, $\mathbb{P}[\text{runtime} \geq 5n\log n] \longrightarrow 0$ as $n \to \infty$

This is not completely satisfactory, we would like to have even sharper tail bounds
And also random variables appearing in the quicksort analysis are not pairwise independence (without being more careful) so we can't use this directly

But for many applications, this bound is good enough

There is also a bound in the other direction $\mathbb{P}[X \leq (1-\delta)\mu] \leq \frac{1}{\delta^2\mu}$

What if everything is fully independent? Then, we have the following exponential moment bound

Exponential Moment Inequality If $X_1, \dots X_n$ are fully independent, then

$$\mathbb{E}[2^X] \le e^{\mu} \quad \& \text{ in general, } \mathbb{E}[\alpha^X] \le e^{(\alpha-1)\mu} \text{ for any } \alpha > 1$$

If you rely on misguided intuition, you might think that

$$\mathbb{E}[2^X] = 2^{\mathbb{E}[X]} \rightarrow \text{This is not true}$$

The above is not true in general, but if $X_1 \dots X_n$ are independent, something close is true as shown by the exponential moment inequality

Proof (of Exponential Moment Inequality) can be found in the lecture notes

Consequences of the Exponential Moment Inequality

$$\mathbb{P}[X \ge 2\mathbb{E}[X]] \le \frac{e^{\mu}}{2^{2\mu}} = \left(\frac{e}{4}\right)^{\mu} \rightarrow \text{If } \mu \gg 1 \text{, then this decays exponentially since } \left(\frac{e}{4}\right) < 1$$

Why? $$\mathbb{P}[X \ge 2\mathbb{E}[X]] = \mathbb{P}[2^X \ge 2^{2\mathbb{E}[X]}]$$

$$\le \frac{\mathbb{E}[2^X]}{2^{2\mu}} \le \frac{e^{\mu}}{2^{2\mu}}$$

Markov's inequality for 2^X

By exponential moment inequality

If we do further manipulations (which can be found in the lecture notes), we get

$$\mathbb{P}[X \ge (1+\delta)\mu] \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \le e^{-\delta^2\mu/3} \text{ if } \delta \in [0,1]$$

Similar inequality holds for $X \le (1-\delta)\mu$

$\boxed{\text{Main message}}$ $\mathbb{P}[X \ge \text{little above its mean}] \le e^{-\text{mean}}$

For example, $\mathbb{P}[X \ge 2\mu] \le e^{-\mu} \le e^{-4n\log n}$ for quicksort

Let's see how to use the exponential tail bounds, which are also called Chernoff bounds, to analyze treaps

Recall that in the last lecture, we showed that if we insert keys into a treap with random priorities, then

$$\mathbb{E}\left[\text{depth}(k)\right] = \sum_{i=1}^{n} \mathbb{P}[\underbrace{i \Uparrow k}_{=\ i \text{ is a proper ancestor of } k}]$$

Let $X = \text{depth}(k)$, then note that $X = \sum_{i=1}^{n} X_i$

where $X_i = \begin{cases} 1 & \text{if } i \Uparrow k \\ 0 & \text{otherwise} \end{cases}$

Thus, the depth can be written as sum of indicator random variables
We now show the following claim :

$\boxed{\text{Claim}}$ For any node k, $X_1, X_2, \dots X_{k-1}$ are mutually independent

and $X_{k+1}, \dots\dots X_n$ are mutually independent

Proof Follows from a careful induction which shows that

$$\mathbb{P}[X_1 = x_1, \dots X_{k-1} = x_{k-1}] = \mathbb{P}[X_1 = x_1] \cdot \mathbb{P}[X_2 = x_2] \cdots \mathbb{P}[X_{k-1} = x_{k-1}]$$

& similarly for $X_{k+1}, \dots X_n$

(Can be found in the lecture notes)

Note The random variables in the two different groups may be dependent
For example,

X_1 & X_n may be dependent if $1 < k < n$

So, we can not use Chernoff bounds directly

So, let us write $X = \underbrace{\sum_{i=1}^{k-1} X_k}_{=\ X_{<k}} + X_k + \underbrace{\sum_{i=k+1}^{n} X_k}_{X_{>k}}$

X_k always 0 since k is never a proper ancestor of k

We showed last time that $\mathbb{E}[X] \le 4 \log n$

So, $\mathbb{P}[X \ge 5 \log n] \le \mathbb{P}\left[\begin{matrix} X_{<k} \ge 5 \log n \\ \text{or} \\ X_{>k} \ge 5 \log n \end{matrix}\right]$

since $X \ge 5 \log n$
$\Rightarrow X_{<k} \ge 5 \log n$
or $X_{>k} \ge 5 \log n$

$$\leq \mathbb{P}[X_{<k} \geq 5\log n] + \mathbb{P}[X_{>k} \geq 5\log n]$$

union bound
$\mathbb{P}[A \vee B] \leq \mathbb{P}[A] + \mathbb{P}[B]$

Sum of mutually independent random variables with $\mathbb{E}[X_{<k}] \leq \log n$

Sum of mutually independent random variables with $\mathbb{E}[X_{<k}] \leq \log n$

Applying Chernoff bounds, we get both are at most $e^{-c(5\log n)}$ for some constant c

Careful computations of the constant give us a bound of $\frac{1}{n^{2.5}}$

$$\text{Thus, } \mathbb{P}[\text{depth}(k) \geq 5\log n] \leq \frac{2}{n^{2.5}}$$

By union bound over all n nodes, we also get that

$$\mathbb{P}[\text{depth}(\text{treap}) \geq 5\log n]$$

$$= \mathbb{P}[\text{depth}(1) \geq 5\log n \text{ or } \text{depth}(2) \geq 5\log n \;\cdots\; \text{or } \text{depth}(n) \geq 5\log n]$$

$$\leq n \cdot \frac{2}{n^{2.5}} \leq \frac{1}{n^{1.5}} \longrightarrow 0 \text{ as } n \to \infty$$

Thus, with high probability, maximum depth is $O(\log n)$ and all treap operations take $O(\log n)$ time