• Randomized Algorithms (18th Jan)

Last time:

- Quicksort (Randomized)

  Expected # comparisons $\leq 2n\mathcal{H}_n. =: q_n.$

- Complexity classes:

  RP, Co-RP, ZPP, BPP.

We are interested in
$\mathbb{P}[Q_n \geq (1+\varepsilon) q_n] \leq n^{-2\varepsilon \ln \ln n}$.
Q_n: Random variable denoting # comparison
[McDiarmid-Hayward]

– Let L be language, i.e., $L \subseteq \{0,1\}^*$
  e.g., PRIMES, CONNECTED GRAPHS

Input: $x \in \{0,1\}^*$

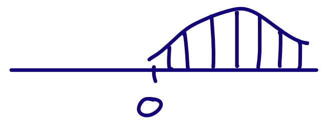
Goal: Determine if $x \in L$ or not.

- An algorithm A is an ϵ-error randomized algorithm for L if $\forall x$, $x \in L$ then $\mathbb{P}[A(x) = \text{Yes}] \geq 1 - \varepsilon$ and if $x \notin L$ then $\mathbb{P}[A(x) = \text{No}] \geq 1 - \varepsilon$.

- **Markov's Inequality:**

If X is a nonnegative random variable and $\mathbb{E}[X] = \mu$, then

$$\mathbb{P}[X \geq a] \leq \frac{\mu}{a} \quad (a > 0).$$



- Alternate version: $\mathbb{P}[X \geq K\mu] \leq \frac{1}{K}$.

Proof: Consider the indicator random variable

$$\mathbb{I}_{\{x \geq a\}} = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$$

Then we claim $\mathbb{I} \leq \frac{X}{a}$

$$\Rightarrow \mathbb{E}[\mathbb{I}] \leq \mathbb{E}[X/a]$$

$$\Rightarrow \boxed{\mathbb{P}[X \geq a] \leq \frac{\mu}{a}.}$$

Story about A. A. Markov (father of Russian(s)) & Markov Brother's. (Inequality in complexity)

Using Markov's ineq. for quicksort we get:

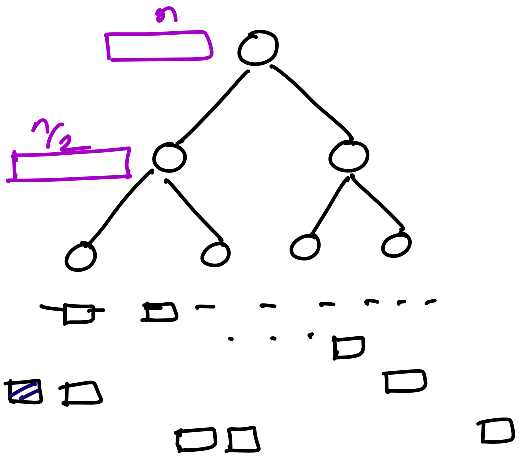
$$\mathbb{P}[Q_n \geq K q_n] \leq 1/K.$$

Another useful corollary:

$$\mathbb{P}[X \geq 1] \leq \mu.$$

useful for integer random variables

Concentration of Quicksort:



Total comparison in each level $= O(n)$.

So we want to bound the number of levels.

Expected size of the array at t'th level $= n/2^t$.

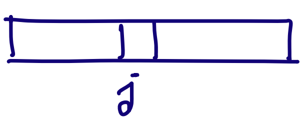
Suppose, if $t \geq 10 \log_2 n$

then this expected size $\leq \frac{n}{n^{10}} = \frac{1}{n^9}$. ← very small.

Q. Can we use union bound?

Can we say there are only n potential branches?

Fix j. define $N_t(j) =$ # other elements in j's array at level t.



$\mathbb{E}[N_2] = 0.8n$ (Note, it is not $n/2$).

Check (HW)

$\mathbb{E}[N_t] \leq n\,(0.8)^t$

$t = (\log_{\frac{10}{8}} n) + k.$

Then $\mathbb{P}[N_t(j) \geq 1] \leq \mathbb{E}[N_t(j)]$

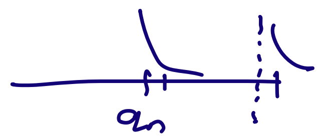
$$= \frac{n}{n\left(\frac{10}{8}\right)^k}$$

$$\mathbb{P}[\exists j : N_t(j) \geq 1] \leq \frac{n}{\left(\frac{10}{8}\right)^k}$$

$$\leq \frac{1}{2^r}$$ by choosing $k = \log_{10/8} n + O(r)$.

Hence, $\mathbb{P}[Q_n \geq 2n \log_{10/8} n + rn]$

$$\leq 2^{-r}.$$



With more adaptation, we can improve further.

Chapter 49

Strong Concentration for Quicksort*

Colin McDiarmid[†] Ryan Hayward[‡]

THEOREM 1.1. *For any $\varepsilon > 0$, as $n \to \infty$,*

$$\Pr\left[\left|\frac{Q_n}{q_n} - 1\right| > \varepsilon\right] = n^{-2\varepsilon(\ln\ln n + O(\ln\ln\ln n))}.$$

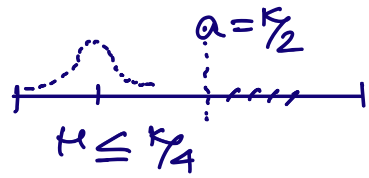
$\mathbb{E}[Q_n]$ #comparison

• Amplification:

Claim: If there is an $(\frac{1}{4})$-error randomized algorithm A for L, then there is an ε-error randomized algorithm for L with running time $O(\boxed{1/\varepsilon}\, t_A(n))$
??

[where $t_A(n)$ is the runtime of alg A on input n].

Idea: Run A, k times independently and go with the majority.



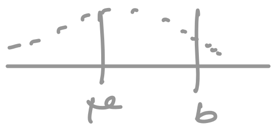
Now using Markov's ineq, we only obtain Prob of error $\leq 1/2$, which is not so strong.

Next class we'll use Chernoff to obtain much stronger bounds.

• **Variance:** $\mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2]$

where $\mu = \mathbb{E}[X]$. $\quad = \mathbb{E}[X^2] - \mu^2$.

Claim: $\mathbb{E}[(X-b)^2] = \mathbb{E}[(X-\mu)^2] + (\mu - b)^2$.



connection w moment of inertia

for $b=0$,

$\mathbb{E}[X^2] - \mu^2 =$ Variance.

• Expectation is the point about which the moment of inertia is minimum.

• Chebyshev's inequality:

$$\mathbb{P}[\,|X-\mu| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}, \qquad (a > 0).$$

Define, $Y := |X-\mu|^2$.

Note Y is a nonnegative RV.

This then directly follows from Markov's ineq.

• Consider a random variable X s.t.

$$X = \underbrace{X_1 + X_2 + \dots + X_k}_{\text{independent}}$$

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[(\sum X_i - \mu)^2] \\
&= \mathbb{E}[(\sum X_i - \sum \mu_i)^2] \\
&= \mathbb{E}[\sum (X_i - \mu_i)^2] \\
&= \mathbb{E}[\sum_{ij} (X_i - \mu_i)(X_j - \mu_j)] \\
&= \sum_{ij} \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \\
&= \sum_i \mathrm{Var}[X_i] \quad \text{(independence).}
\end{aligned}$$

Now, go back to amplification.

$F_1, F_2, \ldots, F_K$, where F_i is the indicator RV for the event that A's i'th run resulted in a wrong answer.

So, total # failures

$$F = \sum F_i.$$

$$\mathbb{E}[F] = \mu \leq (\tfrac{1}{4}) K.$$

Then, $\mathbb{P}[F \geq K/2] \leq \mathbb{P}[\,|F - K/4| \geq K/4]$

$$\leq \frac{\mathrm{Var}[F]}{(K/4)^2} = \frac{\sum_i \mathrm{Var}[F_i]}{(K/4)^2}.$$

Now F_i is a Bernoulli RV with prob $= 3/4$.

Hence the $\text{var}[F_i] = p(1-p) = 3/16$.

Hence, $\mathbb{P}[F \geq K/2] \leq \dfrac{K \cdot 3/16}{K^2/16} = \dfrac{3}{K}$.

We want $\dfrac{3}{K} = \epsilon$, i.e. $\boxed{K = 3/\epsilon.}$

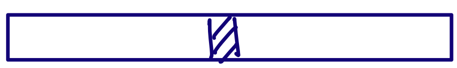
(In fact, later we'll see $O(\log(1/\epsilon))$ runs are sufficient).

Geometric Distribution:

X is $\text{Geom}(p)$. $\mathbb{E}[X] = 1/p$. $\text{Var}[X] = \dfrac{(1-p)}{p^2}$.

HW: what can we say about concentration of coupon collector using Chebyshev.

Randomized Median:

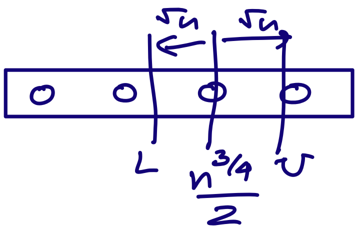
A 

Median (A) is a number z.

$\{i : A[i] \geq z\} \geq n/2$, $\{i : A[i] \leq z\} \geq n/2$.

Consider sorted order : $a_1 \; a_2 \; \cdots | \; a_n$

pick a sample of size $n^{3/4}$.



Sort them & find their median.

Consider two numbers L & U around the median. (may be $\sqrt{n}$ far)

Throw away elements $\leq L$ & $\geq U$.

Run it on the rest ($\simeq n^{3/4}$ elements).

We'll use hypergeometric distribution for sampling. [Different analysis than M-U].

Each time we toss a coin. What is the prob that t'th success happens after median?