6.854 / 18.415: Advanced Algorithms

Last Time: Universal hashing, perfect hashing
(no collisions)

What is a 2-universal hash family? Example?

$$\mathcal{H} = \{ h_{a,b} \mid a \in \{1,2 \ldots p-1\},\ b \in \{0,1 \ldots p-1\} \}$$

$$h_{a,b} = (ax+b \bmod p) \bmod n$$

Today: Load balancing and power of two choices

<u>setup:</u> n equal-sized jobs (arrive online), n machines

single point of failure

How can we distribute load without centralized scheme and without communicating?

<u>Approach #1:</u> Each job chooses a machine randomly

a.k.a. balls in bins

What is the maximum load? Let $Z_i \triangleq$ # balls in bin i

(a) Markov: $\mathbb{E}[Z_i] = 1$, so $\Pr[Z_i \geq k] \leq \frac{1}{k}$

But if we apply union bound

$$\Pr[\exists_i\ Z_i \geq k] \leq \sum_{i=1}^{n} \Pr[Z_i \geq k] \leq \frac{n}{k}$$

doesn't tell us anything

(b) Chebyshev: Z_i is the sum of n i.i.d r.v.s.

$\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$ where Y is indicator of single ball in bin i

Hence $\text{var}(Z_i) = n\left[\frac{1}{n} - \left(\frac{1}{n}\right)^2\right] < 1$

we conclude

σ = standard deviation

$$\Pr[|Z_i - \underset{\mathbb{E}[X_i]}{1}| \geq k] \leq \frac{1}{k^2}$$

and can choose $k = \sqrt{n}$. What if balls choose bins pairwise independently?

(c) direct analysis

$$\Pr[Z_i \geq k] \leq \sum_{\substack{S \subseteq [n] \\ |S| = k}} \underbrace{\Pr\left[\begin{matrix}\text{all balls in} \\ S \text{ sent to } i\end{matrix}\right]}_{\left(\frac{1}{n}\right)^k} \leq \binom{n}{k}\left(\frac{1}{n}\right)^k$$

fact: $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$

can make as large as you want at expense of hidden const

$$\Pr[Z_i \geq k] \leq \left(\frac{e}{k}\right)^k \leq \frac{1}{n^c} \quad \text{for } k = \Omega\left(\frac{\log n}{\log\log n}\right)$$

e.g. ignore e

Now we can union bound

HW: You will show that $\Theta\left(\frac{\log n}{\log\log n}\right)$ is tight

Alternatively can use Chernoff bound

$\ln(1+\delta) \geq \frac{\delta}{1+\frac{\delta}{2}}$

upper tail

Thm: Suppose X_i are independent Bernoulli r.vs with $\mathbb{E}[\sum_i X_i] = \mu$, then

$$\Pr\left[\sum X_i \geq (1+\delta)\mu\right] \leq \left[\frac{e^\delta}{(1+\delta)^{1+\delta}}\right]^\mu \leq e^{\frac{-\delta^2}{2+\delta}\mu}$$

In our setting $\mu = 1$ and $1+\delta = k$, so we recover large deviation bounds at

$$\Theta\left(\frac{\log n}{\log\log n}\right) \quad \text{w.h.p}$$

Approach #2: Each job (arrives sequentially), chooses two machines randomly, picks less loaded

Before I tell you what happens, test intuition: what is the maximum load

(a) $\sqrt{n}$
(b) $\log n / \log\log n$
(c) $\sqrt{\log n/\log\log n}$
* (d) $\log\log n$ * power of two choices [Azar, Broder, Karlin, Upfal]
(e) $O(1)$
(f) Donald Trump

Heuristic analysis: Let β_i be an upperbound on # bins with at least i balls

$$\Pr\left[\begin{array}{c}\text{next ball}\\ \text{lands in bin}\\ \text{with at least } i\end{array}\right] \leq \left(\frac{\beta_i}{n}\right)^2$$

we expect $\beta_{i+1} \leq n\left(\frac{\beta_i}{n}\right)^2$, etc

Now set $\beta_6 = \frac{n}{2e}$, $\beta_7 = n\left(\frac{1}{2e}\right)^2$, $\beta_8 = n\left(\frac{1}{2e}\right)^4$, ...

$\#$ bins with at least 6 $\leq \frac{n}{6} < \beta_6$ $\beta_i = n\left(\frac{1}{2e}\right)^{2^{(i-5)}}$

Why is this analysis heuristic?

(a) we bounded expectation of β_i (soln: use Chernoff)

(b) If we condition on at end, no more than β_i bins with i balls, changes distribution of choices

---

Tool: Stochastic domination

Lemma: If $X_1, Y_2 \ldots Y_n$ are a sequence of r.vs with

$$Y_i \triangleq f_i(X_1, X_2 \ldots X_i)$$

and $\Pr[Y_i = 1 \mid X_1, X_2 \ldots X_{i-1}] \leq p$

then $\Pr[\sum Y_i \geq k] \leq \Pr[\underbrace{\mathrm{Bin}(n,p)}_{\text{flip } n \text{ coins with prob. } p \text{ of heads}} \geq k]$

Some notation:

$\beta_6 = \frac{n}{2e}, \ldots \beta_{i+1} = \frac{e\beta_i^2}{n}$ up to i^* (TBD)

$h(t)$ = height of t^{th} ball

$\nu_i(t) \triangleq$ # bins with at least i balls at time t

$\nu_i(t) \le \mu_i(t)$

$\mu_i(t) \triangleq$ # balls with height at least i "

$E_i \triangleq$ event that $\nu_i(n) \le \beta_i$ * at end *

Proof: Let $Y_t = 1$ iff $h(t) \ge i+1$ and $\nu_i(t-1) \le \beta_i$

$$\Pr\left[Y_t = 1 \mid \text{choices of first } t-1 \text{ balls}\right] \le \frac{\beta_i^2}{n} \triangleq p_i$$

By stochastic domination

$$\Pr\left[\sum_{t=1}^{n} Y_t \ge k\right] \le \Pr\left[B(n, p_i) \ge k\right]$$

If E_i happens, $\sum_{t=1}^{n} Y_t = \underbrace{\mu_{i+1}(n)}_{\text{\# balls ht} \ge i+1} \ge \nu_{i+1}(n)$

$$\Pr\left[\nu_{i+1}(n) \ge k \mid E_i\right] \le \Pr\left[\mu_{i+1}(n) \ge k \mid E_i\right]$$

$$= \Pr\left[\sum_{t=1}^{n} Y_t \ge k \mid E_i\right]$$

$$\le \frac{\Pr\left[\sum_{t=1}^{n} Y_t \ge k\right]}{\Pr[E_i]}$$

$$\le \frac{\Pr\left[B(n, p_i) \ge k\right]}{\Pr[E_i]} \quad \overset{\text{chernoff}}{\le} e^{-p_i n}$$

set $k = e n p_i \Rightarrow$ numerator exponentially small

In first run through, assume always true

$\frac{B_i^2}{n}$ ||

whenever $p_i n \geq 2 \ln n$, we get

induction, denominator is close to 1 anyways

$$\Pr[\nu_{i+1}(n) \geq B_{i+1} \mid E_i] \leq \frac{1}{n^2 \Pr[E_i]}$$

Let i^* be ~~smallest~~ first value of i s.t. check ($i^* = c \log\log n + O(1)$)

$$B_i^2 \leq 2n \ln n$$

$$\Pr[\neg E_{i+1}] \leq \Pr[\neg E_{i+1} \mid E_i] \Pr[E_i] + \Pr[\neg E_i]$$

assume by induction $\Pr[\neg E_i] \leq \frac{i}{n^2} \Rightarrow \Pr[\neg E_{i+1}] \leq \frac{i+1}{n^2}$

$$\Pr[\nu_{i^*+1}(n) \geq 6 \ln n \mid E_{i^*}] \leq \frac{\Pr[\mathrm{Bin}(n, \overbrace{\tfrac{2\ln n}{n}}^{\frac{B_{i^*}^2}{n^2}}) \geq 6 \ln n]}{\Pr[E_{i^*}]}$$

$$\leq \frac{1}{n^2 \Pr[E_{i^*}]}$$

$$\Pr[\nu_{i^*+2}(n) \geq 1] \leq \frac{\Pr[\mathrm{Bin}(n, (\frac{6\ln n}{n})^2) \geq 1]}{\Pr[\mu_{i^*+1} \leq 6 \ln n]}$$

by markov $\leq n\left(\frac{6 \ln n}{n}\right)^2$.

Now if we set $i = c\log\log n + O(1)$ we get $B_i < 1$ and

$$\Pr[\neg E_i] \leq \frac{i}{n^2}$$

and we're done, modulo assumption $p_i n \geq 2 \ln n$

What about three choices?

(a) $(\log \log n)^2$
* (b) $\log \log n$ *
(c) $\sqrt{\log \log n}$
(d) $\log \log \log n$
(e) $O(1)$
(f) Donald Trump