CS7800: Advanced Algorithms

Lecture 17: Randomized Algorithms II

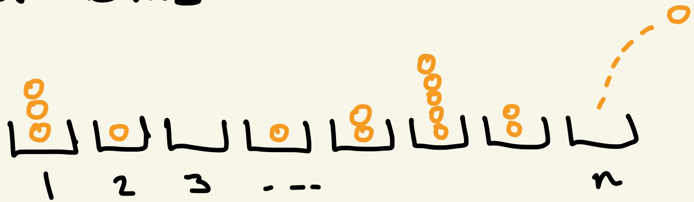
- Finish balls and bins
- String matching

Jonathan Ullman

11-08-2022

Balls and Bins

Throw m balls into n bins <u>independently</u>



$\omega = (1, 8, 19, 23, 6, 2)$

ball 1 to bin 1

ball 2 to bin 8

ball 3 to bin 19

$$\mathbb{P}(\omega_1, \omega_2, \ldots, \omega_m) = \frac{1}{n^m}$$

Questions:

1. How long until bin 1 gets a ball (in expectation)?
2. How long until no bins are empty (in expectation)?
3. What is the most number of balls in any bin (in expectation)?

Maximum Load

want to understand $\mathbb{P}(M \geq t)$ or $\mathbb{E}(M)$

- $M(\omega)$ = maximum load
- M_j = load on bin j $\qquad M = \max\{M_1, M_2, \dots, M_n\}$
- $\mathbb{P}(M \geq t) = \mathbb{P}((M_1 \geq t) \vee (M_2 \geq t) \vee \dots \vee (M_n \geq t))$

$$\leq \sum_{j=1}^{n} \mathbb{P}(M_j \geq t) = n \cdot \underbrace{\mathbb{P}(M_1 \geq t)}$$

"union bound"

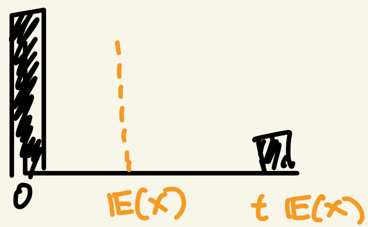
Want to find t so that this is $\ll \frac{1}{n}$

Markov's Inequality

Let X be any random variable such that $X(\omega) \geq 0$

Thm (Markov): $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$

Pf:



$$\mathbb{E}(X) = \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X=i) = \sum_{i=0}^{t-1} i \cdot \mathbb{P}(X=i) + \sum_{i=t}^{\infty} i \cdot \mathbb{P}(X=i)$$

$$\geq 0 + \sum_{i=t}^{\infty} t \cdot \mathbb{P}(X=i)$$

$$= t \cdot \mathbb{P}(X \geq t)$$

Maximum Load

want to understand $\mathbb{P}(M \geq t)$ or $\mathbb{E}(M)$

- $M(\omega)$ = maximum load
- M_j = load on bin j $\qquad M = \max\{M_1, M_2, \dots, M_n\}$
- $\mathbb{P}(M \geq t) = \mathbb{P}((M_1 \geq t) \vee (M_2 \geq t) \vee \dots \vee (M_n \geq t))$

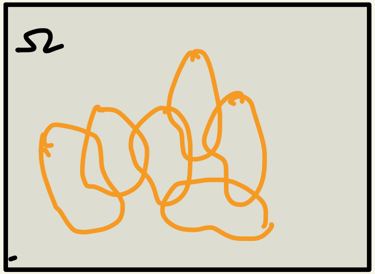
$$\leq \sum_{j=1}^{n} \mathbb{P}(M_j \geq t) = n \cdot \underline{\mathbb{P}(M_1 \geq t)}$$

"union bound"

Want to find t so that this is $\ll \frac{1}{n}$

$$n \cdot \mathbb{P}(M_1 \geq t) \leq n \cdot \frac{\mathbb{E}(M_1)}{t}$$

$$= \frac{n \cdot (m/n)}{t} = \frac{m}{t}$$



Chebyshev's Inequality

Let X be any random variable, let $\mu = \mathbb{E}(X)$

<u>Thm:</u> $\mathbb{P}(|X-\mu| \geq t) \leq \dfrac{\mathbb{E}((X-\mu)^2)}{t^2} = \dfrac{\mathrm{Var}(X)}{t^2}$

<u>Pf:</u> $\mathbb{P}(|X-\mu| \geq t) = \mathbb{P}((X-\mu)^2 \geq t^2) \leq \dfrac{\mathbb{E}((X-\mu)^2)}{t^2}$

Markov

Applying Chebyshev to Balls and Bins

$m = n$

Let M_1 be the load of bin 1 $\qquad \mathbb{E}(M_1) = \frac{m}{n}$ $(=1)$

Let $M_{1,i} = \begin{cases} 1 & \text{if ball } i \text{ goes in bin 1} \\ 0 & \text{otherwise} \end{cases}$

indicator random variable

$\mathbb{E}(I) = \mathbb{P}(I)$

$$\mathbb{E}((M_1-\mu)^2) = \mathbb{E}(M_1^2) - \mathbb{E}(M_1)^2 \leq \frac{m}{n} \quad (\leq 1)$$

$$\mathbb{E}(M_1^2) = \mathbb{E}((M_{1,1} + \dots + M_{1,m})^2) \qquad \mathbb{E}(M_1)^2 = \frac{m^2}{n^2} \quad (=1)$$

$$= \mathbb{E}\left(\sum_{i=1}^{m} M_{1,i}^2 + \sum_{i \neq j} M_{1,i} M_{1,j}\right)$$

$$= m \cdot \frac{1}{n} + m(m-1) \cdot \frac{1}{n^2}$$

$$\leq \frac{m}{n} + \frac{m^2}{n^2} \quad (=2)$$

Applying Chebyshev to Balls and Bins

$m = n$

Let M_1 be the load of bin 1 $\qquad \mathbb{E}(M_1) = \frac{m}{n}$ $(=1)$

Let $M_{1,i} = \begin{cases} 1 & \text{if ball } i \text{ goes in bin } 1 \\ 0 & \text{otherwise} \end{cases}$

indicator random variable

$\mathbb{E}(I) = \mathbb{P}(I)$

$$\mathbb{E}\left((M_1 - \mu)^2\right) = \mathbb{E}(M_1^2) - \mathbb{E}(M_1)^2 \leq \frac{m}{n} \quad (\leq 1)$$

$$\mathbb{P}(M \geq t) \leq n \cdot \mathbb{P}(M_1 \geq t) \leq n \cdot \frac{\mathbb{E}\left(\left(M_1 - \frac{m}{n}\right)^2\right)}{t^2}$$

$$\leq n \cdot \frac{m/n}{t^2} = \frac{m}{t^2}$$

$$\mathbb{P}(M \geq 10 \cdot \sqrt{m}) \leq \frac{1}{100}$$

(Max load is at most $\approx \sqrt{m}$)

Chernoff Bounds

Let $Z_1, \ldots, Z_n$ be independent r.v.'s such that

$$Z_i = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{w.p. } 1-p_i \end{cases}$$

and $Z = Z_1 + \ldots + Z_n$. Let $\mu = \mathbb{E}(Z) = p_1 + \ldots + p_n$

<u>Thm:</u> $\mathbb{P}(Z > (1+\delta)\mu) \leq \left(\dfrac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right)^{\mu}$

Z has mean μ and variance $\sum_{i=1}^{n} p_i(1-p_i) \approx \mu$

$Z \approx N(\mu, \sum_i p_i(1-p_i))$

Chernoff Bound Proof

Let $Z_1, \ldots, Z_n$ be independent r.v.'s such that

$$Z_i = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{w.p. } 1-p_i \end{cases}$$

and $Z = Z_1 + \ldots + Z_n$. Let $\mu = \mathbb{E}(Z) = p_1 + \ldots + p_n$

<u>Thm:</u> $\mathbb{P}(Z > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

$$\mathbb{P}(Z > a\mu) = \mathbb{P}(e^{tZ} > e^{ta\mu})$$

$$\leq e^{-ta\mu} \cdot \mathbb{E}(e^{tZ})$$

$$= e^{-ta\mu} \cdot \prod_{i=1}^{n} \mathbb{E}(e^{tZ_i}) = e^{-ta\mu} \cdot \prod_{i=1}^{n} (p_i e^{t} + 1 - p_i) = e^{-ta\mu} \cdot \prod_{i=1}^{n} (1 + p_i(e^{t}-1))$$

$$\leq e^{-ta\mu} \cdot \prod_{i=1}^{n} e^{p_i(e^{t}-1)} = e^{-ta\mu} \cdot e^{(e^{t}-1) \cdot \sum_{i=1}^{n} p_i} = e^{-ta\mu} \cdot e^{(e^{t}-1)\mu}$$

$$= e^{(e^{t}-1)\mu - ta\mu}$$

$(a = (1+\delta))$ $$= \left(e^{e^{t}-1-ta}\right)^{\mu}$$ $\leftarrow$ Set $a = 1+\delta$, plug in the right value of t

Applying Chernoff to Balls and Bins

M_1 = load on bin 1

$$M_1 = M_{1,1} + \dots + M_{1,m}$$

where $M_{1,i}$ is an indicator

$$\mathbb{E}(M_{1,i}) = \frac{1}{n}$$

Let $Z_1, \dots, Z_n$ be independent r.v.'s such that

$$Z_i = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{w.p. } 1-p_i \end{cases}$$

and $Z = Z_1 + \dots + Z_n$. Let $\mu = \mathbb{E}(Z) = p_1 + \dots$

Thm: $\mathbb{P}(Z > (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$

$$\mathbb{P}(M \geq (1+\delta)\tfrac{m}{n}) \leq n \cdot \mathbb{P}(M_1 \geq (1+\delta)\tfrac{m}{n}) \leq n \cdot \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{\frac{m}{n}}$$

m/n "large"

Assume $\frac{m}{n} \geq 10 \log n$

$$\mathbb{P}(M \geq 3 \cdot \tfrac{m}{n}) \leq n \cdot (.9)^{\frac{m}{n}}$$

$$\leq n \cdot \frac{1}{n^2} = \frac{1}{n}$$

$\frac{m}{n}$ is "small" e.g. $m=n$

if $\delta \approx \log n$

$$\mathbb{P}(M \geq \log n) \leq n \cdot \left(\frac{e^{\log n}}{(\log n)^{\log n}}\right)^1$$

$$\leq \frac{1}{n}$$