$\left\{ \text{Csc 80030} \mid \text{Lecture 3} \right\}$

PROBABILISTIC ANALYSIS &
RANDOMIZED ALGORITHMS

Hoeteck Wee · hoeteck@cs.qc.edu

PART 0

- HOMEWORK: HW1 is due today; HW2 out by next Mon.
- TODAY: quicksort & tail bounds

PART 1 | random permutations & quicksort

Fixed points in permutations

Definition: fixed point

A *fixed point* in a permutation $\pi : [n] \to [n]$ is an input i such that $\pi(i) = i$.

Question

What is the expected number of fixed points in a random permutation $\pi : [n] \to [n]$?

- Write $X = X_1 + \cdots + X_n$ where X_i indicates whether i is a fixed point.
- $\mathrm{E}[X_i] = \Pr[X_i = 1] = \frac{1}{n}$.
- $\mathrm{E}[X] = 1$.

Sorting

SORTING Problem

Input: a list of n distinct values $x_1, \ldots, x_n$ from some domain

Goal: output a sorted list

- Total order on the domain.
- Complexity measure: # of comparisons
- $\log(n!) = \Omega(n \log n)$ comparisons are *necessary*

Quicksort

QUICKSORT Algorithm

Input: A list $S = \{x_1, \ldots, x_n\}$

1. if $|S| = 0$ or $|S| = 1$, return S and halt.
2. Choose a random element x of S as pivot.
3. Let $S_1 = \{y \in S \mid y < x\}$ and $S_2 = \{y \in S \mid y > x\}$
4. Return the list $\text{QUICKSORT}(S_1), x, \text{QUICKSORT}(S_2)$.

- Compute expectation of X, # of comparisons.
- Call $y_1, \ldots, y_n$ the sorted list.
- Fact: any pair $y_i \neq y_j$ is compared at most once.

Analysis of QUICKSORT

1. Write $X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij}$ where X_{ij} indicates y_i and y_j are compared.
2. Compute $\mathrm{E}[X_{ij}]$:
   - Consider the first element x in $S_{ij} = \{y_i, y_{i+1}, \ldots, y_j\}$ to be used as a pivot.
   - If x equals y_i or y_j, then $X_{ij} = 1$.
   - If $x \neq y_i, y_j$, then $X_{ij} = 0$.
   - $\Pr[X_{ij} = 1] = \frac{2}{|S_{ij}|} = \frac{2}{j-i+1}$.
3. Computing

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{|S_{ij}|} \\
&\leq \sum_{i=1}^{n} \left(\frac{2}{2} + \frac{2}{3} + \cdots + \frac{2}{n}\right) \\
&\leq \sum_{i=1}^{n} 2H(n) = 2nH(n) = 2n\ln n + \Theta(n)
\end{aligned}$$

PART 2 | tail bounds

the case against expectation

QUESTION. How early should I arrive at the airport?

STATISTIC. The expected security wait time is 30 mins.

- perhaps... 50% : wait is 5 mins, 50% : wait is 55 mins
- more meaningful: 99% : wait $\leq$ 35 mins

PHILOSOPHY.

“ If you’ve never missed a flight, you’re spending too much time in airports. ” – U. Vazirani

TAIL BOUNDS.

“with high probability, a r.v. X assumes values close to $\mathrm{E}[X]$.”

Tail bounds, I

Markov's Inequality

Let X be a non-negative r.v. Then, for all $a > 0$,

$$\Pr[X \geq a] \leq \mathrm{E}[X]/a$$

- PROOF:

$$\begin{aligned} \mathrm{E}[X] &= \sum_{i=0}^{\infty} i \Pr[X = i] \\ &= \sum_{0 \leq i < a} i \Pr[X = i] + \sum_{i \geq a} i \Pr[X = i] \\ &\geq 0 + \sum_{i \geq a} a \Pr[X = i] = a \cdot \Pr[X \geq a] \end{aligned}$$

- EXAMPLE: $\mathrm{E}[\text{wait}] = 30\,\text{mins} \implies \Pr[\text{wait} \geq 5\,\text{hrs}] \leq \frac{30}{5 \cdot 60} = 0.1$

Tail bounds, II

Chebyshev's Inequality

For any $a > 0$,
$$\Pr\left[\left|X - \mathrm{E}[X]\right| \geq a\right] \leq \mathrm{Var}[X]/a^2$$

- PROOF: apply Markov's to the non-negative r.v. $Y = (X - \mathrm{E}[X])^2$

$$\Pr[Y \geq a^2] \leq \mathrm{E}[Y]/a^2 = \mathrm{Var}[X]/a^2$$

- EXAMPLE: suppose $\mathrm{Var}[\text{wait}] = 5\,\text{mins}^2$. Then,

$$\Pr[|\text{wait} - 30| \geq 10] \leq \frac{5}{10^2} = 0.05$$
$$\implies \quad 95\%\text{: wait between 20 and 40 mins}$$

Example: coin flips

- X : # heads in a sequence of n independent flips of an unbiased coin.
- $X \sim B(n, \frac{1}{2})$, so $\mathrm{E}[X] = \frac{n}{2}$ and $\mathrm{Var}[X] = \frac{n}{4}$.
- By Markov's, $\Pr[X \geq \frac{3n}{4}] \leq \frac{2}{3}$

  $n = 200$: 33% chance # heads less than 150
- By Chebyshev's, $\Pr[|X - \frac{n}{2}| \geq \frac{n}{4}] \leq \frac{n}{4}/(\frac{n}{4})^2 = \frac{4}{n}$.

  $n = 200$: 98% chance # heads between 50 and 150
- In fact, can replace $\frac{4}{n}$ with $2^{-\Omega(n)}$!

  $n = 200$: 99.95% chance # heads between 50 and 150

Tail bounds, III

Chernoff Bounds

Let $X_1, \ldots, X_n$ be *independent* r.v.'s assuming values in $\{0, 1\}$. Let $X = X_1 + X_2 + \cdots + X_n$ and $\mu = \mathrm{E}[X]$. Then,

1. For all $0 < \delta < 1$,

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$

2. For all $0 < \delta < 1$,

$$\Pr[X \leq (1 - \delta)\mu] \leq e^{-\mu\delta^2/2}$$

3. For all $\delta > 0$,

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

Comparison of tail bounds

- GENERALITY: Markov's $\gg$ Chebyshev's $\gg$ Chernoff

  (non-negative $\cdot$ bounded variance $\cdot$ independence)

- "ERROR": Markov's $\ll$ Chebyshev's $\ll$ Chernoff

  (constant $\cdot$ $1/\text{poly}$ $\cdot$ exponential)

- "DEVIATION": Markov's $\ll$ Chebyshev's $=$ Chernoff

  (one-sided $\cdot$ two-sided $\cdot$ two-sided)

THE END | next, parameter estimation