Lecture 8 Randomized Algorithms

Thursday, September 21, 2017 2:51 PM

- relationship between joint probabilities and conditional probabilities

$$Pr[X=i, Y=j] = Pr[X=i] \cdot \boxed{Pr[Y=j \mid X=i]}$$
$$= Pr[Y=j] \cdot Pr[X=i \mid Y=j]$$

$$Pr[Y=j \mid X=i] = \frac{Pr[Y=j] \cdot Pr[X=i \mid Y=j]}{Pr[X=i]}$$

Bayes Law

- prove linearity of expectation

$$E[X+Y] = \sum_k Pr[X+Y=k] \cdot k$$

$$Pr[X+Y=k] = \sum_{\substack{(i,j) \\ i+j=k}} Pr[X=i, Y=j]$$

$$E[X+Y] = \sum_k \sum_{\substack{(i,j) \\ i+j=k}} Pr[X=i, Y=j] \cdot k$$
$$= \sum_k \sum_{\substack{(i,j) \\ i+j=k}} Pr[X=i, Y=j] (i+j)$$

$$\sum_k \sum_{\substack{(i,j) \\ i+j=k}} Pr[X=i, Y=j] \cdot i \qquad \underbrace{\sum_k \sum_{\substack{(i,j) \\ i+j=k}} Pr[X=i, Y=j] \cdot j}_{E[Y]}$$

$$= \sum_{(i,j)} Pr[X=i, Y=j] \cdot i$$
$$= \sum_i \Big(\underbrace{\sum_j Pr[X=i, Y=j]}_{Pr[X=i]}\Big) i$$
$$= \sum_i Pr[X=i] \cdot i$$
$$= E[X]$$

- Quick Sort
  - Worst case example

    suppose we always pick the first number

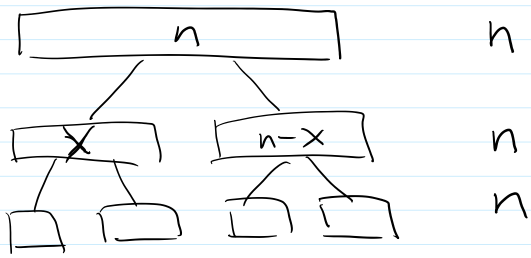
$$\{1, 2, 3, \dots, n\}$$
$$\downarrow$$
$$\{2, 3, \dots, n\}$$
$$\downarrow$$

$\{3, \quad \cdots \quad , n\}$

$\cdots$

running time $= \Theta(n^2)$

- randomness help
  - try to analyze the expected running time of quicksort.



intuition: running time "looks like" $O(n \log n)$

- use induction

Let X_n be running time of quicksort with n numbers

$\mathbb{E}[X_n]$

Induction hypothesis: $\mathbb{E}[X_n] \le \underline{C}\, n \log_2 n$

recursion

$$\mathbb{E}[X_n] = \sum_{i=1}^{n} \underbrace{\Pr[\text{pivot number is } i\text{th smallest}]}_{\frac{1}{n}} \times \underbrace{\mathbb{E}\left[X_n \,\middle|\, \begin{array}{l}\text{pivot number}\\ \text{is } i\text{th smallest}\end{array}\right]}_{\substack{A \cdot n \text{ (split the array)} \\ + X_{i-1} + X_{n-i}}}$$

X_{i-1}: Left array has size $i-1$

X_{n-i}: Right array has size $n-i$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{E}[X_{i-1}] + \mathbb{E}[X_{n-i}]\right) + An$$

Base case: $\mathbb{E}[X_0] = 0 \quad \mathbb{E}[X_1] = 0$

Induction step:

$$\mathbb{E}[X_n] = \frac{1}{n} \sum_{i=1}^{n} \Big(\underbrace{\mathbb{E}[X_{i-1}]}_{0,1,2,\ldots,n-1} + \underbrace{\mathbb{E}[X_{n-i}]}_{n-1,n-2,\ldots,0} \Big) + A \cdot n$$

$\left(\mathbb{E}[X_t] \le C\, t \log_2 t \text{ for all } t < n\right)$

$$= \frac{2}{n} \sum_{i=1}^{n} \mathbb{E}[X_{i-1}] + An$$

$$\le \frac{2}{n} \cdot C \cdot \sum_{i=1}^{n} (i-1) \log_2 (i-1) + An$$

$$\leq \frac{2}{n} \cdot C \cdot \sum_{i=1} (i-1) \log_2 (i-1) + An$$

$$\leq \frac{2}{n} C \cdot \left[\sum_{i=1}^{\frac{n}{2}} (i-1) \log_2 \frac{n}{2} + \sum_{i=\frac{n}{2}+1}^{n} (i-1) \log_2 n \right] + A \cdot n$$

$$= \frac{2}{n} C \left[\sum_{i=1}^{\frac{n}{2}} (i-1) (\log_2 n - 1) + \sum_{i=\frac{n}{2}+1}^{n} (i-1) \log_2 n \right] + A \cdot n$$

$$= \frac{2}{n} C \left[\sum_{i=1}^{n} (i-1) \log_2 n - \sum_{i=1}^{\frac{n}{2}} (i-1) \right] + A \cdot n$$

$$= \frac{2}{n} C \cdot \frac{n(n-1)}{2} \log_2 n - \frac{2}{n} \cdot C \cdot \frac{\frac{n}{2}(\frac{n}{2}-1)}{2} + An$$

ignored for simp

$$\leq C \cdot n \cdot \log_2 n - C \cdot \frac{n}{4} + A \cdot n$$

$$\leq C \cdot n \cdot \log_2 n \quad (\text{when } C \geq 4 \cdot A) \qquad \square$$

Therefore the running time is bounded by $4 \cdot A \cdot \log_2 n$.