Randomness in Computing

CS 537

LECTURE 6

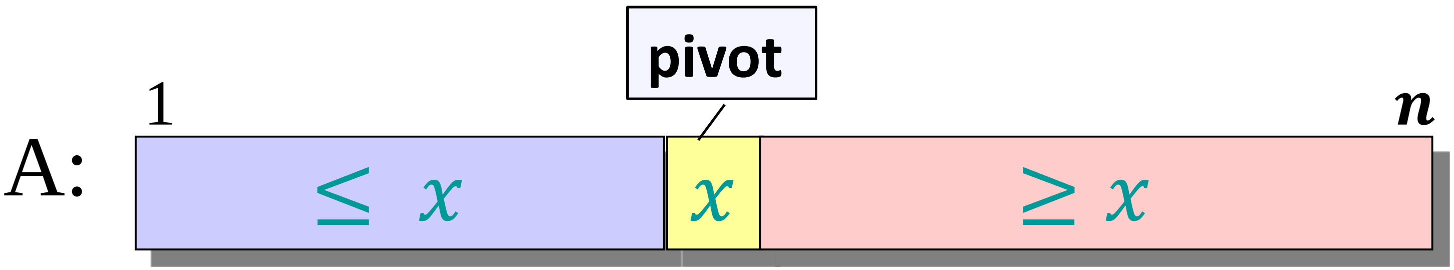
Last time

- Conditional expectation
- Branching process
- Geometric RVs
- Coupon collector problem

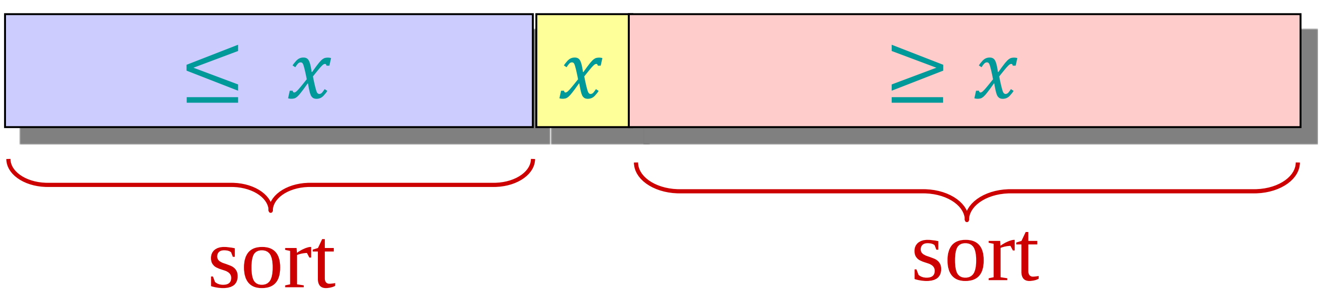
- **Today**
- Randomized quicksort
- Markov’s inequality
- Variance

Quicksort: divide and conquer

- Find a *pivot* element
- Divide: Find the correct position of the pivot by comparing it to all elements.



- Conquer: Recursively sort the two parts, resulting from removing the pivot.



Exercise

How many comparisons does Quicksort perform on sorted array?

Answer:

$$(n-1)+(n-2)+\cdots+2+1=\frac{n(n-1)}{2}=\Omega(n^2)$$

How many comparisons does Quicksort perform if, in every iteration, the pivot splits the array into two halves?

Answer:

Let $C(n)$ be the number of comparisons performed on an array with n elements.

$$C(n)=\Theta(n\log n)$$

Randomized Quicksort

BIG IDEA:

Partition around a ***random*** element.

- Analysis is similar when the input arrives in random order.
- But randomness in the input is unreliable.
- Rely instead on random number generator.

Analysis of Randomized Quicksort

Theorem. If Quicksort chooses each pivot uniformly and independently at random from all possibilities then, for any input, the expected number of comparisons is

$$2n \ln n + O(n).$$

Proof (with an assumption that all elements are distinct):

- Let X be the R.V. for the # of comparisons.
- Let $x_1, x_2, \dots, x_n$ be the input values.
- Let $y_1, y_2, \dots, y_n$ be the input values sorted in increasing order.
- For $i, j \in [n], i < j$, let X_{ij} be the indicator R.V. for the event that y_i and y_j are compared by the algorithm.

$$X = \sum_{i,j[n]:i<j} X_{ij} \text{ and, by linearity of expectation, } \mathbb{E}[X] = \sum_{i,j[n]:i<j} \mathbb{E}\left[X_{ij}\right]$$

Analysis of Randomized Quicksort

Theorem. The expected number of comparisons is $2n \ln n + O(n)$.

Proof (continued):

- Let $y_1, y_2, \ldots, y_n$ be the input values sorted in increasing order.
- For $i, j \in [n], i < j$, let X_{ij} be the indicator R.V. for the event that y_i and y_j are compared by the algorithm.
- $\mathbb{E}\left[X_{ij}\right] = \Pr[\mathrm{X_{ij}} = 1]$
- y_i and y_j are compared iff
either y_i or y_j is the first pivot chosen from $Y_{ij} = \{y_i, \ldots, y_j\}$
- The first time a pivot is chose from Y_{ij}, it is equally likely to be any of $j - i + 1$ elements of Y_{ij}.

Analysis of Randomized Quicksort

Theorem. The expected number of comparisons is $2n \ln n + O(n)$.

Proof (continued):

Markov's Inequality

Theorem. Let X be a RV taking only nonnegative values. Then, for all $a > 0$,

$$\Pr[X \geq a] \leq \frac{\mathbb{E}\,[X]}{a}.$$

Proof: Let a > 0.

Markov's Inequality

Theorem. Let X be a RV taking only nonnegative values. Then, for all $a > 0$,

$$\Pr[X \geq a] \leq \frac{\mathbb{E}\,[X]}{a}.$$

Alternative proof: Let a > 0.

Markov's Inequality

- Theorem. Let X be a RV taking only nonnegative values. Then, for all $a > 0$,

$$\Pr[X \geq a] \leq \frac{\mathbb{E}\,[X]}{a}.$$

- Alternatively: Then, for all $b > 1$,

$$\Pr[X \geq b \cdot \mathbb{E}[X]] \leq \frac{1}{b}.$$

- Example: Show that Randomized Quicksort uses

$$\leq 10n \ln n + O(n)$$

comparisons with probability at least 4/5.

Bernoulli random walk

- Start at position 0.
- At every step go up or down by 1 with probability 1/2 each.
- Let X be the position after n steps.

What is the probability space?

A. Uniform over {0,1}.

B. Uniform over {-1,1}.

C. 2^n.

D. Position after n steps.

E. Uniform over $\{(s_1, \dots, s_n) \mid s_i \in \{-1,1\}$ for $i = 1, \dots, n\}$.

Bernoulli random walk

- Start at position 0.
- At every step go up or down by 1 with probability 1/2 each.
- Let X be the position after n steps.
- What is E[X]?
- How far from the origin should we expect X to be?

A precise answer to this question is the expectation of |X|.
However, it is easier to work with the expectation of X^2.
(It is not the same! But gives us an idea.)

Random variables: variance

- The variance of a random variable X with expectation $E[X] = \mu$ is
$$\text{Var}[X] = E[(X - \mu)^2].$$
- Equivalently, $\text{Var}[X] = E[X^2] - \mu^2$.

- The standard deviation of X is $\sigma[X] = \sqrt{\text{Var}[X]}$.

Variance as a measure of spread

- $X = \begin{cases} -2 \text{ with probability } 1/2 \\ 2 \text{ with probability } 1/2 \end{cases}$
- $Y = \begin{cases} -10 \text{ with probability } 0.001 \\ 0 \text{ with probability } 0.998 \\ 10 \text{ with probability } 0.001 \end{cases}$
- $Z = \begin{cases} -5 \text{ with probability } 1/3 \\ 0 \text{ with probability } 1/3 \\ 5 \text{ with probability } 1/3 \end{cases}$
- Compute the variances and standard deviations of X,Y and Z.

Example from Mathematics: a Discrete introduction by E. Scheinerman