

Analysis of Algorithms CS 3343 Lecture Nine

Prof. William Winsborough
October 21, 2008

Business

- Homework 5 due this Thursday, 10/23
 - Problem 7.1 (p.159)
 - If you are truly unable to turn this in Thursday, I will accept it Tuesday, 10/28.
 - But if you do turn it in Thursday, I plan to return in Tuesday so you can refer to your solutions when I go over the problem in class.
- Read 7.1, 7.2, 5.1, 5.2, 7.3, 7.4

21 October 2008

Winsborough CS 3343 Lecture 9

2

Quicksort (Ch. 7)

- Divide and conquer
 - Divide: partition $A[p..r]$ into two (possibly empty) subarrays $A[p..q-1]$ and $A[q+1..r]$ such that:
 - Each element of $A[p..q-1]$ is less than or equal to $A[q]$, which is less than or equal to each element of $A[q+1..r]$
 - Return q
 - Conquer: recursively sort $A[p..q-1]$ and $A[q+1..r]$
 - Combine: no work needed; the entire array is now sorted
- Worst-case runtime: $O(n^2)$
- Best-case and expected-case runtime: $O(n \lg n)$

21 October 2008

Winsborough CS 3343 Lecture 9

5

Pseudocode for Quicksort

- Refer to p. 146
- What loop invariant of partition enables us to show that the specification on the previous slide is met?

21 October 2008

Winsborough CS 3343 Lecture 9

6

Performance: Worst Case

- What input produces the worst-case runtime?
 - Completely unbalanced divide, eg. if input is already sorted
- $T(n) = T(n-1) + T(0) + \Theta(n)$
 $= T(n-1) + \Theta(n)$
 $= \Theta(n^2)$

21 October 2008

Winsborough CS 3343 Lecture 9

7

Performance: Best Case

- What input produces the best-case runtime?
 - Perfectly balanced divide, eg. if the median value is in the last position of the array
- $T(n) = 2T(n/2) + \Theta(n)$
 $= \Theta(n \lg n)$

21 October 2008

Winsborough CS 3343 Lecture 9

8

Performance: More or Less Balanced Partition

- Suppose that the partition always produces a 9-to-1 split
 - $T(n) \leq T(9n/10) + T(n/10) + cn$
- Refer to recursion tree in fig 7.4, p.151
 - Every level has cost cn down to level $\log_{10}n$
 - The rest of the levels have cost $\leq cn$
 - There are $\log_{10}n = \Theta(\lg n)$ levels in total
 - The total cost is $O(n \lg n)$
- In general, if the split has constant proportionality, the total cost remains $O(n \lg n)$
 - Constants are affected by degree of balance

21 October 2008

Winsborough CS 3343 Lecture 9

9

Intuition about Average Case

- In the average case, partition produces a mix of good and bad splits
- The good splits tend to prevent the bad splits from creating runtime worse than $O(n \lg n)$
- We will study the expected cost of a randomized version of quicksort
 - But first, let's talk about randomized algorithms in general

21 October 2008

Winsborough CS 3343 Lecture 9

10

The Hiring Problem

- Goal: hire the best office assistant while always retaining the best applicant seen so far
 - Employment agency sends one candidate per day: small fee per candidate
 - If the candidate is better than the current employee, fire current and hire candidate: large fee
- Cost is $O(nc_i + mc_h)$
 - Interview n applicants (constant)
 - c_i is cost of interviewing
 - Hire m applicants (depends on input's order)
 - c_h is cost of hiring
- What is the worst case and what is its cost?

21 October 2008

Winsborough CS 3343 Lecture 9

11

Probabilistic Analysis

- There are $n!$ permutations of the n applicants
 - If we can assume that the employment agency selects which candidate to send in a random order, we can assume that each permutation is equally likely
 - Called a *uniform random permutation*
- If we don't know anything about the probability distribution of the input permutations, we can introduce randomness ourselves
 - *Randomized algorithms*
 - Eg., suppose we can tell the employment agency which candidate to send next
 - We can pick the candidate at random, ensuring that the sequence of candidates forms a uniform random permutation

21 October 2008

Winsborough CS 3343 Lecture 9

12

Indicator Random Variables

- Given a sample space S and an event A , the *indicator random variable* $I\{A\}$ associated with even A is defined by:
 - $I\{A\} = 1$ if A occurs
 - $I\{A\} = 0$ if A does not occur
- Eg., coin flip: $S = \{H, T\}$, $\Pr\{H\} = \Pr\{T\} = 1/2$
- Indicator random variable
 - $X_{H} = I\{H\} = 1$ if H occurs; 0 if T occurs

21 October 2008

Winsborough CS 3343 Lecture 9

13

Expected Value

- Expected number of heads obtain in one coin flip:

$$\begin{aligned} E[X_H] &= E[I\{H\}] \\ &= 1 \cdot \Pr\{H\} + 0 \cdot \Pr\{T\} \\ &= 1 \cdot 1/2 + 0 \cdot 1/2 \\ &= 1/2 \end{aligned}$$
- In general, it always holds that $E[X_A] = \Pr\{A\}$

21 October 2008

Winsborough CS 3343 Lecture 9

14

Linearity of Expected Value

- In n coin flips, let the indicator random variable $X_i = I\{\text{the } i\text{'th flip results in heads}\}$
- Suppose X is the random variable denoting the total number of heads in n coin flips
 $X = X_1 + X_2 + \dots + X_n$
- To compute the expected number of heads we can use the linearity of expectation to determine that
 $E[X] = n/2$ (p.96)

21 October 2008

Winsborough CS 3343 Lecture 9

15

Returning to the Hiring Problem

- Assume n applicants arrive in random order
- Let X be the random variable denoting the number of hires
- Let X_i be the random variable associated with hiring the i 'th applicant
- $E[X] = E[X_1] + E[X_2] + \dots + E[X_n]$
- Have: $E[X_i] = \Pr\{\text{candidate } i \text{ is hired}\} = 1/i$
- So $E[X] = 1/1 + 1/2 + \dots + 1/n \leq \lg n + O(1)$

21 October 2008

Winsborough CS 3343 Lecture 9

16