Probabilistic Analysis & Randomized Algorithms

Probabilistic (Average-Case) Analysis and Randomized Algorithms

- Two different but similar analyses
  - Probabilistic analysis of a deterministic algorithm
  - Analysis of a Randomized algorithm
- Tools from probability theory
  - Indicator variables
  - Linearity of expectation
- Example Problems/Algorithms
  - Hiring problem (Chapter 5)
  - Biased coin
  - Sorting/Quicksort (Chapter 7)
  - Hat Check Problem
  - Online Hiring Problem

Probabilistic (Average-Case) Analysis

- Algorithm is deterministic; for a fixed input, it will run the same every time
- Analysis Technique
  - Assume a probability distribution for your inputs
  - Analyze item of interest over the distribution
- Caveats
  - Specific inputs may have much worse performance
  - If the distribution is wrong, analysis may give misleading picture

Randomized Algorithm

- "Randomize" the algorithm; for a fixed input, it will run differently depending on the result of random "coin tosses"
- Randomization examples/techniques
  - Randomize the order that candidates arrive
  - Randomly select a pivot element
  - Randomly select from a collection of deterministic algorithms
- Key points
  - Works well ***with high probability*** on ***every*** input
  - May fail on every input with low probability

Key Analysis Tools

- Indicator variables
  - Suppose we want to study random variable X that represents a composite of many random events
  - Define a collection of "indicator" variables X_i that focus on individual events; typically $X = \Sigma X_i$
- Linearity of expectations
  - Let X, Y, and Z be random variables s.t. X = Y + Z
  - Then E[X] = E[Y+Z] = E[Y] + E[Z]
- Recurrence Relations

Hiring Problem

- Input
  - A sequence of n candidates for a position
  - Each has a distinct quality rating that we can determine in an interview
- Algorithm
  - Current = 0;
  - For k = 1 to n
    - If candidate(k) is better than Current, hire(k) and Current = k;
- Cost:
  - Number of hires
- Worst-case cost is n

Analyze Hiring Problem

- Assume a probability distribution
  - Each of the n! permutations is equally likely
- Analyze item of interest over probability distribution
  - Define random variables
    - Let X = random variable corresponding to # of hires
    - Let X_i = "indicator variable" that i^{th} interviewed candidate is hired
      - Value 0 if not hired, 1 if hired
    - $X = \Sigma_{i = 1 \text{ to } n} X_i$
  - $E[X_i] = ?$
    - Explain why:
  - $E[X] = E[\Sigma_{i = 1 \text{ to } n} X_i] = \Sigma_{i = 1 \text{ to } n} E[X_i] = ?$
    - Key observation: linearity of expectations

Alternative analysis

- Analyze item of interest over probability distribution
  - Let X_i = indicator random variable that the i^{th} best candidate is hired
    - 0 if not hired, 1 if hired
  - Questions
    - Relationship of X to X_i?
    - $E[X_i] = ?$
    - $\Sigma_{i = 1 \text{ to } n} E[X_i] = ?$

Questions

- What is the probability you will hire n times?
- What is the probability you will hire exactly twice?
- Biased Coin
  - Suppose you want to output 0 with probability ½ and 1 with probability ½
  - You have a coin that outputs 1 with probability p and 0 with probability 1-p for some unknown $0 < p < 1$
  - Can you use this coin to output 0 and 1 fairly?
  - What is the expected running time to produce the fair output as a function of p?