Chap 5 – Probability Analysis and Randomized Algorithms

5.1 The hiring problem

5.2 Indicator random variable

5.3 Randomized algorithms

5.4. … further uses of indicator random …

5.1 The hiring problem

- The hiring problem (maximization or minimization)

HIRE-ASSISTANT(n)

$best = 0$ // least-qualified dummy candidate

for $i = 1$ **to** n

  interview candidate i // interview cost c_i

  if candidate i is better than candidate $best$

    $best = i$

    hire candidate i // hire cost c_h

If n candidates, and we hire m of them, the cost is

$O(nc_i + mc_h)$

Assume that $c_h > c_i$, we shall concentrate on analyzing mc_h.

5.1 The hiring problem

- Worst case analysis
  - In the worst case, the candidates appear in increasing order of quality and the hiring cost is $O(nc_h)$.
- Probabilistic analysis (Average case analysis)
  - Depend on probabilistic assumption about input distribution
  - Usually, the assumption is uniform random permutation, i.e. all of the $n!$ permutations are equally likely.
  - The algorithm is deterministic.
- Randomized algorithm
  - The n candidates are permuted uniformly at random to enforce the probabilistic assumption.

5.2 Indicator random variables

- Expected value of a random variable (Appendix C.3)
  - A discrete random variable X is a function

    $X: S \to$ real or integer values

    where S is a finite or countably infinite sample space.
  - The event $X = x$ denotes the set $\{s \in S \mid X(s) = x\}$.
  - The probabilty of occurrence of the event $X = x$ is

    $$\Pr\{X = x\} = \sum_{s \in S : X(s) = x} \Pr\{s\}$$
  - The expectation of X is

    $$\mathrm{E}[X] = \sum_{x} x \Pr\{X = x\}$$

5.2 Indicator random variables

- Expected value of a random variable (Appendix C.3)
  - Example

    Let X be a random variable denoting the number of times number 6 appears in 3 dice casts, then

    $X: S \to \{0,1,2,3\}$, e.g. $X(5,6,6) = 2$

    where $S = \{(d_1, d_2, d_3) | \ 1 \le d_i \le 6\}$ is the sample space

    $X = 2$

    - denotes the event "two of 3 dice casts are 6"
    - denotes $S' = \{(6,6,d), (6,d,6), (d,6,6) | \ 1 \le d \le 5\}$
    - occurs with probability

      $$\Pr\{X = 2\} = \sum \Pr\{s\} = \sum \frac{1}{216} = \frac{15}{216}$$

5.2 Indicator random variables

- Expected value of a random variable (Appendix C.3)
  - Example (Cont'd)

    We also have

    $\Pr\{X = 0\} = 125/216$

    $\Pr\{X = 1\} = 75/216$

    $\Pr\{X = 3\} = 1/216$

    The expected number of times 6 appears in 3 dice casts is

$$\mathrm{E}[X] = \sum_{x=0}^{3} x\Pr\{X = x\}$$

$$= 0 \cdot \frac{125}{216} + 1 \cdot \frac{75}{216} + 2 \cdot \frac{15}{216} + 3 \cdot \frac{1}{216} = \frac{1}{2}$$

5.2 Indicator random variables

- Expected value of a random variable (Appendix C.3)
  - Example

    Let X be a random variable denoting the number of times 6 appears in n dice casts

    Then, $0 \leq X \leq n$

    We have

$$\mathrm{E}[X] = \sum_{x=0}^{n} x \binom{n}{x} \left(\frac{1}{6}\right)^{x} \left(\frac{5}{6}\right)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n}{x} \binom{n-1}{x-1} \left(\frac{1}{6}\right)^{x} \left(\frac{5}{6}\right)^{n-x}$$

5.2 Indicator random variables

- Expected value of a random variable (Appendix C.3)
  - Example (Cont'd)

$$\mathrm{E}[X] = \frac{n}{6}\sum_{x=1}^{n}\binom{n-1}{x-1}\left(\frac{1}{6}\right)^{x-1}\left(\frac{5}{6}\right)^{n-x}$$

$$= \frac{n}{6}\sum_{x=0}^{n-1}\binom{n-1}{x}\left(\frac{1}{6}\right)^{x}\left(\frac{5}{6}\right)^{n-x-1}$$

$$= \frac{n}{6}\left(\frac{1}{6}+\frac{5}{6}\right)^{n-1} = \frac{n}{6}$$

Drawback: Hard to compute

5.2 Indicator random variables

- Expected value of a random variable (Appendix C.3)
  - Example (Cont'd)

    Alternative analysis

    Let X_i be a random variable denoting the number of times 6 appears in the i^{th} dice cast

    Then,

    $X_i = 0 \text{ or } 1$

    $\Pr\{X_i = 0\} = 5/6$

    $\Pr\{X_i = 1\} = 1/6$

    The expected number of times 6 appears in any dice cast is

    $\mathrm{E}[X_i] = 0 \cdot \frac{5}{6} + 1 \cdot \frac{1}{6} = \frac{1}{6} = \Pr\{X_i = 1\}$

5.2 Indicator random variables

- Expected value of a random variable (Appendix C.3)
  - Example (Cont'd)

    Next, linearity of expectation

    $$X = \sum_{i=1}^{n} X_i$$

    $$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{1}{6} = n/6$$

    X_i is called an indicator random variable, i.e.

    $X_i = 1$, if the event "the i^{th} dice cast comes out 6" occurs

    $= 0$, if the event doesn't occur

5.2 Indicator random variables

- Indicator random variables

  Given a sample space and an event A, define the indicator random variable:

  $$\mathrm{I}\{A\} = 1 \qquad \text{if } A \text{ occurs}$$
  $$= 0 \qquad \text{if } A \text{ doesn't occur}$$

  Lemma For an event A, let $X_A = \mathrm{I}\{A\}$. Then, $\mathrm{E}[X_A] = \Pr\{A\}$

  Proof

  Letting $\bar{A}$ be the complement of A, we have

  $$\begin{aligned} \mathrm{E}[X_A] &= \mathrm{E}[\mathrm{I}\{A\}] \\ &= 1 \cdot \Pr\{A\} + 0 \cdot \Pr\{\bar{A}\} \\ &= \Pr\{A\} \end{aligned}$$

5.2 Indicator random variables

- Probability analysis of the hiring problem

Assume that the candidates arrive in a random order.

Define indicator random variables $X_1, X_2, \dots, X_n$, where

$X_i = \{\text{candidate } i \text{ is hired}\}$

Then,

$$\mathrm{E}[X_i] = \Pr\{\text{candidate } i \text{ is hired}\}$$
$$= \Pr\{\text{candidate } i \text{ is the best so far}\} = 1/i$$

$\because$ Assumption that the candidates arrive in a random order

$\Rightarrow$ candidates 1, 2, …, i arrive in random order

$\Rightarrow$ any one of these first i candidates is equally likely to be the best one so far

5.2 Indicator random variables

- Probability analysis of the hiring problem

  Let X be a random variable that equals the number of times we hire a new office assistant. Then,

$$X = \sum_{i=1}^{n} X_i$$

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} 1/i = \ln n + O(1)$$

LEMMA 5.2

Assuming that the candidates are presented in a random order, algorithm HIRE-ASSISTANT has an average hiring cost of $O(c_h \ln n)$.

5.3 Randomized algorithms

- Randomized hiring assistant

  RANDOMIZED-HIRE-ASSISTANT(n)

  randomly permute the list of candidates

  HIRE-ASSISTANT(n)

  LEMMA

  The expected hiring cost of RANDOMIZED-HIRE-ASSISTANT is $O(c_h \ln n)$.

  Proof

  After permuting the input array, we have a situation identical to the probabilistic analysis of deterministic HIRE-ASSISTANT.

5.3 Randomized algorithms

- Deterministic vs randomized algorithms
  - Deterministic algorithms
    - Have best- or worst-case inputs
    - Talk of worst-case or average-case running time
    - For any particular input, the algorithm’s behavior is reproducible.
  - Randomized algorithms
    - All input cases are equal – there are no best- or worst-case inputs, only "lucky or unlucky probability"
    - Talk of expected running time
    - For any particular input, the algorithm’s behavior is not reproducible.

5.3 Randomized algorithms

- Randomly permuting arrays

PERMUTE-BY-SORTING(A)

$n = A.length$

let $P[1..n]$ be a new array // Example

for $i = 1$ **to** n // $A = a_1, a_2, a_3$

    $P[i] =$ RANDOM($1, n^3$) // $P = 17, 26, 5$

sort A, using P as sort keys // $A = a_3, a_1, a_2$

Drawbacks

1 Don't work in-place

2 Take $\Theta(n \lg n)$ time

3 Need more random bits

5.3 Randomized algorithms

- Randomly permuting arrays

A better in-place $\Theta(n)$ algorithm

RANDOMIZE-IN-PLACE(A, n)

for $i = 1$ to n

    swap $A[i]$ with $A[\text{RANDOM}(i, n)]$

LEMMA

RANDOMIZE-IN-PLACE computes a uniform random permutation.

Terminology

Given n elements, a k-permutation is a permutation of k elements chosen from the n elements.

5.3 Randomized algorithms

- Randomly permuting arrays

  LEMMA (Cont'd)

  Loop invariant: Just before the i^{th} iteration, $A[1..i-1]$ contains each possible $(i-1)$-permutation with probability $\frac{(n-i+1)!}{n!}$ $\left(= \frac{1}{n} \times \frac{1}{n-1} \times \cdots \times \frac{1}{n-i+2}\right)$

  Termination: $i = n + 1$

  If the loop invariant is true, then at termination $A[1..n]$ contains each possible n-permutation with probability

  $$\frac{(n-(n+1)+1)!}{n!} = \frac{1}{n!}$$

  as desired.

5.3 Randomized algorithms

- Randomly permuting arrays

  LEMMA (Cont'd)

  Proof of the loop invariant

  Initialization: $i = 1$

  Since $A[1..0]$ is an empty array and a 0-permutation has no elements, at the initialization $A[1..0]$ certainly contains any 0-permutation with probability $1 = (n - 1 + 1)!/n!$

  Maintenance

  $A[1..i-1]$ contains each $(i-1)$-permutation with prob. $(n - i + 1)!/n!$

  $\Rightarrow A[1..i]$ contains each i-permutation with prob. $(n - i)!/n!$

5.3 Randomized algorithms

- Randomly permuting arrays

  LEMMA (Cont'd)

  $A[1..i]$ contains $\langle x_1, x_2, \dots, x_i \rangle$ iff events E_1 and E_2 occur where

  E_1: $A[1..i-1]$ contains $\langle x_1, x_2, \dots, x_{i-1} \rangle$

  E_2: $A[i]$ contains x_i

  By induction hypothesis, we have

  $$\Pr\{E_1\} = \frac{(n-i+1)!}{n!}$$

  Also, since x_i is chosen randomly from $A[i..n]$, we have

  $$\Pr\{E_2 | E_1\} = \frac{1}{n-i+1}$$

5.3 Randomized algorithms

- Randomly permuting arrays

Lemma (Cont'd)

Therefore, $A[1..i]$ contains $\langle x_1, x_2, \ldots, x_i \rangle$ with probability

$$\Pr\{E_2 \cap E_1\} = \Pr\{E_2 | E_1\} \Pr\{E_1\}$$

$$= \frac{1}{n-i+1} \cdot \frac{(n-i+1)!}{n!}$$

$$= \frac{(n-i)!}{n!}$$