The Hiring Problem, Probability Review, Indicator Random Variables

CS255

Chris Pollett

Jan. 30, 2006.

Outline

- The Hiring Problem
- Probability Background
- Indicator Random Variables
- Analysis of the Hiring Problem

The Hiring Problem

We will now begin our investigation of randomized algorithms with a toy problem:

- You want to hire an office assistant from an employment agency.
- You want to interview candidates and determine if they are better than the current assistant and if so replace the current assistant.
- You are going to eventually interview every candidate from a pool of n candidates.
- You want to always have the best person for this job, so you will replace an assistant with a better one as soon as you are done the interview.
- However, there is a cost to fire and then hire someone.
- You want to know the expected price of following this strategy until all n candidates have been interviewed.

Pseudo-Code

```
Hire-Assistant(n)
best <- dummy candidate
for i <- 1 to n
    do interview of candidate i
        if candidate i is better than best
            then best <- i
                hire candidate i
```

Total Cost and Cost of Hiring

- Interviewing has a low cost c_i.
- Hiring has a high cost c_h.
- Let n be the number of candidates to be interviewed and let m be the number of people hired.
- The total cost then goes as $O(n*c_i+m*c_h)$
- The number of candidates is fixed so the part of the algorithm we want to focus on is the $m*c_h$ term.
- This term governs the cost of hiring.

Worst-case Analysis

- In the worst case, everyone we interview turns out to be better than the person we currently have.
- In this case, the hiring cost for the algorithm will be $O(n*c_h)$.
- This bad situation presumably doesn't typically happen so it is interesting to ask what happens in the average case.

Probabilistic analysis

- Probabilistic analysis is the use of probability to analyze problems.
- One important issue is what is the distribution of inputs to the problem.
- For instance , we could assume all orderings of candidates are equally likely.
- That is, we consider all functions rank: [0..n] --> [0..n] where rank[i] is supposed to be the *i*th candidate that we interview. So

  <rank(1), rank(2), …rank(n)> should be a permutation of of <1,…,n>
- There are n! many such permutations and we want each to be equally likely.
- If this is the case, the ranks form a **uniform random permutation**.

Randomized algorithms

- In order to use probabilistic analysis, we need to know something about the distribution of the inputs.
- Unfortunately, often little is known about this distribution.
- We can nevertheless use probability and analysis as a tool for algorithm design by having the algorithm we run do some kind of randomization of the inputs.
- This could be done with a random number generator. i.e.,
- We could assume we have primitive function Random(a,b) which returns an integer between integers a and b inclusive with equally likelihood.
- Algorithms which make use of such a generator are called randomized algorithms.
- In our hiring example we could try to use such a generator to create a random permutation of the input and then run the hiring algorithm on that.

Distributions

- A sample space S will for us be some collection on **elementary events**. For instance, results of coin flips.
- An **event** E is any subset of S.
- For example, if S={HH, TH, HT, TT}, an event might be {TH, HT}
- A probability distribution $Pr_S\{\}$ on S is a mapping from events on S to the real numbers satisfying for any events A and B:
  (a) $Pr_S\{A\} >= 0$
  (b) $Pr_S\{S\} = 1$
  (c) $Pr_S\{A \cup B\} = Pr_S\{A\} + Pr_S\{B\}$ if $A \cap B = \varnothing$
- Notice $1=Pr_S\{S \cup \varnothing\} = Pr_S\{S\} + Pr_S\{\varnothing\} = 1 + Pr_S\{\varnothing\}$. So $Pr_S\{\varnothing\} = 0$.
- Notice $1= Pr_S\{S\}= Pr_S\{A \cup \overline{A}\} = Pr_S\{A\} + Pr_S\{\overline{A}\}$. So $Pr_S\{\overline{A}\}=1-Pr_S\{\overline{A}\}$.

Conditional Probability and Independence

- The **conditional probability** of an event A given an event B is defined to be:
  $Pr\{A|B\} = Pr\{A \cap B\}/Pr\{B\}$.
- Two events are **independent** if
  $Pr\{A \cap B\} = Pr\{A\}Pr\{B\}$
- Given a collection $A_1, A_2, \ldots A_k$ of events we say they are **pairwise independent** if
  $Pr\{A_i \cap A_j\} = Pr\{A_i\}Pr\{A_j\}$ for any i and j.
- They are **mutually independent** if for an subset $A_{i\_1}, A_2, \ldots A_{i\_m}$ of then
  $Pr\{A_{i\_1} \cap \ldots A_{i\_m}\} = Pr\{A_{i\_1}\}* *Pr\{A_{i\_m}\}$

Discrete Random Variables

- A **discrete random variable** X is a function from a finite sample space S to the real numbers.
- Given such a function X we can define the **probability density function** for X as:

$$f(x) = \Pr\{X = x\}$$

where the little x is a real number.

Expectation and Variance

- The **expected value** of a random variable X is defined to be:

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

- The variance of X, Var[X] is defined to be:

  $E[(X- E(X))^2]= E[X^2] -(E[X])^2$

- The standard deviation of X, σ_X, is defined to be the $(Var[X])^{1/2}$.

Indicator Random Variables

- In order to analyze the hiring problem we need a convenient way to convert between probabilities and expectations.
- We will use indicator random variables to help us do this.
- Given a sample space S and an event A. Then the **indicator random variable** I{A} associated with event A is define as:

$$I\{A\} = \begin{cases} 1 & \text{if A occurs ,} \\ 0 & \text{if A does not occur .} \end{cases}$$

Example

- Suppose our sample space S={H,T} with Pr{H}=Pr{T}=1/2.
- We can define an indicator random variable X_H associated with the coin coming up heads:

$$X_H = I\{H\} = \begin{cases} 1 & \text{if H occurs ,} \\ 0 & \text{if T occurs .} \end{cases}$$

- The expected number of heads in one coin flip is then

$$\begin{aligned} E[X_H] &= E[I\{H\}] \\ &= 1 \cdot Pr\{H\} + 0 \cdot Pr\{T\} \\ &= 1 \cdot (1/2) + 0 \cdot (1/2) \\ &= 1/2 \end{aligned}$$

Lemma 5.1

Given a sample space S and an event A in S, let X_A=I{A}. Then E[X_A]=Pr{A}.

Proof: E[X_A] = E[I{A}] = 1*Pr{A}+ 0*Pr{A} =Pr{A}.

More Indicator Variables

- Indicator random variables are more useful if we are dealing with more than one coin flip.
- Let X_i be the indicator that indicates whether the result of the *i*th coin flip was a head.
- Consider the random variable: $X = \sum_{i=1}^{n} X_i$
- The the expected number of head in *n* tosses is

$$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} 1/2 = n/2$$

Analysis of the Hiring Problem

- Let X_i be the indicator random variable which is 1 if candidate *i* is hired and 0 otherwise.
- Let

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

- By our lemma $E[X_i]$ = Pr{candidate *i* is hired}
- Candidate *i* will be hired if *i* is better than each of candidates 1 through *i*-1.
- As each candidate arrives in random order, any one of the first candidate i is equally likely to be the best candidate so far. So $E[X_i] = 1/i$.

More analysis of hiring problem

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

Lemma Assume that the candidates are presented in random order, then algorithm Hire-Assistant has a hiring cost of $O(c_h \ln n)$

Proof. From before hiring cost is $O(m*c_h)$ where m is the number of candidate hired. From the lemma this is O(ln n).