Probabilistic Analysis

Yu-Tai Ching
Department of Computer Science
National Chiao Tung University

Hiring Problem

- Hire new office assistant through employment agency.
- Agency provides one candidate a day, interview, decide either to hire or not.
- Cost
  - Pay the agency a small fee to interview an applicant,
  - To actually hire an applicant is more costly, fire the current office assistant, pay substantial hiring fee to the agency.
- Committed to have the best possible at all time, i.e., a new one is better than current one, then hire the new one.

- What the price will be?
- Assume that candidates number 1 through n.
- Interview has low cost c_i, hiring has higher cost c_h.
- If m people hired, total cost $O(c_i \cdot n + c_h \cdot m)$. $c_i \cdot n$ is fixed, we analyze $c_h \cdot m$.
- Worst case, hire everyone interviewed, candidates come in strictly increasing order of quality. $O(c_h \cdot n)$.

Probabilistic Analysis

- Must use knowledge of the distribution of the input, averaging total cost over all possible input, average case analysis.
- For the hiring problem, assume that the applicatns come in a random order.
- Meaning
  - Assume there is a total order for the qualification,
  - i.e., candidate can be numbered $1, 2, \ldots, n$ and $rank(i)$ the rank of application i,
  - $< rank(1), rank(2), \ldots, rank(n) >$ a permutation of $< 1, 2, \ldots, n >$.
  - Applicants come in random order is equivalent to saying that this list of ranks is equally likely to be any one of $n!$ permutation of $1, 2, \ldots, n$,
  - or the ranks from a uniform random permutation.

Indicator Random Variable

- A convenient method for converting between probabilities and expectation.
- Given a sample space S and an event A, indicator random variable $I\{A\}$ associated with event A defined as

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

- Expected number of heads obtained when flipping a fair coin
  - sample space $\{H, T\}$, $Pr\{H\} = Pr\{T\} = 1/2$,
  - define an indicator random variable X_H Associated with the coin coming up head (event H),

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

- Expected number of heads obtained in one flip = the expected value of the indicator variable X_H,

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

- Expected value of an indicator random variable assocoated with an event A = the probability that A occurs.

Lemma 5.1 Given a sample space S and an event A in the sample space S, let $X_A = I\{A\}$. Then $E[X_A] = Pr\{A\}$

Proof By the definition of the indicator random variable

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

and the definition of the expected value, we have

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

where $\overline{A}$ denotes $S - A$, the complement of A.

- Indicator random variable useful for analyzing situations in which we perform repeated random trial.
- Let X_i be the indicator random variable associated with the event in which ith flip comes up head,
- $X_i = I\{\text{the } i\text{th flip results in the event } H\}$.
- Let X be the random variable denoting the total number of heads in the n coin flips,

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

- Expected number of heads

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

By the linearity of expectation,

$$
\begin{aligned}
E[X] &= E\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} E\left[X_i\right] \\
&= \sum_{i=1}^{n} \frac{1}{2} \\
&= \frac{n}{2}.
\end{aligned}
$$

Now the Hiring Problem

- Average cost $=$ the expected number of times that we hire a new assistant.
- Assume that the candidates arrive in random order,
- Let X be the random variable whose value equals the number of times we hire a new assistant, then

$$E[X] = \sum_{x=1}^{n} x \cdot Pr\{X = x\}.$$

- Difficult to compute $E[X]$.

Use Indocator Random Variable

- Define n variables related to whether or not each particular candidate is hired.
- Let X_i be the indicator random variable associated with event in which the ith candidate is hired.

$$
\begin{aligned}
X_i &= I\{\text{candidate } i \text{ is hired}\} \\
&= \begin{cases} 1 & \text{if candidate } i \text{ is hired,} \\ 0 & \text{if candidate } i \text{ is not hired.} \end{cases}
\end{aligned}
$$

- and $X = X_1 + X_2 + \ldots + X_n$.

- By Lemma 5.1, $E[X_i] = Pr\{\text{Candidate } i \text{ is hired.}\}$
- What is the probability of candidate i is hired?
- i is hired iff i is better that each of candidates 1 through $i - 1$.
- Since candidates come arrive in random order, first i candidates have appear in random order,
- and any one of the first i candidates is equally likely to be the best-qualify so far,
- candidate i has a probability of $1/i$ of being better quality that candidates 1 through $i - 1$.
- and thus a probability of $1/i$ of being hired.

$$E[X_i] = \frac{1}{i}.$$

$$
\begin{aligned}
E[X] &= E\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} E\left[X_i\right] \\
&= \sum_{i=1}^{n} \frac{1}{i} \\
&= \ln n + O(1)
\end{aligned}
$$

Thus the average cost for hiring office assistant is $O(c_h \cdot \ln n)$