Average Case Analysis

- We need to know instance distribution
  - Still have to model for some distribution
  - Sometimes we make ideal assumption that all instances of any given sizes are equally distributed
    - Randomize algorithms
- Probabilistic Analysis for average-case or expect time

The hiring problem

- You are to hire a new office assistant
- A recruiter send you one out of *n* candidates each day
- You interview the candidate being sent
  - The interview cost is c_i, which is small
- You fire current assistant and hire the candidate if he/she is the best so far
  - The firing/hiring cost is c_h, which is high
- You like to know the expect "cost" (not execution time here) of the hiring

The algorithm for the hiring problem

```
HIRE-ASSISTANT(n)
{
    best = 0;
    for (i=1; i<=n; i++) {
      interview candidate i;                      <------ cost c_i
      if (candidate i is better than candidate best) {
           fire candidate best;    }
           best = i;               }  <------ cost c_h
           hire candidate i;       }
      }
    }
}
```

What is the worst case cost? Best case?
What is the expected cost?

Probabilistic Analysis and Randomized Algorithm

- Probabilistic analysis
  - Use probability in the analysis
  - Need to know to make assumption the distribution of inputs/instances
  - Computing expected running time
    - Averaging over all possible inputs
- Randomized algorithm
  - Randomize the input so the behavior of algorithm depends on the input as well as the random number generator

The randomized algorithm for the hiring problem

```
HIRE-ASSISTANT(n)
{
    best = 0;
    for (i=1; i<=n; i++) {
       randomly to choose a candidate c to interview;
       if (candidate c is better than candidate best) {
            fire candidate best;
            best = c;
            hire candidate c;
       }
    }
}
```

Indicator random variables

- Given a sample space S and an event A. The indicator variable $I\{A\}$ associated with event A is

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

- Example: flip a fair coin. $S = \{H, T\}$.
  - X_H (I{H}) is the indicator variable associate with the event H

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

Expected value of indicator variable

- Theorem:
  - Given a sample space S and an event A, let $X_A = I\{A\}$ Then $E[X_A] = \Pr\{A\}$.
- Example:
  - X_H is the indicator variable associate with the event H (head) $E[X_H] = \Pr(H) = 1/2$
  - What is the expected number of heads in n coin flips
    - Let X be the random variable denoting the total number of head in the n coin flips, X_i =I{the i_{th} flip results in the event H}

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

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

Analysis of the hiring problem

- Let X be the random variable denoting the total number of times a new assistant is hired
- X_i =I{the i_{th} candidate is hired}

$$E[X_i] = \Pr\{i_{th} \text{ candidate is hired}\} = \frac{1}{i}$$

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

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

- Assuming that the candidates are presented in a random order, algorithm HIRE_ASSISTANT has a total hiring cost $O(c_h \ln n)$

Insertion sort

- Assume that
  - all n! permutations of the array elements are equally likely
    - Or we randomize the array in the beginning
  - The n elements of the array are distinct

```
void insertionSort(int A[0..n-1])
{
  int i, j, tmp;

   for (i=1; i<n; i++) {
     tmp=A[i];
     j = i-1;
     while (j>=0 && tmp<A[j]) {
       A[j+1] = A[j];
       j--;
     }
     A[j+1] = tmp;
   }
}
```

Average case analysis of insertion sort

- Choose the while loop condition check as the barometer
- Define the *partial rank* of A[i] be the position of A[i] if the sub array A[0..i] is sorted
  - The partial rank of A[i] is equally likely to take any value between 0 and i
- Note that A[0..i-1] is sorted. Let the partial rank of A[i] be k. The barometer executes i-k+1times. The average number of times of the test is
  $c_i = \frac{1}{i+1}\sum_{k=0}^{i}(i-k+1) = \frac{i}{2}+1$
- The total average is $\sum_{i=1}^{n-1} c_i = \sum_{i=1}^{n-1}(\frac{i}{2}+1) = \frac{(n-1)(n+4)}{4}$

Analysis using indicator variable

- Let X_{ij} be the indictor variable that associated with event A[i]<A[j]
- Let X_i be the number of times of the test at iteration *i (from 1 to n-1)*, we have $X_i = \sum_{j=0}^{i-1} X_{ij} + 1$

$$E[X_{ij}] = \frac{1}{2}$$

$$E[X_i] = E[\sum_{j=0}^{i-1} X_{ij}] + 1 = \frac{i}{2} + 1$$