Due: November 5, 1992

Consider the following algorithm to sort \(A[1..n]\), where it is assumed that the elements of \(A\) are distinct real numbers independently distributed uniformly in the interval \([0..m]\). Assume \(n=c^*m\), for some constant \(c \geq 1\).

**MultiInsertionSort:**

\[
\text{for } i:=0 \text{ to } m-1 \text{ do } \text{MakeEmptyList}(L_i) \\
\text{for } j:=1 \text{ to } n \text{ do } \text{Insert}(A[j], L \setminus L_i) \\
\text{for } i:=0 \text{ to } m-1 \text{ do } \text{InsertionSort}(L_i) \\
\text{for } i:=0 \text{ to } m-2 \text{ do } \text{Concatenate the tail of } L_i \text{ to the head of } L_{i+1}
\]

Since the first, second and fourth steps of MultiInsertionSort can be executed quickly, the bulk of the time will be taken by the third step. We want to find the expected number of swaps done in the third step

\[
\text{for } i:=0 \text{ to } m-1 \text{ do } \text{InsertionSort}(L_i)
\]

A) In HW#2, we computed the expected number of swaps to InsertionSort an array of 3 elements for a particular distribution over the set of 3!=6 inputs. What is the expected number of swaps to InsertionSort an array of \(k\) elements for a uniform distribution over the set of \(k!\) inputs?

B) Find the expected number of swaps done in the third step of MultiInsertionSort. **Hint:** In solving the general problem, you will probably want to compute \(\Pr\{L_i \text{ has } k \text{ elements}\}, 0 \leq k \leq n, 0 \leq i \leq m-1\). In doing this, you may want to think in the following way:

-What is the probability that \(k\) designated elements are in \(L_i\)?
-What is the probability that the other \(n-k\) designated elements are not in \(L_i\)?
-In how many ways can these \(k\) designated elements in \(L_i\) be chosen?
C.S.504
Solution for H.W. #6

A) When element $i$, $1 \leq i \leq k$, moves up in InsertionSort, we showed in class that it can be expected to move up a distance $\frac{i-1}{2}$. So the expected number of moves (swaps) for $k$ elements is

$$\sum_{1 \leq i \leq k} \frac{i-1}{2} = \frac{1}{2} \left( \sum_{1 \leq i \leq k} i - \sum_{1 \leq i \leq k} 1 \right) = \frac{1}{2} \left( \frac{k(k+1)}{2} - k \right) = \frac{k(k-1)}{4}$$

B) Let random variable $X_i$, $0 \leq i \leq m-1$, denote the number of swaps in InsertionSorting $L_i$, and let $X = \sum_{0 \leq i \leq m-1} X_i$. We note that $E[X]$ is the answer we seek. Since the elements are uniformly distributed over the lists, for any fixed element $A[j]$ and any fixed list $L_i$,

$\Pr\{A[j] \to L_i\}=1/m$. Fixing any $k$ elements of $A$, the probability that they all go to list $L_i$ is $(1/m)^k$. The probability that none of the other $n-k$ elements go to $L_i$ is $(m-1/m)^{n-k}$. Since these $k$ of $n$ elements could be chosen in $\binom{n}{k}$ ways, $\Pr\{L_i \text{ has } k \text{ elements}\}=\binom{n}{k} \left(\frac{1}{m}\right)^k \left(\frac{m-1}{m}\right)^{n-k}$.

$$E[X] = E\left[ \sum_{0 \leq i \leq m-1} X_i \right] = \sum_{0 \leq i \leq m-1} E[X_i] = \sum_{0 \leq i \leq m-1} \sum_{0 \leq i \leq k \leq n} \frac{k(k-1)}{4} \Pr\{L_i \text{ has } k \text{ elements}\} = \sum_{0 \leq i \leq m-1} \sum_{0 \leq k \leq n} \frac{k(k-1)}{4} \binom{n}{k} \left(\frac{1}{m}\right)^k \left(\frac{m-1}{m}\right)^{n-k}$$

Applying absorption/extraction to the first term, and using an identity from class to simplify the second term, yields

$$= \frac{m}{4} \sum_{0 \leq k \leq n} nk \left(\frac{n-1}{k-1}\right) \left(\frac{1}{m}\right)^k \left(\frac{m-1}{m}\right)^{n-k} - \frac{n}{m}$$

Replacing $k$ by $k+1$ yields

$$= \frac{mn}{4} \sum_{0 \leq k+1 \leq n} \left(\frac{n-1}{k+1}\right) \left(\frac{1}{m}\right)^{k+1} \left(\frac{m-1}{m}\right)^{n-k-1} - \frac{n}{4}$$

$$= \frac{mn}{4} \left[ \sum_{-1 \leq k \leq n-1} k \left(\frac{n-1}{k}\right) \left(\frac{1}{m}\right)^{k+1} \left(\frac{m-1}{m}\right)^{n-k-1} + \sum_{0 \leq k+1 \leq n} \left(\frac{n-1}{k}\right) \left(\frac{1}{m}\right)^{k+1} \left(\frac{m-1}{m}\right)^{n-k-1} \right] - \frac{n}{4}$$

Applying absorption/extraction to the first term, and using the fact that $\binom{n}{-1}=0,$
Replacing $k$ by $k+1$ in the first term, and applying the binomial theorem to the second term yields
\[
\frac{nm(n-1)}{4} \sum_{0 \leq k \leq n-1} \binom{n-2}{k+1} \left(\frac{1}{m} \right)^{k+1} \left(\frac{m-1}{m} \right)^{n-k-1} + \frac{1}{m} \sum_{0 \leq k = n-1} \binom{n-1}{k} \left(\frac{1}{m} \right)^{k} \left(\frac{m-1}{m} \right)^{(n-1)-k} - \frac{n}{4}
\]

Using the binomial theorem,
\[
\frac{n(n-1)}{4m} \sum_{0 \leq k \leq n-2} \binom{n-2}{k} \left(\frac{1}{m} \right)^{k} \left(\frac{m-1}{m} \right)^{(n-2)-k}
\]

Using the binomial theorem, \[
\frac{n(n-1)}{4m}. \quad \text{Since } n=cm, \quad E[X] = \frac{cm(cm-1)}{4m} = \frac{c^2}{4}m - \frac{1}{4}, \quad \text{and}
\]

MultiInsertionSort (a.k.a. Interpolation Sort and Bucket Sort) takes expected linear time when the keys are uniformly distributed.