COS 341, October 5, 1998
Handout Number 3

The Secretary Problem

In the Secretary Problem with parameters n, k, the probability space U is the set of all permutations of $\{1, 2, \cdots, n\}$. It is easy to see that a permutation $u = (i_1, i_2, \cdots, i_n) \in U$ is in T if and only if the following conditions are satisfied:

C1: $n \in \{i_{k+1}, i_{k+2}, \cdots, i_n\}$;
C2: Let $i_j = n$ where $k < j \leq n$, then the maximum of $i_1, i_2, \cdots, i_{j-1}$ is among the first k of these numbers.

Let T_j denote the set of all u's satisfying C1, C2 with j being the j in C2. Then T is the disjoint union of such T_j. By the Addition Principle, we have

$$|T| = \sum_{k<j\leq n} |T_j|. \tag{1}$$

We assert that for each $k < j \leq n$,

$$|T_j| = \frac{k}{j-1}(n-1)!. \tag{2}$$

To see this, note that T_j is itself the disjoint union of $T_{j,1}, T_{j,2}, \cdots, T_{j,k}$, where $T_{j,s}$ consists of $a \in T_j$ with the maximum of $\{i_1, i_2, \cdots, i_{j-1}\}$ occurring at i_s. Now each element of $T_{j,s}$ can be specified by first choosing an $(n-j)$-permuation of $\{1, 2, 3, \cdots, n-1\}$ (to fix $i_{j+1}, i_{j+2}, \cdots, i_n$), and then a $(j-2)$-permutation of the set $\{1, 2, 3, \cdots, n-1\} - \{i_{j+1}, i_{j+2}, \cdots, i_n\}$ minus its maximum element (to fix $(i_1, i_2, \cdots, i_{j-1})$). Therefore, by the Multiplication Principle, we have

$$\begin{aligned} T_{j,s} &= P(n-1, n-j) \cdot (j-2)! \\ &= \frac{(n-1)!}{(n-1-(n-j))!}(j-2)! \\ &= \frac{(n-1)!}{j-1}. \end{aligned}$$

This proves (2). (Alternatively, one can argue that it is equally likely for a *random* u with $i_j = n$ to have the maximum of $i_1, i_2, \cdots, i_{j-1}$ to occur at s for any $1 \leq s \leq j-1$. Since there are $(n-1)!$ u's with $i_j = n$, the number of such permutations with the minimum occurring in the first k locations is equal to $\frac{k}{j-1} \cdot (n-1)!$.)

It follows from (1) and (2) that

$$\begin{aligned}|T| &= \sum_{k<j\le n} \frac{k}{j-1}\cdot (n-1)! \\ &= n!\frac{k}{n}(\frac{1}{k}+\frac{1}{k+1}+\cdots+\frac{1}{n-1}).\end{aligned}$$

Since $|U| = n!$, we obtain the following result.

Theorem 1

$$\Pr\{T\} = \frac{k}{n}(\frac{1}{k}+\frac{1}{k+1}+\cdots+\frac{1}{n-1}).$$

This completes the analysis of the probability of success under the strategy used with parameters n, k.

For example, if $n = 7, k = 4$,

$$\Pr\{T\} = \frac{4}{7}(\frac{1}{4}+\frac{1}{5}+\frac{1}{6}) = 37/105 = 0.3524.$$

We now address the following question: Given n, what is the best k to use? In other words, let $p_{n,k}$ denote the value $\Pr\{T\}$ for parameters n, k, what is the k with the largest $p_{n,k}$? To simplify the notation, we regard n to be fixed in what follows. Let $a_k = p_{n,k}$.

Recall that

$$a_k = \frac{k}{n}(\frac{1}{k}+\frac{1}{k+1}+\cdots+\frac{1}{n-1}). \tag{3}$$

We show that the sequence $a_1, a_2, \cdots, a_{n-1}$ is *unimodal* in that, it first increases until reaching maximum and then decreases. Precisely, let k_0 be the largest integer k such that $1 - \sum_{k\le j\le n-1}\frac{1}{j} < 0$. (Note that k_0 depends on n.)

Theorem 2

$$a_1 < a_2 < \cdots < a_{k_0} \ge a_{k_0+1} > a_{k_0+2} > \cdots > a_{n-1}.$$

To prove Theorem 2, observe that from Theorem 1 we have

$$\begin{aligned}a_{k-1} - a_k &= \frac{k-1}{n}\sum_{k-1\le j\le n-1}\frac{1}{j} - \frac{k}{n}\sum_{k\le j\le n-1}\frac{1}{j} \\ &= \frac{1}{n}(1 - \sum_{k\le j\le n-1}\frac{1}{j}). \qquad (4)\end{aligned}$$

It follows that $a_{k-1} - a_k < 0$ for $2 \le k \le k_0$. Also, from the definition of k_0, we have $1 - \sum_{k\le j\le n-1}\frac{1}{j} \ge 0$ for $k = k_0 + 1$, and $1 - \sum_{k\le j\le n-1}\frac{1}{j} \ge 0$ for $k \ge k_0 + 2$. Thus, by (4) we conclude that $a_{k_0} \ge a_{k_0+1}$ and $a_{k-1} > a_k$ for $k_0 + 2 \le k \le n - 1$. This completes the proof of Theorem 2.