

A SELECTION PROBLEM WITH COST — ‘SECRETARY PROBLEM’ WHEN UNLIMITED RECALL IS ALLOWED

Sigeiti Morigati
Professor Emeritus, University of Tokyo

(Received February 17, 1992; Revised July 6, 1992)

Abstract We want to select a good one from among n candidates. Taking the cost of interview into consideration, we determine an optimum number i , take a random sample of size i and choose the best one out of the sample. The probability distribution of the absolute rank r of the chosen individual is obtained. After a fundamental theory is developed, the asymptotic theories are considered when n becomes very large, for the cases of high, medium, and low cost.

1. Introduction

Suppose there are n candidates, which can be ranked from 1 (the best) to n (the poorest) and there is no tie. We take a random sample of size i and select the best one — the one with the relative rank 1 — in the sample. Let r be the absolute rank — the rank in the whole population — of the individual thus selected. The expected value $E(r)$ of r is a criterion which we would like to minimize. On the other hand, let the cost of interview be k per individual, then it takes ki to interview i individuals. So, the problem is now to minimize the total cost

$$(1.1) \quad t = E(r) + ki$$

by taking a proper sample size i , and we are also interested in the distribution of r in that case.

This is a variation of the “classical secretary problem” (see Chow et al. (1964), Freeman (1983), Ferguson (1989), and many references cited therein) in the sense that going back to any observed candidate is fully permitted.

2. Distribution of the Rank

Sampling i items at random from a population of size n , we select the best one of the i items in the sample. Let p_{ir} denote the probability that the rank of the selected item in the whole population is r . The cumulative probability is

$$(2.1) \quad F_{ir} = p_{i1} + \cdots + p_{ir},$$

and the rest is

$$(2.2) \quad Q_{ir} = p_{i,r+1} + \cdots + p_{in}.$$

Of these, the easiest one to obtain is Q_{ir} . It is the probability that all the sampled i items are below rank r . Hence, in terms of the combinatorics, it is the ratio of the number

of combinations of i items taken out of $n - r$ items to the number of combinations of i items taken out of n items. That is

$$(2.3) \quad Q_{ir} = \binom{n-r}{i} / \binom{n}{i}$$

It can be expressed as

$$(2.4) \quad Q_{ir} = \frac{(n-r)!(n-i)!}{(n-r-i)!n!}$$

or, using the notation

$$(2.5) \quad x^{(m)} = x(x-1) \cdots (x-m+1),$$

as

$$(2.6) \quad Q_{ir} = \frac{(n-r)^{(i)}}{n^{(i)}}.$$

Incidentally, from (2.4), we note the ‘‘symmetry’’

$$(2.7) \quad Q_{ir} = Q_{ri}.$$

The form (2.6) is convenient when we take ‘‘difference’’ or ‘‘sum’’, moving the index r .

First, it is observed from (2.2) that p_{ir} is obtained by ‘‘difference’’

$$(2.8) \quad p_{ir} = Q_{i,r-1} - Q_{ir},$$

which requires

$$(2.9) \quad \begin{aligned} & (n-r+1)^{(i)} - (n-r)^{(i)} \\ &= (n-r+1)(n-r)^{(i-1)} - (n-r)^{(i-1)}(n-r-i+1) \\ &= i(n-r)^{(i-1)} \end{aligned}$$

and gives

$$(2.10) \quad p_{ir} = \frac{i(n-r)^{(i-1)}}{n^{(i)}}.$$

Next, the expectation of r can be obtained by

$$(2.11) \quad \begin{aligned} E(r) &= 1 \cdot p_{i1} + 2 \cdot p_{i2} + \cdots + n \cdot p_{in} \\ &= Q_{i0} + Q_{i1} + \cdots + Q_{in}, \end{aligned}$$

where $Q_{i0} = 1$. In order to evaluate (2.11), we use the formula

$$(2.12) \quad (n-r+1)^{(i+1)} - (n-r)^{(i+1)} = (i+1)(n-r)^{(i)}$$

which can be obtained replacing i in (2.9) by $i+1$. Thus we get

$$(2.13) \quad \begin{aligned} E(r) &= \sum_{r=0}^n Q_{ir} = \sum_{r=0}^{n-i} \frac{(n-r)^{(i)}}{n^{(i)}} \\ &= \frac{1}{n^{(i)}} \sum_{r=0}^{n-i} \frac{(n-r+1)^{(i+1)} - (n-r)^{(i+1)}}{i+1} \\ &= \frac{1}{n^{(i)}} \frac{(n+1)^{(i+1)} - 0}{i+1} \\ &= \frac{n+1}{i+1}. \end{aligned}$$

Cumulative probability (2.1) is $1 - Q_{ir}$. In order to compute Q_{ir} , it is advisable to start with $Q_{i0} = 1$ and use the recurrence formula

$$(2.14) \quad Q_{i,r+1} = Q_{ir} \frac{n-r-i}{n-r} \quad (r = 0, 1, 2, \dots)$$

which can be derived from (2.6). Also, it is to be noted that

$$(2.15) \quad Q_{ir} = 0 \quad \text{for } r \geq n - i,$$

so that one can truncate computing at $r = n - i$.

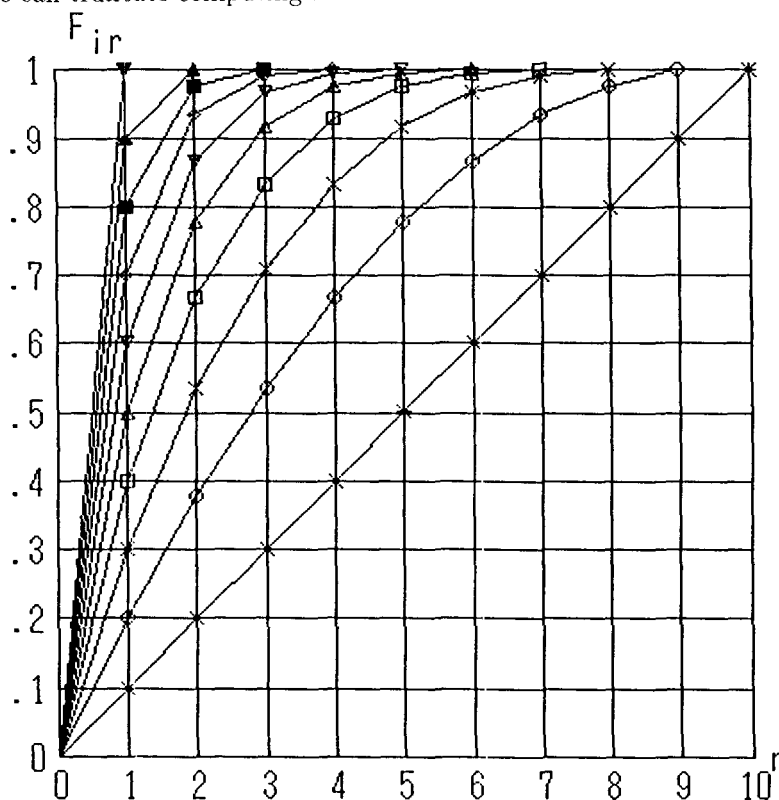


Fig. 1 Cumulative probability F_{ir} (case $n = 10$)

Fig. 1 shows the result of such computation in the case of $n = 10$. Here the lines connecting adjacent points have no meaning beyond linking the points for the same value of i . This simple example illustrates what will happen in general, including asymptotic cases of $n \rightarrow \infty$.

3. Optimal Number of Observations.

As shown in the previous Section, the expected rank of the selected individual is

$$(3.1) \quad E(r) = \frac{n+1}{i+1},$$

so that the objective function (1.1) is now

$$(3.2) \quad t = \frac{n+1}{i+1} + ki.$$

If we make i larger, then the first term decreases, but the second term increases. Let k_i denote such value of k that, for given i , the two terms on the right-hand side of (3.2) are equal. The value of (3.2) is then t_i , where

$$(3.3) \quad k_i = \frac{n+1}{i(i+1)},$$

and

$$(3.4) \quad t_i = 2\frac{n+1}{i+1}.$$

Drawing a broken line connecting $(k_i, t_i) (i = 1, 2, \dots, n)$ on the kt -plane, we get a concave graph (Fig. 2).

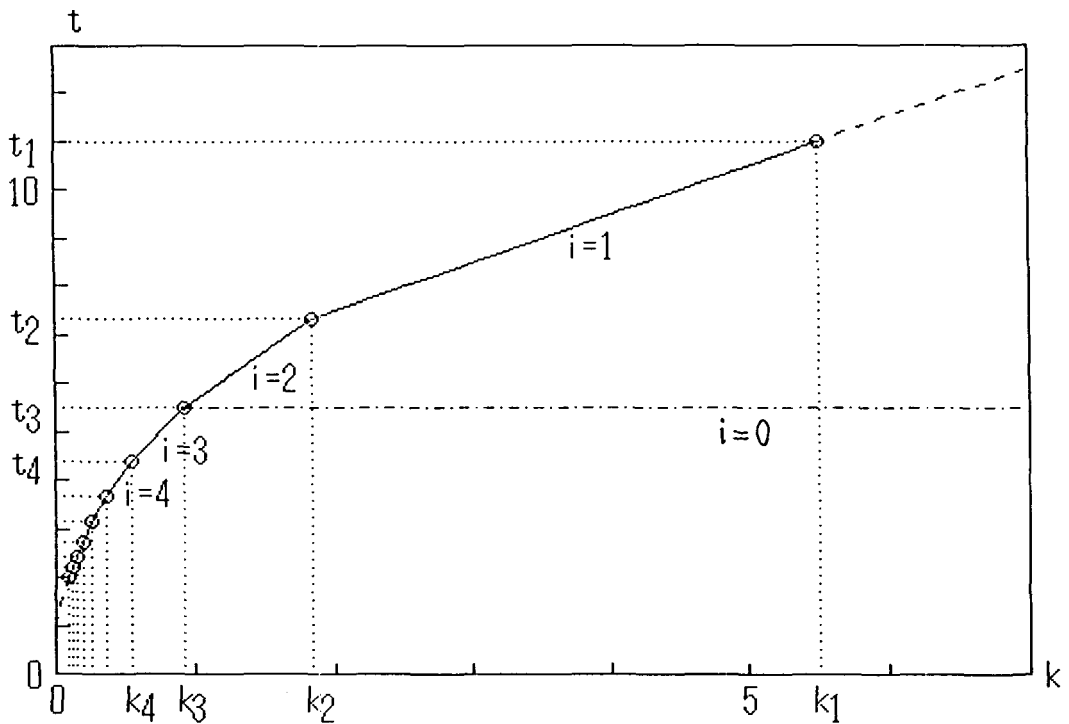


Fig. 2 Graph of t vs. k (case $n = 10$)

It is easy to prove that, for a value of k which is in the interval

$$(3.5) \quad k_{i+1} \leq k \leq k_i,$$

i is the optimal value. Hence we get a graph like Fig. 3 which gives the optimal i for given k .

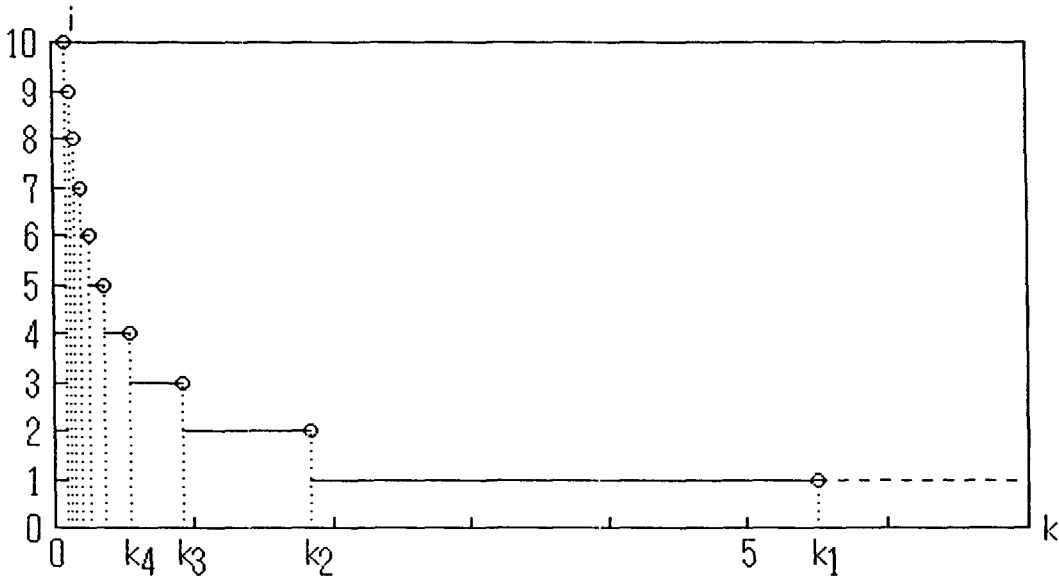


Fig. 3 Optimal i for given k (case $n = 10$)

However, a very small value of k and a very large value of k must be treated separately. For $k < k_n, i = n$ is optimal and the total cost then becomes

$$(3.6) \quad t = 1 + kn,$$

which would be self explanatory.

For very large k , we must consider the case where we take $i = 0$, i. e. selecting one candidate at random without any observation. In that case, the expected rank will be

$$(3.7) \quad t = \frac{n + 1}{2}.$$

As t_3 is equal to (3.7), $i = 0$ is optimal for any $k > k_3$.

On the other hand, if we think that one observation will be done even if choosing one candidate at random, then the minimum of i will be 1, whence $i = 1$ is optimal for $k > k_1$ also, and the objective function becomes

$$(3.8) \quad t = \frac{n + 1}{2} + k$$

in that case.

4. Asymptotic Theory for Low Cost.

Fundamental theory having been completed, we now turn to the asymptotic behavior or the solution when n tends to infinity. In this Section, we consider the case where

$$(4.1) \quad k = O(1/n).$$

In other words, the cost of observing all the candidates

$$(4.2) \quad K = kn$$

is $O(1)$, which means the cost of one observation is very low.

In this case, using (4.2) in (3.2), and letting n tend to infinity keeping the “sampling rate”

$$(4.3) \quad u = i/n$$

constant, we get the expression of the total cost in the limit:

$$(4.4) \quad t = 1/u + Ku.$$

The value of u which minimizes this expression is

$$(4.5) \quad u = 1/\sqrt{K},$$

which gives the total cost

$$(4.6) \quad t = 2\sqrt{K}.$$

These formulae are valid when $K \geq 1$, because the sampling rate must be ≤ 1 . For the value $K < 1$, $u = 1$ is optimal, and the total cost then is $t = 1 + K$.

As for the probability distribution of the achieved rank, we change (3.4) into

$$(4.7) \quad Q_{ir} = (n - i)^{(r)} / n^{(r)},$$

and, substituting un for i , make n tend to infinity. Thus we get

$$(4.8) \quad Q_{ir} = (1 - u)^r.$$

It is easy to derive from this the cumulative probability

$$(4.9) \quad F_{ir} = 1 - (1 - u)^r,$$

and the probability of each r

$$(4.10) \quad p_{ir} = u(1 - u)^{r-1} \quad (r = 1, 2, \dots).$$

The rank r is “geometrically” distributed with the initial term u and the common ratio $1 - u$.

Example.

If $K = kn = 9$, then the optimal sampling rate is $u = 1/3$. Hence it is optimal to take one third of the whole population as a sample, and select the best among them. Then the distribution of the absolute rank will be as follows: you hit the very best ($r = 1$) with probability $1/3 = 0.333$, the second best ($r = 2$) with probability $(1/3)(2/3) = 2/9 = 0.222$, the third best ($r = 3$) with probability $4/27 = 0.148, \dots$. Although the expected value of the attained rank is equal to 3, the distribution is skewed, and the probability of hitting one of the best three is $F_{i3} = 1 - (2/3)^3 = 1 - 8/27 = 0.704$. The cost of observation is $9 \times (1/3) = 3$, and it is not mere coincidence that this equals the expected value of the rank.

5. Asymptotic Theory for Medium Cost.

Next let us consider the case where

$$(5.1) \quad k = O(1).$$

Now we keep

$$(5.2) \quad v = i/\sqrt{n},$$

$$(5.3) \quad \rho = r/\sqrt{n}$$

constant and make n tend to infinity.

Since the total cost t becomes $O(\sqrt{n})$, setting

$$(5.4) \quad T = t/\sqrt{n},$$

we get

$$(5.5) \quad T = 1/v + kv.$$

Its minimum is given by

$$(5.6) \quad v = 1/\sqrt{k}$$

as

$$(5.7) \quad T_{\min} = 2\sqrt{k}.$$

The sampling rate is then $1/\sqrt{kn}$, and the total cost t is $2\sqrt{kn}$.

The probability distribution of the attained rank is, as will be obtained from (4.8) by substituting $r = \rho\sqrt{n}$, $u = v/\sqrt{n}$ and making n tend to infinity,

$$(5.8) \quad Q_{ir} = \exp(-v\rho) \quad (0 < \rho < \infty)$$

In this case, ρ should be considered to be a continuous random variable and its cdf and pdf are given by

$$(5.9) \quad F(\rho) = 1 - \exp(-v\rho),$$

$$(5.10) \quad f(\rho) = v \exp(-v\rho) \quad (0 < \rho < \infty).$$

Namely, the rank approaches a continuous distribution of $O(\sqrt{n})$, its shape being exponential and the parameter v . The expected value of the rank is \sqrt{n} times \sqrt{k} .

Remark.

One might be interested to generalize the arguments by setting

$$(5.11) \quad k = Cn^p$$

with a constant C , and p between -1 and 1 .

In this case, setting

$$(5.12) \quad q = (1 - p)/2, \quad q' = 1 - q,$$

and

$$(5.13) \quad v = i/n^q,$$

$$(5.14) \quad \rho = r/n^{q'},$$

the total cost becomes $O(n^{q'})$. So, in the limit,

$$(5.15) \quad T = t/n^{q'} = 1/v + Cv$$

is to be minimized. The minimum is given by $v = 1/\sqrt{C}$, so that the optimal sampling ratio is $1/\sqrt{kn}$. The cdf and pdf of ρ will turn out to be the same as (5.9) and (5.10) above.

Thus for $-1 < p < 1$ we get qualitatively the same result. But for $p = -1$ we have still another situation as we shall see in the next Section.

6. Asymptotic Theory for High Cost.

When the cost of observing one item is $O(n)$, it will be reasonable to observe only a few of the candidates. Now, let us put

$$(6.1) \quad \kappa = k/n,$$

$$(6.2) \quad \tau = t/n.$$

Then from (3.2), we get

$$(6.3) \quad \tau = 1/(i+1) + \kappa i$$

by making n tend to infinity. This is what we are to minimize in the limit.

From this point on, discussion goes just as in Section 3. Thus, let

$$(6.4) \quad \kappa_i = 1/(i(i+1)),$$

$$(6.5) \quad \tau_i = 2/(i+1).$$

Then, in general, i is the optimal number of observations for κ in the interval

$$(6.6) \quad \kappa_{i+1} < \kappa \leq \kappa_i,$$

and the total cost is given by (6.3).

When κ is greater than $\kappa_3 = 1/12 = 0.08333$, it is necessary to consider either the random selection with $\tau_0 = 1/2$ or the minimum number of observation $i = 1$ with $\tau = 1/2 + \kappa$.

The distribution of attained rank is to be treated with

$$(6.7) \quad \rho = r/n$$

as a continuous variate. Using this in (2.6), and making n tend to infinity, we get

$$(6.8) \quad Q_{ir} = (1 - \rho)^i.$$

Therefore the cdf of ρ is

$$(6.9) \quad F(\rho) = 1 - (1 - \rho)^i,$$

and its pdf is

$$(6.10) \quad f(\rho) = i(1 - \rho)^{i-1} \quad (0 < \rho < \infty).$$

Example.

Consider the case where $\kappa = k/n = 0.04$, that is, the cost of observing one candidate is equivalent to losing the expected rank by 4%. This value of κ is between $\kappa_4 = 1/20 = 0.050$ and $\kappa_5 = 1/30 = 0.033$, so that the optimal number of observations is $i = 4$. The expected rank is then $n/5 = 0.20n$, and the total cost is n times $\tau = 0.20 + 0.04 \times 4 = 0.36$. In order to get a 30 point better result than a random sampling, we pay 16 point as the cost of observation.

The distribution of the attained rank in this case is given by (6.9) and (6.10). For instance, the probability of getting the expected rank $E(\rho) = 1/5 = 0.20(20\%)$ or better is

$$F(0.20) = 1 - (1 - 0.20)^4 = 0.5904$$

so that we get that outcome 6 times out of 10.

Remark.

To determine the optimal i for given κ , we may solve the quadratic equation

$$(6.11) \quad i(i + 1) - 1/\kappa = 0$$

with respect to i and take the integral part of the solution. It gives

$$(6.12) \quad i = \text{trunc}[\sqrt{1/\kappa + 1/4} - 1/2].$$

For instance, for the above Example, we get

$$i = \text{trunc}[\sqrt{1/0.04 + 1/4} - 1/2] = \text{trunc}[4.525] = 4$$

which is the right value.

7. Concluding Remarks.

The above argument, be it an example of elementary theory of probability, will serve as a guideline in dealing with similar problems, for instance the “classical secretary problems” ([1], [2], [3]) in which recall is not permitted. In particular, it will be useful to summarize the results in the form of Table 1.

Table 1. Summary of the results

Case	Low cost	Medium cost	High cost
cost of one observation	$k = K/n$	k	$k = \kappa n$
transformation of variables	$i = un$ $r = r$ $t = t$	$i = v\sqrt{n}$ $r = \rho\sqrt{n}$ $t = T\sqrt{n}$	$i = i$ $r = \rho n$ $t = \tau n$
optimal sample size	$u = 1/\sqrt{K}$	$v = 1/\sqrt{k}$	$i = \text{trunc}[\sqrt{1/\kappa + 1/4} - 1/2]$
expected rank	$E(r) = 1/u$	$E(\rho) = 1/v$	$E(\rho) = 1/(i + 1)$
total cost	$t = 2\sqrt{K}$	$T = 2\sqrt{k}$	$\tau = 1/(i + 1) + \kappa i$
distribution of attained rank	geometric $F_{ir} = 1 - (1 - u)^r$ $p_{ir} = u(1 - u)^{r-1}$	exponential $F(\rho) = 1 - \exp(-v\rho)$ $f(\rho) = v \cdot \exp(-v\rho)$	power $F(\rho) = 1 - (1 - \rho)^i$ $f(\rho) = i(1 - \rho)^{i-1}$

As a general lesson, we note the following:

If the cost of observation k is very low and is $O(1/n)$, then the optimal number of observations will be $i = O(n)$, and the attained rank will be $r = O(1)$. If k is very high and is $O(n)$, then i will be $O(1)$, with $r = O(n)$. In between, there are cases where $k = O(1)$, and i and r will both be $O(\sqrt{n})$. It is to be expected from the fact that $i \cdot E(r) = O(1)$ and the fact that the optimal i is determined by the condition $ki = E(r)$.

References

- [1] Chow, Y. S., Moriguti, S., Robbins, H., and Samuel, S.(1964) "Optimal selection based on relative rank." *Israel Jour. Math.* 2, 81–90.
- [2] Freeman, P. R.(1983) "The secretary problem and its extensions: A review." *International Statistical Review* 51, 189–206.
- [3] Ferguson, T. S.(1989) "Who solved the secretary problem?" *Statistical Science* 4, 282–296.

Sigeiti Moriguti
Syoon 2-16-10
Suginami-ku
Tokyo 167, JAPAN