

ASYMPTOTIC THEORY OF SELECTION BY RELATIVE RANK WITH LOW COST

Sigeiti Moriguti
Professor Emeritus, University of Tokyo

(Received September 14, 1992; Revised January 6, 1993)

Abstract Selection from among n objects by relative rank with no recall – the “secretary problem” – in the asymptotic case when $n \rightarrow \infty$ is considered, assuming that k , the cost ratio, is $O(1/n)$, i.e. $K = k \cdot n$ is a finite constant. Starting from the special case of $K = 0$, the situation changes smoothly as K grows and eventually approaches the “medium cost” case. Thus, the ratio of expected cost of observation to the expected rank goes smoothly from 0 to 1 as K goes from 0 to ∞ .

1. Introduction

The “classical secretary problem” (cf. Chow et al. (1964)) assumes that the cost of observation is 0. This might be considered to be too optimistic in many practical cases, especially when n , the total number of objects, is very large. So Moriguti (1993a) discussed the basic theory of the “secretary problem” with cost. Subsequently, Moriguti (1993b) dealt with asymptotic behavior with “medium cost”. There the cost ratio:

$$(1.1) \quad k = \frac{\text{cost of observing one object}}{\text{loss of getting one lower expected rank}}$$

was kept to be a finite constant while $n \rightarrow \infty$.

It turns out that there is a considerable gap between the “classical” case of $k = 0$ and the “medium cost” case of $k = O(1/n)$. To bridge the gap, we discuss in this paper the “low cost” case, assuming that

$$(1.2) \quad K = k \cdot n$$

is kept constant when $n \rightarrow \infty$.

The “classical” (no cost) case is also included, as a special case, in this paper. In fact, the results of that case are found to be useful for numerical computations in the non-zero “low cost” case.

As in the previous paper on the “medium cost” case (Moriguti (1993b)), not only the expected value of the attained absolute rank r , but also the distribution of r is also given here. The latter result is perhaps new even in the “classical” case.

2. Notations and fundamental formulae.

Let us introduce the following notations:

$$(2.1) \quad u = i/n,$$

$$(2.2) \quad T(u) = t_i,$$

$$(2.3) \quad C(u) = c_i,$$

$$(2.4) \quad \bar{E}(u) = e_i/n,$$

$$(2.5) \quad u_s = i_s/n,$$

$$(2.6) \quad \bar{Q}(u) = Q(i).$$

Then, as $n \rightarrow \infty$ with s finite, we get

$$(2.7) \quad \bar{Q}(u) = u_1 u_2 \cdots u_s / u^s \quad (u_s \leq u \leq u_{s+1}),$$

from (2.15) of Moriguti (1993a). $\bar{Q}(u)$ is continuous at $u = u_{s+1}$, as easily proved. This is the probability that no stopping occurs at $i = un$ or before.

The expected number of observations until stopping is given by $n\bar{E}(0)$, where

$$(2.8) \quad \bar{E}(0) = 2u_1 + u_1 \log_e \frac{u_2}{u_1} - \sum_{s=3}^{\infty} \frac{u_1 u_2 \cdots u_{s-1}}{(s-1)(s-2)u_s^{s-2}},$$

from (3.4) of Moriguti (1993a).

The expected rank of the selected object is given, from (5.4) of Moriguti (1993a) and (2.7) above, by

$$(2.9) \quad \begin{aligned} C(0) &= \sum_{s=1}^{\infty} \int_{u_s}^{u_{s+1}} \frac{u_1 u_2 \cdots u_s}{u^{s+2}} \frac{s(s+1)}{2} du \\ &= \sum_{s=1}^{\infty} u_1 u_2 \cdots u_s \frac{s}{2} \left[\frac{1}{u_s^{s+1}} - \frac{1}{u_{s+1}^{s+1}} \right] \\ &= \frac{1}{2} \left[\frac{1}{u_1} + \sum_{s=2}^{\infty} \frac{u_1 u_2 \cdots u_{s-1}}{u_s^s} \right]. \end{aligned}$$

The recurrence formula for the conditional total loss t_i , (6.7) of Moriguti (1993a), becomes now the differential equation

$$(2.10) \quad -\frac{dT(u)}{du} = \frac{s(s+1)}{2u^2} + K - \frac{s}{u}T(u) \quad (u_s \leq u \leq u_{s+1}).$$

(Note (1.2) above.) Using the integration factor $1/u^s$, it can be transformed into

$$(2.11) \quad -\frac{1}{u^s} \frac{dT(u)}{du} + \frac{s}{u^{s+1}} T(u) = \frac{s(s+1)}{2u^{s+2}} + \frac{K}{u^s},$$

which gives

$$(2.12) \quad -\frac{1}{u^s} T(u) = -\frac{s}{2u^{s+1}} - \frac{K}{s-1} \frac{1}{u^{s-1}} + A_s \quad (u_s \leq u \leq u_{s+1})$$

for $s \geq 2$, and

$$(2.13) \quad -\frac{1}{u} T(u) = -\frac{1}{2u^2} + K \log_e u + A_1 \quad (u_1 \leq u \leq u_2)$$

for $s = 1$, where A_1, A_2, A_3, \dots are constant each in the corresponding interval.

The equation (6.8) of Moriguti (1993a) which gives the optimum criterion now becomes

$$(2.14) \quad s = u_s T(u_s) \quad (s = 1, 2, \dots)$$

in the limit as $n \rightarrow \infty$. Using (2.14) together with (2.13) and (2.12), we get

$$(2.15) \quad \frac{1}{u_1^2} - \frac{2}{u_2^2} = \frac{1}{2} \left(\frac{1}{u_1^2} - \frac{1}{u_2^2} \right) + K \log_e \frac{u_2}{u_1} \quad (s = 1),$$

$$(2.16) \quad \frac{s}{u_s^{s+1}} - \frac{s+1}{u_{s+1}^{s+1}} = \frac{s}{2} \left(\frac{1}{u_s^{s+1}} - \frac{1}{u_{s+1}^{s+1}} \right) + \frac{K}{s-1} \left(\frac{1}{u_s^{s-1}} - \frac{1}{u_{s+1}^{s-1}} \right) \quad (s \geq 2).$$

Now, let us rewrite (2.12) in the form

$$(2.17) \quad T(u) = \frac{s}{2u} + \frac{Ku}{s-1} - A_s u^s \quad (u_s \leq u \leq u_{s+1}, \quad s \geq 2),$$

whence

$$(2.18) \quad T'(u) = -\frac{s}{2u^2} + \frac{K}{s-1} - A_s s u^{s-1} \quad (u_s \leq u \leq u_{s+1}, \quad s \geq 2),$$

$$(2.19) \quad T''(u) = \frac{s}{u^3} - A_s s(s-1) u^{s-2} \quad (u_s \leq u \leq u_{s+1}, \quad s \geq 2).$$

Also, (2.13) gives us

$$(2.20) \quad T(u) = \frac{1}{2u} - K u \log_e u - A_1 u \quad (u_1 \leq u \leq u_2),$$

$$(2.21) \quad T'(u) = -\frac{1}{2u^2} - K(\log_e u + 1) - A_1 \quad (u_1 \leq u \leq u_2),$$

$$(2.22) \quad T''(u) = \frac{1}{u^3} - \frac{K}{u} \quad (u_1 \leq u \leq u_2).$$

These formulae will be used in Appendix 2.

So far, the derived formulae are almost exactly the same as those in Section 2 of Moriguti (1993b). But from here on, we shall see more or less different pictures.

3. Case of no cost.

Now let us consider the case where $k = 0$, namely the so-called “classical secretary problem”.

Without the term containing K , the formula (2.15) becomes the same as (2.16) with $s = 1$, and (2.16) can now be transformed into

$$(3.1) \quad \frac{s}{u_s^{s+1}} = \frac{s+2}{u_{s+1}^{s+1}} \quad (s \geq 1),$$

which is equivalent to

$$(3.2) \quad \frac{u_{s+1}}{u_s} = \left(\frac{s+2}{s} \right)^{\frac{1}{s+1}} \quad (s \geq 1).$$

(2.1) shows that, as i goes from 0 to n , u goes from 0 to 1. Hence it is natural to assume that

$$(3.3) \quad \lim_{s \rightarrow \infty} u_s = 1.$$

From (3.2) and (3.3), we can get

$$(3.4) \quad \frac{1}{u_1} = \prod_{s=1}^{\infty} \left(\frac{s+2}{s} \right)^{\frac{1}{s+1}}.$$

On the other hand, $T(u) = C(u)$ when $K = 0$, and

$$(3.4) \quad C(0) = C(u_1) = T(u_1) = 1/u_1$$

as seen from (5.6) of Moriguti (1993a) and (2.14) above. Hence we get an important formula

$$(3.5) \quad C(0) = \prod_{s=1}^{\infty} \left(\frac{s+2}{s} \right)^{\frac{1}{s+1}}.$$

The infinite product on the right-hand side is convergent and has the value 3.86952 (for the details, see Appendix 1).

Starting with $u_1 = 1/C(0) = 0.25843$, we can get the values of u_2, u_3, \dots using (3.2) successively. The first 100 values are listed in Table 1.

Table 1. First 100 values of u_s for $K = 0$.

s	u(s)	s	u(s)	s	u(s)	s	u(s)	s	u(s)
1	.2584	21	.9112	41	.9529	61	.9680	81	.9758
2	.4476	22	.9149	42	.9540	62	.9685	82	.9760
3	.5640	23	.9184	43	.9551	63	.9690	83	.9763
4	.6408	24	.9216	44	.9561	64	.9695	84	.9766
5	.6949	25	.9246	45	.9570	65	.9699	85	.9769
6	.7350	26	.9273	46	.9579	66	.9704	86	.9771
7	.7658	27	.9299	47	.9588	67	.9708	87	.9774
8	.7903	28	.9322	48	.9596	68	.9712	88	.9777
9	.8101	29	.9344	49	.9604	69	.9716	89	.9779
10	.8265	30	.9365	50	.9612	70	.9720	90	.9781
11	.8403	31	.9385	51	.9619	71	.9724	91	.9784
12	.8521	32	.9403	52	.9626	72	.9728	92	.9786
13	.8623	33	.9420	53	.9633	73	.9732	93	.9788
14	.8711	34	.9437	54	.9640	74	.9735	94	.9791
15	.8789	35	.9452	55	.9646	75	.9739	95	.9793
16	.8858	36	.9467	56	.9652	76	.9742	96	.9795
17	.8920	37	.9481	57	.9658	77	.9745	97	.9797
18	.8975	38	.9494	58	.9664	78	.9748	98	.9799
19	.9025	39	.9506	59	.9669	79	.9752	99	.9801
20	.9070	40	.9518	60	.9675	80	.9755	100	.9803

The cumulative distribution function of the number of observations is $1 - \bar{Q}(u)$, where $\bar{Q}(u)$ is given by (2.7) above.

The curve of $\bar{Q}(u)$ vs. u is shown in Fig. 1, together with points showing $Q(i)$ vs. i/n for some finite values of n . (The latter points were obtained using (2.5) of Moriguti (1993a).)

The dot-dash vertical line corresponds to the expected number of observations (divided by n):

$$(3.6) \quad \bar{E}(0) = 0.50647$$

which is given by (2.8) above.

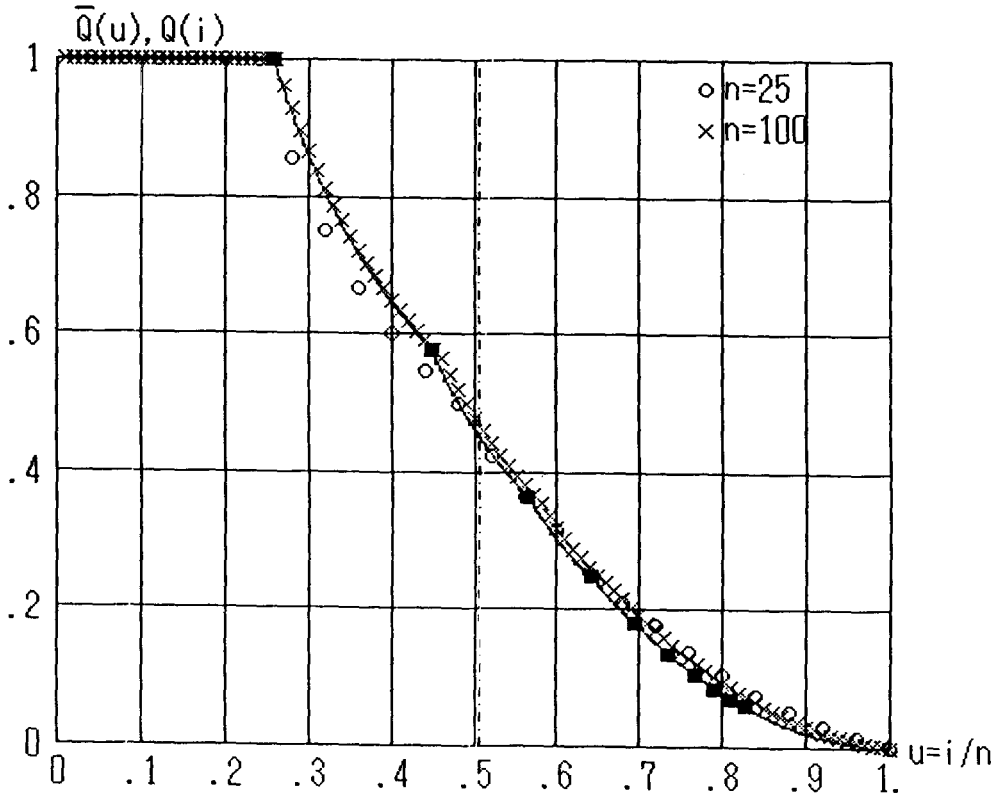


Fig. 1. Curve of $\bar{Q}(u)$ for $K = 0$ with points for $n = 25$ and 100.

4. Distribution of the absolute rank.

The basic formulae for the probability distribution of the absolute rank r of the selected object was developed in Section 4 of Moriguti (1993a). The corresponding asymptotic formulae for low-cost cases are to be developed here. Since they are applicable to both zero-cost and non-zero-cost cases, we change the Section. But since actual numerical values will be given in this Section only for the zero-cost case, the reader may take this Section as the continuation of Section 3, for the time being.

The distribution of r is a discrete one. In particular, the probability that $r = 1$ is derived from (4.13) of Moriguti (1993a) as

$$(4.1) \quad f(1) = \bar{E}(0) - u_1,$$

and for the no-cost case, its numerical value is

$$(4.2) \quad f(1) = 0.50647 - 0.25843 = 0.24804.$$

For a general value of r , its probability is given by (4.11) of Moriguti (1993a), and $f(i, r)$ there is different from $f(i, 1)$ only for such i that $s_i < r$ (compare (4.12) and (4.14) there). For such i , (4.15) there approaches, as $n \rightarrow \infty$,

$$(4.3) \quad \begin{aligned} g(i, r) &\equiv f(i, r) - f(i, r+1) \\ &= \frac{Q(un-1)}{un} \binom{r-1}{s_i-1} \frac{(n-r-1)!}{(un-s_i-1)!(n-un-r+s_i)!} \cdot \frac{(un)!(n-un)!}{n!} \\ &\sim \frac{\bar{Q}(u)}{un} \binom{r-1}{s_i-1} \frac{(un)^{s_i+1}(n-un)^{r-s_i}}{n^{r+1}} \\ &= \frac{\bar{Q}(u)}{un} \binom{r-1}{s_i-1} u^{s_i+1}(1-u)^{r-s_i} \end{aligned}$$

Using (2.7) above, we get

$$(4.4) \quad \begin{aligned} g(i, r) &\sim \frac{u_1 u_2 \cdots u_s}{u^{s+1} n} \binom{r-1}{s-1} u^{s+1} (1-u)^{r-s} \\ &= \frac{u_1 u_2 \cdots u_s}{n} \binom{r-1}{s-1} (1-u)^{r-s} \quad (u_s \leq u \leq u_{s+1}). \end{aligned}$$

Hence

$$(4.5) \quad \sum_{i \in I_s} g(i, r) \sim u_1 u_2 \cdots u_s \binom{r-1}{s-1} \int_{u_s}^{u_{s+1}} (1-u)^{r-s} du,$$

and the integral in (4.5) is given by $[(1-u_s)^{r-s+1} - (1-u_{s+1})^{r-s+1}]/(r-s+1)$.

Using these results, we can derive

$$(4.6) \quad \begin{aligned} f(r) - f(r+1) &= \sum_{s=1}^r \sum_{i \in I_s} g(i, r) \\ &\sim \sum_{s=1}^r u_1 u_2 \cdots u_s \binom{r-1}{s-1} \int_{u_s}^{u_{s+1}} (1-u)^{r-s} du \end{aligned}$$

Therefore, starting with (4.1), we can get $f(r)$ ($r = 2, 3, \dots$) using (4.6) successively. After that, the c. d. f. $F(r)$ can be obtained by cumulative summation of $f(r)$.

Table 2 lists $f(r)$ and $F(r)$ for $r = 1, \dots, 40$ in the zero-cost case.

Fig. 2 shows the graph of $F(r)$ as a broken line, together with the points showing $F(r)$ for some finite values of n , all in the zero-cost case. The vertical dot-dash line shows the expected value (3.5).

5. Case of non-zero cost.

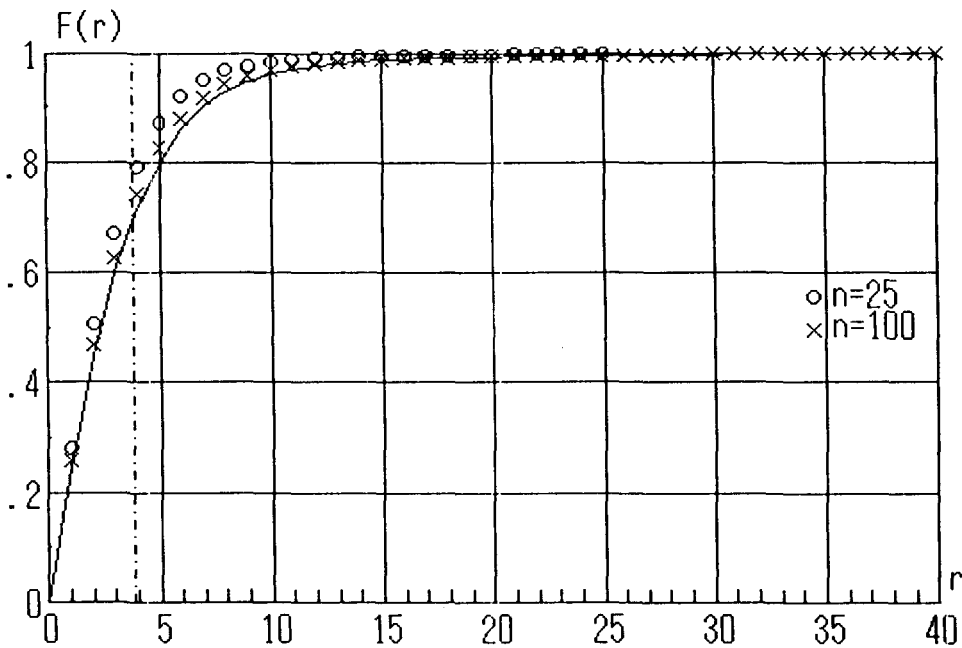
Now, let us proceed to the case where $K > 0$.

For $s \geq 2$, from (2.12) we get

$$(5.1) \quad uT(u) = \frac{s}{2} + \frac{Ku^2}{s-1} - A_s u^{s+1} \quad (u_s \leq u \leq u_{s+1})$$

Table 2. Distribution of the rank r of the selected object ($K = 0$).

r	$f(r)$	$F(r)$	r	$f(r)$	$F(r)$	r	$f(r)$	$F(r)$	r	$f(r)$	$F(r)$
1	.2480	.2480	11	.0091	.9694	21	.0007	.9939	31	.0002	.9974
2	.1991	.4472	12	.0065	.9758	22	.0006	.9945	32	.0002	.9976
3	.1541	.6012	13	.0047	.9805	23	.0005	.9950	33	.0002	.9977
4	.1151	.7164	14	.0034	.9839	24	.0004	.9955	34	.0001	.9979
5	.0835	.7998	15	.0026	.9865	25	.0004	.9959	35	.0001	.9980
6	.0590	.8588	16	.0020	.9885	26	.0003	.9962	36	.0001	.9981
7	.0409	.8998	17	.0016	.9901	27	.0003	.9965	37	.0001	.9982
8	.0281	.9279	18	.0013	.9914	28	.0003	.9968	38	.0001	.9983
9	.0192	.9471	19	.0010	.9924	29	.0002	.9970	39	.0001	.9984
10	.0132	.9602	20	.0008	.9932	30	.0002	.9972	40	.0001	.9985

Fig. 2. Curve of $F(r)$ for $K = 0$ with points for $n = 25$ and 100 .

and, from (2.13),

$$(5.2) \quad uT(u) = \frac{1}{2} - Ku^2 \log_e u - A_1 u^2 \quad (u_1 \leq u \leq u_2)$$

Using (2.14) with (5.1), we have

$$(5.3) \quad s = \frac{s}{2} + \frac{Ku_s^2}{s-1} - A_s u_s^{s+1},$$

whence

$$(5.4) \quad A_s u_s^{s+1} = \frac{Ku_s^2}{s-1} - \frac{s}{2} \quad (s \geq 2).$$

Also, using (2.14) replacing s with $s + 1$, we have

$$(5.5) \quad s + 1 = \frac{s}{2} + \frac{Ku_{s+1}^2}{s-1} - A_s u_{s+1}^{s+1},$$

whence

$$(5.6) \quad A_s u_{s+1}^{s+1} = \frac{Ku_{s+1}^2}{s-1} - \frac{s+2}{2} \quad (s \geq 2).$$

From (5.4) and (5.6), we get an instrumental formula

$$(5.7) \quad \frac{2Ku_s^2 - s(s-1)}{u_s^{s+1}} = \frac{2Ku_{s+1}^2 - (s-1)(s+2)}{u_{s+1}^{s+1}} \quad (s \geq 2).$$

Similarly, we can derive, from (5.2) and (2.14),

$$(5.8) \quad \frac{1}{2Ku_1^2} = \frac{3}{2Ku_2^2} + \log_e \frac{u_2}{u_1}.$$

Formula (5.7) implies:

Proposition[5a] If $2Ku_s^2 = s(s-1)$ then $2Ku_{s+1}^2 = (s-1)(s+2)$, and vice versa;

Proposition[5b] If $2Ku_{s+1}^2 > (s-1)(s+2)$ then $2Ku_s^2 > s(s-1) > (s-2)(s+1)$;

Proposition[5c] If $2Ku_s^2 < s(s-1)$ then $2Ku_{s+1}^2 < (s-1)(s+2) < s(s+1)$.

It can be proved (see Appendix 2) that:

Proposition[5d] The function $T(u)$ has a minimum at one and only point \bar{u} which lies in the interval $(u_1, 1)$;

Proposition[5e] If \bar{u} is in the interval $[u_s, u_{s+1})$ and $s \geq 2$, then

$$(5.9) \quad \frac{2Ku_s^2 - s(s-1)}{2K\bar{u}^2 - s(s-1)} = \frac{1}{s} \left(\frac{u_s}{\bar{u}} \right)^{s+1},$$

$$(5.10) \quad \frac{2Ku_{s+1}^2 - (s+2)(s-1)}{2K\bar{u}^2 - s(s-1)} = \frac{1}{s} \left(\frac{u_{s+1}}{\bar{u}} \right)^{s+1}.$$

Proposition[5f] If \bar{u} is in the interval (u_1, u_2) , then

$$(5.11) \quad \frac{1}{2Ku_1^2} + \log_e u_1 = \frac{1}{2K\bar{u}^2} + 1 + \log_e \bar{u}.$$

$$(5.12) \quad \frac{3}{2Ku_2^2} + \log_e u_2 = \frac{1}{2K\bar{u}^2} + 1 + \log_e \bar{u}.$$

Proposition[5g] If \bar{u} is in the interval $[u_{s^*}, u_{s^*+1})$ then

$$2Ku_s^2 \geq s(s-1) \quad \text{for } s = s^*, s^* - 1, \dots, 1, \text{ and}$$

$$2Ku_s^2 < s(s-1) \quad \text{for } s = s^* + 1, s^* + 2, \dots.$$

Proposition[5h] The s^* defined in [5g] is determined by the formula

$$(5.13) \quad s^* = \text{int}[1/2 + \sqrt{(1/4 + 2K\bar{u}^2)}],$$

where $\text{int}[x]$ denotes the greatest integer not greater than x .

For computational purpose, we may use a temporary variable

$$(5.14) \quad v = \sqrt{(2K)} \cdot u,$$

and the notations $\bar{v} = \sqrt{(2K)} \cdot \bar{u}$, $v_s = \sqrt{(2K)} \cdot u_s$ ($s = 1, 2, \dots$). Then all the expressions like $2Ku^2$ in the above formulae should be changed into corresponding v^2 .

Given \bar{v} , we proceed as follows:

Step 1. Determine s^* by (5.13), and obtain v_{s^*} and v_{s^*+1} by (5.9), (5.10) if $s^* \geq 2$, or by (5.11), (5.12) if $s^* = 1$.

Step 2. If $s^* > 2$, then use (5.7) to get the value of u_s from u_{s+1} for $s = s^* - 1, s^* - 2, \dots, 2$, and finally use (5.8) to get u_1 from u_2 . In each step, successive approximation is to be used, using for example

$$(5.15) \quad v_s = \sqrt{s(s-1) + v_{s+1}^2 [v_{s+1}^2 - (s-1)(s+2)] / v_{s+1}^{s+1}}$$

for $s \geq 2$.

Step 3. Use (5.7) to get the value of u_{s+1} from u_s for $s = s^* + 1, s^* + 2, \dots$. In this case, Newton's method is more efficient than the simple iteration like (5.15). Theoretically, this step could go indefinitely. In practice, however, it is wise to store the value of

$$(5.16) \quad d_s = \log_e \frac{v_{s+1}}{v_s} - \frac{1}{s+1} \log_e \frac{s+2}{s}$$

and stop at such s that $|d_s|$ is sufficiently small. The idea is that the series

$$(5.17) \quad D = d_1 + d_2 + \dots$$

converges much faster than the series

$$(5.18) \quad \log_e \frac{v_\infty}{v_1} = \sum_{s=1}^{\infty} \log_e \frac{v_{s+1}}{v_s}$$

and the numerical value of (5.18) can easily be obtained by adding

$$(5.19) \quad \sum_{s=1}^{\infty} \frac{1}{s+1} \log_e \frac{s+2}{s} = 1.353130$$

(cf. Appendix 1) to the value of (5.17).

Step 4. Once the value of (5.18) is obtained, it is easy to get (using the value of v_1 which is already known)

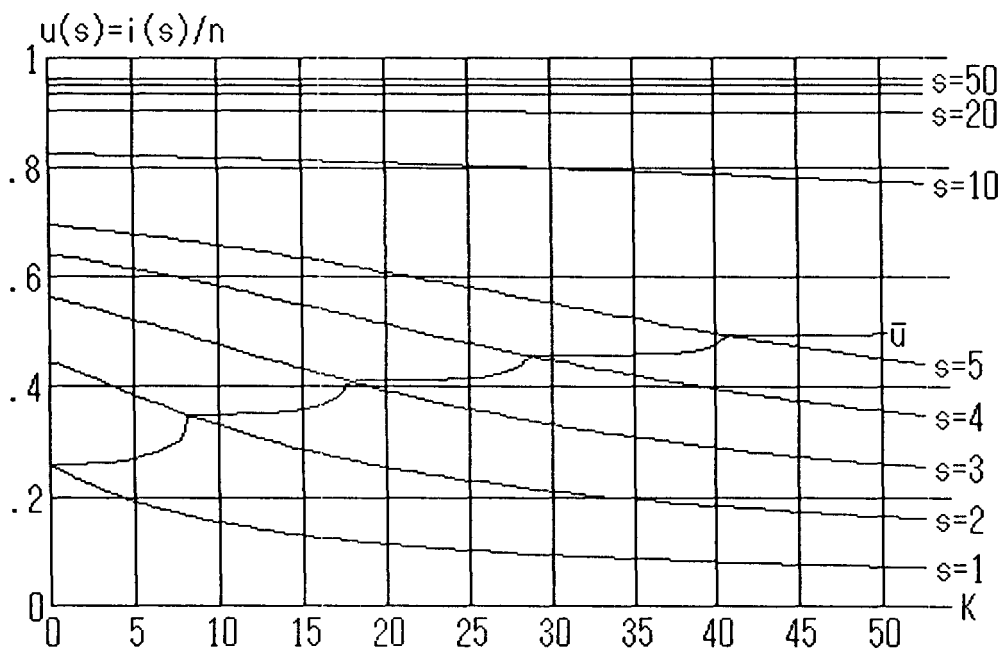
$$(5.20) \quad v_\infty = \sqrt{(2K)}.$$

Step 5. Compute

$$(5.21) \quad K = v_\infty^2 / 2,$$

values of $\bar{u} = \bar{v} / \sqrt{(2K)}$ and

$$(5.22) \quad u_s = v_s / \sqrt{(2K)} \quad (s = 1, 2, \dots).$$

Fig. 3. Curves of u_s (for some values of s) and \bar{u} vs. K .Table 3. K for 50 values of \bar{v} .

\bar{v}	K	\bar{v}	K	\bar{v}	K	\bar{v}	K	\bar{v}	K
.1	.0748	1.1	7.1385	2.1	16.2780	3.1	26.7814	4.1	38.5485
.2	.2989	1.2	7.7130	2.2	16.9860	3.2	27.5953	4.2	39.4608
.3	.6708	1.3	8.0522	2.3	17.4398	3.3	28.1316	4.3	40.0686
.4	1.1874	1.4	8.1717	2.4	17.6488	3.4	28.3977	4.4	40.3813
.5	1.8424	1.5	9.1971	2.5	18.4362	3.5	29.0770	4.5	41.0009
.6	2.6238	1.6	10.4433	2.6	19.9703	3.6	30.8553	4.6	43.0045
.7	3.5106	1.7	11.7231	2.7	21.5098	3.7	32.6205	4.7	44.9801
.8	4.4677	1.8	13.0010	2.8	23.0158	3.8	34.3299	4.8	46.8815
.9	5.4412	1.9	14.2274	2.9	24.4368	3.9	35.9285	4.9	48.6502
1.0	6.3587	2.0	15.3411	3.0	25.7121	4.0	37.3556	5.0	50.2231

This completes the procedure to get K , \bar{u} , and u_1, u_2, \dots for given \bar{v} . Fig. 3 shows the values of $u_s (s = 1, 2, \dots, 5; 10, 20, \dots, 50)$ and \bar{u} plotted against K .

If one wants to know the value of \bar{v} corresponding to a desired value of K , then e.g. a bi-section method can easily be programmed, starting from two values of \bar{v} giving values of K one of which is too small and the other too large. Table 3 will be useful in getting those starting values. For example, we get $\bar{v} = 0.85445, 1.56486, 2.60193, 4.28564$ for $K = 5, 10, 20, 40$ respectively.

For a given \bar{v} (or K), using the computed values $u_s (s = 1, 2, \dots)$ and the formulae (2.8), (2.9), it is easy to get the values of expected interview cost $K \cdot \bar{E}(0)$, expected absolute rank $C(0)$, and the expected total "loss"

$$(5.23) \quad T(0) = C(0) + K \cdot \bar{E}(0).$$

It can also be computed by the formula

$$(5.24) \quad T(0) = 1/u_1 + Ku_1,$$

and the equality of (5.23) and (5.24) can be used as a check.

Fig. 4 plots computed values of $C(0)$, $K \cdot \bar{E}(0)$, and $T(0)$ against K . It is interesting to note the tendency that $C(0)$ and $K \cdot \bar{E}(0)$ approach each other as K becomes larger and larger. This indicates smooth transition to the “medium cost” case treated in Moriguti (1993b).

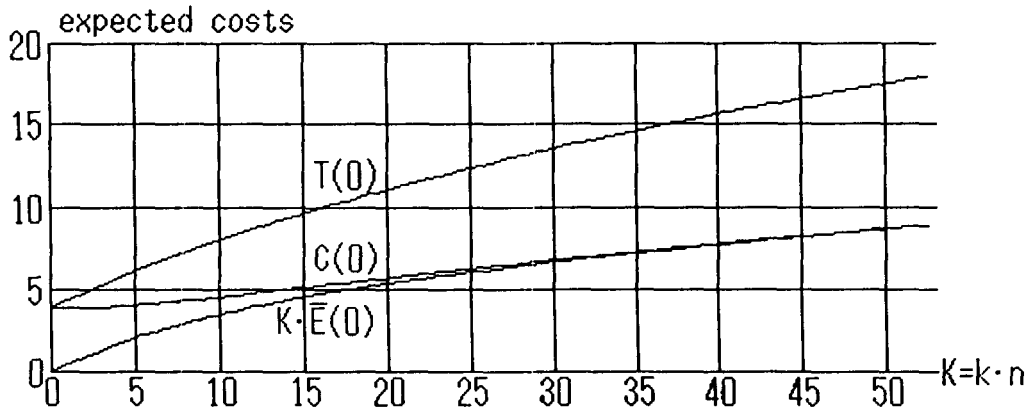


Fig. 4. Curves of $C(0)$, $K \cdot \bar{E}(0)$, and $T(0)$ vs. K .

6. Conditional expected costs.

In this Section, let us observe the behavior of the functions $C(u)$, $K \cdot \bar{E}(u)$, and $T(u)$. These are the expected rank, expected interview cost, and the expected total “loss”, respectively, under the condition that stopping does not occur at $i = un$ or before.

The recurrence formula (3.6) of Moriguti (1993a) becomes now the differential equation

$$(6.1) \quad -\frac{d\bar{E}(u)}{du} = -\frac{s}{u}\bar{E}(u) + 1 \quad (u_s \leq u \leq u_{s+1}).$$

Its general solution is

$$(6.2) \quad \begin{aligned} \bar{E}(u) &= u/(s-1) + E_s u^s \quad (u_s \leq u \leq u_{s+1}, s \geq 2), \\ &= -u \cdot \log_e u + E_1 u \quad (u_1 \leq u \leq u_2), \\ &= -u + E_0 \quad (0 \leq u \leq u_1), \end{aligned}$$

where each constant E_s is to be determined successively, starting from $\bar{E}(0)$ discussed in Section 5.

Fig. 5 shows the result of computation along this line for $K = 0, 5, 10, 20$, and 40 . As K increases, the curve comes to resemble the case of “medium cost”, especially in the right part of the graph (i. e. as $u \rightarrow 1$).

The expected rank of the selected object under the condition that stopping does not occur at $i = un$ or before is represented by $C(u)$. The recurrence formula (5.6) of Moriguti (1993a) becomes now the differential equation

$$(6.3) \quad -\frac{dC(u)}{du} = \frac{s(s+1)}{2u^2} - \frac{s}{u}C(u) \quad (u_s \leq u \leq u_{s+1}).$$

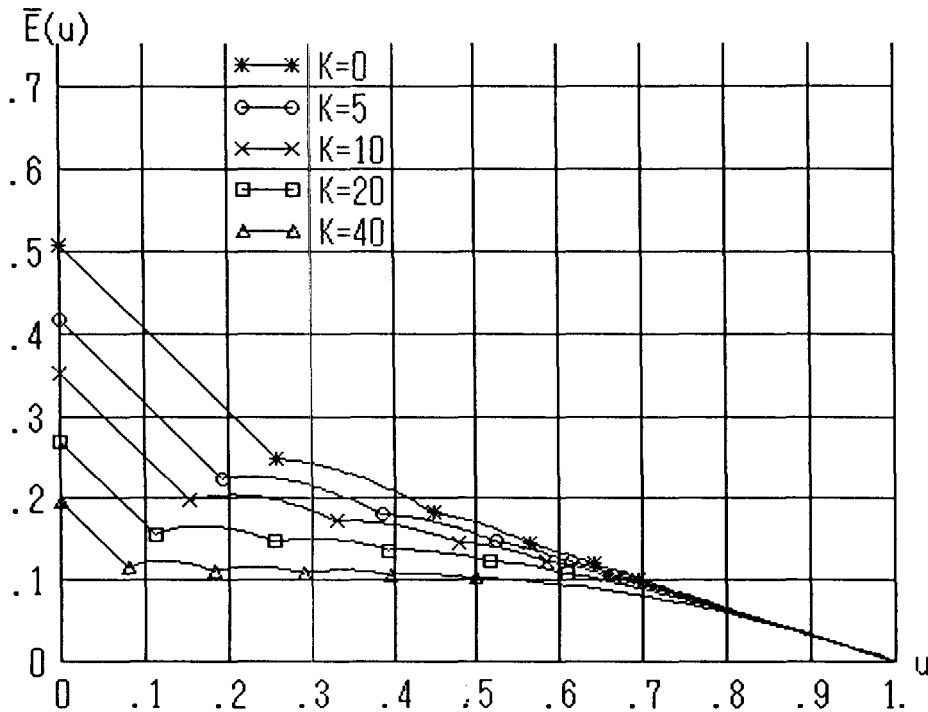


Fig. 5. Curves of $\bar{E}(u)$ for several values of K .

Its general solution is

$$(6.4) \quad \begin{aligned} C(u) &= \frac{s}{2u} - C_s u^s \quad (u_s \leq u \leq u_{s+1}, s \geq 1), \\ &= C_0 \quad (0 \leq u \leq u_1), \end{aligned}$$

where each constant C_s is to be determined successively, starting from $C(0)$ given by (2.9) above.

Fig. 6 shows the result of computation along this line for $K = 0, 5, 10, 20$, and 40 . Here again, we see the resemblance to the “medium cost” case in the right part.

The conditional total “loss”

$$(6.5) \quad T(u) = C(u) + K \cdot \bar{E}(u)$$

can be computed from the values of $C(u)$ and $\bar{E}(u)$. On the other hand, we can also compute $T(u)$ starting from (5.24) and utilizing

$$(6.6) \quad T(u) = T(0) - K \cdot u \quad (0 \leq u \leq u_1),$$

$$(6.7) \quad A_1 = \frac{1}{2u_1^2} + K \cdot \log_e u_1,$$

$$(6.8) \quad T(u) = \frac{1}{2u} - K \cdot u \log_e u - A_1 u \quad (u_1 \leq u \leq u_2),$$

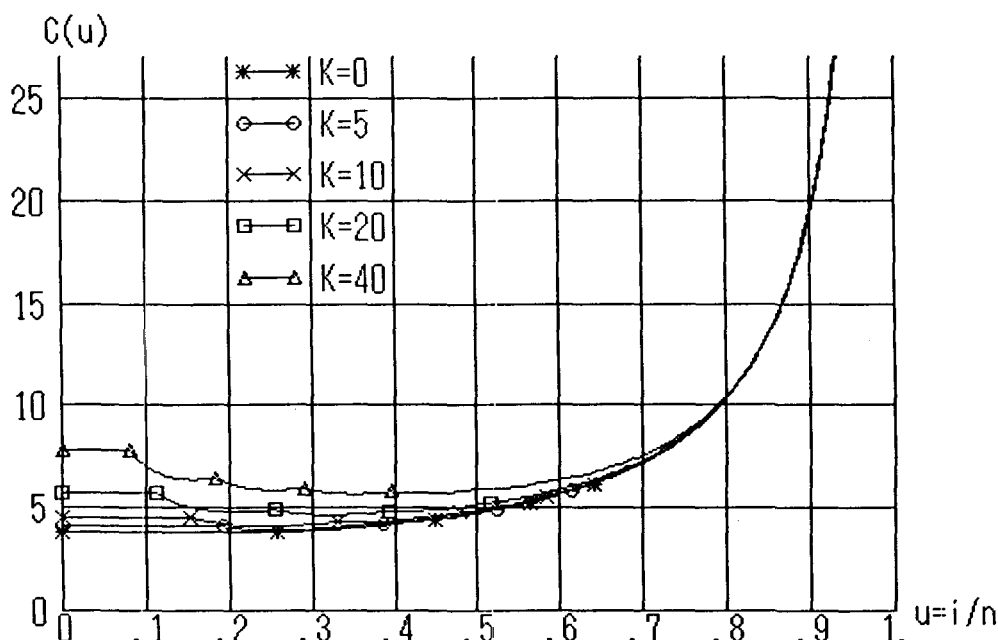


Fig. 6. Curves of $C(u)$ for several values of K .

and

$$(6.9) \quad A_s = \left(\frac{K u_s^2}{s-1} - \frac{s}{2} \right) \frac{1}{u_s^{s+1}},$$

$$(6.10) \quad T(u) = \frac{s}{2u} + \frac{Ku}{s-1} - A_s u^s \quad (u_s \leq u \leq u_{s+1})$$

successively for $s = 2, 3, \dots$. (For these formulae, see (5.1), (5.2), and (5.4) in Section 5.)

Fig. 7 shows the result of computation along the latter line for $K = 0, 5, 10, 20$, and 40 . The vertical dotted line shows the point \bar{u} of the unique minimum of $T(u)$ for each $K > 0$. (Cf. Proposition [5d].)

7. Cumulative distribution of u and r .

Cumulative distribution function of u , the number of observations divided by n , is $1 - \bar{Q}(u)$, where $\bar{Q}(u)$ is given by (2.7). The curve of $\bar{Q}(u)$ vs. u for the case $K = 0$ has already been shown (see Fig. 1). Fig. 8 plots the curves of $\bar{Q}(u)$ vs. u for $K = 0, 5, 10, 20, 40$.

Comparison of Fig. 8 here with Fig. 2 of Moriguti (1993b) suggests the smooth transition to the "medium cost" case. This point is made clearer in Fig. 9, in which the abscissa is changed to $u/\sqrt{(2K)}$ that corresponds to $u/\sqrt{(2k)}$ in Moriguti (1993b).

Cumulative distribution function $F(r)$ of r , the absolute rank of the selected object can be computed with (4.1), (4.6), and cumulative summation of $f(r)$ thus obtained. Fig. 10 plots the $F(r)$ vs. r for $K = 0, 5, 10, 20, 40$. (The adjacent points are connected by a line segment for the sake of convenience.)

Comparison of Fig. 10 here with Fig. 3 of Moriguti (1993b) suggests the smooth transition to the "medium cost" case. It is seen more clearly in Fig. 11, in which the abscissa is changed to $r/\sqrt{(2K)}$ that corresponds to $\rho/\sqrt{(2k)}$ in Moriguti (1993b).

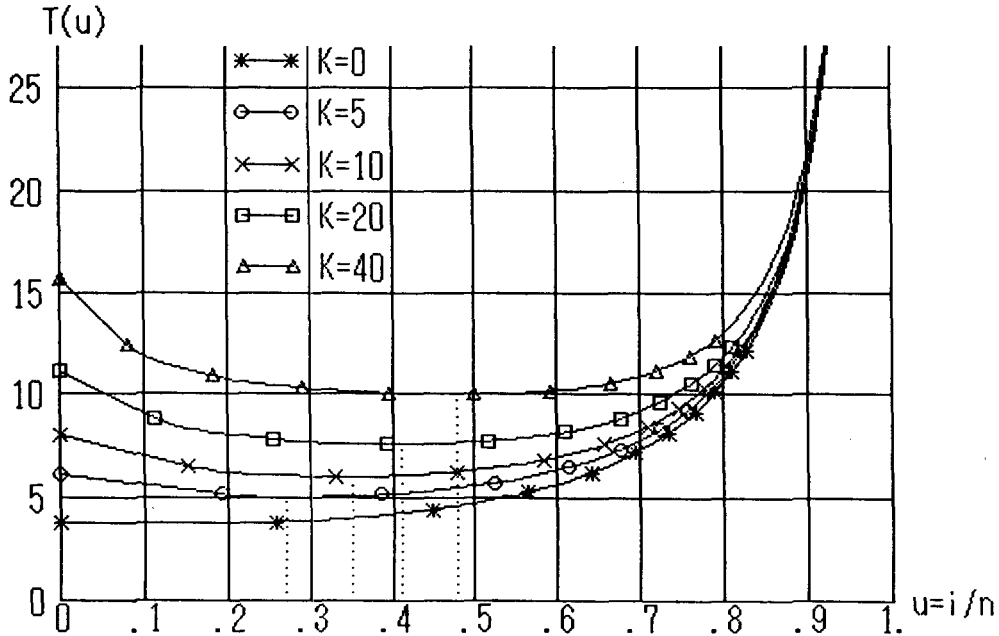


Fig. 7. Curves of $T(u)$ for several values of K .

8. Final remarks.

The author believes that this paper is in line with the comment by Herbert Robbins on Ferguson (1989).

The gap between the “zero-cost” case and the “medium cost” case has been satisfactorily bridged by this paper. Arguments on “Phase 3” and “Phase 4” in Moriguti (1993b) are still applicable, but the author does not think it necessary to repeat the same arguments here.

The results of Moriguti (1992) suggests that “high cost” case will require a little different approach from the other side of the “medium cost” case. We shall leave it to another paper which will follow soon.

Appendix 1. Convergence and The Value of (3.5).

In order to compute the value of (3.5), it is convenient to compute its logarithm first. That is

$$(A1.1) \quad \log_e C(0) = \sum_{s=1}^{\infty} \frac{1}{s+1} \log_e \frac{s+2}{s} = \sum_{s=2}^{\infty} \frac{1}{s} \log_e \frac{s+1}{s-1}.$$

Incidentally, it is easy to show that (A1.1) is < 2 , using the inequality $\log_e[(s+1)/(s-1)] = \log_e[1 + 2/(s-1)] < 2/(s-1)$ for $s \geq 2$.

Each term in the right-hand side of (A1.1) can be expanded into power series as

$$(A1.2) \quad \frac{1}{s} \log_e \frac{s+1}{s-1} = \frac{2}{s} \left(\frac{1}{s} + \frac{1}{3s^3} + \frac{1}{5s^5} + \cdots \right) = 2 \sum_{n=1}^{\infty} \frac{1}{(2n-1)s^{2n}}.$$

Hence the series (A1.1) converges like $\sum 1/s^2$, which is known to be notoriously slow. So we have to sum the first $N-1$ terms, say, directly and compute the remainder in a clever way.

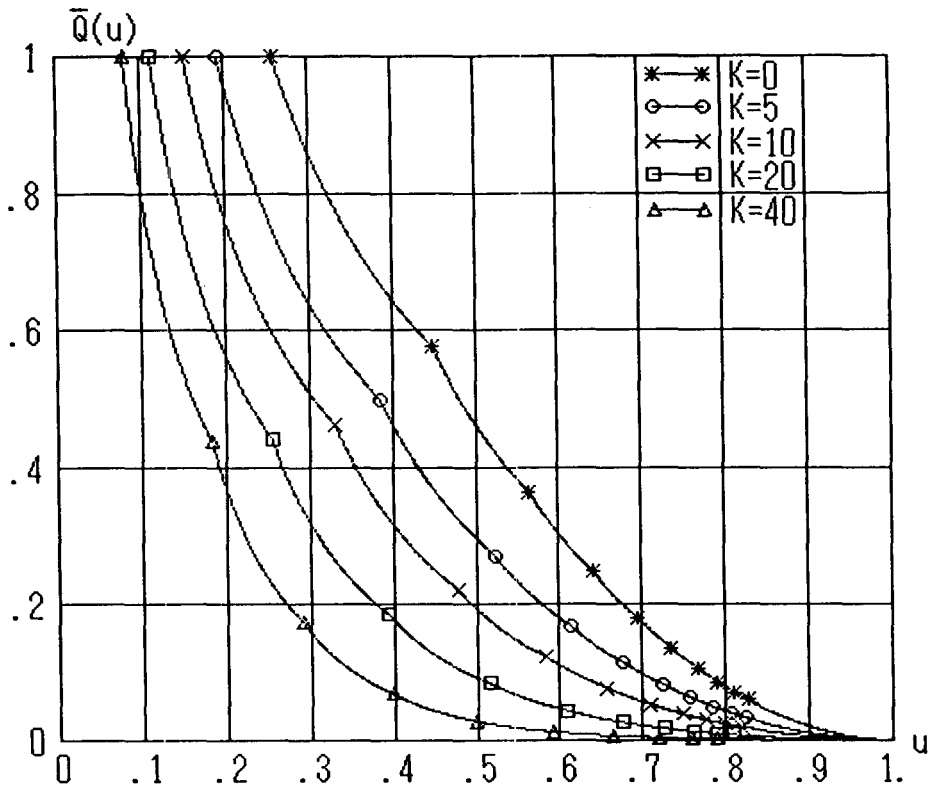


Fig. 8. Curves of $\bar{Q}(u)$ for several values of K .

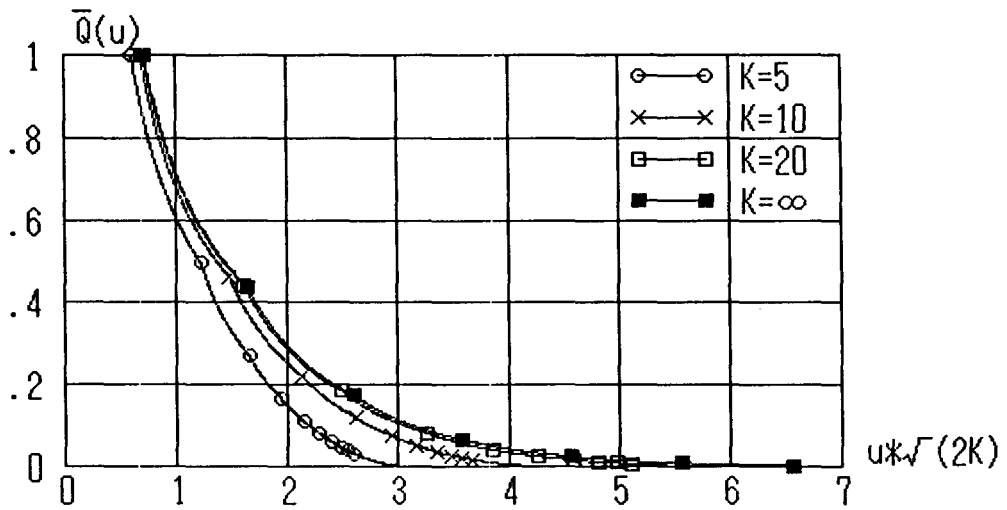
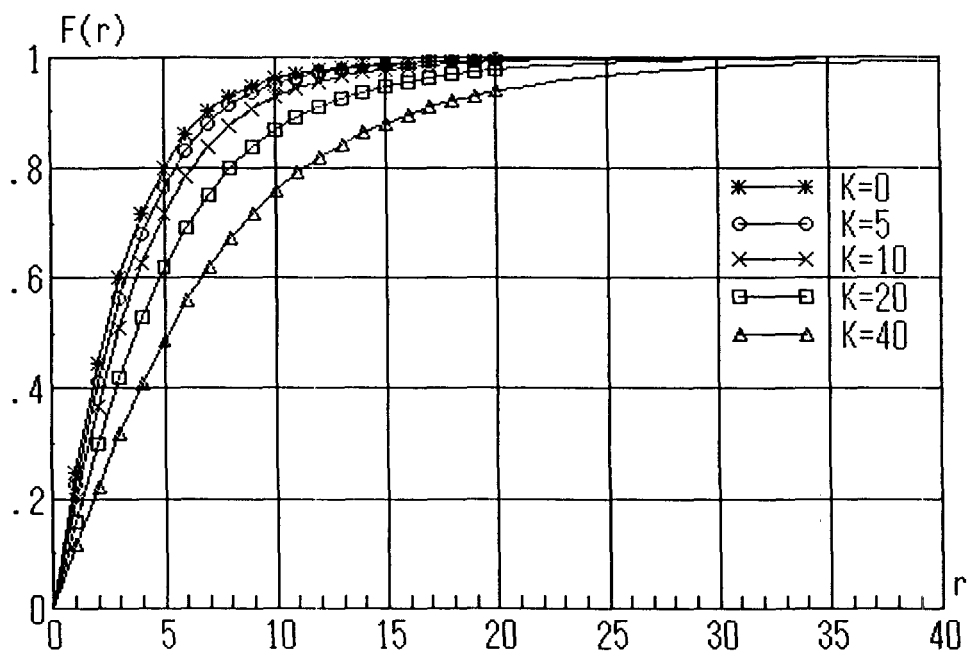
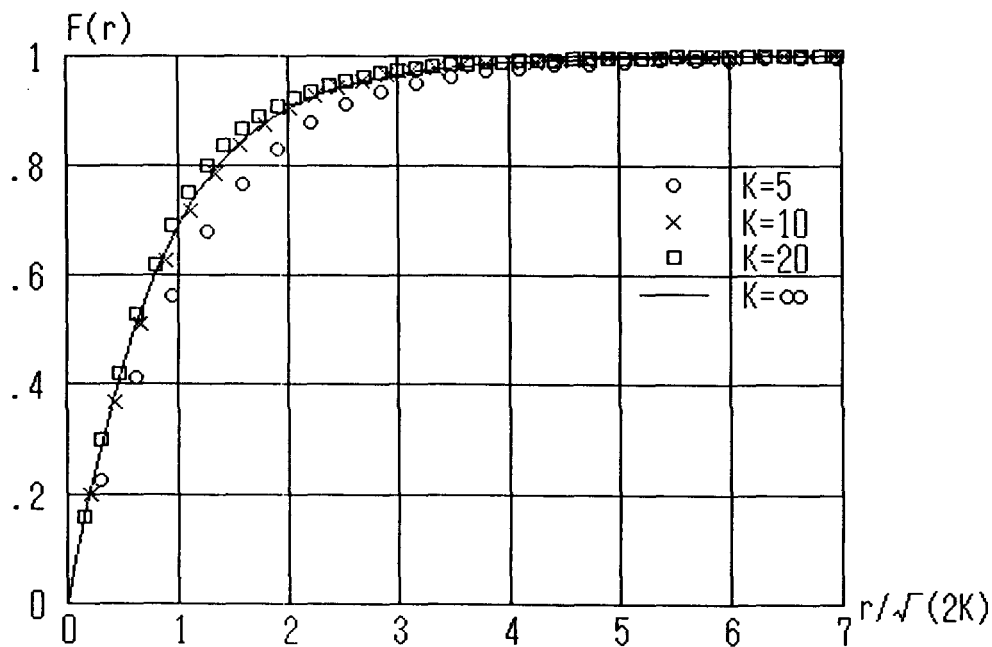


Fig. 9. Curves of $\bar{Q}(u)$ vs. $u * \sqrt{(2K)}$ for several values of K .

Fig. 10. Curves of $F(r)$ for several values of K .Fig. 11. Curves of $F(r)$ vs. $r/\sqrt{(2K)}$ for several values of K .

Thus we have

$$(A1.3) \quad \log_e C(0) = \sum_{s=2}^N \frac{1}{s} \log_e \frac{s+1}{s-1} + R_N,$$

where

$$(A1.4) \quad R_N = \sum_{s=N+1}^{\infty} \frac{1}{s} \log_e \frac{s+1}{s-1}.$$

Using the well-known formula

$$(A1.5) \quad \int_0^{\infty} e^{-sx} x^{m-1} dx = \frac{1}{s^m} \Gamma(m) = \frac{(m-1)!}{s^m},$$

R_N can be expressed as

$$(A1.6) \quad \begin{aligned} R_N &= \sum_{s=N+1}^{\infty} \int_0^{\infty} e^{-sx} \sum_{n=1}^{\infty} \frac{2}{2n-1} \cdot \frac{x^{2n-1}}{(2n-1)!} dx \\ &= \int_0^{\infty} \sum_{n=1}^{\infty} \frac{2}{2n-1} \cdot \frac{x^{2n-1}}{(2n-1)!} \sum_{s=N+1}^{\infty} e^{-sx} dx \\ &= \int_0^{\infty} \sum_{n=1}^{\infty} \frac{2x^{2n-1}}{(2n-1)!(2n-1)} \frac{e^{-(N+1)x}}{1-e^{-x}} dx \\ &= \int_0^{\infty} \sum_{n=1}^{\infty} \frac{2x^{2n-1}}{(2n-1)!(2n-1)} \frac{e^{-Nx}}{e^x-1} dx. \end{aligned}$$

But

$$(A1.7) \quad \begin{aligned} \frac{x}{e^x-1} &= 1 - \frac{x}{2} + \frac{B_1}{2!}x^2 - \frac{B_2}{4!}x^4 + \frac{B_3}{6!}x^6 - \dots \\ &= 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} + \frac{x^6}{30240} - \dots, \end{aligned}$$

where B_1, B_2, \dots are Bernoulli numbers.

Contribution of the second term on the right-hand side of (A1.7) to R_N is

$$(A1.8) \quad \begin{aligned} &\int_0^{\infty} e^{-Nx} \sum_{n=1}^{\infty} \frac{2x^{2n-2}}{(2n-1)!(2n-1)} \cdot \frac{-x}{2} dx \\ &= -\frac{1}{2} \sum_{n=1}^{\infty} \int_0^{\infty} e^{-Nx} \frac{2x^{2n-1}}{(2n-1)!(2n-1)} dx \\ &= -\frac{1}{2} \frac{1}{N} \log_e \frac{N+1}{N-1}. \end{aligned}$$

Contribution of the rest is

$$(A1.9) \quad \begin{aligned} &\int_0^{\infty} e^{-Nx} \sum_{n=1}^{\infty} \frac{2x^{2n-2}}{(2n-1)!(2n-1)} \left[1 + \frac{x^2}{12} - \frac{x^4}{720} + \frac{x^6}{30240} - \dots \right] dx \\ &= 2 \int_0^{\infty} e^{-Nx} \left[1 + \frac{x^2}{18} + \frac{x^4}{600} + \dots \right] \left[1 + \frac{x^2}{12} - \frac{x^4}{720} + \dots \right] dx \\ &= 2 \int_0^{\infty} e^{-Nx} \left[1 + \frac{5}{36}x^2 + \frac{53}{10800}x^4 + \dots \right] dx \\ &= \frac{2}{N} + \frac{5}{9} \frac{1}{N^3} + \frac{53}{225} \frac{1}{N^5} + \dots \end{aligned}$$

Adding (A1.8) and (A1.9), we get

$$(A1.10) \quad R_N = -\frac{1}{2} \cdot \frac{1}{N} \log_e \frac{N+1}{N-1} + \frac{2}{N} + \frac{5}{9} \frac{1}{N^3} + \frac{53}{225} \frac{1}{N^5} + \dots$$

Using the expansion in (A1.3), we obtain a workable formula. For $N = 10, 20, 50$, and 100 , it yields the same result

$$(A1.11) \quad \log_e C(0) = 1.3531303,$$

so that we get finally

$$(A1.12) \quad C(0) = \exp(1.3531303) = 3.8695192.$$

Appendix 2. Proof of Propositions [5d] Through [5h]

Among the propositions stated in Section 5, Proposition [5d] is indeed the main theorem and all the others are lemmas and corollaries.

First, let us establish the continuity of $T'(u)$. It is obviously continuous in every interval $[u_s, u_{s+1}) (s = 1, 2, \dots)$. From (2.10), we get

$$(A2.1) \quad T'(u) = -\frac{s(s+1)}{2u^2} - K + \frac{s}{u^2} u T(u) \quad (u_s \leq u < u_{s+1}).$$

Considering (2.14), we obtain

$$(A2.2) \quad T'(u_s + 0) = \frac{-s(s+1)}{2u_s^2} - K + \frac{s^2}{u_s^2} = \frac{s(s-1)}{2u_s^2} - K,$$

and

$$(A2.3) \quad T'(u_{s+1} - 0) = \frac{-s(s+1)}{2u_{s+1}^2} - K + \frac{s(s+1)}{u_{s+1}^2} = \frac{s(s+1)}{2u_{s+1}^2} - K.$$

Thus we have

$$(A2.4) \quad T'(u_s + 0) = T'(u_s - 0) \quad (s = 1, 2, \dots),$$

which means that $T'(u)$ is continuous also at the boundary points. So it is continuous everywhere in the interval $(0, 1)$.

Since $T'(u_1) = -K < 0$ and $T'(u_s) > 0$ for sufficiently large value of s (e.g. for $s = 1 + \sqrt{(2K)}$), there must be at least one point \bar{u} in the interval $(u_1, 1)$ where $T'(\bar{u}) = 0$. This is a part of the "Theorem" [5d].

If \bar{u} lies in the interval $[u_s, u_{s+1})$ where $s \geq 2$, then (2.10) gives us

$$(A2.5) \quad T(\bar{u}) = \frac{\bar{u}}{s} \left[\frac{s(s+1)}{2\bar{u}^2} + K \right],$$

and substituting it into (2.12), we get

$$(A2.6) \quad -\frac{1}{\bar{u}^s} \frac{\bar{u}}{s} \left[\frac{s(s+1)}{2\bar{u}^2} + K \right] = -\frac{s}{2\bar{u}^{s+1}} - \frac{K}{s-1} \frac{1}{\bar{u}^{s-1}} + A_s,$$

that is

$$(A2.7) \quad A_s = \frac{1}{2s(s-1)\bar{u}^{s+1}}[2K\bar{u}^2 - s(s-1)].$$

On the other hand, (2.12) and (2.14) gives us

$$(A2.8) \quad A_s = -\frac{s}{u_s^{s+1}} + \frac{s}{2u_s^{s+1}} + \frac{K}{s-1} \frac{1}{u_s^{s-1}} = \frac{1}{2(s-1)u_s^{s+1}}[2Ku_s^2 - s(s-1)],$$

and

$$(A2.9) \quad A_s = -\frac{s+1}{u_{s+1}^{s+1}} + \frac{s}{2u_{s+1}^{s+1}} + \frac{K}{s-1} \cdot \frac{1}{u_{s+1}^{s-1}} = \frac{1}{2(s-1)u_{s+1}^{s+1}}[2Ku_{s+1}^2 - (s+2)(s-1)].$$

The ratio of (A2.8) to (A2.7) gives us (5.9), and the ratio of (A2.9) to (A2.7) gives us (5.10). Proposition [5e] is thus proved. Proposition [5f] is similarly proved, using (2.13) in place of (2.12).

Now we have come to the point to prove the remaining part of "Theorem" [5d], that $T(u)$ has a unique minimum at \bar{u} .

If \bar{u} lies in the interval $[u_s, u_{s+1})$ and $s \geq 2$, then substituting (A2.7) into (2.19) we get

$$(A2.10) \quad T''(u) = \frac{s}{u^3} - \frac{u^{s-2}}{2\bar{u}^{s+1}}[2K\bar{u}^2 - s(s-1)].$$

On the other hand, since $s = \text{int}[\bar{u}T(\bar{u})]$, we have

$$(A2.11) \quad s \leq \bar{u}T(\bar{u}) < s+1,$$

and (A2.5) together with this gives us

$$(A2.12) \quad s(s-1) \leq 2K\bar{u}^2 < s(s+1).$$

At \bar{u} , (A2.10) becomes

$$(A2.13) \quad \begin{aligned} T''(\bar{u}) &= \frac{s}{\bar{u}^3} - \frac{\bar{u}^{s-2}}{2\bar{u}^{s+1}}[2K\bar{u}^2 - s(s-1)] \\ &= \frac{1}{2\bar{u}^3}[2s - \{2K\bar{u}^2 - s(s-1)\}] = \frac{1}{2\bar{u}^3}[s(s+1) - 2K\bar{u}^2], \end{aligned}$$

and we get $T''(\bar{u}) > 0$ because of (A2.12). Hence if $T'(\bar{u}) = 0$, then $T(u)$ has a minimum there.

If \bar{u} happens to coincide with u_s , then we have to carry out more manipulation.

Since u_s is at the lower end of the interval $[u_s, u_{s+1})$, the above argument tells us only that if $T'(u_s) = 0$ then

$$(A2.14) \quad T''(u_s + 0) > 0.$$

Considering the limit, as $u \rightarrow u_s - 0$, of (A2.13) after replacing s by $s-1$, we get when $\bar{u} = u_s$

$$(A2.15) \quad T''(u_s - 0) = \frac{1}{2u_s^3}[(s-1)s - 2Ku_s^2] = 0.$$

This is short of what we would hope for. But, replacing s by $s - 1$ in (2.19) and using (A2.5), we get

$$(A2.16) \quad A_{s-1} = \frac{1}{(s-2)u_s^{s+1}} > 0.$$

Then, from (5.1) and (A2.9), we have

$$(A2.17) \quad \begin{aligned} T'''(u) &= -\frac{3(s-1)}{u^4} - A_{s-1}(s-1)(s-2)(s-3)u^{s-4} \\ &< 0 \quad (u_{s-1} \leq u < u_s, s \geq 3). \end{aligned}$$

Therefore, at least for sufficiently small $h > 0$, we obtain

$$\begin{aligned} T'(u_s - h) &= T'(u_s) - hT'''(u_s) + \frac{h^2}{2!}T'''(u_s - \theta h) \\ &< 0, \end{aligned}$$

where θ is between 0 and 1. This result, together with (A2.14), will assure us that, even if $\bar{u} = u_s$, $T(u)$ takes a minimum at \bar{u} .

When \bar{u} lies in the interval $[u_1, u_2]$, then we must use (2.20), (2.21), (2.22) in place of (2.17), (2.18), (2.19). But (A2.1) still holds here. Therefore, if

$$(A2.18) \quad T'(\bar{u}) = -\frac{1}{\bar{u}^2} - K + \frac{1}{\bar{u}^2}\bar{u}T(\bar{u}) = 0,$$

then $1 = \text{int}[\bar{u}T(\bar{u})]$ implies

$$(A2.19) \quad 1 \leq \bar{u}T(\bar{u}) = 1 + K\bar{u}^2 < 2,$$

and consequently

$$(A2.20) \quad 1 - K\bar{u}^2 > 0.$$

Now (2.22) gives us

$$(A2.21) \quad T''(\bar{u}) = \frac{1}{\bar{u}^3}(1 - K\bar{u}^2) > 0.$$

Since $T'(u_1 - 0) = -K < 0$, there is no problem even if $\bar{u} = u_1$.

Only remaining point is what will happen if $\bar{u} = u_2$. In that case, (A2.14) still holds when $s = 2$. In stead of (A2.13) with s replaced by $s - 1$, we have to use here (2.22) which gives

$$(A2.22) \quad T''(u_2 - 0) = \frac{1}{u_2^3}(1 - Ku_2^2) = 0$$

On the other hand, we get from (2.21) that, if $T'(u_2) = 0$ then

$$(A2.23) \quad A_1 = -\frac{1}{2u_2^2} - K(\log_e u_2 + 1),$$

whence

$$(A2.24) \quad T'(u) = -\frac{1}{2u^2} + \frac{1}{2u_2^2} + K \log_e \frac{u}{u_2} < 0 \quad \text{if } u < u_2.$$

This will suffice to complete the argument above when $\bar{u} = u_2$.

Now we are ready to complete the proof of the remaining part of "Theorem" [5d], that is the uniqueness of the point \bar{u} where $T'(\bar{u}) = 0$. Through a rather lengthy arguments, we have established the "Lemma":

$$(A2.25) \quad \text{If } T'(\bar{u}) = 0 \text{ then } T(u) \text{ has a \underline{minimum} at } \bar{u}.$$

As $T'(u)$ exists and is continuous in the whole interval $(0,1)$, at a point where $T(u)$ has a minimum or a maximum, $T'(u)$ must be zero. But (A2.25) prohibits the existence of a maximum. If there were more than one minima, then there would be a maximum between two adjacent minima, which is impossible. Therefore there can be only one minimum. This completes the proof of "Theorem" [5d].

Let the unique minimum of $T(u)$ be in the interval $[u_{s^*}, u_{s^*+1})$, then (A2.12) tells us that $2K\bar{u}^2 - s^*(s^* - 1) \geq 0$, and consequently $2Ku_s^2 - s(s - 1) \geq 0$ for $s = s^*$ thanks to (5.9). Now, [5b] will produce

$$(A2.26) \quad 2Ku_s^2 \geq s(s - 1) \quad \text{for } s = s^*, s^* - 1, \dots, 1.$$

Next, (A2.12) tells us that $2K\bar{u}^2 - s^*(s^* + 1) < 0$, and consequently $2Ku_s^2 < (s + 1)(s - 2)$ for $s = s^* + 1$ thanks to (5.9). Then [5c] will produce

$$(A2.27) \quad 2Ku_s^2 \leq (s + 1)(s - 2) < s(s - 1) \quad \text{for } s = s^* + 1, s^* + 2, \dots.$$

This completes the proof of Proposition [5g].

Proposition [5h] comes from the fact that s^* is the greatest integer satisfying $s(s - 1) \leq 2K\bar{u}^2$.

REFERENCES

- [1] Chow, Y.S., Moriguti, S., Robbins, H. and Samuels, S.M. (1964) "Optimal selection based on relative rank (the 'secretary problem')", *Israel Journal of Mathematics* Vol. 2, 81-90.
- [2] Ferguson, T.S. (1989) "Who solved the secretary problem", *Statistical Science* Vol. 4, 282-296.
- [3] Moriguti, S. (1992) "A selection problem with cost — 'secretary problem' when unlimited recall is allowed", *J. of Operations Research Society of Japan* Vol. 35, 373-382.
- [4] Moriguti, S. (1993a) "Basic theory of selection by relative rank with cost", *J. of Operations Research Society of Japan* Vol. 36, 46-61.
- [5] Moriguti, S. (1993b) "Asymptotic theory of selection by relative rank with medium cost", *J. of Operations Research Society of Japan* Vol. 36, 102-117.

Sigeiti Moriguti
 Syoan 2-16-10
 Sugunami-ku
 Tokyo 167, JAPAN