RAIRO. RECHERCHE OPÉRATIONNELLE

Z. GOVINDARAJULU

# The secretary problem : optimal selection with batch-interviewing and cost

*RAIRO. Recherche opérationnelle*, tome 26, nº 4 (1992), p. 391-407

<http://www.numdam.org/item?id=RO_1992__26_4_391_0>

Recherche opérationnelle/Operations Research
(vol. 26, n° 4, 1992, p. 391 à 407)

# THE SECRETARY PROBLEM: OPTIMAL SELECTION WITH BATCH-INTERVIEWING AND COST (*)

by Z. GOVINDARAJULU ([1])

Communicated by J. ABADIE

Abstract. — *We consider the secretary problem when the candidates are interviewed in batches of two at each stage and cost of interviewing is present. Closed form asymptotic results are derived and are compared with case when there is no recall of the candidates. There is a small reduction in the expected loss when the interview cost is a power function which is nonconstant.*

Keywords : Interview in batches; Interview cost; Optimal selection.

Résumé. — *Nous considérons le « problème de la secrétaire » où les candidates sont interviewées deux par deux, avec un coût d'interview. Nous donnons une formule asymptotique exacte qui est comparée avec le cas où il n'y a pas de rappel des candidates. Il y a une légère réduction du coût moyen lorsque le coût de l'interview est une fonction puissance non constante.*

Mots clés : Interview par lots, coût d'interview, sélection optimale.

## 1. INTRODUCTION

Chow, Moriguti, Robbins and Samuels (CMRS) (1964) consider the secretary problem and obtain significant results. Here we consider the same problem with interview cost present and the candidates are interviewed in batches of two. The added advantage is that we will be able to recall the immediately preceding candidate. In other words, we consider the following version of the secretary problem.

An executive puts an ad in the paper regarding a certain vacant position. $n$ candidates apply for the position. Assume that $n=2m$ where $m$ is a positive integer. He interviews candidates 1 and 2 at stage 1. If he does not hire one of these candidates, he interviews candidates 3 and 4 at stage 2. If he does

(*) Received November 1991, accepted April 1992.
A.M.S. Subject Classification: Primary 62L15; Secondary 60G40.
([1]) University of Kentucky, Lexington, KY 40506, U.S.A..

not hire one of these candidates, he moves on to stage 3. He has to hire someone by stage $m$. If he stops at stage $i$, he hires the better of the $(2i-1)$st and $2i$th candidates. If he stops at stage $i$, the loss incurred is

$$l(i, X_{2i-1}, X_{2i}) = \begin{cases} X_{2i-1} + h(2i) & \text{if } (2i-1)\text{st candidate is chosen,} \\ X_{2i} + h(2i) & \text{if } (2i)\text{th candidate is chosen,} \end{cases}$$

where $(X_1, X_2, \ldots, X_n)$ denotes the true ranks of the candidates and $h(j)$ denotes the cumulative cost of interviewing $j$ candidates.

## 2. CERTAIN RESULTS

Let $X_1, \ldots, X_n$ denote the random permutation of the integers $1, \ldots, n$ and assume that all $n!$ permutations are equally likely. The rank 1 corresponds to the best candidate, ..., and $n$ to the worst candidate. For any $i = 1, \ldots, n$, let $Y_i$ be the relative rank of the $i$th candidate to be interviewed (*i.e.*, $Y_i = 1 + \text{number of } X_1, \ldots, X_{i-1}$ which are less than $X_i$). Then it is well known that random variables $Y_1, \ldots, Y_i$ are independent with

$$P(Y_i = j) = \frac{1}{i}, \qquad j = 1, \ldots, i. \tag{2.1}$$

Also, from Govindarajulu [2], we have

$$E(X_{i-1} \mid Y_1, \ldots, Y_i) \tag{2.2}$$

$$= E(X_{i-1} \mid Y_{i-1}, Y_i) = \begin{cases} \dfrac{n+1}{i+1} Y_{i-1} & \text{if } Y_{i-1} < Y_i \\ \dfrac{n+1}{i+1}(1 + Y_{i-1}) & \text{if } Y_{i-1} \geqq Y_i, \end{cases}$$

and from CMRS [1], Eqn. (3)), we have

$$E(X_i \mid Y_i, \ldots, Y_i) = E(X_i \mid Y_i) = \left(\frac{n+1}{i+1}\right) Y_i. \tag{2.3}$$

Let $\tau$ denote the stage at which the executive stops interviewing $(\tau = 1, \ldots, m)$. Then

$$El(\tau, X_{2\tau-1}, X_{2\tau}) = EE\{l(\tau, X_{2\tau-1}, X_{2\tau}) \mid \tau\}.$$

Consider

$$(2.4)\quad E\{l(i, X_{2i-1}, X_{2i}) \mid \tau=i\} = EE\{l(i, X_{2i-1}, X_{2i}) \mid \tau=i, Y_{2i-1}, Y_{2i}\}$$

$$= \begin{cases} E\left(\dfrac{n+1}{2i+1} Y_{2i-1} + h(2i) \mid \tau=i, Y_{2i-1}, Y_{2i}\right) & \text{if } Y_{2i-1} < Y_{2i} \\ E\left(\dfrac{n+1}{2i+1} Y_{2i} + h(2i) \mid \tau=i, Y_{2i-1}, Y_{2i}\right) & \text{if } Y_{2i} \leqq Y_{2i-1} \end{cases}$$

$$= E\left\{\left(\frac{n+1}{2i+1}\right) \min(Y_{2i-1}, Y_{2i}) + h(2i) \mid \tau=i, Y_{2i-1}, Y_{2i}\right\}.$$

Suppose you are at stage $m-1$. If you stop now and select the better of the candidates $2m-3$, $2m-2$, your conditional expected loss is

$$(2.5)\quad \frac{n+1}{2m-1} \min(Y_{2m-3}, Y_{2m-2}) + h(2m-2)$$

$$= \frac{n+1}{n-1} \min(Y_{n-3}, Y_{n-2}) + h(n-2).$$

If instead you continue to the $m$th stage (last stage), you do so knowing the values of $Y_{n-3}$, $Y_{n-2}$, and what you expect to have to pay is the conditional expectation given $Y_{n-3}$ and $Y_{n-2}$ of the best you can do at the last stage, namely

$$(2.6)\quad C_{m-1}(Y_{n-3}, Y_{n-2}) = E\left\{\frac{n+1}{n+1} \min(Y_{n-1}, Y_n) + h(n) \mid Y_{n-2}, Y_{n-3}\right\}$$

$$(2.7)\quad = E\{\min(Y_{n-1}, Y_n)\} + h(n)$$

$$(2.8)\quad = \frac{n+1}{3} + h(n),$$

since $E\{\min(Y_{n-1}, Y_n)\} = (n+1)/3$ (*see* Govindarajulu [2], Eq. (2.10)). That is, $C_{m-1}(Y_{n-3}, Y_{n-2})$ is free of $Y_{n-3}$ and $Y_{n-2}$.

So, one should stop at stage $m-1$ if (2.5) is smaller than (2.6). That is, stop at stage $m-1$ if

$$(2.9)\quad \min(Y_{n-3}, Y_{n-2}) < \frac{n-1}{3} + \frac{n-1}{n+1}\{h(n) - h(n-2)\}.$$

This problem persists backward in time. $C_{i-1}$ is given by

$$(2.10)\quad \left\{ \begin{array}{l} C_{i-1}=E\left[\min\left\{\dfrac{n+1}{2i+1}\min(Y_{2i-1}, Y_{2i})+h(2i);\ C_i\right\}\right], \\ \qquad i=m, \ldots, 1. \end{array} \right.$$

Note that $C_i=C_i(n)$ denotes the minimal possible expected loss if we confine ourselves to stopping rules $\tau$ such that $\tau \geqq i+1$. We would like to find the value of $C_0$.

We can rewrite (2.10) as

$$(2.11)\quad C_{i-1}-h(2i) = \left(\frac{n+1}{2i+1}\right)E\left[\min\left\{\min(Y_{2i-1}, Y_{2i}), \left(\frac{2i+1}{n+1}\right)(C_i-h(2i))\right\}\right].$$

Let

$$(2.12)\quad s_i=\left[\left(\frac{2i+1}{n+1}\right)(C_i-h(2i))\right], \qquad i=1, \ldots, m-1=\frac{n}{2}-1,$$

where [.] denotes the largest integer contained in (.).

The optimal stopping rule, which is implicit in (2.11) is: If you are at stage $i$, stop if $\min(Y_{2i-1}, Y_{2i}) \leqq s_i$ and select the better of the $(2i-1)$st and $(2i)$th candidates.

Now let

$$(2.13)\quad D_i=\left\{\left(\frac{2i+1}{n+1}\right)(C_i-h(2i))\right\}, \qquad i=0, \ldots, m-1.$$

Then (2.11) becomes

$$\left(\frac{2i+1}{n+1}\right)(C_{i-1}-h(2i))=\frac{1}{(2i-1)2i}\sum_{j=1}^{2i-1}\sum_{k=1}^{2i}\min(j, k, D_i).$$

That is,

$$\begin{aligned} &2i(2i-1)(2i+1)(C_{i-1}-h(2i))(n+1)^{-1} \\ &\quad= \sum_{i\leqq j\leqq k\leqq 2i-1}\sum \min(j, D_i)+\sum_{1\leqq k<j\leqq 2i-1}\sum \min(k, D_i) \\ &\quad+\sum_{j=1}^{2i-1}\min(j, 2i, D_i)=2\sum_{j=1}^{2i-1}(2i-j)\min(j, D_i) \end{aligned}$$

$$= 2 \sum_{j=1}^{s_i} (2i-j)j + 2 \sum_{j=1+s_i}^{2i-1} (2i-j) D_i$$

$$= 2 i s_i (1+s_i) - \frac{1}{3} s_i (1+s_i)(2 s_i + 1) + D_i (2i - s_i - 1)(2i - s_i)$$

$$= \frac{1}{3} s_i (1+s_i)(6i - 2 s_i - 1) + \left(\frac{2i+1}{n+1}\right)(2i - s_i - 1)(2i - s_i)(C_i - h(2i)).$$

Thus,

$$(2.14) \quad C_{i-1} - h(2i) = \frac{s_i (1+s_i)(6i - s_i - 1)(n+1)}{6i(2i-1)(2i+1)} + \frac{(2i - s_i - 1)(2i - s_i)}{2i(2i-1)}(C_i - h(2i)),$$

or, alternatively,

$$(2.15) \quad C_{i-1} = \frac{(2i - s_i - 1)(2i - s_i)}{2i(2i-1)} C_i + \frac{h(2i)}{2i(2i-1)} \{ s_i (4i - s_i - 1) \} + \frac{(n+1)}{6i(2i-1)(2i+1)} s_i (1+s_i)(6i - 2 s_i - 1) \qquad i = m-1, \ldots, 1.$$

If $h(i) \equiv 0$, $C_0$ gives the expected rank of the candidate chosen by the optimal rule. Via (2.6) and (2.14), we can successively compute $C_{m-1}$, $s_{m-1}$, $C_{m-2}, \ldots, s_1$ and $C_0$.

We consider the following interview cost function

$$(2.16) \qquad h(i) = a\left(\frac{i}{n+1}\right)^r, \qquad a > 0, \qquad r = 0, 1, 2, \ldots$$

## Expected number of stages

It can easily be shown that

$$(2.17)\quad E\tau=\sum_{i=0}^{n-1} P(\tau>i)=1+\sum_{i=1}^{n-1} P(\tau>i)$$

$$=1+\sum_{i=1}^{n-1} P(\min(Y_{2j-1}, Y_{2j})>s_j, \qquad \text{for } j\leqq i)$$

$$=1+\sum_{i=1}^{n-1}\prod_{j=1}^{i}\left(\frac{2j-1-s_j}{2j-1}\right)\left(\frac{2j-s_i}{2j}\right).$$

For example, when $n=14$ and $a=r=1$, we have

| $i$ | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|
| $C_{i-1}$ | 5 93 | 4 55 | 3 84 | 3 43 | 3 14 | 3 14 | 3 14 |
| $s_i$ | | 4 | 2 | 1 | 1 | 0 | 0 |

$E\tau=4\ 40$

## 3. ASYMPTOTIC SOLUTIONS

In this section following the "clever" method of Robbins [4], we derive a closed-form asymptotic expression for $C_0$ when the cost of interviewing is zero and an approximate solution when the cost of interviewing is a power function given by (2.16). Now (2.15) can be rewritten as

$$(3.1)\quad C_{i-1}-C_i=\frac{\{-s_i(4i-1)+s_i^2\}}{2i(2i-1)}\{C_i-h(2i)\}+\frac{(n+1)s_i(1+s_i)(6i-2s_i-1)}{6i(2i-1)(2i+1)}.$$

Divide both sides of (3.1) by $1/(n+1)$, let $n$ and $i$ tend to infinity in such a way that $\frac{i}{(n+1)}\to t$, and obtain

$$\lim_{\substack{i\to\infty\\ n\to\infty}}(C_{i-1}-C_i)(n+1)=-f'(t)=\frac{-s}{t}f(t)+\frac{H(t)s}{t}+\frac{s(1+s)}{4t^2}$$

where

$$\lim_{\substack{i \to \infty \\ n \to \infty}} h(2i) = H(t).$$

That is,

$$f'(t) - \frac{s}{t} f(t) = \frac{-s}{t} H(t) - \frac{s(1+s)}{4t^2}, \tag{3.2}$$

which is valid for all $t$ such that

$$s \leqq 2tg(t) = 2t\{f(t) - H(t)\} \leqq s+1. \tag{3.3}$$

Notice that (3.3) follows from the definition of $s_i$ given by (2.12). We can rewrite (3.2) as

$$g'(t) - \frac{s}{t} g(t) = \frac{-s(1+s)}{4t^2} - H'(t). \tag{3.4}$$

where

$$g(t) = f(t) - H(t).$$

Next, define the sequence of $t$'s,

$$0 = t_0 < t_1 < \ldots < \frac{1}{2}, \tag{3.5}$$

by the equations

$$2t_s g(t_s) = s, \qquad s = 0, 1, 2, \ldots, \tag{3.6}$$

so that the differential Equation (3.4) holds in the interval

$$t_s < t < t_{s+1}.$$

In (3.4), multiplying by the integrating factor $t^{-s}$ and integrating both sides, we have

$$g(t)t^{-s} = \frac{-s(1+s)}{4} \int t^{-2-s} dt - \int H'(t) t^{-s} dt + \frac{1}{2} A_s$$

$$= \frac{s}{4} t^{-1-s} - \int H'(t) t^{-s} dt + \frac{1}{2} A_s.$$

So,

$$(3.7) \qquad 2\,tg(t)=\frac{s}{2}-2\,t^{1+s}\int H'(t)\,t^{-s}\,dt+A_s\,t^{1+s}.$$

Next, we consider some special cases for $H(t)$.

*Case* 1: Let $H(t)=a$. Then (3.7) becomes

$$2\,tg(t)=\frac{s}{2}+A_s\,t^{1+s}.$$

Hence,

$$2\,t_s\,g(t_s)=s=\frac{s}{2}+A_s\,t_s^{1+s},$$

and

$$2\,t_{s+1}\,g(t_{s+1})=s+1=\frac{s}{2}+A_s\,t_{s+1}^{1+s}.$$

Eliminating $A_s$, we have

$$\left(\frac{t_{s+1}}{t_s}\right)^{1+s}=\frac{s+2}{s}$$

or

$$t_{s+1}=t_1\prod_{i=1}^{s}\left(1+\frac{2}{i}\right)^{1/(1+i)}$$

However, $t_s \to 1/2$ as $s \to \infty$. Hence,

$$t_1=\frac{1}{2}\prod_{i=1}^{\infty}\left(1+\frac{2}{i}\right)^{-(1/(1+i))}$$

Now from (3.7) for $s=0$, $g(t_1)=g(0)=\lim_{n\to\infty} C_0$. Thus, since $t_1\,g(t_1)=1/2$, we have

$$(3.8) \qquad \lim_{n\to\infty} C_0=g(t_1)=\frac{1}{2}\,t_1^{-1}=\prod_{i=1}^{\infty}\left(1+\frac{2}{i}\right)^{1/(1+i)},$$

which coincides with the asymptotic expression obtained by CMRS [1] when interviewing is one candidate at each stage and the cost of interviewing is zero.

*Remark* 3.1: $g(t_1)$ gives the expected rank of the candidate selected by the optimal rule.

*Case* 2: Let $H(t)=at^r$ for some $a>0$ and $r>0$. Then (3.7) becomes

$$2\,tg(t)=\frac{s}{2}+\frac{2\,ar}{s-r}t^{r+1}+A_s\,t^{1+s}.$$

Again eliminating $A_s$ from the equations,

$$2\,t_s\,g(t_s)=\frac{s}{2}+\frac{2\,ar}{s-r}t_s^{r+1}+A_s\,t_s^{1+s}=s$$

and

$$2\,t_{s+1}\,g(t_{s+1})=\frac{s}{2}+\frac{2\,ar}{s-r}t_{s+1}^{r+1}+A_s\,t_{s+1}^{1+s}=s+1,$$

we obtain

$$\frac{4\,ar}{s-r}t_{s+1}^{r+1}\left\{1-\left(\frac{t_{s+1}}{t_s}\right)^{s-r}\right\}+s\left(\frac{t_{s+1}}{t_s}\right)^{1+s}=s+2. \tag{3.9}$$

For large $s$, $t_{s+1}$ will be close to 1/2 and $t_{s+1}/t_s$ will be close to unity. So, we can ignore the first term on the left side of (3.9). Solving the rest of the Equation, we obtain

$$t_{s+1}\doteq t_1\prod_{i=1}^{s}\left(1+\frac{2}{i}\right)^{1/(1+i)}$$

Since $t_{s+1}\to 1/2$ as $s\to\infty$, we have

$$t_1=\frac{1}{2}\prod_{i=1}^{\infty}\left(1+\frac{2}{i}\right)^{-((1/1+i))}$$

and

$$t_{s+1}=\frac{1}{2}\prod_{i=s+1}^{\infty}\left(1+\frac{2}{i}\right)^{-(1/(1+i))}$$

Hence

$$\left(\frac{t_{s+1}}{t_s}\right)^{s-r}=\left(1+\frac{2}{s}\right)^{(s-r)/(1+s)}, \tag{3.10}$$

and

$$\ln t_{s+1}=-\ln 2-\sum_{i=s+1}^{\infty}(i+1)^{-1}\ln\left(1+\frac{2}{i}\right)=-\ln 2-2$$
$$\times\sum_{s+1}^{\infty}\frac{1}{i(i+1)}=-\ln 2-2\sum\left(\frac{1}{t}-\frac{1}{i+1}\right)=-\ln 2-\frac{2}{1+s}.$$

Using this in (3.9) we have (after setting $t_{s+1}=(1/2)e^{-2/(1+s)}$)

$$\frac{4\,ar}{s(s-r)}2^{-(r+1)}\left\{1-\left(1+\frac{2}{s}\right)^{(s-r)/(s+1)}\right\}e^{-2(r+1)/(1+s)}+\left(\frac{t_{s+1}}{t_s}\right)^{1+s}=\frac{s+2}{s}.$$

Thus,

$$\left(\frac{t_{s+1}}{t_s}\right)^{1+s}=\frac{s+2}{s}+\frac{2\,ar}{2^r s(s-r)}\left\{\left(1+\frac{2}{s}\right)^{(s-r)/(s+1)}-1\right\}e^{-2(r+1)/(1+s)} \tag{3.11}$$
$$\doteq\frac{s+2}{s}+\frac{2\,ar}{2^r s(s-r)}\cdot\frac{(s-r)}{s+1}\cdot\frac{2}{s}e^{-2(r+1)/(1+s)}\ (s\geqq 2)$$
$$=\frac{s(s+1)(s+2)+ar\,2^{2-r}e^{-2(r+1)/(1+s)}}{s^2(1+s)}.$$

Hence,

$$t_{s+1}\doteq t_2\prod_{i=2}^{s}\left\{\frac{i(i+1)(i+2)+ar\,2^{2-r}e^{-2(r+1)/(1+s)}}{i^2(i+1)}\right\}^{1/(1+i)}$$

Since $t_{s+1}\to 1/2$ as $s\to\infty$ and $2\,t_2\,g(t_2)=2$,

$$g(t_2)=t_2^{-1}=2\prod_{i=2}^{\infty}\left\{\frac{i(i+1)(i+2)+ar\,2^{2-r}e^{-2(r+1)/(i+1)}}{i^2(i+1)}\right\}^{1/(1+i)} \tag{3.12}$$

Setting $s=1$ in (3.9), we have

$$\left(\frac{t_2}{t_1}\right)^2=3+\frac{4\,ar}{1-r}t_2^{r+1}\left\{\left(\frac{t_2}{t_1}\right)^{1-r}-1\right\}\doteq 3+\frac{4\,ar}{1-r}t_2^{r+1}(3^{(1-r)/2}-1),$$

after using (3.10).

Hence,

$$t_1 \doteq t_2 \left\{ 3 + \frac{4ar}{1-r} t_2^{1+r} (3^{(1-r)/2} - 1) \right\}^{-1/2}$$

Also, $2 t_1 g(t_1) = 1$ implies that $g(t_1) = \left(\frac{1}{2}\right) t_1^{-1}$, and hence

$$(3.13) \quad g(t_1) = \frac{1}{2} t_2^{-1} \left\{ 3 + \frac{4ar}{1-r} t_2^{1+r} (3^{(1-r)/2} - 1) \right\}^{1/2}$$

$$= \frac{1}{2} g(t_2) \left\{ 3 + \frac{4ar}{1-r} g^{-(1+r)}(t_2)(3^{(1-r)/2} - 1) \right\}^{1/2}$$

$$= \frac{1}{2} \{ g(t_2) \}^{(1-r)/2} \left\{ 3 g^{1+r}(t_2) + \frac{4ar}{1-r} (3^{(1-r)/2} - 1) \right\}^{1/2}.$$

Now (3.12) and (3.13) will yield $g(t_1)$.

*Remark* 3.2: When $r = 1$,

$$(3.14) \qquad g(t_1) = \frac{1}{2} \{ 3 . g^2(t_2) + 2a \ln 3 \}^{1/2}$$

since $\lim_{r \to 1} (A^{1-r} - 1)/(1-r) = \ln A$ for any $A$.

In the next section we shall obtain similar results for the case when we cannot recall the immediately preceding candidate.

## 4. ASYMPTOTIC SOLUTION WHEN ONE CANDIDATE IS INTERVIEWED AT EACH STAGE

The optimal selection rule is implemented via the constants $C_{n-1}$, $C_{n-2}, \ldots, C_0$ which can be computed from the following equations which are analogous to (2.8) and (2.15).

$$(4.1) \qquad C_{n-1} = \frac{n+1}{2} + h(n),$$

$$(4.2) \qquad C_{i-1} = \left(1 - \frac{s_i}{i}\right) C_i + \frac{h(i) s_i}{i} + \frac{(n+1)}{2i(i+1)} s_i (1 + s_i)$$

and

$$(4.3) \qquad s_i = \left[ \left( \frac{i+1}{n+1} \right) \{ C_i - h(i) \} \right], \qquad i = 1, \ldots, n-1$$

with $s_n = n$ and $[\cdot]$ denotes the largest integer contained in $(\cdot)$.

The optimal stopping rule is: If you are at the $i$-th candidate, stop and select the $i$-th candidate if $Y_i \leqq s_i$ $(i = 1, \ldots, n)$.

**Expected stopping time**

If $\tau$ denotes the stopping time, proceeding as in (3.2), we have

$$(4.4) \qquad E\tau = 1 + \sum_{i=1}^{n-1} \prod_{j=1}^{i} \left( \frac{i - s_i}{i} \right).$$

For example, when $n = 14$ and $a = r = 1$, we obtain

| $i$ | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $C_{i-1}$ | 8 43 | 6 67 | 5 66 | 5 00 | 4 52 | 4 15 | 3 87 | 3 65 | 3 45 | 3 34 | 3 32 | 3 32 | 3 32 | 3 32 |
| $s_i$ | | 7 | 5 | 3 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

$E\tau = 7\ 26$

One can rewrite (4.2) as

$$C_{i-1} - C_i = \frac{-s C_i}{i} + h(i) \frac{s_i}{i} + \frac{(n+1)}{2 i (i+1)} s_i (1 + s_i).$$

Now proceeding as in Section 3, we obtain the differential equation

$$(4.5) \qquad f'(t) - \frac{s}{t} f(t) = \frac{-s}{t} H(t) - \frac{s(1+s)}{2t^2}$$

where

$$f'(t) = \lim_{\substack{n \to \infty \\ i \to \infty}} (C_i - C_{i-1})(n+1)$$

and

$$\lim_{\substack{n \to \infty \\ i \to \infty}} h(i) = H(t).$$

It should be noted that (4.5) is valid for all $t$ such that

$$s \leqq tg(t) = t(f(t) - H(t)) \leqq s+1$$

which follows from the definition of $s_i$. We can rewrite (4.5) as

$$g'(t) - \frac{s}{t} g(t) = \frac{-s(1+s)}{2t^2} - H'(t). \tag{4.6}$$

Define the sequence of $t'_s$ $0 = t_0 < t_1 < \ldots < 1$ by the equations

$$t_s g(t_s) = s, \qquad s = 0, 1, \ldots, \tag{4.7}$$

so that the differential equation holds in the interval

$$t_s < t < t_{s+1}.$$

Solving (4.6), we obtain

$$tg(t) = \frac{s}{2} - t^{s+1} \int H'(t)\, t^{-s}\, dt + A_s t^{s+1}. \tag{4.8}$$

Now, let us consider special cases for $H(t)$.

*Case* 3: If $H(t) = a$, then one can obtain

$$\lim_{n \to \infty} C_0 = g(t_1) = \prod_{i=1}^{\infty} \left(1 + \frac{2}{i}\right)^{1/(1+i)} \tag{4.9}$$

which coincides with (3.8). That is, asymptotically, there is no difference when you interview candidates in a batch of 2 or one at each stage.

*Case* 4: Let $H(t) = at^r$, $a > 0$ and $r > 0$. Then (4.8) becomes

$$tg(t) = \frac{s}{2} + \frac{art^{r+1}}{s-r} + A_s t^{s+1}. \tag{4.10}$$

Eliminating $A_s$ from the two equations,

$$t_s g(t_s) = \frac{s}{2} + \frac{art_s^{r+1}}{s-r} + A_s t_s^{s+1} = s,$$

$$t_{s+1} g(t_{s+1}) = \frac{s}{2} + \frac{art_{s+1}^{r+1}}{s-r} + A_s t_{s+1}^{s+1} = s+1,$$

we obtain

$$(4.11)\quad 2\,ar t_{s+1}^{r+1}\left[1-\left(\frac{t_{s+1}}{t_s}\right)^{s-r}\right]+s(s-r)\left(\frac{t_{s+1}}{t_s}\right)^{s+1}=(s+2)(s-r).$$

For large $s$, $t_{s+1}/t_s$ will be close unity. So ignore the first term in (4.11) and obtain

$$t_{s+1}=t_1\ \overset{s}{\underset{i=1}{\pi}}\left(1+\frac{2}{i}\right)^{1/(1+i)}.$$

Since $t_{s+1}\to 1$ as $s\to\infty$, we have

$$t_1=\prod_{i=1}^{\infty}\left(1+\frac{2}{i}\right)^{-1/(1+i)}$$

and

$$t_{s+1}=\prod_{i=s+1}^{\infty}\left(1+\frac{2}{i}\right)^{-1/(1+i)}.$$

Hence,

$$(4.12)\qquad \frac{t_{s+1}}{t_s}=\left(1+\frac{2}{s}\right)^{1/(s+1)},$$

and

$$\ln t_{s+1}=-\sum_{s+1}^{\infty}(i+1)^{-1}\ln\left(1+\frac{2}{i}\right)=-2\sum_{s+1}^{\infty}\frac{1}{i(i+1)}=\frac{-2}{s+1}.$$

Using this in (4.11) (and after setting $t_{s+1}=e^{-2/(s+1)}$), we have

$$\left(\frac{t_{s+1}}{t_s}\right)^{s+1}\doteq\frac{s+2}{s}+\frac{2\,ar}{s(s-r)}\left[\left(1+\frac{2}{s}\right)^{(s-r)/(s+1)}-1\right]e^{-2(r+1)/(s+1)}$$

$$\doteq\frac{s+2}{s}+\frac{2\,ar}{s(s-r)}\cdot\frac{(s-r)}{s+1}\cdot\frac{2}{s}e^{-2(r+1)/(s+1)}\ (s\geqq 2)$$

$$=\frac{s(s+1)(s+2)+4\,ar\,e^{-2(r+1)/(s+1)}}{s^2(1+s)}.$$

Hence

$$t_{s+1} \doteq t_2 \prod_{i=2}^{s} \left\{ \frac{i(i+1)(i+2)+4\,ar\,e^{-2(r+1)/(s+1)}}{i^2(i+1)} \right\}^{1/(i+1)}$$

Since $t_{s+1} \to 1$ as $s \to \infty$, we have

$$t_2 = \prod_{i=2}^{\infty} \left\{ \frac{i(i+1)(i+2)+4\,ar\,e^{-2(r+1)/(s+1)}}{i^2(i+1)} \right\}^{-1(1+i)}$$

Since $t_2\, g(t_2)=2$, we obtain

$$g(t_2)=2 \prod_{i=2}^{\infty} \left\{ \frac{i(i+1)(i+2)+4\,\mathrm{ar}\,e^{-2(r+1)/(i+1)}}{i^2(i+1)} \right\}^{1/(1+i)}. \tag{4.13}$$

Setting $s=1$ in (4.11) we have

$$\left(\frac{t_2}{t_1}\right)^2 = 3+\frac{2\,ar}{1-r}\,t_2^{1+r}\left\{\left(\frac{t_2}{t_1}\right)^{1-r}-1\right\} \doteq 3+\frac{2\,ar}{1-r}\,t_2^{1+r}\left\{3^{(1-r)/2}-1\right\}$$

after using (4.12).

So,

$$t_1 = t_2\left\{3+\frac{2\,ar}{1-r}\,t_2^{1+r}(3^{(1-r)/2}-1)\right\}^{-1/2}$$

$$= t_2^{(1-r)/2}\left\{3\,t_2^{-(1+r)}+\frac{2\,ar}{1-r}(3^{(1-r)/2}-1)\right\}^{-1/2}$$

and

$$g(t_1)=t_1^{-1}=t_2^{(r-1)/2}\left\{3\,t_2^{-(1+r)}+\frac{2\,ar}{1-r}(3^{(1-r)/2}-1)\right\}^{1/2} \tag{4.14}$$

$$=\frac{1}{2}g^{(1-r)/2}(t_2)\left\{3\,g^{1+r}(t_2)+2^{2+r}\frac{ar}{1-r}(3^{(1-r)/2}-1)\right\}^{1/2}.$$

*Remark* 4.1: If $r=1$,

$$g(t_1)=\frac{1}{2}\{3\,g^2(t_2)+4\,a\ln 3\}^{1/2}. \tag{4.15}$$

TABLE 4.1

*Giving $C_0$ and for selected values of $n$.*

| $r$ sample size | $a$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|
| 14 | 0 | 2.52 (2.78) | 2.52 (2.78) | 2.52 (2.78) | 2.52 (2.78) |
| | 0.5 | 3.02 (3.28) | 2.83 (3.06) | 2.73 (2.95) | 2.67 (2.90) |
| | 1.0 | 3.52 (3.78) | 3.13 (3.21) | 2.93 (3.11) | 2.81 (3.01) |
| | 2.0 | 4.52 (4.78) | 3.63 (3.80) | 3.21 (3.38) | 3.04 (3.19) |
| | 3.0 | 5.52 (5.78) | 4.04 (4.28) | 3.42 (3.64) | 3.17 (3.35) |
| 24 | 0 | 2.90 (3.10) | 2.90 (3.10) | 2.90 (3.10) | 2.90 (3.10) |
| | 0.5 | 3.40 (3.60) | 3.17 (3.36) | 3.07 (3.26) | 3.01 (3.20) |
| | 1.0 | 3.90 (4.10) | 3.44 (3.62) | 3.24 (3.41) | 3.13 (3.31) |
| | 2.0 | 4.90 (5.10) | 3.99 (4.12) | 3.56 (3.70) | 3.34 (3.50) |
| | 3.0 | 5.90 (6.10) | 4.44 (4.59) | 3.82 (3.95) | 3.53 (3.67) |
| 50 | 0 | 3.28 (3.41) | 3.28 (3.41) | 3.28 (3.41) | 3.28 (3.41) |
| | 0.5 | 3.78 (3.91) | 3.56 (3.67) | 3.45 (3.57) | 3.40 (3.51) |
| | 1.0 | 4.28 (4.41) | 3.81 (3.92) | 3.60 (3.72) | 3.50 (3.61) |
| | 2.0 | 5.28 (5.41) | 4.32 (4.42) | 3.89 (4.40) | 3.69 (3.79) |
| | 3.0 | 6.28 (6.41) | 4.79 (4.89) | 4.16 (4.25) | 3.87 (3.96) |
| 100 | 0 | 3.53 (3.60) | 3.53 (3.60) | 3.53 (3.60) | 3.53 (3.60) |
| | 0.5 | 4.03 (4.10) | 3.79 (3.86) | 3.68 (3.75) | 3.63 (3.70) |
| | 1.0 | 4.53 (4.60) | 4.04 (4.11) | 3.83 (3.90) | 3.72 (3.79) |
| | 2.0 | 5.53 (5.60) | 4.54 (4.60) | 4.11 (4.17) | 3.90 (3.97) |
| | 3.0 | 6.53 (6.60) | 5.01 (5.07) | 4.37 (4.43) | 4.08 (4.13) |
| ∞ | 0 | 3.86 (3.86) | 3.86 (3.86) | 3.86 (3.86) | 3.86 (3.86) |
| | 0.5 | 3.86 (3.86) | 3.93 (3.99) | 3.89 (3.98) | 3.88 (3.95) |
| | 1.0 | 3.86 (3.86) | 3.99 (4.11) | 3.92 (4.09) | 3.89 (4.03) |
| | 2.0 | 3.86 (3.86) | 4.11 (4.34) | 3.98 (4.30) | 3.91 (4.19) |
| | 3.0 | 3.86 (3.86) | 4.23 (4.57) | 4.03 (4.50) | 3.93 (4.34) |

Values in parentheses are for the case when one candidate is interviewed at each stage.

*Remark* 4.2: Comparing (4.12) and (4.14) with (3.12) and (3.13), we infer that there will be a reduction in the expected loss when we interview in batches of 2 and when the interview cost is present.

## Discussion of Table 4.1

The values of $C_0$ for the case of interviewing in batches of two are consistently smaller than the corresponding values for the case of interviewing one candidate at a time with no recall. Notice that when $r=0$, the values of $C_0$ are comparable to the values of $g(t_1)+a$, whereas when $r>0$, the values of $C_0$ are comparable to those of $g(t_1)$. The values of $g(t_1)$ are based on a certain iteration method and taking only the first two terms in the binomial expansion of $(1+2/s)^{(s-r)/(s+1)}$. The asymptotic values of $g(t_1)$ for both the schemes are slightly smaller than their counterparts for $n=100$. This may be attributed to the fact that we use lower bounds for $t_{s+1}$ (namely, $t_{s+1}=(1/2)e^{-2/(1+s)}$ and $t_{s+1}=e^{-2/(1+s)}$). Even in the asymptotic case, there are small differences between the two minimal risk functions. The author has obtained asymptotic solutions that are based on taking the first three terms in the binomial expansion of $(1+2/s)^{(s-r)(1+s)}$. Then the numerical values of the minimal risk functions are slightly smaller than those given in Table 4.1. However, the differences are not appreciable. Thus, the asymptotic formulae for $g(t_2)$ and the numerical values based on this approximation are not presented here.

Lorenzen [3] considers the classical secretary problem (*i.e.*, no recall) with interview cost. In particular, when the interview cost is linear, he obtains numerical values of the optimal risk for large $n$. For instance, when $r=a=1$, he obtains the value 4.37 which is not far from our value of (4.11). He approximates the finite secretary problem by the infinite secretary problem and solves a single differential equation in order to obtain the asymptotic solution.

It is surmised that if the size of the batch is increased, especially if it is $m=[np]$ for some small positive $p$, there will be a significant difference between the two minimal risk functions.

## ACKNOWLEDGEMENTS

I thank Professor Thomas Ferguson for some discussions and Mahibbur Rahman for his assistance in the computations of numerical values in Table 4.1.

## REFERENCES

1. Y. S. Chow, S. Moriguti, H. Robbins and S. M. Samuels, Optimum Selection Based on Relative Rank (the "secretary problem"), *Israel J. Math.*, 1964, *2*, p. 81-90.
2. Z. Govindarajulu, The Secretary Problem: Optimal Selection with Interview Cost and Recall II. Unpublished manuscript, 1990.
3. T. J. Lorenzen, Optimal Stopping with Sampling Cost: the Secretary Problem, *Ann. Prob.*, 1981, *9*, p. 167-172.
4. H. Robbins, Remarks on the Secretary Problem. Robbins Volume of *Amer*, *J. Math. and Mgt. Sci.*, 1991, *11*, p. 25-37.