Finding the $\kappa$ best out of $n$ rankable objects. A consecutive thresholds Algorithm

F. Thomas Bruss∗ Guy Louchard †

January 28, 2013

Abstract

The objective of this paper is to find in a setting of $n$ sequential observations of objects a good online policy to select the $\kappa$ best of these $n$ uniquely rankable objects. Selection is without recall, and the idea is to use thresholds which are functions of selections made so far. Our main interest lies in the asymptotic behaviour of these thresholds as $n \to \infty$ and in the corresponding asymptotic performance of the threshold algorithm. This is motivated by the fact that it is hard to find closed form solutions of optimal strategies for general $\kappa$ and $n$.

Keywords: Multiple stopping rule, secretary problem, monotonocity, unimodality, relative ranks, record-values, optimality.

2010 Mathematics Subject Classification: 60G40 (68W27).

1 Introduction

The so-called secretary problem is a classical problem of optimal stopping on a sequence of rankable observations. It is the problem of choosing the best of $n$ rankable options appearing in random arrival order with probability $1/n!$ each. It was first solved by Lindley [5]. See e.g. Ferguson [4] for an interesting review. This is the special case $\kappa = 1$ of the problem we consider here, namely to find with one stop without recall the best of $n$ uniquely rankable objects or options. Arguably the shortest solution to this problem is provided by the more general odds-theorem of optimal stopping (Bruss [1]).

It is well known that the optimal strategy is to wait until the threshold index $s$ being the smallest $k$ such that $\sum_{j=k}^{n} r_j \geq 1$ where $r_k = (1/k)/(1-1/k) = 1/(k-1)$ and to accept the first relative rank from $s$ onwards (if any), and that with $s = s_n$, $\lim_{n \to \infty} s_n/n = 1/e$. The corresponding win probability is then given by

$$V_n = \prod_{j=s}^{n} \left( \frac{j-1}{j} \right) \sum_{j=s}^{n} \frac{1}{j-1} = \frac{s-1}{n} \sum_{j=s-1}^{n-1} \frac{1}{j}$$

which has the same limit as the function $s_n/n$ namely $1/e$.

The problem of the $\kappa$ best is to find an efficient stopping rule to select the $\kappa$ best objects online from a sequence of $n$ objects. This is the objective of the present paper. If $\kappa > 1$ then the optimal solution as well as the limiting behaviour of the thresholds are difficult, and this difficulty attracts our special attention.

Related work

Platen [6] studies the problem of choosing a fixed number of objects (secretaries) from a sequence of $n$ but his objective function is different. He thinks of the rank $k$ as generating the cost $k$ and considers

∗Université Libre de Bruxelles, Département de Mathématique, CP 210 Boulevard du Triomphe, B-1050 Bruxelles, Belgium, tbruss@ulb.ac.be

†Université Libre de Bruxelles, Département d’Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, louchard@ulb.ac.be
the problem of minimizing the sum of the cost of all selections. Notice that this is different from the problem of minimizing the expected rank as studied by Chow et al [3] and also again quite different from our problem of maximizing the expected payoff of one if and only if all the $\kappa$ best objects are chosen, and zero otherwise. Platen also considers more general payoff (cost) functions, but, motivated by thinking of secretary problems, limits his interest to additive functions, which do not apply to our problem.

A more weakly related problem has been studied by Rose [7] for the case $\kappa = 2$. Rose introduces in the same model two decision makers, one of which is defined as dominant. The objective is to select the two best objects, the best one being assigned to the dominant decision maker. Rose considers a strategy as successful if and only if both decision makers reach their goal under the online assignment constraint. Hence here the optimal success probability should be (and is, as we shall see) smaller than in our problem without assignments. Note also that the concurrent assignment constraint gives Roses’ problem quite a different character.

Wilson [10] generalizes the model of Rose for general $\kappa$, also with simultaneous online assignment of ranks. The paper displays important optimality conditions for general $\kappa$, but closed-form solutions seem now to become hard.

Tamaki [8] looks at the same problem for $\kappa = 3$ and shows, interestingly, that in contrast to the case $\kappa = 1$ (secretary problem) and the case $\kappa = 2$ the optimal strategy is now no longer simple. Here “simple” means that, for the $\kappa$ selections, the strategy waits until a threshold indice $s_k$ and selects then consecutively the next $k$ (if possible) record candidates. Tamaki’s observation of the non-simplicity of the optimal rule for $\kappa = 3$ (and $\kappa \geq 3$) is important in this context because it gives additional motivation to the work of the approach presented in this paper as well as to earlier work of Vanderbei [9] described below.

Vanderbei [9] studies the corresponding problem without online assignments with an original approach by formulating the problem as a control problem defined on a Markov chain. An explicit solution is found for $n$ even and $\kappa = n/2$. The corresponding optimal win probability equals $\frac{1}{\ell+1}$. For general $\kappa$, this article confirms its interest to studying a simple sub-optimal solution. Thus Vanderbei studies the same problem as ours but the approach is different. As far as we understand, Vanderbei’s sufficient condition (see the last line of [9] p.481) is not a sufficient condition in the set of strategies which we consider. Indeed, we need in general also information about 2-records, 3-records, up to $(\kappa - 1)$-records. Having said this we see by Theorem 5.1 of [9] that the sufficient condition we require is implicit by the definition of “marginal”, and the definition of the sets $T$ and $S'$, so that our strategies may coincide. A closer comparison between approaches is difficult, however, since Vanderbei’s approach leads to a difference-differential equation (see (5.3) in [9]) whereas our approach is purely algorithmic, and, as we think, an attractive alternative.

1.1 Algorithm

Our algorithm consists of two parts: firstly, define a suitable strategy, secondly, compute the necessary parameters. We are interested in the asymptotic ($n \to \infty$) version of our algorithm.

Our strategy

Our strategy, based on thresholds, is defined as follows:

we use $\kappa$ thresholds: $1 \leq j_1 < j_2 < \ldots < j_\kappa \leq n$. Following Vanderbei’s terminology, we define two kinds of candidates:

- **compulsory candidates** i.e. candidates whom we must retain, given the ranks of the candidates we have already chosen. For instance, if we have chosen so far $\ell$ candidates, ranked 1 < 2 < \ldots < $\ell$ (rank 1 corresponds to the best candidate, etc..), we must retain any candidate with rank $R$ (after relabelling) ≤ $\ell$.

- **marginal candidates** i.e. candidates whose relative ranks are less or equal to $\ell + 1$.
According to our strategy, we first choose the first record (if any) located at position \( u_1 \geq j_1 \). This is done for all possible values of \( j_1 \) where \( j_1 \) is some index between 1 and \( n \). We then compute for each possible position \( u_1 \) an optimal threshold \( j_2(u_1) \geq u_1 \). The reasoning is as follows. We have two possibilities, namely

- We observe a compulsory candidate at position \( u_2 < j_2 \). We take this candidate and we start again the strategy at position \( u_2 \), with a new threshold \( j_3(u_2) \). The strategy iterates then this procedure accordingly.
- We do not observe any compulsory candidate before \( j_2 \). Then, from \( j_2 \) on, we retain a compulsory candidate or a marginal candidate (if any) at position \( u_2 \). We now start again the strategy at position \( u_2 \), with a new threshold \( j_3(u_2) \) and, as before, we iterate the procedure.

For instance, we define a \((j_1,j_2)\)-strategy, respectively \((j_1,j_2,j_3)\)-strategy as follows

\( \kappa = 2 \)

- wait until index \( j_1 \) without any action;
- select, from \( j_1 \) onwards, the first record (if any):
  - 1. if we select only one record value up to a certain index \( j_2 \) then we accept a record value or 2-record value after \( j_2 \)
  - or
  - 2. we select two consecutive records up to \( j_2 \) (if any)

\( \kappa = 3 \)

The definition of a \((j_1,j_2,j_3)\)-strategy if \( \kappa = 3 \) is analogous.

The corresponding optimal thresholds are denoted by \( j_1^*, j_2^*, j_3^* \). The order of computation is backwards.

We can summarize our results in the following theorem

**Theorem 1.1**

- The optimal thresholds \( j_1^* \) and \( j_2^* \) for the case \( \kappa = 2 \) satisfy the asymptotic relationship

\[
j_1^*/n \rightarrow -e^{-1/2}W(-\exp(-3 + e^{1/2})) = 0.2291147286 \ldots \text{ where } W(x) \text{ is the Lambert’s function },
\]

\[
j_2^*/n \rightarrow e^{-1/2} = 0.6065306596 \ldots,
\]

and the asymptotic success probability of the \((j_1^*, j_2^*)\)-strategy equals 0.2254366561 \ldots

- In the case \( \kappa = 3 \), the corresponding asymptotic relationship is

\[
j_1^*/n \rightarrow .1666171752 \ldots,
\]

\[
j_2^*/n \rightarrow -e^{-1/3}W(-\exp(-5/2 + e^{1/3})) = .4369818602 \ldots,
\]

\[
j_3^*/n \rightarrow e^{-1/3} = .7165313106 \ldots,
\]

and the asymptotic success probability of the \((j_1^*, j_2^*, j_3^*)\)-strategy equals .1625200069 \ldots

Our computational technique can be extended to any \( \kappa \).

We will illustrate the optimal thresholds computation and analyze the performance, i.e. the success probability of our algorithm for the cases \( \kappa = 2, \kappa = 3 \).
2 The case $\kappa = 2$

Let us first compute $j_2(u_1) \geq u_1$, given $u_1$. In order to simplify our expressions, we will introduce a function $G(a, b, c)$:

$$G(a, b, c) := \prod_{k=a}^{b} \left( 1 - \frac{c}{k} \right) = \frac{(a - 1)!}{b! c}. $$

We have two possibilities: $u_1 < j_2 \leq u_2$, or, alternatively, $u_1 < u_2 < j_2$.

The probability of success of the first case is given by

$$\sum_{u_2=j_2}^{n} \prod_{k=u_1+1}^{u_2} \left( 1 - \frac{1}{k} \right) \prod_{\ell=j_2}^{u_2-1} \left( 1 - \frac{2}{\ell} \right) \frac{2}{u_2} \prod_{s=u_2+1}^{n} \left( 1 - \frac{2}{s} \right)$$

$$= \sum_{u_2=j_2}^{n} G(u_1 + 1, j_2 - 1, 1)G(j_2, u_2 - 1, 2) \frac{2}{u_2} G(u_2 + 1, n, 2)$$

$$= \sum_{u_2=j_2}^{n} \frac{u_1}{n} \frac{j_2 - 2}{u_2 - 2} n - 1 =: u_1 F_1(u_1, j_2) \text{ say} .$$

To see this note that after index $j_2$, we have two possible types of candidates, a compulsory or a marginal candidate at position $u_2$. After $u_2$, we must exclude any compulsory candidate.

We note that $u_1 F_1(u_1, j_2)$ is unimodal in $j_2$ because $u_1 F_1(u_1, j_2) - u_1 F_1(u_1, j_2 - 1)$ yields after a few steps of simplification

$$\frac{1}{n(n - 1)} 2u_1 \left( 1 - \sum_{u_2=j_2+1}^{n} \frac{1}{u_2 - 2} \right)$$

which changes sign at most once.

The probability of success of the second case is given by

$$\sum_{u_2=1}^{j_2-1} \prod_{k=u_1+1}^{u_2} \left( 1 - \frac{1}{k} \right) \frac{1}{u_2} \prod_{\ell=u_2+1}^{n} \left( 1 - \frac{2}{\ell} \right)$$

$$= \sum_{u_2=1}^{j_2-1} G(u_1 + 1, u_2 - 1, 1) \frac{1}{u_2} G(u_2 + 1, n, 2)$$

$$= \sum_{u_2=1}^{j_2-1} \frac{u_1}{n(n - 1)} =: u_1 F_2(u_1, j_2) \text{ say} .$$

Here we note that after $u_1$ and before $j_2$, we have one possible compulsory candidate at position $u_2$. After $u_2$, we must exclude any compulsory candidates. $u_1 F_2(u_1, j_2)$ is clearly monotone in $j_2$ because it is, as seen in the last sum term, affine linear in $j_2$.

Consequently $u_1 F_1(u_1, j_2) + u_1 F_2(u_1, j_2)$ is unimodal in $j_2$ because the sum of a unimodal function and an affine linear function is unimodal.

Of course, $F_2(u_1, j_2)$ is quite simple here, but, viewing general $\kappa$, we prefer to keep a general notation.

As $n \to \infty$, we use the Euler Maclaurin formula to replace sums by integrals (see also the continuous model Tamaki [8]). Moreover, we use continuous variables $u'_1 := u_1/n, u'_2 := u_2/n, \ldots$. By a slight abuse of notations, (to simplify our expressions), we will continue to use $u_1, u_2, \ldots$ as continuous time variables.

Since both parts contributing to the win probability have the same factor $u_1$, it now suffices to maximize the sum $F_1(u_1, j_2) + F_2(u_1, j_2)$, conditioning on $u_1$, that is we must maximize the asymptotic expression

$$V_1(u_1, j_2) := \int_{u_2=j_2}^{1} \frac{2 j_2}{u_2} du_2 + \int_{u_2=u_1}^{j_2} 1 du_2 = -u_1 + j_2 - 2 j_2 \ln(j_2).$$  \hspace{1cm} (1)
This gives \( j_2^* = e^{-1/2} = 0.6065306596 \ldots \).

We note that \( V_1(u_1, j_2) \) can be seen as the sum of two functions \( \varphi_1(u_1) \) and \( \varphi_2(j_2) \) namely \( \varphi_1(u_1) + \varphi_2(j_2) \). Since \( \frac{d\varphi_2(j_2)}{dj_2} = 1 - 2 \ln(j_2) - 2 \) is monotone in \( j_2 \), the function \( \varphi_2 \) is unimodal. A plot of \( \varphi_2(j_2) \) is given in Figure 1.

\[ \begin{array}{c}
\text{Figure 1: } \varphi_2(j_2) \\
\end{array} \]

We must now compute \( j_1 \). But we note that the first record can occur after \( j_2^* \). It is clear from the unimodality of \( \varphi(j_2) \), that \( j_2(u_1) \) is then exactly given by \( u_1 \). Actually

\[
j_2(u_1) = j_2^*, \quad u_1 < j_2^*,
\]

\[
= u_1, \quad u_1 \geq j_2^*.
\]

So we obtain the following success probability \( P(j_1) = P_1(j_1) + P_2(j_1) \), with, on the one hand

\[
\begin{align*}
& u_1 < j_2^* : P_1(j_1) = \sum_{u_1=j_1}^{j_2^*-1} \prod_{i=j_1}^{u_1-1} \left( 1 - \frac{1}{i} \right) \frac{1}{u_1} u_1 [F_1(u_1, j_2^*) + F_2(u_1, j_2^*)] \\
& \quad = \sum_{u_1=j_1}^{j_2^*-1} \frac{j_1 - 1}{u_1 - 1} [F_1(u_1, j_2^*) + F_2(u_1, j_2^*)], \text{ which is unimodal in } j_1,
\end{align*}
\]

and on the other hand

\[
\begin{align*}
& j_2^* \leq u_1 < u_2 : P_2(j_1) = \sum_{u_1=j_2^*}^{n-1} \sum_{u_2=u_1+1}^{n} \prod_{i=j_1}^{u_2-1} \left( 1 - \frac{1}{i} \right) \frac{1}{u_1} \prod_{\ell=u_1+1}^{u_2-1} \left( 1 - \frac{2}{\ell} \right) \frac{2}{u_2} \prod_{s=u_2+1}^{n} \left( 1 - \frac{2}{s} \right) \\
& \quad = \sum_{u_1=j_2^*}^{n-1} \sum_{u_2=u_1+1}^{n} \frac{2}{u_2 - 2} \frac{j_1 - 1}{n} \frac{1}{n - 1}, \text{ which is again monotone in } j_1.
\end{align*}
\]

Therefore \( P(j_1) = P_1(j_1) + P_2(j_1) \) is unimodal in \( j_1 \). This gives the final asymptotic success probability, namely

\[
V_2(j_1) = j_1 \int_{u_1=j_1}^{j_2^*} \frac{1}{u_1} V_1(u_1, j_2^*) du_1 + j_1 \int_{u_1=j_2^*}^{1} \int_{u_2=u_1}^{1} \frac{2}{u_2} du_2 du_1
\]
\[ j_1 = j_1 - 2 \ln(j_1) e^{-1/2} - 5e^{-1/2} + 2. \]

Maximizing \( V_2(j_1) \) then yields the asymptotically optimal threshold

\[ j_1^* = -e^{-1/2} W(-\exp(-3 + e^{1/2})) = 0.2291147286 \ldots \]

where \( W(x) \) is the Lambert function, that is the solution of

\[ W(x)e^{W(x)} = x. \]

We use its principal branch which is analytic at 0. This leads from the above to \( V_2(j_1^*) = 0 \). A plot of \( V_2(j_1) \) is given in Figure 2.

Figure 2: \( V_2(j_1) \)

We have compared, for \( n = 500 \), \( V_2(j_1) \) with \( P(j_1) \). This is shown in Figure 3 (line=\( V_2(j_1) \), box= \( P(j_1) \)). This is quite satisfactory. The relative error at \( j_1^* \) is given by \( \frac{V_2(j_1^*)}{P(j_1^*)} - 1 \).

3 The case \( \kappa = 3 \)

The strategy for general \( \kappa \) proceeds as follows. Assume that we are at time \( u \in [j_k^*, j_{k+1}^*) \),

- if we have already choosen \( k - 1 \) candidates or less, we select a marginal candidate (if any) in \( [u, j_{k+1}^*) \);
- if we have already choosen more than \( k - 1 \) candidates, we select a compulsory candidate (if any) in \( [u, j_{k+1}^*) \).

The initial condition is the following: for \( u \in [j_1^*, j_2^*) \), the marginal candidate is the first record, by definition.

Let us now consider the case \( \kappa = 3 \). We must first compute \( j_3(u_2) \geq u_2 \), given \( u_2 \). Again, we have two possibilities: \( u_2 < j_3 \leq u_3 \), or, alternatively, \( u_2 < u_3 < j_3 \). From now on, we will omit the \( G \)-function expressions.
Figure 3: Comparison, for \( n = 500 \) of \( V_2(j_1) \) (line) with \( P(j_1) \) (box).

The probability of success of the first case is given by

\[
\sum_{u_3=j_3}^{n} \prod_{s=u_2+1}^{j_3-1} \left( 1 - \frac{2}{s} \right) \prod_{t=j_3}^{u_3-1} \left( 1 - \frac{3}{t} \right) \frac{3}{u_3} \prod_{v=u_3+1}^{n} \left( 1 - \frac{3}{v} \right)
\]

\[
= \sum_{u_3=j_3}^{n} \frac{3}{u_3-3} \prod_{s=u_2+1}^{n} \left( 1 - \frac{2}{s} \right) \frac{j_3-3}{n-2} =: \prod_{s=u_2+1}^{n} \left( 1 - \frac{2}{s} \right) F_3(u_2,j_3) \text{ say,}
\]

Note that this expression is unimodal in \( j_3 \) if and only if

\[
g(j_3) := \frac{3}{n-2} \sum_{u_3=j_3}^{n} \frac{1}{u_3-3} (j_3 - 3)
\]

is unimodal in \( j_3 \). This is true if

\[
h(j) := \left( \sum_{u=j}^{n} \frac{j-3}{u-3} \right) - \left( \sum_{u=j+1}^{n} \frac{(j-3) + 1}{u-3} \right)
\]

changes, as a function of \( j \), at most once its sign. An easy calculation shows that

\[
h(j) = 1 - \sum_{u=j+1}^{n} \frac{1}{u - 3}
\]

so that \( h(j) \) is monotone increasing in \( j \).

The probability of success of the second case is given by

\[
\sum_{u_3=u_2+1}^{j_3-1} \prod_{s=u_2+1}^{u_3-1} \left( 1 - \frac{2}{s} \right) \frac{2}{u_3} \prod_{t=u_3+1}^{n} \left( 1 - \frac{3}{t} \right)
\]
\[
\sum_{u_3 = u_2 + 1}^{j_3 - 1} \frac{2}{n - 2} \prod_{s = u_2 + 1}^{n} \left(1 - \frac{2}{s}\right) =: \prod_{s = u_2 + 1}^{n} \left(1 - \frac{2}{s}\right) F_4(u_2, j_3) \text{ say,}
\]

which is clearly monotone in \(j_3\).

Asymptotically, from \(F_3(u_2, j_3) + F_4(u_2, j_3)\), conditioning on \(u_2\), we must maximize, with respect to the index \(j_3\), the function

\[
V_3(u_2, j_3) := \int_{u_3 = j_3}^{1} \frac{3 j_3}{u_3} du_3 + \int_{u_3 = u_2}^{j_3} 2du_3 = -2u_2 + 2j_3 - 3j_3 \ln(j_3).
\]

Note the similarity with \(V_1\) given in equation (1). This gives \(j_3^* = e^{-1/3} = .7165313106 \ldots\) Again we note that \(V_3(u_2, j_3)\) has the form \(\varphi_3(u_2) + \varphi_4(j_3)\), and that \(\varphi_4(j_3)\) is unimodal in \(j_3\).

We must now compute \(j_2(u_1) \geq u_1\) given \(u_1\). But we note that the relevant position \(u_2\) can occur after \(j_3^*\). It is then clear, from the unimodality of \(\varphi_4(j_3)\), that \(j_3(u_2)\) is then exactly given by \(u_2\).

Actually,

\[
j_3(u_2) = j_3^*, u_2 < j_3^*,
\]

\[
= u_2, u_2 \geq j_3^*.
\]

We have now to consider the following cases:

\[
\begin{align*}
&u_1 < j_2 \leq u_2 < j_3^*, \\
&u_1 < j_2 < j_3^* < u_2 < u_3, \\
&u_1 < u_2 \leq j_2 < j_3^*.
\end{align*}
\]

This leads respectively to

\[
P_3(u_1, j_2) = \sum_{u_2 = j_2}^{j_3 - 1} \prod_{k = u_2 + 1}^{j_3 - 1} \left(1 - \frac{1}{k}\right) \prod_{\ell = j_2}^{u_2 - 1} \left(1 - \frac{2}{\ell}\right) \prod_{s = u_2 + 1}^{n} \left(1 - \frac{2}{s}\right) [F_3(u_2, j_3^*) + F_4(u_2, j_3^*)]
\]

\[
= \sum_{u_2 = j_2}^{j_3 - 1} \frac{2}{u_2 - 2} \frac{u_1 j_2 - 2}{u_3 - 3n - 1} [F_3(u_2, j_3^*) + F_4(u_2, j_3^*)],
\]

Note the similarity with equation (2),

\[
P_4(u_1, j_2) = \sum_{u_2 = j_2^*}^{n - 1} \sum_{u_3 = u_2 + 1}^{n} \prod_{k = u_2 + 1}^{u_2 - 1} \left(1 - \frac{1}{k}\right) \prod_{\ell = j_2}^{u_2 - 1} \left(1 - \frac{2}{\ell}\right) \prod_{s = u_2 + 1}^{n} \left(1 - \frac{3}{s}\right) \prod_{t = u_3 + 1}^{n} \left(1 - \frac{3}{t}\right)
\]

\[
= \sum_{u_2 = j_2^*}^{n - 1} \sum_{u_3 = u_2 + 1}^{n} \frac{6 u_1}{u_3 - 3n - (n - 1)(n - 2)},
\]

Note the similarity with equation (3),

\[
P_5(u_1, j_2) = \sum_{u_2 = u_1 + 1}^{j_2 - 1} \prod_{k = u_2 + 1}^{u_2 - 1} \left(1 - \frac{1}{k}\right) \prod_{s = u_2 + 1}^{n} \left(1 - \frac{2}{s}\right) [F_3(u_2, j_3^*) + F_4(u_2, j_3^*)]
\]

\[
= \sum_{u_2 = u_1 + 1}^{j_2 - 1} \frac{u_1}{n(n - 1)} [F_3(u_2, j_3^*) + F_4(u_2, j_3^*)].
\]

Asymptotically, conditioning on \(u_1\), we must then maximize

\[
V_4(u_1, j_2) = \int_{u_2 = j_2}^{j_3} \frac{2 j_2}{u_2} V_3(u_2, j_3^*) du_2 + \int_{u_2 = j_2}^{j_3} \int_{u_3 = u_2}^{1} \int_{u_3 = u_2}^{j_2} \int_{u_3 = u_2}^{j_3} \frac{6 j_2}{u_3} du_3 du_2 + \int_{u_2 = u_1}^{j_2} V_3(u_2, j_3^*) du_2
\]

\[
= -3e^{-1/3} u_1 + u_1^2 + 3j_2^2 + 6j_2 - 6e^{-1/3} j_2 \ln(j_2) - 11j_2 e^{-1/3}.
\]
This gives \( j_2^* = -e^{-1/3}W(-\exp(-5/2 + e^{1/3})) \approx 0.4369818602 \ldots \) Again we note that \( V_4(u_1, j_2) \) has the form \( \varphi_5(u_1) + \varphi_6(j_2) \) and \( \varphi_6(j_2) \) is unimodal in \( j_2 \).

We now compute \( j_1 \). We note that \( u_1 \) can occur after \( j_2^* \). It is then clear, from the unimodality of \( \varphi_6(j_2) \), that \( j_2(u_1) \) is then exactly given by \( u_1 \). Actually now the split in the index \( j_2^* \) is given by

\[
\begin{align*}
j_2(u_1) &= j_2^*, u_1 < j_2^*, \\
&= u_1, u_1 \geq j_2^*.
\end{align*}
\]

We obviously need a concise notation for all cases we will consider. This is related with the Bose-Einstein statistic, where, here, we throw \( \kappa \) indistinguishable balls into \( \kappa \) ordered urns labelled 1, 2, \ldots, \( \kappa \). The number of possible cases is given by \((2^{\kappa-1})\) i.e. 10 cases for \( \kappa = 3 \). We next label the balls in increasing order. Denote then by \((\alpha_1, \alpha_2, \alpha_3)\) the urn labels of balls (1, 2, 3). In our model, balls correspond to chosen candidates and urns to intervals \([j_k, j_{k+1}] - 1\). For instance, the case \((1, 3, 3)\) corresponds to \( j_1 \leq u_1 < j_2 < j_3 \leq u_2 < u_3 \) and the corresponding probability is given by

\[
\begin{align*}
\sum_{u_1=j_1}^{j_2-1} \sum_{u_2=u_1+1}^{j_3-1} \sum_{u_3=u_2+1}^{j_4-1} \prod_{i=j_1}^{u_1-1} \left(1 - \frac{1}{i}\right) \prod_{k=u_1+1}^{j_2-1} \left(1 - \frac{1}{k}\right) \prod_{\ell=j_2}^{u_2-1} \left(1 - \frac{2}{\ell}\right) \frac{2}{u_2} \times \\
\times \prod_{s=u_2}^{u_3-1} \left(1 - \frac{3}{s}\right) \prod_{t=u_3}^{n} \left(1 - \frac{3}{t}\right).
\end{align*}
\]

We will not detail the tedious, but not particularly complicated computations corresponding to our 10 cases. We provide only the final asymptotic integrals.

For this we must maximize the success probability \( V_5(j_1) = j_1[K_1 + K_2 + K_3 + K_4 + K_5] \) with

\[
\begin{align*}
cases \ (1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 2, 2), (1, 2, 3), (1, 3, 3) : \\
K_1 &= \int_{u_1=j_1}^{j_2} \frac{1}{u_1} V_4(u_1, j_2^*) du_1, \\
case \ (2, 2, 3) : \\
K_2 &= \int_{u_1=j_1}^{j_2} \int_{u_2=u_1}^{j_3} \int_{u_3=u_2}^{j_4} 6 \frac{j_3}{u_2 u_3} du_3 du_2 du_1, \\
case \ (2, 3, 3) : \\
K_3 &= \int_{u_1=j_1}^{j_2} \int_{u_2=u_1}^{j_3} \int_{u_3=u_2}^{j_4} 6 \frac{1}{u_3} du_3 du_2 du_1, \\
case \ (2, 2, 2) : \\
K_4 &= \int_{u_1=j_1}^{j_2} \int_{u_2=u_1}^{j_3} \int_{u_3=u_2}^{j_4} 4 \frac{1}{u_2} du_3 du_2 du_1, \\
case \ (3, 3, 3) : \\
K_5 &= \int_{u_1=j_1}^{j_2} \int_{u_2=u_1}^{j_3} \int_{u_3=u_2}^{j_4} 6 \frac{1}{u_3} du_3 du_2 du_1.
\end{align*}
\]

Maximizing \( V_5(j_1) \) leads to an equation of type

\[
C_1 j_1^2 + C_2 j_1 + C_3 \ln(j_1) + C_4 = 0,
\]

where \( C_i \) depends on \( j_2^*, j_3^* \). This equation does not have an explicit solution.

But if we introduce a “generalized” Lambert function \( W_1(x, y) \) as being defined by

\[
W_1 e^{y W_1 + W_1^2} = x,
\]
we can express \( j_1^* \) as
\[
  j_1^* = W_1(x, y),
  \quad x = e^{-C_4/C_3/C_5},
  \quad y = C_2 C_5/C_3,
  \quad C_5 = \sqrt{C_3/C_1}.
\]

Numerically, this gives \( j_1^* = .1666171752 \ldots \) and the success probability \( V_5(j_1^*) = .1625200069 \ldots \) Again, \( V_5(j_1) \) is unimodal in \( j_1 \) as well as all relevant discrete expressions in this section.

We note that the easier problem of predicting an index from which onwards we have exactly three record values has an optimal value of \( 0.224042 \ldots \) See Bruss and Paindaveine [2], Sec. 6, equation (13):
\[
  V(\ell) = \frac{\ell^\ell}{\ell! e^\ell},
  \quad \ell = 3 : V(3) = \frac{9}{2} e^{-3} = 0.22404 \ldots .
\]

Hence getting the three very best is actually not that so much harder.

4 Conclusion

Using simple probabilistic arguments, we can compute a set of asymptotic thresholds and the final success probability for any \( \kappa \). Further work would be to build a computer algebra system which could mechanically compute the different cases and the corresponding asymptotic probabilities. Also two open problems remain:

PROBLEM 1: Is our procedure for finding the \( \kappa \) best secretaries out of \( n \) candidates optimal for all \( \kappa \)?

PROBLEM 2 (which is related with PROBLEM 1) : Can one bring all the relevant involved functions in a sufficient tractable form to prove the unimodality for all \( \kappa \) in the asymptotic continuous case?

These problems will be the objective of future work.

References


