An optimal algorithm for stopping on the element closest to the center of an interval

Malgorzata Sulkowska, Ewa Kubicka, Grzegorz Kubicki, Malgorzata Kuchta

Highlights

An Optimal Algorithm for Stopping on the Element Closest to the Center of an Interval

Ewa M. Kubicka[l], Grzegorz Kubicki[l], Małgorzata Kuchta[w], and Małgorzata Sulkowska[w]

- Version of the sectretary problem with the goal to stop on the central element
- Recursive construction of an optimal stopping rule
- Optimal stopping algorithm has very irregular stopping region
- Class of algorithms with rectangular stopping region and the same asymptotic behavior
- Asymptotic performance of the optimal stopping algorithm is of order $n^{-1/2}$

An Optimal Algorithm for Stopping on the Element Closest to the Center of an Interval

Ewa M. Kubicka[l], Grzegorz Kubicki[l], Małgorzata Kuchta[w], and Małgorzata Sulkowska[w]

[l] *University of Louisville, Department of Mathematics, Louisville, KY 40292, USA;*
[w] *Wrocław University of Science and Technology, Faculty of Fundamental Problems of Technology, Department of Fundamentals of Computer Science, Wrocław, Poland*

Abstract

A decision maker observes a sequence of n independent realizations from the uniform distribution on the unit interval. However, he does not observe the precise values of these realizations, but only their ranks relative to those that have appeared previously. The goal of the decision maker is to select the realization whose value is closest to $\frac{1}{2}$. A realization can only be selected at the moment of its appearance. We derive a stopping rule which maximizes the probability of achieving this goal, together with the asymptotic probability of success.

Keywords:
Combinatorial optimization, Optimal stopping, Secretary problem.

---

1. Introduction

In many real life situations, for example when controlling the rate of inflation, finding a safe investment with a satisfactory return, or maintaining an appropriate level of sugar or minerals in one's body, we want to be as far as possible from extreme values and therefore the most desirable choice is the value closest to the middle. This article considers a problem of this form where a decision maker must make an online decision within a fixed discrete time horizon. The decision maker cannot measure observations precisely, but is able to rank each successive observation with respect to the previous ones. This manuscript derives an optimal strategy for selecting the most central observation.

More precisely, consider the following online problem: n numbers, labelled $x_1, x_2, \ldots, x_n$, are randomly selected from the interval $[0, 1]$. A decision maker cannot observe these values precisely, but on observing x_k, $1 \leq k \leq n$, can rank the first k observations. We ignore the possibility that some of these numbers are equal (this event has a probability of zero). Our goal is to stop on the appearance of the number x_k which is the closest to $\frac{1}{2}$, the center of the interval, amongst all of the n numbers. We will construct an optimal stopping algorithm and show that for large values of n the probability of success under this algorithm is of order $\frac{1}{\sqrt{n}}\sqrt{\frac{2}{\pi}}$.

This problem is a new relative of the classical secretary problem. In the classical secretary problem, the goal is to choose the best of n linearly ordered objects. In our model, this corresponds to choosing the object whose value is closest to 1. The classical secretary problem, whose solution was derived by Lindley [8], has attracted a lot of attention and various modifications have been considered. A thorough review of this research is given by Ferguson [2]. Many generalizations of the classical problem have been studied, for example problems in which linear orders have been replaced by partial orders (Morayne [9]; Preater [10]; Freij & Wästlund [3]; Georgiou, Kuchta, Morayne & Niemiec [4]; Stadje [13]), or by a graph or digraph structure (Kubicki & Morayne [7]; Goddard, Kubicka, & Kubicki [6]; Sulkowska [14]; Benevides & Sulkowska [1]). The optimal solution of the classical secretary problem itself was also deeply analyzed. E.g., Rogerson [11] derived the probability that the optimal algorithm for choosing the best candidate returns j^{th} candidate.

Nevertheless, there are still very natural questions referring to the classical secretary problem that remain unanswered. The optimal algorithm for choosing online the k^{th} candidate (instead of the best one) out of n linearly ordered is still not known. The case $k = 1$ is the classical secretary problem and the case $k = 2$, known as the postdoc problem, was solved independently by Rose [12] and Vanderbei [15]. The solution for any $k \geq 3$ (to the best of our knowledge) has not been given yet. It seems that the case when $k = \lfloor n/2 \rfloor$ (i.e., choosing the middle rank element) is the hardest one. The problem we study in this paper is similar to choosing the middle rank element, but it is not exactly the same. In our case, stopping on the element of middle rank does not guarantee that this number would be closest to $\frac{1}{2}$. Also, stopping on elements other than the one of middle rank gives a nonzero probability of success. Another difference is that since we refer to a specific value, namely $\frac{1}{2}$, we have to make some assumption about the distribution of incoming numbers whose ranks we observe. The most natural one seems to be the uniform distribution.

This paper is organized as follows. In Section 2, using recursion, we construct an optimal stopping algorithm and derive a formula for the probability of success. Unsurprisingly, this algorithm prescribes stopping only on numbers which do not appear very early and have ranks not far from the middle. We provide an example of how this algorithm works for $n = 10$ and what the stopping region looks like. The asymptotic performance of the algorithm is analyzed in Section 3. First, we construct an algorithm that is not optimal, but has a more regular stopping region. This enables us to estimate the asymptotic performance of our algorithm from below. Then we consider a slightly easier problem for which it is simple to calculate the asymptotic performance of the optimal strategy. This provides an upper bound. Since these bounds are identical, this proves that the asymptotic probability of success under the optimal stopping algorithm is of order $\frac{1}{\sqrt{n}}\sqrt{\frac{2}{\pi}}$.

2. Optimal Stopping Algorithm

Assume that n numbers $x_1, x_2, ..., x_n$ are chosen from the uniform distribution on the interval [0, 1] and presented to a decision maker in sequence. The decision maker knows n in advance, but after observing the first t numbers, $1 \leq t \leq n$, knows only their relative ranks, not their values. Let us relabel these numbers such that, at moment t, $y_1^{(t)} < y_2^{(t)} < ... < y_t^{(t)}$. Assume that the rank of x_t is r, i.e. $x_t = y_r^{(t)}$. Our goal is to stop on the number x_t such that $|x_t - \frac{1}{2}| \leq |x_i - \frac{1}{2}|$ for all i, $1 \leq i \leq n$, i.e. x_t is the closest to the midpoint of the interval. We will call such an event "x_t is the best".

Before constructing the optimal stopping algorithm (denoted by $\mathcal{A}_n$), we need two results providing formulas for the probability that a number with a specific rank is the best.

Theorem 2.1. *If $y_1 < y_2 < ... < y_r < ... < y_n$ are the ranked numbers at time n, then*

$$Pr(y_r \text{ is the best}) = \binom{n-1}{r-1} \cdot \frac{1}{2^{n-1}}.$$

Proof. We have

$$\begin{aligned}\Pr(y_r \text{ is the best}) &= \Pr\Big(\Big(y_r < \frac{1}{2} < y_{r+1}\Big) \text{ and } \Big(|y_r - \frac{1}{2}| \leq |y_{r+1} - \frac{1}{2}|\Big)\Big)\\ &+ \Pr\Big(\Big(y_{r-1} < \frac{1}{2} < y_r\Big) \text{ and } \Big(|y_{r-1} - \frac{1}{2}| \geq |y_r - \frac{1}{2}|\Big)\Big).\end{aligned}$$

If Z_i denotes the distance between y_i and $1/2$, then $Z_1, Z_2, \ldots, Z_n$ are independent random variables drawn from the uniform distribution on the interval $[0, 1/2]$. Therefore,

$$\begin{aligned}&\Pr\Big(\Big(|y_r - \frac{1}{2}| \leq |y_{r+1} - \frac{1}{2}|\Big)\Big|\Big(y_r < \frac{1}{2} < y_{r+1}\Big)\Big)\\ &\qquad = \Pr\Big(\min\{Z_1, Z_2, \ldots, Z_r\} < \min\{Z_{r+1}, Z_{r+2}, \ldots, Z_n\}\Big) = \frac{r}{n}.\end{aligned}$$

Analogously, we obtain

$$\Pr\Big(\Big(|y_{r-1} - \frac{1}{2}| \geq |y_r - \frac{1}{2}|\Big)\Big|\Big(y_{r-1} < \frac{1}{2} < y_r\Big)\Big) = \frac{n-r+1}{n}$$

and, finally,

$$\begin{aligned}\Pr(y_r \text{ is the best}) &= \binom{n}{r} \cdot \frac{1}{2^n} \cdot \frac{r}{n} + \binom{n}{r-1} \cdot \frac{1}{2^n} \cdot \frac{n-r+1}{n}\\ &= \frac{1}{2^n}\left[\frac{(n-1)!}{(r-1)!(n-r)!} + \frac{(n-1)!}{(r-1)!(n-r)!}\right] = \binom{n-1}{r-1} \cdot \frac{1}{2^{n-1}}.\end{aligned}$$

□

Theorem 2.2. *If $y_1^{(t)} < y_2^{(t)} < ... < y_r^{(t)} < ... < y_t^{(t)}$ are the ranked numbers at time t, then*

$$Pr(y_r^{(t)} \text{ will be the best}) = \frac{1}{2^{n-1}} \sum_{j=0}^{n-t} \binom{n-1}{r-1+j} \binom{n-t}{j} \frac{r^j (t+1-r)^{n-t-j}}{(t+1)^{n-t}}. \tag{1}$$

Proof. Since $n-t$ additional numbers will appear, the rank r of the number $y_r^{(t)}$ will increase by some j, where $0 \le j \le n-t$. Each number following $y_r^{(t)}$ will fall into one of the intervals $(0,\ y_1^{(t)})$, $(y_1^{(t)},\ y_2^{(t)})$, ... , $(y_t^{(t)},\ 1)$, independently and with the same probability $\frac{1}{t+1}$. Every time a number falls into one of the first r intervals, the rank of $y_r^{(t)}$ increases by 1. Therefore, the probability that after the appearance of all n numbers, the rank of $y_r^{(t)}$ will be $r+j$ is $\binom{n-t}{j} \frac{r^j (t+1-r)^{n-t-j}}{(t+1)^{n-t}}$. Thus, from Theorem 2.1,

$$\text{Pr } (y_r^{(t)} \text{ will be the best} \mid \text{its rank is } r+j) = \binom{n-1}{r-1+j} \frac{1}{2^{n-1}}$$

and formula (1) follows from the law of total probability. □

From now on, $\Pr(y_r^{(t)}$ will be the best) will be abbreviated to $\mathrm{P}_r^{(t)}$. Also, we denote the optimal algorithm from the set of algorithms that stop only in rounds $t,\ t+1, ..., n-1$, or n by $\mathcal{A}_n^{(t)}$ (i.e. such algorithms never stop before time t). We now construct an optimal stopping algorithm $\mathcal{A}_n$ using recursion. Note that $\mathcal{A}_n = \mathcal{A}_n^{(1)}$.

$\mathcal{A}_n^{(n)}$ is the algorithm that stops only on the number that comes in the last round, thus $\Pr(\mathcal{A}_n^{(n)}$ succeeds) $= \frac{1}{n}$. Algorithm $\mathcal{A}_n^{(n-1)}$ stops only in rounds $n-1$ or n. Therefore, it stops on number $y_r^{(n-1)}$ in the $(n-1)^{\text{th}}$ round if and only if $\mathrm{P}_r^{(n-1)} \ge \frac{1}{n}$. Using the formula from Theorem 2.2 with $t = n-1$, we obtain the inequality $\frac{1}{2^{n-1}} \left[\binom{n-1}{r-1} \frac{n-r}{n} + \binom{n-1}{r} \frac{r}{n} \right] \ge \frac{1}{n}$, which is equivalent to

$$\binom{n-2}{r-1} \ge \frac{2^{n-2}}{n-1}.$$

Solving this inequality for $r-1$ gives a symmetric interval from the $(n-2)^{\text{th}}$ row of the Pascal triangle, namely $r-1 \in [z_1,\ n-2-z_1]$ for some z_1, or,

setting $r_1 = z_1 + 1$, $r \in [r_1,\ n - r_1]$.
Therefore, algorithm $\mathcal{A}_n^{(n-1)}$ stops in round $n-1$ if and only if the rank of the number that appears in that round is from the stopping interval $[r_1,\ n-r_1]$. Of course,

$$\Pr(\mathcal{A}_n^{(n-1)} \text{ succeeds}) = \sum_{r=r_1}^{n-r_1} \frac{1}{n-1} \mathrm{P}_r^{(n-1)} + \frac{2(r_1 - 1)}{n-1} \frac{1}{n},$$

where the two terms represent the probabilities of winning if the rank of the $(n-1)^{\text{th}}$ number is in $[r_1,\ n-r_1]$ or is outside of that interval, respectively.

In general, assume that for $k = t+1, t+2, \ldots, n$ we know the probabilities $\Pr(\mathcal{A}_n^{(k)}$ succeeds) and the stopping region in round k, the interval $[r_{n-k},\ k+1-r_{n-k}]$. The optimal algorithm $\mathcal{A}_n^{(t)}$ stops on the number $y_r^{(t)}$ in round t if and only if its rank r satisfies the inequality

$$\mathrm{P}_r^{(t)} \geq \Pr(\mathcal{A}_n^{(t+1)} \text{ succeeds}). \tag{2}$$

If inequality (2) has a solution, then the solution set, which is a symmetric interval $[r_{n-t},\ t+1-r_{n-t}]$, is the stopping region for $\mathcal{A}_n^{(t)}$ in round t and

$$\Pr(\mathcal{A}_n^{(t)} \text{ succeeds}) = \sum_{r=r_{n-t}}^{t+1-r_{n-t}} \frac{1}{t} \mathrm{P}_r^{(t)} + \frac{2(r_{n-t} - 1)}{t} \Pr(\mathcal{A}_n^{(t+1)} \text{ succeeds}). \tag{3}$$

If there is no r satisfying inequality (2), then the algorithm $\mathcal{A}_n^{(t)}$ never stops in round t and $\Pr(\mathcal{A}_n^{(t)}$ succeeds) $= \Pr(\mathcal{A}_n^{(t+1)}$ succeeds). Recall that the optimal algorithm for our decision problem is $\mathcal{A}_n = \mathcal{A}_n^{(1)}$. This optimal strategy and the corresponding set of $\Pr(\mathcal{A}_n^{(t)}$ succeeds), $t = 1, 2, \ldots, n$ can be calculated by recursion based on Equation (3).

Algorithm 1: Implementation of $\mathcal{A}_n^{(1)}$ - the optimal strategy

Data: $x_1, x_2, \ldots, x_n$ - numbers chosen uniformly at random from $[0, 1]$; $P_r^{(t)}$ for $t = 1, 2, ..., n-1$ as calculated from Equation (1); $\Pr(\mathcal{A}_n^{(t)}$ succeeds) for $t = 2, 3, ..., n$ calculated recursively as described above.

Result: candidate for the number closest to 1/2 among $x_1, x_2, \ldots, x_n$ via online search.

```
begin
    for t = 1, 2, ..., n − 1 do
        r := rank of element x_t based on x_1, ..., x_t (x_t = y_r^(t))
        if rank r satisfies P_r^(t) ≥ Pr(A_n^(t+1) succeeds) then
            return x_t
    return x_n
```

Recall that $y_1^{(t)} < y_2^{(t)} < ... < y_t^{(t)}$ are the ordered values of $x_1, x_2, \ldots, x_t$, i.e., the ordering of the numbers appearing not later than moment t.

Implementation of the algorithm $\mathcal{A}_n$ is straightforward and our next example illustrates what the optimal stopping strategy looks like for $n = 10$.

Example. The optimal algorithm $\mathcal{A}_{10}$ never stops in rounds 1, 2, and 4. It stops in round 3 only on a number which has current rank 2. The stopping region is shaded in Figure 1. The number in bold in Table 1 is $\Pr(\mathcal{A}_{10}^{(1)}$ succeeds), the probability of success under $\mathcal{A}_{10}$.

t	$10-t$	r_{10-t}	stopping interval	$Pr(\mathcal{A}_{10}^{(t)}$ succeeds)
1	9			**0.1893**
2	8			0.1893
3	7	2	$\{2\}$	0.1893
4	6			0.1858
5	5	3	$\{3\}$	0.1858
6	4	3	$[3,4]$	0.1798
7	3	3	$[3,5]$	0.1701
8	2	4	$[4,5]$	0.1585
9	1	4	$[4,6]$	0.1378
10	0	1	$[1,10]$	0.1

Table 1: Stopping intervals at time t and probabilities that the algorithm $\mathcal{A}_n^{(t)}$ succeeds for $n = 10$.

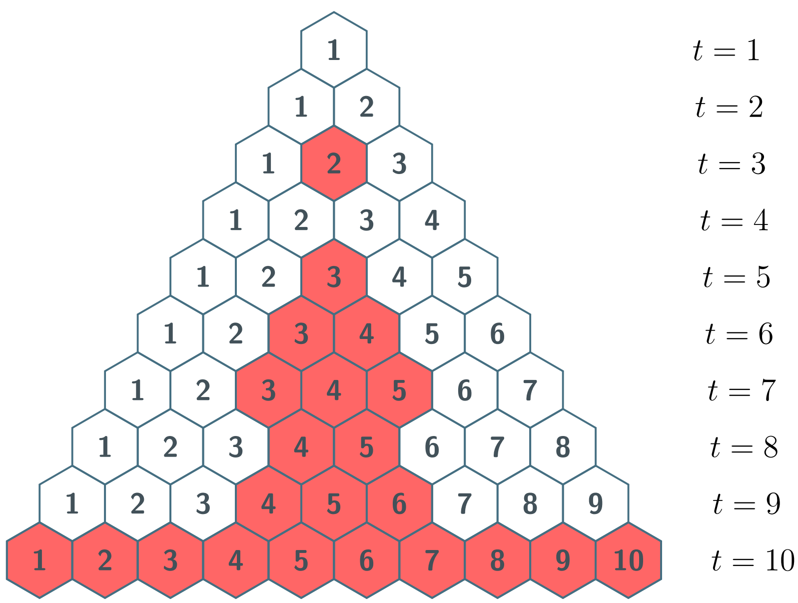

Figure 1. The stopping regions for the optimal algorithm $\mathcal{A}_{10}$.

As can be seen from this example, the stopping region for our algorithm $\mathcal{A}_n$ is rather irregular and the recursive formulas used to calculate

$\Pr(\mathcal{A}_n$ succeeds) give little hope of finding a closed formula for this probability. Despite these difficulties, in the next section we will derive the asymptotic performance of the optimal algorithm $\mathcal{A}_n$.

3. Asymptotics

Throughout this section we use the standard notation:

$$f(n) \sim g(n) \text{ if } \tfrac{f(n)}{g(n)} \xrightarrow[n\to\infty]{} 1 \text{ and } f(n) = o(g(n)) \text{ if } \tfrac{f(n)}{g(n)} \xrightarrow[n\to\infty]{} 0.$$

Also, when z is not a natural number, the binomial coefficient $\binom{n}{z}$ is defined as

$$\binom{n}{z} = \frac{\Gamma(n+1)}{\Gamma(z+1)\Gamma(n-z+1)},$$

where $\Gamma(z)$ is the gamma function.

The example from the previous section for $n = 10$ might be misleading, because for large values of n the stopping region under the optimal algorithm is relatively small. Based on computer simulations carried out for $n \leq 5000$, we found that, for large values of n, the algorithm $\mathcal{A}_n$ does not stop until it reaches round $\lceil n - n^{2/3}\sqrt{\ln n}\rceil$ and in round $n-1$ stops only on elements whose ranks are close to the middle. In fact, for large values of n, $r_1 \sim \frac{n}{2} - \frac{1}{2}\sqrt{n \ln \frac{2n}{\pi}}$. This result is proved in Corollary 3.5. Prior to this proof, we need several simple lemmas.

Lemma 3.1. *For large values of n, $\binom{n}{n/2} \sim \frac{\sqrt{2}\cdot 2^n}{\sqrt{\pi n}}$.*

This formula follows easily from Stirling's approximation.

Lemma 3.2. *If $g(n) \xrightarrow[n\to\infty]{} \infty$ and $g(n) = o(f(n))$, then*

$$\left[\frac{\left(1+\frac{1}{f(n)}\right)^{f(n)}}{e}\right]^{g(n)} \xrightarrow[n\to\infty]{} 1.$$

Proof. Since

$$1 \geq \left[\frac{\left(1+\frac{1}{f(n)}\right)^{f(n)}}{e}\right]^{g(n)} \geq \left[\left(1+\frac{1}{f(n)}\right)^{f(n)}\left(1-\frac{1}{f(n)}\right)^{f(n)}\right]^{g(n)} = \left[\left(1-\frac{1}{f^2(n)}\right)^{f^2(n)}\right]^{\frac{g(n)}{f(n)}},$$

the result follows from the sandwich theorem since the lower bound approaches 1. $\square$

Lemma 3.3. *If $s = s(n)$ and $w = w(n)$ are positive sequences such that $s(n) \xrightarrow[n\to\infty]{} \infty$ and $w = o\left(s^{2/3}\right)$, then $\frac{\binom{2s}{s}}{\binom{2s}{s-w}} \sim e^{\frac{w^2}{s}}$.*

Proof. The ratio $\frac{\binom{2s}{s}}{\binom{2s}{s-w}}$ simplifies to

$$\frac{(s+1)(s+2)...(s+w-1)(s+w)}{(s-w+1)(s-w+2)...(s-1)s} = (1+\frac{w}{s-w+1})(1+\frac{w}{s-w+2})...(1+\frac{w}{s}).$$

Therefore, $(1+\frac{w}{s})^w \leq \frac{\binom{2s}{s}}{\binom{2s}{s-w}} \leq (1+\frac{w}{s-w+1})^w$ or, equivalently,

$$\frac{[(1+\frac{1}{s/w})^{s/w}]^{w^2/s}}{e^{w^2/s}} \leq \frac{\frac{\binom{2s}{s}}{\binom{2s}{s-w}}}{e^{w^2/s}} \leq \frac{[(1+\frac{1}{(s-w+1)/w})^{\frac{s-w+1}{w}}]^{\frac{w^2}{s-w+1}}}{e^{w^2/s}}.$$

From the assumption that $w = o\left(s^{2/3}\right)$, we get $\frac{w}{s^{2/3}} \to 0$, so $\frac{w^2}{s^{4/3}} \to 0$ implying that $\frac{w^2}{s} = o\left(s^{1/3}\right)$. From $\frac{w}{s^{2/3}} = \frac{s^{1/3}}{s/w} \to 0$, we obtain $s^{1/3} = o\left(\frac{s}{w}\right)$. Therefore, $\frac{w^2}{s} = o\left(\frac{s}{w}\right)$ and applying Lemma 3.2 with $f(n) = \frac{s}{w}$ and $g(n) = \frac{w^2}{s}$, we conclude that both lower and upper bounds approach 1. $\square$

From Lemma 3.1 and Lemma 3.3, we immediately obtain the following result.

Corollary 3.4. *If $s = s(n) \xrightarrow[n\to\infty]{} \infty$ and $w = o\left(s^{2/3}\right)$, then*

$$\binom{s}{\frac{s}{2}-w} \sim \frac{\sqrt{2}\cdot 2^s}{\sqrt{\pi s}\cdot e^{\frac{2w^2}{s}}}.$$

Corollary 3.5. *Under the assumption that $r \leq \frac{n}{2}$, the asymptotic solution of the inequality $\binom{n-2}{r-1} \geq \frac{2^{n-2}}{n-1}$ is $r \geq \frac{n}{2} - \frac{1}{2}\sqrt{n \ln \frac{2n}{\pi}}$.*

Proof. Let $s = n-2$ and $r-1 = \frac{s}{2} - w$. We want to find w for which $\binom{s}{\frac{s}{2}-w} \geq \frac{2^s}{s+1}$. Assuming that $w = o(s^{2/3})$ and using Corollary 3.4, we get the inequality

$$\frac{\sqrt{2}\cdot 2^s}{\sqrt{\pi s}\cdot e^{\frac{2w^2}{s}}} \geq \frac{2^s}{s+1}, \text{ which is equivalent to } e^{\frac{2w^2}{s}} \leq \frac{\sqrt{2}(s+1)}{\sqrt{\pi s}}.$$

Hence, $\frac{2w^2}{s} \leq \ln \frac{\sqrt{2}(s+1)}{\sqrt{\pi s}}$ or, equivalently, $w \leq \sqrt{\frac{s}{2}} \cdot (\ln \frac{\sqrt{2}(s+1)}{\sqrt{\pi s}})^{1/2} \sim \frac{1}{2}\sqrt{s \ln \frac{2s}{\pi}}$. Therefore, $r = \frac{s}{2} + 1 - w \geq \frac{n}{2} - \frac{1}{2}\sqrt{n \ln \frac{2n}{\pi}}$. $\square$

The rest of this section covers the derivation of the exact asymptotics of the probability that $\mathcal{A}_n$ succeeds. First, we define the algorithm $\mathcal{A}(h_n, w_n)$, which is not optimal, but has a more regular stopping region than the optimal algorithm $\mathcal{A}_n$. This will be helpful in finding a reasonable lower bound for the performance of $\mathcal{A}_n$.

The stopping region of the algorithm $\mathcal{A}(h_n, w_n)$ is defined by two natural numbers h_n and w_n. This algorithm never stops before time h_n. For $t \geq h_n$, it stops on x_t if and only if x_t falls between $y^{(t-1)}_{\lceil \frac{t}{2} \rceil - w_n}$ and $y^{(t-1)}_{\lfloor \frac{t}{2} \rfloor + w_n}$, where $y^{(t-1)}_1 < y^{(t-1)}_2 < ... < y^{(t-1)}_{t-1}$ are the ordered numbers at time $t-1$. If this never happens, $\mathcal{A}(h_n, w_n)$ stops at x_n.

Algorithm 2: Implementation of $\mathcal{A}(h_n, w_n)$

Data: $x_1, x_2, \ldots, x_n$ - numbers chosen uniformly at random from $[0, 1]$; h_n, w_n - parameters describing the stopping region

Result: candidate for the number closest to $1/2$ among $x_1, x_2, \ldots, x_n$ via online search

begin
- **for** $t = 1, 2, \ldots, h_n - 1$ **do**
  - rank x_t with respect to previous observations
- **for** $t = h_n, h_n + 1, \ldots, n - 1$ **do**
  - rank x_t with respect to previous observations
  - **if** $x_t \in [y^{(t-1)}_{\lceil \frac{t}{2} \rceil - w_n}, y^{(t-1)}_{\lfloor \frac{t}{2} \rfloor + w_n}]$ **then**
    - **return** x_t
- **return** x_n

Recall that $y^{(t)}_1 < y^{(t)}_2 < \ldots < y^{(t)}_t$ are the ordered values of $x_1, x_2, \ldots, x_t$, i.e., the ordering of the numbers appearing not later than moment t.

The following technical lemma will be used in the next theorem to estimate the performance of $\mathcal{A}(h_n, w_n)$.

Lemma 3.6. *Let* $a_s = \frac{1}{2^s}\binom{s}{\frac{s}{2} - w}$. *For* $m > 2w^2 - 1$, *the sequences* $\{a_{2m}\}_{m \geq 0}$ *and* $\{a_{2m+1}\}_{m \geq 0}$ *are decreasing.*

Proof. Consider the sequence $\{a_{2m+1}\}_{m \geq 0}$. Using the fact that $\Gamma(z+1) = z\Gamma(z)$, we obtain

$$\begin{aligned}
\frac{a_{2m+3}}{a_{2m+1}} &= \frac{2^{2m+1}}{2^{2m+3}} \cdot \frac{\binom{2m+3}{\frac{2m+3}{2}-w}}{\binom{2m+1}{\frac{2m+1}{2}-w}} \\
&= \frac{1}{4} \cdot \frac{\Gamma(2m+4)}{\Gamma(m-w+\frac{5}{2})\Gamma(m+w+\frac{5}{2})} \cdot \frac{\Gamma(m-w+\frac{3}{2})\Gamma(m+w+\frac{3}{2})}{\Gamma(2m+2)} \\
&= \frac{1}{4} \cdot \frac{(2m+3)(2m+2)}{(m-w+\frac{3}{2})(m+w+\frac{3}{2})} = \frac{2m^2+5m+3}{2m^2+6m-2w^2+\frac{9}{2}}.
\end{aligned}$$

It follows that the ratio $\frac{a_{2m+3}}{a_{2m+1}}$ is smaller than 1 and thus the sequence $\{a_{2m+1}\}_{m\geq 0}$ is decreasing when $m > 2w^2 - \frac{3}{2}$.

The fact that the sequence $\{a_{2m}\}_{m\geq 0}$ is decreasing for $m > 2w^2 - 1$ can be shown in an analogous manner. □

Figure 2 illustrates the rectangular stopping region for the algorithm $\mathcal{A}(h_n, w_n)$. Note that $n - h_n + 1$ and $2w_n$ can be interpreted as the height and width of this stopping region, respectively.

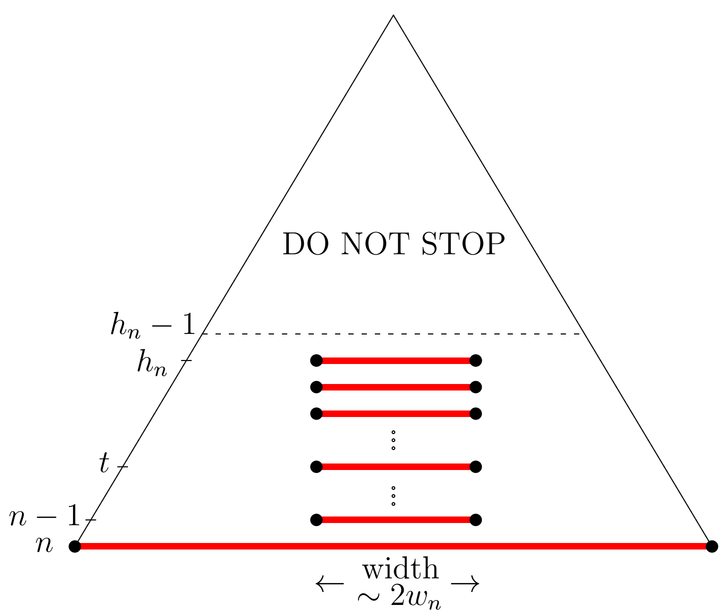

Figure 2. The stopping region for the algorithm $\mathcal{A}(h_n, w_n)$.

Corollary 3.7. *Let s, w, and n be natural numbers such that $n > s$ and $s > 4w^2$. Then*

$$\frac{1}{2^{s-1}}\binom{s-1}{\lceil \frac{s}{2}\rceil - w} \geq \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2} - w}.$$

Proof. Whenever n and s are both odd or both even, from Lemma 3.6 we immediately obtain

$$\frac{1}{2^{s-1}}\binom{s-1}{\lceil \frac{s}{2}\rceil - w} \geq \frac{1}{2^{s-1}}\binom{s-1}{\frac{s-1}{2} - w} \geq \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2} - w}.$$

Now, assume that s is even and n is odd. Then, from Lemma 3.6, since $n > s$

$$\frac{1}{2^{s-1}}\binom{s-1}{\lceil \frac{s}{2}\rceil - w} = \frac{1}{2^{s-1}}\binom{s-1}{\frac{s}{2} - w} \geq \frac{1}{2^{s-1}}\cdot\frac{1}{2}\cdot\binom{s}{\frac{s}{2} - w} \geq \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2} - w}.$$

Finally, assume that s is odd and n is even. Then, analogously

$$\begin{aligned}\frac{1}{2^{s-1}}\binom{s-1}{\lceil \frac{s}{2}\rceil - w} &= \frac{1}{2^{s-1}}\binom{s-1}{\frac{s+1}{2} - w} \geq \frac{1}{2^{s-1}}\cdot\frac{1}{2}\cdot\binom{s}{\frac{s+1}{2} - w} \\ &\geq \frac{1}{2^{s}}\cdot\binom{s}{\frac{s}{2} - w} \geq \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2} - w}.\end{aligned}$$

□

Theorem 3.8. *For any sequences h_n and w_n of natural numbers such that $h_n \leq n$ and $4w_n^2 < n$, we have*

$$Pr(\mathcal{A}(h_n, w_n) \text{ succeeds}) \geq v(h_n, w_n),$$

where $v(h_n, w_n)$ is a function such that for $w_n \xrightarrow[n\to\infty]{} \infty$

$$v(h_n, w_n) \sim \frac{h_n}{n2^{n-1}}\binom{n-1}{\frac{n-1}{2} - w_n}\cdot\left(1 - \left(1 - \frac{2w_n}{h_n}\right)^{n-h_n}\right).$$

Proof. For $s \in \{1, 2, \ldots, n\}$, let B_s be the event that the best element arrives at time s; this means that x_s is the closest element to $\frac{1}{2}$. Of course,

$\Pr(B_s) = \frac{1}{n}$. Hence,

$$
\begin{aligned}
\Pr(\mathcal{A}(h_n, w_n) \text{ succeeds}) &= \sum_{s=1}^{n} \Pr(\mathcal{A}(h_n, w_n) \text{ succeeds } \big| B_s) \cdot \Pr(B_s) \\
&= \frac{1}{n} \sum_{s=h_n}^{n} \Pr(\mathcal{A}(h_n, w_n) \text{ succeeds } \big| B_s),
\end{aligned}
$$

since our algorithm never stops before time h_n.

In order for $\mathcal{A}(h_n, w_n)$ to succeed, the numbers $x_{h_n}, x_{h_n+1}, ..., x_{s-1}$ must fall outside of the stopping intervals and the number x_s must fall into the interval $\big[y^{(s-1)}_{\lceil \frac{s}{2} \rceil - w_n}, y^{(s-1)}_{\lfloor \frac{s}{2} \rfloor + w_n}\big]$. These events are independent and

$$
\Pr\big(x_t \text{ falls outside } \big[y^{(t-1)}_{\lceil \frac{t}{2} \rceil - w_n}, y^{(t-1)}_{\lfloor \frac{t}{2} \rfloor + w_n}\big] \big| B_s\big) = 1 - \tfrac{1}{t}\big(\lfloor \tfrac{t}{2} \rfloor + w_n - \lceil \tfrac{t}{2} \rceil + w_n\big),
$$

because the expression in parenthesis (which is at most $2w_n$) counts the number of intervals (out of t intervals) that x_t cannot fall into (due to the stopping criterion). It follows that

$$
\Pr\big(x_t \text{ falls outside } \big[y^{(t-1)}_{\lceil \frac{t}{2} \rceil - w_n}, y^{(t-1)}_{\lfloor \frac{t}{2} \rfloor + w_n}\big] \big| B_s\big) \geq 1 - \frac{2w_n}{t}. \tag{4}
$$

For the event that x_s falls into the interval $\big[y^{(s-1)}_{\lceil \frac{s}{2} \rceil - w_n}, y^{(s-1)}_{\lfloor \frac{s}{2} \rfloor + w_n}\big]$ to occur given B_s, note that all of the $s-1$ numbers that came before x_s must be outside the interval $\big[\frac{1}{2} - d, \frac{1}{2} + d\big]$, where $d = \big|x_s - \frac{1}{2}\big|$. Suppose j numbers out of $s-1$ fall into the interval $\big[0, \frac{1}{2} - d\big)$. In order for x_s to have a rank such that the algorithm will stop on it, j must satisfy the inequality $\lceil \frac{s}{2} \rceil - w_n \leq j \leq \lfloor \frac{s}{2} \rfloor + w_n - 1$. Therefore,

$$
\begin{aligned}
\Pr\big(x_s \text{ falls into } \big[y^{(s-1)}_{\lceil \frac{s}{2} \rceil - w_n}, y^{(s-1)}_{\lfloor \frac{s}{2} \rfloor + w_n}\big] \big| B_s\big) &= \sum_{j=\lceil \frac{s}{2} \rceil - w_n}^{\lfloor \frac{s}{2} \rfloor + w_n - 1} \binom{s-1}{j} \frac{1}{2^{s-1}} \\
&\geq \frac{1}{2^{s-1}} \cdot (2w_n - 1) \cdot \binom{s-1}{\lceil \frac{s}{2} \rceil - w_n}.
\end{aligned} \tag{5}
$$

Using inequalities (4) and (5), we obtain

$$\begin{aligned}\Pr\big(\mathcal{A}(h_n,w_n)\text{ succeeds}\big) &\geq \frac{1}{n}\sum_{s=h_n}^{n}\left(\prod_{t=h_n}^{s-1}(1-\frac{2w_n}{t})\right)\frac{2w_n-1}{2^{s-1}}\binom{s-1}{\lceil\frac{s}{2}\rceil-w_n}\\ &\geq \frac{1}{n}\frac{2w_n-1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2}-w_n}\sum_{s=h_n}^{n}(1-\frac{2w_n}{h_n})^{s-h_n},\end{aligned}$$

where the last inequality uses Corollary 3.7 regarding the monotonicity of $\frac{1}{2^{s-1}}\binom{s-1}{\lceil\frac{s}{2}\rceil-w_n}$. After changing the index of summation, we obtain

$$\begin{aligned}\Pr\big(\mathcal{A}(h_n,w_n)\text{ succeeds}\big) &\geq \frac{2w_n-1}{n2^{n-1}}\binom{n-1}{\frac{n-1}{2}-w_n}\sum_{s=0}^{n-h_n}\Big(1-\frac{2w_n}{h_n}\Big)^{s}\\ &= \frac{2w_n-1}{n2^{n-1}}\binom{n-1}{\frac{n-1}{2}-w_n}\frac{1}{\frac{2w_n}{h_n}}\left(1-\Big(1-\frac{2w_n}{h_n}\Big)^{n-h_n+1}\right)\\ &= v(h_n,w_n).\end{aligned}$$

Whenever $w_n \xrightarrow[n\to\infty]{} \infty$, we obtain

$$v(h_n,w_n)\sim\frac{h_n}{n2^{n-1}}\binom{n-1}{\frac{n-1}{2}-w_n}\left(1-\Big(1-\frac{2w_n}{h_n}\Big)^{n-h_n}\right).$$

□

Corollary 3.9. *If $h_n=\lceil n\big(1-\frac{\sqrt{\ln(n)}}{n^{1/3}}\big)\rceil$ and $w_n=\lceil n^{1/3}\rceil$, then*

$$\sqrt{n}\ Pr\big(\mathcal{A}(h_n,w_n)\ \textit{succeeds}\big)\geq\sqrt{n}\ v(h_n,w_n)\xrightarrow[n\to\infty]{}\sqrt{\frac{2}{\pi}},$$

and thus the inequality $Pr\big(\mathcal{A}(h_n,w_n)\ \textit{succeeds}\big)\geq\sqrt{\frac{2}{\pi n}}$ holds for sufficiently large n.

Proof. From Theorem 3.8, it follows that for sufficiently large n

$$\sqrt{n}\ \Pr\big(\mathcal{A}(h_n,w_n)\text{ succeeds}\big)\geq\sqrt{n}\ v(h_n,w_n)$$

and

$$\sqrt{n}\ v(h_n,w_n)\sim\sqrt{n}\frac{h_n}{n2^{n-1}}\binom{n-1}{\frac{n-1}{2}-w_n}\left(1-\Big(1-\frac{2w_n}{h_n}\Big)^{n-h_n}\right).$$

The last factor can be rewritten as

$$1-\left[\left(1-\frac{1}{\frac{h_n}{2w_n}}\right)^{\frac{h_n}{2w_n}}\right]^{\frac{2w_n(n-h_n)}{h_n}}$$

and since $\frac{h_n}{2w_n} \sim \frac{1}{2}n^{2/3}\big(1-\frac{\sqrt{\ln(n)}}{n^{1/3}}\big) \xrightarrow[n\to\infty]{} \infty$ and
$\frac{2w_n(n-h_n)}{h_n} \sim \frac{2n^{1/3}n^{2/3}\sqrt{\ln(n)}}{n\left(1-\frac{\sqrt{\ln(n)}}{n^{1/3}}\right)} = \frac{2\sqrt{\ln(n)}}{1-\frac{\sqrt{\ln(n)}}{n^{1/3}}} = \frac{2n^{1/3}\sqrt{\ln(n)}}{n^{1/3}-\sqrt{\ln(n)}} \sim 2\sqrt{\ln(n)}$,
the last factor approaches 1 as $n \to \infty$.
For the remaining factors, using Corollary 3.4, we obtain

$$\frac{\sqrt{n}h_n}{n2^{n-1}}\binom{n-1}{\frac{n-1}{2}-w_n} \sim \frac{\sqrt{n}n\big(1-\frac{\sqrt{\ln(n)}}{n^{1/3}}\big)}{n2^{n-1}}\frac{\sqrt{2}\cdot 2^{n-1}}{\sqrt{\pi(n-1)}\cdot e^{2n^{2/3}/(n-1)}} \sim \sqrt{\frac{2}{\pi}},$$

and the result follows. □

The choice of h_n in Corollary 3.9 was not accidental. It equals $\lceil n - n^{2/3}\sqrt{\ln n}\rceil$, which is the number of rounds for which the algorithm $\mathcal{A}_n$ never accepts an observation according to our numerical calculations - see the beginning of this section. Choosing the stopping region on the basis of these values of w_n and h_n, the probability of success under the algorithm $\mathcal{A}(h_n, w_n)$ is bounded from below by a function which asymptotically behaves like $\frac{1}{\sqrt{n}}\sqrt{\frac{2}{\pi}}$. Since the optimal algorithm $\mathcal{A}_n$ is not worse, this lower bound also applies to $\mathcal{A}_n$. It remains to show that this function of n is also asymptotically an upper bound for the performance of $\mathcal{A}_n$.

Theorem 3.10. *For the online decision problem considered here, the optimal stopping algorithm $\mathcal{A}_n$ has asymptotic performance*

$$Pr\big(\mathcal{A}_n \textit{ succeeds}\big) \sim \frac{1}{\sqrt{n}}\sqrt{\frac{2}{\pi}}.$$

Proof. From the analysis of the algorithm $\mathcal{A}(h_n, w_n)$ (Corollary 3.9), we know that the performance of $\mathcal{A}_n$ may be bounded from below by a function which asymptotically behaves like $\frac{1}{\sqrt{n}}\sqrt{\frac{2}{\pi}}$.

To find an upper bound, we consider an online decision problem that is easier than the problem in question. Suppose that a decision maker observes the ranks of n numbers which are independent realizations from the uniform distribution on the interval $[0, 1]$. The decision maker must choose one of these numbers with the aim of maximizing the probability of choosing the element which is closest to $\frac{1}{2}$. The optimal strategy is simple. From Theorem 2.1, we know that we have to select the number of rank r such that the binomial coefficient $\binom{n-1}{r-1}$ takes its maximum value. This happens if $r - 1 = \lfloor \frac{n-1}{2} \rfloor$ or $r - 1 = \lceil \frac{n-1}{2} \rceil$. Thus

$$\Pr\big(x_r \text{ is the best }\big) = \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor} \cdot \frac{1}{2^{n-1}}$$

and using Lemma 3.1, we obtain

$$\Pr\big(x_r \text{ is the best }\big) \sim \binom{n-1}{\frac{n-1}{2}} \cdot \frac{1}{2^{n-1}} \sim \frac{\sqrt{2} \cdot 2^{n-1}}{\sqrt{\pi n}} \cdot \frac{1}{2^{n-1}} = \frac{1}{\sqrt{n}} \sqrt{\frac{2}{\pi}},$$

which gives an asymptotic upper bound for the performance of $\mathcal{A}_n$. $\square$

In Figures 3 and 4, we present the asymptotic behavior of the performance of $\mathcal{A}_n$. We have

$$\tilde{v}(h_n, w_n) = \frac{h_n}{n 2^{n-1}} \binom{n-1}{\frac{n-1}{2} - w_n} \cdot \left(1 - \left(1 - \frac{2w_n}{h_n}\right)^{n-h_n}\right)$$

which, from Theorem 3.8, is the function describing the asymptotic behavior of the function $v(h_n, w_n)$. The choices of h_n and w_n are $h_n = \lceil n\big(1 - \frac{\sqrt{\ln n}}{n^{1/3}}\big) \rceil$ and $w_n = \lceil n^{1/3} \rceil$.

4. Final remarks

If the interval $[0, 1]$ is replaced by the interval $[a, b]$, where $a < b$, and the goal is to stop on the element closest to the interval's midpoint, then the optimal stopping algorithm is identical to our algorithm $\mathcal{A}_n$.

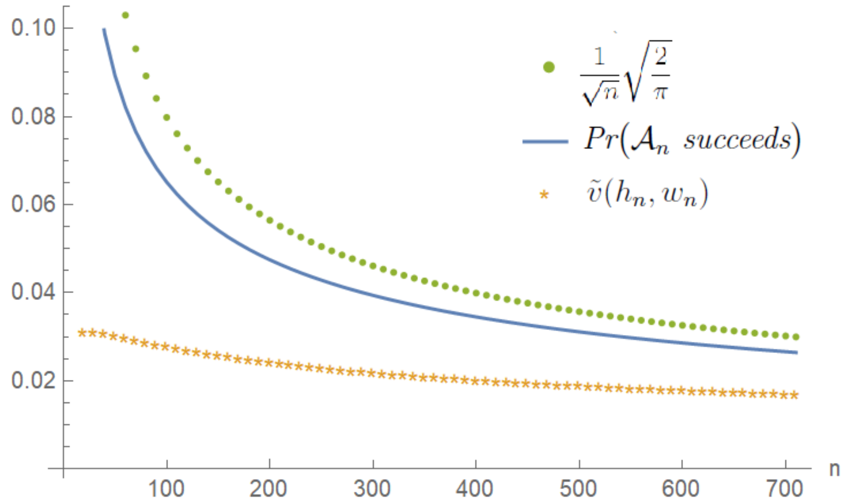

Figure 3. Performance of $\mathcal{A}_n$ with its asymptotic bounds for $h_n = \lceil n\big(1 - \frac{\sqrt{\ln n}}{n^{1/3}}\big)\rceil$ and $w_n = \lceil n^{1/3} \rceil$.

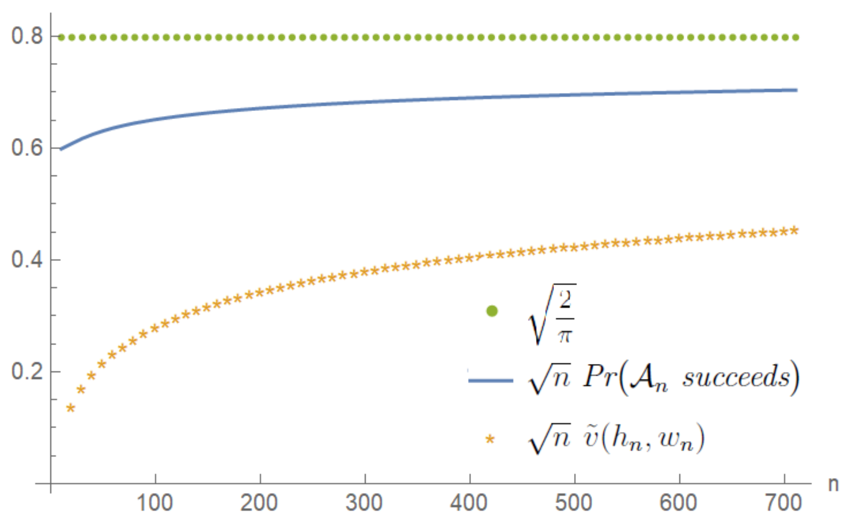

Figure 4. Function $\sqrt{n}\ \Pr\big(\mathcal{A}_n \text{ succeeds}\big)$ with its asymptotic bounds for $h_n = \lceil n\big(1 - \frac{\sqrt{\ln n}}{n^{1/3}}\big)\rceil$ and $w_n = \lceil n^{1/3} \rceil$.

How does the situation change if we sequentially observe n numbers from the interval $[0, 1]$, but we are informed about the value of each number drawn?

Since we now know whether the revealed number is greater or smaller than $\frac{1}{2}$, by replacing each x_k greater than $\frac{1}{2}$ by $1-x_k$, we obtain a problem equivalent to finding the maximum element of a sequence of n numbers. This problem was solved by Gilbert & Mosteller [5] and the optimal strategy in what they called 'the full-information game' has an asymptotic probability of success approximately equal to 0.580164. On the other hand, if our aim is to minimize the expected difference between the number selected and $\frac{1}{2}$, then we should adopt another stopping algorithm from (Gilbert & Mosteller [5]) which gives an expected difference of order $\frac{1}{n}$.

Funding: This research was partially supported by NCN Grant DEC-2015/17/B/ST6/01868.

References

[1] F.S. Benevides, M. Sulkowska
Percolation and best choice problem for powers of paths
Journal of Applied Probabability, 54, no. 2(2017), 343–362.

[2] T.S. Ferguson
Who solved the secretary problem?
Statistical Science, 4 (1989), 282–296.

[3] R. Freij, J. Wästlund
Partially ordered secretaries
Electronic Communications in Probabability, 115 (2010), 504–507.

[4] N. Georgiou, M. Kuchta, M. Morayne, J. Niemiec
On a universal best choice algorithm for partially ordered sets
Random Structures and Algorithms, 32 (2008) 263–273.

[5] J. P. Gilbert, F. Mosteller
Recognizing the maximum of a sequence
Journal of the American Statistical Association, 61 (1966), 35–73.

[6] G. Goddard, E. Kubicka, G. Kubicki
An efficient algorithm for stopping on a sink in a directed graph
Operations Research Letters, 41 (2013), 238–240.

[7] G. Kubicki, M. Morayne
Graph-theoretic generalization of the secretary problem: the directed path case
SIAM Journal on Discrete Mathematics, 19, no. 3 (2005), 622–632.

[8] D.V. Lindley
Dynamic programming and decision theory
Applied Statistics, 10 (1961), 39–51.

[9] M. Morayne
Partial-order analogue of the secretary problem; the binary tree case
Discrete Mathematics, 184 (1998), 165–181.

[10] J. Preater
The best-choice problem for partially ordered objects
Operations Research Letters, 25 (1999), 187–190.

[11] P.A. Rogerson
Probabilities of choosing applicants of arbitrary rank in the secretary problem
Journal of Applied Probabability, 224, no. 2 (1987), 527–533.

[12] J.S. Rose
A problem of optimal choice and assignment
Operations Research, 30, no. 1 (1982), 172-–181.

[13] W. Stadje
Efficient stopping of a random series of partially ordered points
Lecture Notes in Economics and Mathematical Systems, 177 (1980), 430–447.

[14] M. Sulkowska
The best choice problem for upward directed graphs
Discrete Optimization, 9 (2012), 200–204.

[15] R.J. Vanderbei
The postdoc variant of the secretary problem
http://www.princeton.edu/ rvdb/tex/PostdocProblem/PostdocProb.pdf
(2012), unpublished.