

# Dynamic Threshold Strategy for Universal Best Choice Problem

Jakub Kozik

► **To cite this version:**

Jakub Kozik. Dynamic Threshold Strategy for Universal Best Choice Problem. 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10), 2010, Vienna, Austria. pp.439-452. hal-01185565

**HAL Id: hal-01185565**

**<https://hal.inria.fr/hal-01185565>**

Submitted on 20 Aug 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# *Dynamic Threshold Strategy for Universal Best Choice Problem*

Jakub Kozik <sup>†</sup>

*Theoretical Computer Science, Jagiellonian University,  
ul. Łojasiewicza 6, 30-348 Krakow, Poland*

---

We propose a new strategy for universal best choice problem for partially ordered sets. We present its partial analysis which is sufficient to prove that the probability of success with this strategy is asymptotically strictly greater than  $1/4$ , which is the value of the best universal strategy known so far.

**Keywords:** partial order, secretary problem, best choice

---

## Introduction

The secretary problem is one of the most commonly known best-choice problems. It models a situation of  $n$  candidates applying for a single secretary position. The candidates are interviewed in a random order. After each interview an irrevocable decision to accept or reject the current candidate is made. Once a candidate is accepted, no more interviews are arranged. All candidates are linearly ordered according to their qualifications, but the qualifications of the candidate are not known before the interview. The objective is to choose the best candidate.

The classical result states that the best strategy guarantees a probability of success greater than  $e^{-1}$ . The strategy ignores the first  $r_n$  candidates, and then chooses the first one better than all the candidates interviewed so far. The number  $r_n$  can be easily computed and it asymptotically behaves like  $e^{-1} \cdot n$ . For precise definitions and historical view on the classical secretary problem consult [2].

Many variants of this classical problem have been considered (e.g. [2] for a brief overview and bibliography). The problem can be generalized by relaxing the assumption that the candidates are linearly ordered. One of the natural extensions allows for several criteria, each one ordering the candidates linearly. In such a case the candidates are, in fact, partially ordered and the order has dimension not greater than the number of criteria. The objective is to choose one of the maximal elements (comp. [4]). Similar research has been done for other classes of partial orders (comp. [5]).

In the most general variant the candidates are partially ordered with no constraints imposed on the order. The number of candidates is still known beforehand, and the objective is to choose one of the candidates maximal in the partial order. It was an interesting question, whether there exists a strategy,

---

<sup>†</sup>This work was supported by grant number N206 3761 37 funded by Polish Ministry of Science and Higher Education.

with a probability of success greater than some positive constant (on every partial order). First such strategy was presented by Preater [6]. The analysis provided in the paper showed that the probability of success exceeds  $1/8$ . Georgiou, Kuchta, Morayne and Niemiec [3] proved that, for a variant of Preater's strategy, this probability is no smaller than  $1/4$ , and that, for every  $\epsilon > 0$ , there exists a poset for which this probability is smaller than  $1/4 + \epsilon$ .

The classical secretary problem shows that the asymptotic probability of success for the general problem cannot exceed  $e^{-1}$ . The question what is the probability for the best possible strategy for the general problem remains open.

In the present paper we propose a new strategy called The Dynamic Threshold Strategy, with a probability of success greater than  $1/4 + \epsilon$  for some  $\epsilon > 0$  and for all sufficiently large posets. As long as it does not compromise our objective to show that the probability is greater than  $1/4$ , we favor simplicity of the presentation over the precision of given estimations. However, the actual asymptotic value of this probability seems to be much bigger.

## 1 The universal best choice problem (UBC)

For the general definition of the universal strategy as a random variable which is some stopping time see e.g. [6]. For our purposes it will be sufficient to consider specific subclass, namely stateless deterministic strategies.

A *pointed poset* is a poset with a one distinguished element. A *stateless deterministic strategy* for UBC is a (computable) function, that given a pointed poset  $P$  and a number  $n$  which is not smaller than the size of  $P$ , returns *accept* or *reject*. The interpretation is straightforward – pointed poset is an induced poset on the candidates presented so far, and the distinguished vertex is the current candidate. The value of the function determines whether to accept or reject the current candidate.

Let  $P$  be a poset of size  $n$ . We assume that its nodes are labeled by distinct numbers from  $\{1, \dots, n\}$ . For a permutation  $\pi$  of  $\{1, \dots, n\}$  let  $P^\pi = (P_1^\pi, \dots, P_n^\pi)$  denote a sequence of pointed posets such that,  $P_j^\pi$  is a subposet of  $P$  induced by elements  $\{\pi(1), \dots, \pi(j)\}$ , with distinguished vertex  $\pi(j)$ . Poset  $P_j^\pi$  is called *the induced poset at time  $j$* . When the time  $j$  is clear from the context we call it just *the induced poset*.

Let  $S$  be a stateless deterministic strategy. For every poset  $P$  of size  $n$ , and permutation of its elements  $\pi$  (we call it a *presentation order*), we define the chosen element as an element  $\pi(j)$  (i.e. the  $j$ -th element that comes) for a minimal  $j$  for which  $S(P_j^\pi, n) = \text{accept}$ , and if there is no such  $j$  we put  $j = n$ . A choice was a success if the chosen element is maximal in  $P$ . We put  $\text{val}_S(P, \pi) = 1$  if a choice was success and  $\text{val}_S(P, \pi) = 0$  otherwise. For a poset  $P$  of size  $n$  we define *the value of a strategy  $S$  on  $P$*  as:

$$p_S(P) = \frac{\sum_{\pi \in S_n} \text{val}_S(P, \pi)}{n!}$$

i.e. the probability of success when a presentation order is chosen uniformly at random ( $S_n$  denotes the set of all permutations of  $\{1, \dots, n\}$ ).

Let  $\mathbb{P}_n$  be the set of all partial orders on the set  $\{1, \dots, n\}$ . For a fixed  $n \in \mathbb{N}$  we define the value of a strategy as a minimal value among all the values on the posets of size  $n$ :

$$p_S(n) = \min_{P \in \mathbb{P}_n} p_S(P).$$

An asymptotic value of a strategy  $S$  is defined as

$$p_S = \liminf_{n \rightarrow \infty} p_S(n).$$

The general problem is to determine the value  $p_S$  for the best possible strategy. It is known to be not greater than  $e^{-1}$ , since no strategy can be better than the classical threshold strategy for the secretary problem (i.e. on linear orders).

**Remark 1.1** *Preater's Policy B ([6, 3]) is not stateless deterministic strategy. It needs both internal state and randomization. Analogous definitions for strategies with internal state and randomization are straightforward. The asymptotic value of the Policy B equals  $1/4$ .*

## 1.1 Dynamic Threshold Strategy

When the poset of candidates has many maximal elements, it seems reasonable not to choose the first one that comes. Intuitively, the longer we wait, the more information about the poset we get, and the chances of making a mistake get smaller. If, on the other hand, we wait too long, we can miss the last maximal element, and lose. If we knew the number of maximal elements  $m$ , we could calculate the expected time of arrival of the last maximal element, and focus on the elements that come around this time. Unfortunately, we do not know  $m$ . However, when the number of maximal elements is large, we can get some estimation of it by observing the number of maximal elements in the induced poset at the time  $n/2$  and multiplying it by 2 (where  $n$ , as usual, denotes the number of candidates). Although this value is very likely to be overestimated, the probability of essential underestimation is rather low (we expect that about half of maximal elements have already come). In most cases relying on it would make us wait longer than we should. However, we can still observe the elements that come and actualize our estimation accordingly. When the last maximal element comes, our estimation should be quite close to the real value. Unfortunately, at that moment any overestimation results in a failure.

Our strategy can be summarized as: try to estimate the number of maximal elements, wait until some threshold time  $tr_n(m)$  ( $tr_n(m)$  depends on the estimated value), and pick the first maximal element after that time.

**Definition 1.2 (Dynamic Threshold Strategy)** *For a poset with size  $n$ , pick an element  $v$  at the step  $t$  if and only if:*

- $v$  is maximal in the induced poset at time  $t$
- and  $t > n \cdot e^{-1/m}$ , where  $m$  is the number of maximal elements of the **induced** poset at time  $t$ .

**Theorem 1.3 (Main result)** *There exists  $\varepsilon > 0$  such that the asymptotic value of the Dynamic Threshold Strategy exceeds  $1/4 + \varepsilon$ .*

## 2 Analysis of the strategy

The following sections are devoted to the analysis of the Dynamic Threshold Strategy (*the strategy*). The analysis is divided in two parts. The first one deals with posets that have large number of maximal elements. The second one tackles the posets with number of maximal elements smaller than some constant. Due to the limited space, we present the first part in detail, and sketch the proofs of the second one.

## 2.1 Preliminaries

We define  $tr(m) = e^{-1/m}$  and, for a positive integer  $n$ , we put

$$tr_n(m) = \lfloor tr(m) \cdot n \rfloor = \lceil tr(m) \cdot n - 1 \rceil.$$

Note that, since  $tr(m)$  is never rational for integer  $m \geq 1$ , using floor (resp. ceiling) function does not change the described strategy. We put also

$$ltr(m) = 1 - tr(m) \quad \text{and} \quad ltr_n(m) = n - tr_n(m).$$

For a fixed poset  $P$ , we usually denote by  $m$  the number of its maximal elements, and by  $n$  the number of all its elements. The interval  $[tr(m), 1]$  is called *the game interval*. For a presentation order  $\pi$ , we say that some element  $v$  comes in the interval  $[a, b]$  if  $a \leq \frac{\pi^{-1}(v)}{n} \leq b$ .

Let  $\mathbb{P}(n, m)$  denote the set of all posets on the set  $\{1, \dots, n\}$  with  $m$  maximal elements. We put:

$$p(n, m) = \min_{P \in \mathbb{P}(n, m)} p_S(P),$$

where  $p_S(P)$  denotes the probability of success of our strategy on the poset  $P$ .

The following observation is crucial for the presented analysis.

**Observation 2.1** *For every time moment  $t$ , if the strategy did not pick an element before  $t$ , and  $k$  maximal elements of  $P$  came before  $t$ , then it will not pick an element before the moment  $tr_n(k + 1)$ .*

Clearly, if  $k$  maximal elements came already, they can not be dominated by any elements that will come. Therefore, the observed number of maximal elements will never be smaller than  $k$ . Moreover, if some element comes and it is maximal in the induced poset, the observed number of maximal elements at that time is not smaller than  $k + 1$ . Hence it can not be picked unless it comes after  $tr_n(k + 1)$ . Note also, that we usually do not know the actual value of  $k$ .

### 2.1.1 Hypergeometric distribution

In the presented analysis, in many places we meet random variables with hypergeometric distribution. For nonnegative integers  $r, g, b$  the distribution  $HD(r, g, b + g)$  can be defined as a distribution of number of green balls in a random sample of size  $r$ , taken without repetitions from the urn containing  $g$  green balls and  $b$  black balls. It is easy to see that probability of having  $k$  green balls in the sample is  $\binom{g}{k} \frac{g^k \cdot b^{r-k}}{(b+g)^r}$  where  $a^m = a \cdot (a - 1) \cdot \dots \cdot (a - (m - 1))$ . E.g. for a poset of size  $n$  with  $m$  maximal elements the number of maximal elements that come before  $tr_n(1)$  follows distribution  $HD(m, tr_n(1), n)$ . For a random variable  $X$  with distribution  $HD(r, g, b + g)$  we have

$$E(X) = \frac{rb}{b+g}, \quad Var(X) = \frac{rbg}{(b+g)^2} \left( 1 - \frac{r-1}{b+g-1} \right)$$

(see e.g. [1] for the proofs). Note that the variance is smaller than the variance of analogous binomial distribution (i.e. for sampling with repetitions).

## 2.2 Large number of maximal elements

In this section we are going to prove that, for all sufficiently large  $m$  and  $n$ , the probability  $p(n, m)$  is greater than 0.27.

First, we prove that for every  $\gamma \in (0, 1)$  and for all sufficiently large  $n$  and  $m$ , we do not pick an element before time  $tr_n(\gamma \cdot m)$  with probability which is arbitrarily close to 1.

**Lemma 2.2** *For every  $\gamma \in (0.5, 1)$  and  $\epsilon > 0$  there exist  $M, N \in \mathbb{N}$  such that for all  $m, n$  such that  $M < m \leq n > N$  we do not pick an element before time  $tr_n(\gamma \cdot m)$  with probability exceeding  $1 - \epsilon$ .*

**Proof:** Let us fix  $\gamma \in (0.5, 1)$ , and consider three events:

1. at least one maximal element comes before  $tr_n(1)$ ,
2. at least half of maximal elements come before  $tr_n(2)$ ,
3. at least  $\gamma \cdot m$  maximal elements come before  $tr_n(m/2)$ .

According to Observation 2.1 the first event guarantees that we do not pick an element before  $tr_n(2)$ . The second that we do not pick an element in the interval  $[tr_n(2), tr_n(m/2))$ , and the third – in the interval  $[tr_n(m/2), tr_n(\gamma \cdot m))$ . Therefore, if all these events occur, we can be sure of not picking an element before  $tr_n(\gamma \cdot m)$ .

It is not hard to see that for sufficiently large  $m$  and  $n$  the probability of each of the events is arbitrarily close to 1. Indeed, the first is trivial. The expected number of maximal elements before  $tr_n(2)$  is asymptotically equal to  $m \cdot tr(2) \approx m \cdot 0.605..$ , and the distribution of the number of elements before  $tr_n(2)$  is concentrated around the mean, since the standard deviation grows like  $\sqrt{m}$ . The application of the Chebyshev's inequality shows that the probability of the second event goes to 1 with  $m \rightarrow \infty$ . The expected number of maximal elements that come after  $tr_n(m/2)$  converges to 2 (with  $m$ ), and  $(1 - \gamma) \cdot m$  goes with  $m$  to infinity. It remains to apply the Markov's inequality to see that the probability of the third event goes to 1 with  $m \rightarrow \infty$ .  $\square$

We analyze the behavior of  $p(n, m)$  in two regions. Let us put  $g(n) = n^{2/3} \log n$ . First, we analyze all the pairs  $(m, n)$  for which  $m > g(n)$ , and  $n$  is greater than some fixed  $N$  (we call it Region 1). The remaining part, i.e. all the pairs  $(m, n)$  for which  $M < m \leq g(n)$  and  $n > N$ , for some fixed  $M, N$  is called Region 2.

## 2.3 Region 1 ( $m > g(n)$ )

We omit the analysis of the cases, when  $m > n/2$ , where it is easy to show that the probability of success is much bigger than 0.25.

According to Lemma 2.2, for sufficiently large  $n$ , with arbitrarily large probability we can almost reach the game interval without picking an element. We focus on the interval consisting of the last  $\lfloor 2 \cdot n \cdot ltr(m) \rfloor$  steps, which we call *the extended game interval*. It turns out that for large enough  $n$  with probability arbitrarily close to 1, there are no "false maximal elements" in that interval.

**Lemma 2.3** *For every  $\epsilon > 0$ , for all sufficiently large  $n$  and for all  $m > g(n)$  the probability that some not-maximal element comes in the extended game interval and it is maximal in its induced order, is smaller than  $\epsilon$ .*

**Proof:** We partition the set of not-maximal elements of the poset into  $m$  disjoint classes in such a way, that  $j$ -th class is a subset of the strict lower cone of  $j$ -th maximal element (we take any such partition).

Let  $l$  be the length of the extended game interval. Clearly, we have at most  $l$  maximal elements and at most  $l$  not-maximal elements in the extended game interval. Let us consider the construction of the presentation order such that: first we choose positions for the maximal elements, then we distribute not-maximal elements in the remaining positions and finally we distribute maximal elements in the chosen positions. Before the last step, at most  $l$  classes have their representatives in the extended game interval, we call their maximal elements forbidden elements. If none of the last  $l$  maximal elements in constructed presentation order is forbidden, then we can be sure that there are no not-maximal elements in the extended game interval which are maximal in the induced order. The probability of such an event is greater than:

$$\frac{(m-l)^l}{m^l} > \left(\frac{m-2l}{m-l}\right)^l = \left(1 - \frac{l}{m-l}\right)^l,$$

and it is a matter of simple calculation to show that the assumption on the growth of  $m$  is sufficient for the probability above to tend to 1, with  $n \rightarrow \infty$ , uniformly for all  $m > g(n)$ .  $\square$

According to the previous lemma and Lemma 2.2, asymptotically almost surely, we do not pick an element before the extended game interval, and no "false maximal" comes in that interval. In such a case if at least one maximal element comes in the game interval, we can be sure of success. The probability, that there is no maximal element that comes in the game interval is:

$$\frac{tr_n(m)^m}{n^m} \leq \left(\frac{tr_n(m)}{n}\right)^m \leq tr(m)^m = e^{-1}.$$

**Corollary 2.4** For every  $\epsilon > 0$ , there exists  $N$  such that, for all  $n > N$  and all  $m > g(n)$  we have  $p(n, m) > 1 - e^{-1} - \epsilon$ .

## 2.4 Region 2 ( $m \leq g(n)$ )

Let us start with the observation that, since  $g(n)$  is sublinear, we have

$$tr_n(m) \sim_n n \cdot tr(m)$$

uniformly, for all  $m \leq g(n)$ . For any functions  $f(n, m), g(n, m)$  writing  $f(n, m) \sim_n g(n, m)$  we mean asymptotic equivalence for  $n \rightarrow \infty$ , which is uniform for all  $m$  in currently considered range of  $m$ .

We are going to give a lower bound for  $p(n, m)$  by calculating the probability that we succeed and exactly one maximal element comes in the game interval. We proceed in three steps. First we estimate the probability that exactly one maximal element comes in the game interval. Then, that we do not pick an element before the game interval. Finally, that we pick a maximal element in the game interval.

It is easy to see that the probability that exactly one maximal element comes in the game interval depends only on the size of the poset  $n$ , and number of its maximal elements  $m$ , but not on its structure. We denote this number by  $pm(n, m)$ .

**Lemma 2.5** For every  $\epsilon > 0$  and for all sufficiently large  $m, n$  we have  $pm(n, m) \geq e^{-1} - \epsilon$ .

The purely technical (but simple) proof is omitted.

### 2.4.1 Probability of playing

According to Lemma 2.2 for every  $\gamma \in (0, 1)$  we do not pick an element before  $tr_n(\gamma m)$  with probability  $1 - \epsilon_1$  (for sufficiently large  $m$  and  $n$ ).

It is also easy to check that for every  $\epsilon_2 > 0$  we can choose  $\gamma \in (0, 1)$  sufficiently large to ensure that the probability of having no maximal element in the interval  $[tr_n(\gamma m), tr_n(m)]$  is greater than  $1 - \epsilon_2$  (for sufficiently large  $m$  and  $n$ ). (The fact that  $g(n)$  is sublinear is used in the proof).

Finally, Lemma 2.5 says that with probability exceeding  $e^{-1} - \epsilon_3$  we have exactly one maximal element in the game interval.

If all three events occur we have  $(m - 1)$  maximal elements before  $tr_n(\gamma m)$ , which guarantees that we do not pick an element before the game interval. Therefore, for sufficiently large  $n, m$ , with probability exceeding  $e^{-1} - \epsilon_1 - \epsilon_2 - \epsilon_3$  exactly one maximal element comes in the game interval, and we do not pick an element before the game interval. It remains to estimate the conditional probability of winning in the game interval.

## 2.5 Submaximal elements

For a maximal element  $v$  we say that an element  $w$  is *submaximal* for  $v$  if and only if the strict upper cone of  $w$  equals  $\{v\}$ . Suppose that  $v$  is the unique maximal element that comes in the game interval. Let  $l$  denote the number of submaximal elements for  $v$ . We are going to show that, independently of the value of  $l$ , the probability of taking the maximal element in the game interval is greater than 0.75. All the probabilities in this section are conditioned on the event  $E$  defined as follows: we do not pick an element before the game interval and the only maximal element that comes in the game interval is  $v$ .

### 2.5.1 Small number of submaximal elements

If  $l$  is small it is very likely that all of the submaximal elements for  $v$  come before the game interval. The (conditional) probability of this event equals

$$pAll(n, m, l) = \frac{(tr_n(m) - m + 1)^l}{(n - m)^l},$$

and for all  $l < cm$  (for any constant  $c$ ) we have

$$pAll(n, m, l) \gtrsim_n \exp\left(-\frac{l}{m-1}\right).$$

It shows that for all  $l \leq m/4$  (and sufficiently large  $m$  and  $n$ ) the probability of success in the game interval is asymptotically greater than  $\exp(-1/4) \approx 0.779\dots$

### 2.5.2 Large number of submaximal elements

When the number of submaximal elements is large in comparison to  $m$ , we usually see a lot of submaximal elements in the beginning of the game interval, which postpones the decision of picking.

We assume that  $l \geq 4m$ . Suppose that we see  $j$  submaximal elements at the beginning of the game interval. They appear as maximal, therefore we will not take any elements before  $tr_n(m - 1 + j)$  unless it is the last maximal element. Therefore, we succeed with probability exceeding

$$1 - \frac{ltr_n(m - 1 + j)}{ltr_n(m)}.$$



Let us fix any  $\gamma \in (0, 1)$ . It is not hard to show that, for sufficiently large  $m$ , with probability exceeding  $1 - \epsilon$ , we observe at least  $\gamma l$  submaximal elements at the beginning of the game interval. Then we can estimate from below the probability of success by

$$1 - \frac{ltr_n(m-1+\gamma l)}{ltr_n(m)} - \epsilon.$$

And one can easily check that we can chose  $\gamma$  such that, for all sufficiently large  $m$  and all  $l \geq 4m$ , the probability above exceeds 0.75.

### 2.5.3 Moderate number of submaximal elements

For the values of  $l$  between  $m/4$  and  $4m$  neither of two considered events is sufficient to grant the probability of success exceeding 0.75. In this region we consider three disjoint events (for a fixed  $\gamma \in (0, 1)$ ):

**A – all submaximal elements come before the game interval.** The probability of this event is denoted by  $pl_0(n, m, l)$ . In this case success is guaranteed.

**B – all but one submaximal elements come before the game interval.** The probability is denoted by  $pl_1(n, m, l)$ . Then we can be sure of success if the last maximal element or the last submaximal element comes before  $tr_n(m-1+l)$  – (conditional) probability of success is denoted by  $pls_1(n, m, l)$

**C – at least  $\gamma \cdot l$  but no more than  $l - 2$  submaximal elements come before  $tr_n(m)$ .** The probability is denoted by  $pl_\gamma(n, m, l)$ . Success is guaranteed if the last maximal element comes before  $tr_n(m + \gamma \cdot l - 1)$  – (conditional) probability of success is denoted by  $pls_\gamma(n, m, l)$

We give lower bounds for the described probabilities. By asymptotic inequality  $f(n) \gtrsim_n s(n)$  we mean that there is a function  $h$  such that  $f(n) \geq h(n) \sim_n s(n)$ .

**Proposition 2.6** *All of the following asymptotic (in)equalities w.r.t.  $n$  hold uniformly for all  $m$  larger than some constant  $M$  and  $m < g(n)$ , and all  $l$  such that  $m/4 < l < 4m$ . Asymptotic equalities w.r.t.  $m$  hold uniformly for all  $l$  satisfying  $m/4 < l < 4m$ .*

1.  $pl_0(n, m, l) \gtrsim_n \exp(-\frac{l}{m-1}) \sim_m \exp(-\frac{l}{m})$ ,
2.  $pl_1(n, m, l) \gtrsim_n ltr(m) \cdot l \cdot \exp(\frac{l-1}{m-1}) \sim_m \frac{l}{m} \exp(-\frac{l}{m})$
3.  $pls_1(n, m, l) \sim_n 1 - \left(\frac{ltr(m+l-1)}{ltr(m)}\right)^2 \sim_m 1 - \left(\frac{m}{m+l}\right)^2$
4.  $pl_\gamma(n, m, l) \sim_n 1 - pl_0(n, m, l) - pl_1(n, m, l) - \epsilon(m)$ , and  $\epsilon(m) \rightarrow_m 0$ .
5.  $pls_\gamma(n, m, l) \sim_n 1 - \frac{ltr(m-1+\gamma l)}{ltr(m)} \sim_m 1 - \frac{m}{m+\gamma l}$

**Proof:** The first estimation is given in Section 2.5.1. In an analogous way we derive the second one. The third, and the fifth are straightforward. In the fourth, we have to show that the asymptotic probability of having more than  $(1 - \gamma) \cdot l$  submaximal elements in the game interval goes to 0 with  $m \rightarrow \infty$ .

The number of submaximal elements in the game interval follows the hypergeometric distribution  $HD(l, ltr_n(m) - 1, n - m)$ . Therefore, the expected number of elements in the game interval is eventually smaller than some positive constant. Since  $\gamma \cdot l \rightarrow_{m \rightarrow \infty} \infty$  the Markov's inequality finishes the proof.  $\square$

The total probability of success, (conditioned on the event  $E$ ), is

$$pl_0(n, m, l) + pl_1(n, m, l) \cdot pls_1(n, m, l) + pl_\gamma(n, m, l) \cdot pls_\gamma(n, m, l). \quad (1)$$

Hence, according to Proposition 2.6, we have the following asymptotic lower bound:

$$\exp(-l/m) \left(1 + \frac{l}{m} \cdot \left(1 - \left(\frac{m}{m+l}\right)^2\right)\right) + (1 - \exp(-l/m) \cdot (1 + l/m)) \cdot \left(1 - \frac{m}{m + \gamma l}\right)$$

Replacing  $l$  by  $\alpha m$ , we obtain:

$$\exp(-\alpha) \cdot \left(1 + \alpha \left(1 - \left(\frac{1}{1+\alpha}\right)^2\right)\right) + (1 - \exp(-\alpha) \cdot (1 + \alpha)) \cdot \left(1 - \frac{1}{1 + \gamma\alpha}\right).$$

We are interested in  $\alpha \in (1/4, 4)$  that minimizes the formula above. We used a software for symbolic manipulation Mathematica [7], which was able to verify symbolically that the for sufficiently large  $\gamma \in (0, 1)$  the value of the formula above is greater than 0.75 for all values of  $\alpha > 1/4$ . From numerical approximations we know that the minimal value is obtained for  $\alpha$  close to 1.57661.. and it is roughly 0.769487..

Three previous sections show that, independently of the number of submaximal elements, the conditional probability of success in the game interval asymptotically exceeds 0.75. Additionally, we know that the asymptotic probability that we do not pick before the game interval and that  $m - 1$  maximal elements come before the game interval exceeds  $e^{-1}$ . This justifies the following statement:

**Proposition 2.7** *For any  $\epsilon > 0$  and all sufficiently large  $n$  and  $m$ , such that  $m < g(n)$ , we have  $p(n, m) > 0.75 \cdot e^{-1} - \epsilon$ .*

Additionally, we have  $0.75 \cdot e^{-1} \approx 0.2759\dots$  For the next section let us fix some  $\epsilon < 0.01$  and let  $M$  be the minimal value for which there exists  $N$  such that for all  $m, n$  fulfilling  $n > N$  and  $M < m < g(n)$  we have  $p(n, m) > 0.75 \cdot e^{-1} - \epsilon$ .

### 3 Small number of maximal elements

In this section we consider the posets with number of maximal elements not greater than  $M$  from the previous section. For a small number of maximal elements, the submaximal elements play important role. For maximal elements  $v_1, \dots, v_m$  let  $l_1, \dots, l_m$  be their corresponding numbers of submaximal elements. For every maximal element  $v_i$  we are going to give a lower bound for the probability of success, conditioned on the event that  $v_i$  is the maximal element that comes last. This lower bound depends only on  $m$  and  $l_i$  and is denoted by  $p_s(m, l_i)$ . Then, the total probability of success is greater than the average value of all  $p_s(m, l_i)$  (for  $i = 1, \dots, m$ ). Moreover, it turns out that even the minimal value is greater than 0.25. Since the lower bound, that we are going to use, depends only on the number  $l_i$  (and  $m$ ) but not on the structure of a poset, we can assume that all maximal elements in our poset have the same number  $l$  of submaximal elements – the number that minimizes  $p_s(m, l)$ .

It is easy to see that for every  $\epsilon$  and for large enough  $l$ , the number of submaximal elements that come before  $tr_n(1)$  with arbitrary large probability is large enough that we will not pick a not maximal element before the time  $(1 - \epsilon) \cdot n$ . Then, if the last maximal element comes in the interval  $(tr(m), (1 - \epsilon))$  we can be sure of success. The number  $m$  is bounded, therefore choosing large enough  $l$  we can make sure that asymptotic probability of success is greater than  $e^{-1}$ . In the rest of this section we assume that  $l$  is smaller than some constant  $L$ .

The situation, when the numbers  $m$  and  $l$  are bounded by constants, and we are interested in asymptotic probability of success (i.e. when  $n$  goes to infinity), allows for the change of probabilistic model. We give a lower bound for the probability of success which is based only on the times of arrival of maximal and submaximal elements. The total number of maximal and their submaximal elements is bounded, therefore assuming that for every  $n$  their arrival times are independent and uniformly distributed on the set  $\{1, \dots, n\}$ , will not change the asymptotic value of our lower bound. Further we use  $m + l \cdot m$  independent random variables  $X_1, \dots, X_m, Y_{1,1}, \dots, Y_{1,l}, \dots, Y_{l,1}, \dots, Y_{l,l}$ , with uniform distributions on the unit interval. Variable  $X_i$  describes the normalized time of arrival of  $i$ -th maximal element, and variable  $Y_{i,j}$  the normalized time of arrival of  $j$ -th submaximal element of the  $i$ -th maximal element. For  $\beta \in [0, 1]$  we define  $I_\beta$  as follows:

$$I_\beta = \#\{i \in \{1, \dots, m\} : X_i \leq \beta\} + \#\{i \in \{1, \dots, l\} : Y_{k,i} \leq \beta\}$$

where  $k$  is the index of the maximal element that comes last ( $I_\beta$  is the number of maximal elements and elements which are submaximal for the last maximal, that come before  $\beta$ ). We define an event  $F$  as:

- there exists a time moment  $\beta < \max_{i=1, \dots, m} X_i$  such that  $\beta > tr(I_\beta + 1)$
- or  $\max_{i=1, \dots, m} X_i < tr(m)$ .

The first condition says that there is a time moment before the arrival of the last maximal element, in which any incoming element which is maximal in the induced poset would be picked. If this event does not occur, we can be sure of not picking any element before the arrival of the last maximal. The second condition corresponds to the situation when no maximal element arrives in the game interval. The event complementary to  $F$ , abbreviated by  $S$  will be called a success. Now it is easy to see, that for all  $l < L$ , and all  $m < M$  the probability of the event  $S$  gives a lower bound for the asymptotic probability of success on all posets with  $m$  elements, in which all maximal elements have  $l$  submaximal elements. We are going to show that this probability is greater than 0.25. Due to the limited space we present only brief description of the methods.

### 3.0.4 Minimizing $M$

We look for the minimal  $m$  for which the arguments from the previous section work. Therefore, in our new model we repeat the reasoning from the section 2.4 with several modifications. In particular, for probabilities of exactly one element in the game interval and not picking before the game interval, we use lower bounds which are increasing with  $m$ . We also take into account submaximal elements while estimating the probability of not picking before the game interval. Then we obtain a lower bound for the probability of exactly one element in the game interval and not picking before the game interval, by evaluating our lower bound at  $m = 30$ . The last step – estimating the probability of success in the game interval – is analogous to the developments of Section 2.5.3, i.e. we get a lower bound dependent only on the quotient  $l/m$ , and show that it is always greater than 0.7. All these developments are rather technical

and we made an extensive use of a system of symbolic manipulation Mathematica [7]. We were able to show that the probability of winning exceeds 0.252 for all  $m \geq 30$ .

Then, we repeat the previous scenario once again, this time using more exact lower bounds (in particular nonmonotonic). Since we have only few values of  $m$  left to be analyzed, we can afford to check them all. In particular we use the exact value of the probability of having exactly one maximal element in the game interval, which is decreasing with  $m$ , and was previously estimated from below by  $e^{-1}$ . In that way we show that:

1. if there are no submaximal elements, then the probability of success is greater than 0.26 for all  $m \geq 8$ ,
2. if there is at least one submaximal element, but no more than  $3 \cdot m$ , then the probability of success is greater than 0.26 for all  $m \geq 6$ ,
3. if there are more than  $3 \cdot m$  submaximal elements, then the probability of success is greater than 0.26 for all  $m \geq 3$ ,

### 3.0.5 Remaining cases

In the remaining cases the probability of not picking before the game interval needs a special care, since the estimations used so far are very inaccurate for low values of  $m$ .

Let us fix a poset with  $m$  maximal elements. Let  $r$  be a positive integer, and suppose that the last maximal element comes after  $tr(r)$ . All maximal elements but the last one and all submaximal elements are called *semimaximal elements*. Having  $k$  semimaximal elements, the probability that we do not pick an element before the time  $tr(r)$ , and that all  $k$  semimaximal elements comes before  $tr(r)$  is greater than the value of function  $G(k, r)$  defined recursively as follows:

$$G(k, r) = \begin{cases} 0, & k + 1 < r \\ tr(1)^k, & r = 1 \\ \sum_{j=r-1, \dots, k} \binom{k}{j} \cdot G(j, r-1) \cdot (tr(r) - tr(r-1))^{k-j}, & r > 1. \end{cases}$$

The first two cases are obvious: for  $r > 1$  we can be sure of not picking before  $tr(r)$  if we did not pick before  $tr(r-1)$  and there were at least  $r-1$  semimaximal elements present at time  $tr(r-1)$ .

Let  $l$  denote the number of submaximal elements. Suppose that the last maximal element comes in the interval  $(tr(k), tr(k+1))$ , for  $k \geq m$ . Then we can estimate the probability of success by:

$$P_1(m, l, k) = \sum_{j=k, \dots, m-1+l} G(j, k) \cdot \binom{l}{j-m-1} \cdot ltr(k)^{l-(j-m+1)}.$$

Indeed, we need at least  $k$  semimaximal elements at  $tr(k)$  to be sure of not picking not maximal element before  $tr(k+1)$ . If we have  $j$  semimaximal elements at  $tr(k)$ , then exactly  $m-1$  of them must be maximal elements. Therefore, we have  $\binom{l}{j-m-1}$  possibilities to choose the remaining submaximal elements. Additionally, all  $j$  elements that come before  $tr(k)$  must be distributed in such a way that it ensures not picking an element before  $tr(k)$  (this happens with probability greater than  $G(j, k)$ ). Finally, all the remaining submaximal elements must come after  $tr(k)$  (probability  $ltr(k)^{l-(j-m+1)}$ ).

As a result, the probability that the last maximal element comes in the interval  $(tr(k), tr(k + 1))$  and we succeed is greater than:

$$P_2(m, l) = \sum_{k=m, \dots, m-1+l} m \cdot (tr(k + 1) - tr(k)) \cdot P_1(m, l, k).$$

By a construction, the probability  $P_2(m, l)$  is increasing with  $l$ . For each  $m = 1, 2, 3, 4, 5$  the smallest  $l$  for which  $P_2(m, l) > 0.25$  is resp. 3, 4, 5, 6, 7.

For the remaining cases we consider another possibility of winning. The case when the last maximal element comes after  $tr(m + l)$ , has not been included in  $P_2(m, l)$ . In such a case we can be sure of success if we do not pick before  $tr(m + l - 1)$  and all semimaximal elements come before that time. The probability of this event is greater than

$$P_3(m, l) = m \cdot ltr(m + l) \cdot G(m - 1 + l, m - 1 + l).$$

Evaluating  $P_2(m, l) + P_3(m, l)$  for the remaining cases shows that the probability is greater than 0.25 for all except  $m = 1, l = 1$ . In this case it is necessary for a success (in a sense of  $S$ ), that the submaximal element comes before  $tr(1)$  and maximal one comes after. The probability of this event is 0.2325... However, in that case very simple analysis concerning subsubmaximal elements (i.e. elements whose strict upper cones contains exactly maximal and submaximal element) shows that the probability is greater than 0.26.

## 4 Concluding remarks

We have shown that asymptotic probability of success using Dynamic Threshold Strategy is strictly greater than 0.25. It is an easy task to improve the given estimations, however at a cost of bigger complexity of the calculations (and less clear presentation). Although the improvement might be considerable, we do not think that it is possible to approach the value  $e^{-1}$  in this way. However, in many cases, when the estimated value was close to 0.25, we took into account only the event when there was exactly one maximal element in the game interval. Considering the cases of two or more maximal elements in the game interval would probably improve our lower bound in an essential way. The other idea of improvement could be taking into account subsubmaximal elements (and subsubsubmaximal etc.) Finally, it is important to remark that, unlike the results of [6, 3] concerning Preater's strategy, our results deals only with asymptotic value of the strategy. However it is possible to alternate the strategy, by adding some randomization, to obtain analogous results for all posets.

## Acknowledgement

The author wish to acknowledge fruitful discussions with Piotr Micek.

## References

- [1] William Feller, *An introduction to probability theory and its applications, volume 1*, Wiley, January 1968.
- [2] T. Ferguson, *Who solved the secretary problem?*, *Statistical Science* **4** (1989), 282–289.

- [3] Nicholas Georgiou, Malgorzata Kuchta, Michal Morayne, and Jaroslaw Niemiec, *On a universal best choice algorithm for partially ordered sets*, *Random Struct. Algorithms* **32** (2008), no. 3, 263–273.
- [4] A. V. Gnedin, *Multicriteria extensions of the best choice problem: sequential selection without linear order*, *Contemp. Math* **125** (1992), 153–172.
- [5] Michal Morayne, *Partial-order analogue of the secretary problem: the binary tree case*, *Discrete Mathematics* **184** (1998), no. 1-3, 165–181.
- [6] J. Preater, *The best-choice problem for partially ordered objects*, *Oper. Res. Lett.* **25** (1999), no. 4, 187–190.
- [7] Inc. Wolfram Research, *Mathematica edition: Version 7.0*, Wolfram Research, Inc., 2008.

