An Analysis of the Secretary Problem with Competing Hypothesis

Matei Gabriel Cosa
Supervisors: prof. Andrea Celli, prof Marek Elias

December 2023

1 Introduction

In theoretical computer science, decision theory and mathematics, optimal stopping theory is concerned with problems where one must choose a time to take a particular action in some optimal way. A famous example of such a problem is the secretary problem.
An intuitive description of this problem goes as follows: suppose one is a manager that interviews a group of n candidates for a particular job. The skill of each candidate is measured by some real number unknown to the interviewer initially. The interviews take place sequentially and after each interview, the score of that candidate becomes available to the manager. There is a catch, however, in that the manager can only choose the last interviewed candidate. Therefore, the manager cannot go back to a candidate she has seen multiple timestamps ago. In the most basic formulation, the manager wants to maximize the score of the candidate chosen for the position.
For the original setting, the best known algorithm is the odds algorithm proposed by Bruss (2000) [1] which guarantees that the best candidate will be chosen with probability at least $1/e$. This success probability is close to 36% and represents a relatively surprising result with an elegant proof.
Following the emergence of machine learning as a field with a remarkable potential for a multitude of applications, theoretical computer science has started to analyse the impact of predictions for the construction of algorithms [2]. These predictions can be seen as ML algorithms that provide additional information to the decision-maker. However, not all predictions might be accurate. Therefore, it is necessary to devise an approach to use the information coming from predictions in an efficient way. Such a setting is similar to experts problem in the field of online learning.

1.1 Problem formulation

We consider a set of n candidates with random ordering. This means each sequence of candidates σ can be seen as a permutation π over the set of candidates $C = \{1, 2, \ldots, n\}$. We evaluate this candidates by applying the function $v : C \to \mathbb{R}$

which is injective by assumption, meaning no two candidates may have the same score. The goal of the decision-maker is to solve the following problem:

$$i^* = \underset{c \in C}{\text{argmax}}\ v(c)$$

We let $h : C \to \mathbb{R}$ be a hypothesis. If σ represents the sequence of candidates, then $v(\sigma(i))$ represents the true score of the candidate at position i. If $v(h(i)) = v(\sigma(i))\ \forall i$, then h is the true hypothesis. We will denote this hypothesis with h^* and assume it always belongs to our set of hypothesis $\mathcal{H}$. Any other hypothesis $h' \in \mathcal{H}$ contains exactly k elements $j_1, \cdots, j_k$ for which $v(h'(j_i)) \neq v(\sigma(j_i))$. This k differing elements allow us to eliminate false hypothesis.

1.2 Preliminaries

Our proofs require some basic elements of combinatorics. We will briefly cover the main mathematical tools in this sub-section.
The following results are assumed to be known:

- The number of permutations of n distinct elements is $n!$;
- There are $\binom{n}{k}$ ways of picking k elements from a set of n distinct elements without order;
- There are $\frac{n!}{(n-k)!}$ ways of choosing k elements from a set of n distinct elements with order;

Other mathematical results will be stated when used for specific proofs.

2 Two hypothesis

We first begin with a simplified version of the problem. In this setting, there are only two competing hypothesis, the true one and a false one. This formulation will be useful for deriving some fundamental results and gaining a deeper understanding of the problem. We will later generalise to the case with multiple competing hypothesis. We simplify the problem even further in the first theorem by assuming $k = 1$:

Theorem 2.1. *With two competing hypothesis and a single differing element, the probability of success is:*

$$P(success) = \frac{1}{2}$$

Proof. Let j be the position of the differing element. Then if i^* comes before j, we are guaranteed to eliminate the false hypothesis and choose the best candidate. Therefore:

$$P(success) = P[\pi(i^*) < \pi(j)] = \sum_{i=1}^{n} \frac{1}{n} \cdot \frac{(n-i)}{n} = \frac{1}{2}$$

□

We now move to the case when the false hypothesis differs from the true one for k elements $j_1, \cdots j_k$. The following theorem is true:

Theorem 2.2. *With two competing hypothesis and k differing elements, the probability of success is:*

$$P(success) = \frac{k}{k+1}$$

Proof. Firstly, we note that:

$$P(success) = 1 - P[\{\pi(i^*) < \pi(j_1)\} \cup \ldots \cup \{\pi(i^*) < \pi(j_k)\}]$$

To compute such a probability, we first make a series of observations for $1 \leq k < n$:

- There are $n!$ total permutations;
- The $k+1$ elements of interest i^*, j_1, …, j_k can be arranged (without order) in $\binom{n}{k+1}$ ways;
- Since we imposed $\pi(i^*) < \pi(j_1)$ and … and $\pi(i^*) < \pi(j_k)$, i^* must be the first of the $k+1$ elements, while the remaining k may be shuffled around in $k!$ ways;
- The remaining $n-k-1$ elements can be permuted in $(n-k-1)!$ ways.

Combining the previous observations, we derive:

$$\begin{aligned}
P[\{\pi(i^*) < \pi(j_1)\} \cup \ldots \cup \{\pi(i^*) < \pi(j_k)\}] &= \frac{\binom{n}{k+1}}{n!} \cdot k! \cdot (n-k-1)! \\
&= \frac{n!}{(k+1)!(n-k-1)!} \cdot \frac{k! \cdot (n-k-1)!}{n!} \\
&= \frac{1}{k+1},
\end{aligned}$$

from which the result follows trivially. □

For $k = 1$, we obtain a probability of $\frac{1}{2}$, while for large k (approaching n), the probability approaches 1. The formula holds for $1 \leq k < n$.

3 Multiple hypothesis

We now approach the general setting with l copeting hypothesis. Let h^* be the true hypothesis and i^* be the position of the true best candidate. We assume that each of the $l-1$ wrong hypothesis differs from the true h^* in exactly k places: j_1, …, j_k.

3.1 Deterministic strategy

The deterministic strategy we will follow always picks the hypothesis with the highest promised best and sticks with it until we either find this value, or we find an element which is different than what the hypothesis predicted. In this latter case, we switch to the hypothesis that promises the second highest score, and so on.
If $b : \{h_1, h_2, \ldots, h_l\} \to \mathbb{R}$ is the function that outputs the promised highest score predicted by each hypothesis, let us assume without loss of generality that $h_1, h_2, \ldots, h_l$ are arranged such that $b(h_1) > b(h_2) > \cdots > b(h_l)$. Then let us assume $h^* = h_m$, i.e. the true hypothesis is at position m in the sequence of ordered hypothesis.

Lemma 3.1. *The following are true for this setting:*

1. *By employing the deterministic strategy, our algorithm never choose hypothesis h_i with $i > m$;*
2. *For a given i^*, successfuly finding the true best candidate requires eliminating all hypothesis h_i with $i < m$ before reaching position i^*;*
3. *The order in which we eliminate hypothesis h_i with $i < m$ is irrelevant;*
4. *The hypothesis $h_1, h_2, \ldots, h_l$ are independent of each other in the sense that eliminating h_i before position k does not depend on whether h_j was eliminated or not before position k for all $j \neq i$.*

Proof. We will provide a short proof of each claim:

1. Suppose our algorithm chooses $h^* = h_m$. Then at each element in the sequence of secretaries, the prediction will coincide with the true value and the algorithm would never switch again;
2. Suppose we reach position i^* and our algorithm chooses h_i with $i < m$. Then we will miss the true best candidate located in position i^* by not selecting her and waiting for a higher value promised by h_i;
3. Let j_i with $i < m$ represent the positions of the first differing element from each hypothesis. Then at position k, the algorithm will choose h_p such that $j_i < k \quad \forall i < p$. Clearly, this result requires no assumption on the relative ordering of the j_i's;
4. Since the algorithm observes the predictions of all hypothesis, at each step k, the hypothesis h_i such that $j_i = k$ are eliminated. This iterative elimination procedure makes no assumption on the previously eliminated or currently available hypothesis.

□

We first restrict ourselves to the scenario with two competing hypothesis, meaning there exists a true hypothesis and one with k differing elements. This is effectively the scenario explored in the previous section, but approached from a different perspective. The following theorem is true:

Theorem 3.2. *For a hypothesis h with k differing elements $j_1, j_2, \ldots, j_k$ w.r.t. to the true h^*, the probability of successfully rejecting h before position i^* is as follows:*

$$P(success|i^* = i) = 1 - \frac{(n-i)! \cdot (n-k-1)!}{(n-1)! \cdot (n-k-i)!}$$

Proof. For a given $i^* = i$, we first compute the probability of failing to reject h before position i:

$$P(failure|i^* = i) = P[\{i < \pi(j_1)\} \cup \ldots \cup \{i < \pi(j_k)\}]$$

We now make the following remarks:

- By fixing i^* in position i, there are only $n-1$ places for the remaining elements, so $(n-1)!$ total permutations;
- There are $\binom{n-i}{k}$ ways of picking $j_1, j_2, \ldots, j_k$ without order;
- For any configuration of $j_1, j_2, \ldots, j_k$ within the sequence, we can permute these elements in $k!$ ways;
- The remaining $n-k-1$ elements can be permuted in $(n-k-1)!$ ways.

Therefore:

$$P(failure|i^* = i) = \frac{\binom{n-i}{k} \cdot k! \cdot (n-k-1)!}{(n-1)!} = \frac{(n-i)! \cdot (n-k-1)!}{(n-1)! \cdot (n-k-i)!},$$

from which the final result follows. □

To verify that this is in fact correct, we show that using this result we reach the same overall probability of failure for two hypothesis as with the previous approach:

$$\begin{aligned}
P(failure) &= \frac{1}{n} \sum_{i=1}^{n-k} P(failure|i^* = i) \\
&= \frac{1}{n} \sum_{i=1}^{n-k} \frac{(n-i)! \cdot (n-k-1)!}{(n-1)! \cdot (n-k-i)!} \\
&= \frac{1}{n} \cdot \frac{k! \cdot (n-k-1)!}{(n-1)!} \sum_{i=1}^{n-k} \binom{n-1}{k} \\
&= \frac{1}{n} \cdot \frac{k! \cdot (n-k-1)!}{(n-1)!} \cdot \frac{n}{k+1} \cdot \binom{n-1}{k} \\
&= \frac{1}{k+1},
\end{aligned}$$

which coincides with our previous result. We now generalise the result to the case of multiple competing hypothesis by stating the following theorem:

Theorem 3.3. *Let l competing hypothesis $h_1, \dots, h_m, \dots, h_l$ be in decreasing order of their promised best candidate value. If each $h_i \neq h_m$ has exactly k differing elements with respect to h^* the probability of successfully choosing $h^* = h_m$ is:*

$$P(success) = \frac{k}{n} + \frac{1}{n} \sum_{i=1}^{n-k} \left[1 - \frac{\binom{n-i}{k}}{\binom{n-1}{k}} \right]^{(m-1)}$$

Proof. We begin the proof with the following equality:

$$P(success) = \frac{1}{n} \sum_{i=1}^{n} P(\{success(h_1)|i^* = i\} \cap \dots \cap \{success(h_{m-1})|i^* = i\}),$$

where $P(\{success(h_j)|i^* = i)$ represents the probability of successfully rejecting $h_j \neq h_m$ before position i. Then using 3.1.4 (independence), we can write the joint probability in the following way:

$$\begin{aligned} P(success) &= \frac{1}{n} \cdot \sum_{i=1}^{n} P(success(h_1)|i^* = i)^{m-1} \\ &= \frac{k}{n} + \sum_{i=1}^{n-k} P(success(h_1)|i^* = i)^{m-1} \\ &= \frac{k}{n} + \frac{1}{n} \cdot \sum_{i=1}^{n-k} \left[1 - \frac{(n-i)! \cdot (n-k-1)!}{(n-1)! \cdot (n-k-i)!} \right]^{(m-1)}, \end{aligned}$$

from which the final result can be derived by using binomial coefficients. □

We would now like to understand how this probability depends on the number of hypothesis, differing elements and number of candidates. To conduct this analysis, we state the following lemma:

Lemma 3.4. *Let $f : \mathbb{Z}_+ \to [0, 1]$ with $f(k) = P(success)$, where $P(success)$ is the result from theorem 3.3. Then f is a non-decreasing function of k.*

Proof. Let $j^* = \min\{\pi(j_1), \dots, \pi(j_k)\}$. Then we note that:

$$P(success(h)|i^* = i) = P(j^* < i)$$

Inserting a new j_{k+1} into the sequence of candidate scores predicted by h can be done in the following ways:

- if $\pi(j_{k+1}) < j^*$, then $P(success(h)|i^* = i)$ increases, because we have a greater chance of rejecting h at an earlier step;
- if $\pi(j_{k+1}) > j^*$, then $P(success(h)|i^* = i)$ remains unchanged.

□

We further state without proof a calculus result which will prove useful in our next theorem:

Lemma 3.5. *Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be non-decreasing and continuous and $S = \sum_{i=1}^{n} f(i)$. Then:*

$$f(1) + \int_1^n f(x)dx \leq S \leq f(n) + \int_1^n f(x)dx$$

We are now ready to state the main result of this subsection:

Theorem 3.6. *Let there be l competing hypothesis over a sequence of n candidates. For any false hypothesis, there are exactly k differing elements with respect to the true sequence. Then:*

$$P(success) \geq \frac{1}{m},$$

where m represents the position of the correct hypothesis in the descending ordering of hypothesis according to best candidate promised value.

Proof. Using lemma 3.4., $P(success)$ is greater or equal to the value obtained when $k = 1$:

$$\begin{aligned} P(success) &\geq \frac{1}{n} + \frac{1}{n} \cdot \sum_{i=1}^{n-1} \left[1 - \frac{n-i}{n-1}\right]^{(m-1)} \\ &\geq \frac{1}{n} + \frac{1}{n \cdot (n-1)^{m-1}} \cdot \sum_{i=1}^{n-1} (i-1)^{(m-1)} \\ &\geq \frac{1}{n} + \frac{\frac{(n-2)^m}{m}}{n \cdot (n-1)^{m-1}}, \end{aligned}$$

where the last inequality comes from using lemma 3.5 with $f(i) = (i-1)^{(m-1)}$. Then, we compute the limit as the sequence length tends to infinity:

$$\lim_{n \to \infty} P(success) = \frac{1}{m}$$

By noting that $P(success)$ is decreasing in n, we obtain the final result. □

3.2 Random strategy

Since randomization often helps in online learning and optimal stopping problems, we also propose a random algorithm. More specifically, suppose now that before we see a new candidate, we pick a hypothesis at random among the ones that have not been rejected yet.
In this new setting, the probability of success, as well as other combinatorial results are stochastic variables themselves. This meas we will be interested in computing expectations with respect to the distribution of candidates within the hypothesis sequences. We thus introduce the following definition:

Definition 3.1. *Let R_i be a random variable that counts the number of hypothesis rejected before position i. Furthermore, let $R_{i,j}$ be a random variable that captures whether h_j is rejected before position i. This translates to:*

$$R_{i,j} = \begin{cases} 1, if\ h_j\ rejected\ before\ position\ i \\ 0, otherwise \end{cases}$$

Using fundamental probility theory, we derive the following result:

Lemma 3.7. *The random variables R_i have the following distribution:*

$$R_i \sim Binomial(l-1, p_i),$$

where $p_i = p(success(h_j)|i^* = i) = 1 - \frac{(n-i)! \cdot (n-k-1)!}{(n-1)! \cdot (n-k-i)!}$ *from theorem 3.2.*

Proof. By definition 3.1, $R_{i,j} \sim Bernoulli(p_i)$. Furthermore, by lemma 3.1.4, we have that $R_{i,j}$ are independent, meaning $R_{i,j}$ are i.i.d. Bernoulli variables. Also by definition 3.1, $R_i = \sum_{j:h_j \neq h^*} R_{i,j} \sim Binomial(l-1, p_i)$, which concludes the proof. □

In this setting, the probability of success is given by the following lemma:

Lemma 3.8. *For a given sequence of numbers r_i representing the number of hypothesis rejected before position i, the probability of success is:*

$$P(success) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{l - r_i}$$

Proof. Firstly, for a given $i^* = i$ representing the position of the true best candidate, we have:

$$P(success|i^* = i) = \frac{1}{l - r_i},$$

since the algorithm picks one hypothesis at random among the ones that have not been reject until position i. It thus follows that:

$$P(success) = \frac{1}{n} \sum_{i=1}^{n} P(success|i^* = i) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{l - r_i}$$

□

We are now ready to compute the expected probability of success. This is given by the following theorem:

Theorem 3.9. *Let us consider l competing hypothesis $h_1, \ldots, h_m, \ldots, h_l$ with each $h_i \neq h_m$ having exactly k differing elements with respect to h^*. Then:*

$$E[P(success)] = \frac{k}{n} + \frac{1}{n \cdot l} \sum_{i=1}^{n-k} \frac{1 - p_i^l}{1 - p_i}$$

Proof. By lemmas 3.7 and 3.8, we have the following:

$$\begin{aligned}
E[P(success)] &= E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{l - R_i}\right] \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} E\left[\frac{1}{l - R_i}\right] \\
&= \frac{k}{n} + \frac{1}{n} \cdot \sum_{i=1}^{n-k} \sum_{j=0}^{l-1} \frac{1}{l-j} \binom{l-1}{j} p_i^j (1-p_i)^{l-j-1} \\
&= \frac{k}{n} + \frac{1}{n} \cdot \sum_{i=1}^{n-k} \sum_{j=0}^{l-1} \frac{1}{l \cdot (1-p_i)} \cdot \binom{l}{j} p_i^j (1-p_i)^{l-j} \\
&= \frac{k}{n} + \frac{1}{n \cdot l} \cdot \sum_{i=1}^{n-k} \frac{1 - p_i^l}{1 - p_i}
\end{aligned}$$

□

Similarly to the setting in which we employed a deterministic algorithm, we are interested in the behavior of the expected probability of success when the sequence length tends to infinity. We thus derive the following result:

Theorem 3.10. *Let us consider l competing hypothesis $h_1, \ldots, h_m, \ldots, h_l$ with each $h_i \neq h_m$ having exactly k differing elements with respect to h^*. Then:*

$$E[P(success)] \geq \frac{1}{l}$$

Proof.

$$\begin{aligned}
E[P(success)] &= \frac{k}{n} + \frac{1}{n \cdot l} \sum_{i=1}^{n-k} (1 + p_i + \cdots + p_i^{l-1}) \\
&= \frac{k}{n} + \frac{1}{n \cdot l} (n - k + 1 + \ldots)
\end{aligned}$$

By taking the $n \to \infty$ and noting that $\sum_{i=1}^{n-k} p_i^q < 1$ we obtain:

$$\lim_{n \to \infty} E[P(success)] = \frac{1}{l}$$

Since $E[P(success)]$ is decreasing in n, we obtain the final result. □

An interesting remark has to do with the dependence of k on n. In the previous computations, we assumed k is fixed. If, however, $k = \alpha n$ with $\alpha \in [0, 1]$, then the expected probability of success would be at least α.

4 Conclusion

Our analysis provided several interesting findings. Firstly, with two competing hypothesis, the success probability is very high and depends on the number of wrongly predicted elements. For this setting, we can see a clear improvement over the classical algorithm. This performance gain can be directly attributed to the power of predictions.
When it comes to case of several competing hypothesis, the results are not as impressive. Our results are show that on average, the deterministic algorithm performs better than randomly selecting a hypothesis. However, when taking k is equal to a fraction of n, one obtains relatively good performance.
Perhaps the general takeaway has to do with the fact that the benefits of predictions are outweighed by having a large number of false hypothesis. This is intuitive, but the nature of this trade-off deserves more careful attention.
Another further research direction could investigate the average quality of the chosen candidate instead of computing the success probability. This setting could be more realistic and perhaps more interesting results.

References

[1] F. Thomas Bruss. Sum the odds to one and stop. *The Annals of Probability*, 28(3):1384 – 1391, 2000.

[2] Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. *CoRR*, abs/2006.09123, 2020.