Knowing when to stop

Tommaso Green[*] & Nicola Gulmini[†]
University of Padua
Department of Information Engineering

Abstract

The purpose of this study is to investigate the secretary problem using a simple but important theorem that provides the optimal stopping rule and the related algorithm for all problems in the same class. We also analyzed an extension of the secretary problem, called the hiring problem. In this scenario we studied the behaviour of some strategies and their differences, focusing on the comparison of two main strategies which expect to hire candidates who are above the mean or the median.

Despite the similarity of these two problems, the obtained results are quite different and non-trivial, and lead us to make an interesting discussion and to think about more realistic extensions.

Keywords: optimal stopping; odds theorem; secretary problem; hiring problem

Imagine an administrator who wants to hire the best secretary out of n rankable applicants for a position. On the day of the interview, the applicants show up one by one in a random order. If the decision can be deferred to the end, the problem becomes trivial, because it would be sufficient to wait for the n-th applicant and select the best overall. The difficulty is that the decision about each particular applicant needs to be made immediately after the interview. Once rejected, an applicant cannot be recalled. During the interview, the administrator gains information sufficient to rank the applicant among all interviewed applicants, but is unaware of the quality of yet unseen applicants. The question is about the strategies the employer can use in order to hire the best candidate out of the n available. Said strategies are called *stopping rules*, since they prescribe when it is best to stop in the sequence. The quest is to find the optimal stopping rule, which is the one that maximizes the probability of picking the best candidate.

Formally, let $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_n$ be indicators of independent events $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$, where n is known. We observe the events sequentially, and we call the event "$\mathcal{I}_j = 1$" a success. We want to find a stopping rule that maximizes $\mathbb{P}(\mathcal{I}_t = 1, \mathcal{I}_{t+1} = \mathcal{I}_{t+2} = \cdots = \mathcal{I}_n = 0)$ over all t and the value of this probability. In other words, we want to stop at the index that maximizes the probability of being the last success. We will see that theorem 1.1 provides us with an optimal strategy for this class of problems. Note that several different sequential decision problems can be modelled in this way, especially if we admit some modifications (random number of events, i.e. n unknown; more of one secretary to hire...).

In this work we considered two main papers: in Bruss (2000) the author provides a simple algorithm that can find the optimal strategy and the value of the associated probability, while in Broder et al. (2010) the *hiring problem* is presented, similar in spirit to the secretary problem but at the same time quite different in the hypotheses. Two strategies, called *Lake Wobegon strategies*, are deeply analyzed and compared.

[*] tommaso.green@studenti.unipd.it
[†] nicola.gulmini@studenti.unipd.it

1 The Odds Theorem

Theorem 1.1 (Odds-theorem, Bruss (2000))**.** *Let $\{\mathcal{I}_j\}_{j=1}^n$ be a sequence of independent Bernoulli random variables with parameters $p_j = \mathbb{E}(\mathcal{I}_j) = 1 - q_j$ and let $r_j = p_j/q_j$ be the* odds*. The optimal rule to stop on the last success (if any) exists and comprises the two following steps:*

1. *wait for an amount of time equal to*

$$s := \sup\left\{1, \sup\left\{k \in [1,n] : \sum_{j=k}^{n} r_j \geq 1\right\}\right\} \tag{1}$$

   with $\sup(\varnothing) := -\infty$*;*

2. *stop on the first $k \geq s$ such that $\mathcal{I}_k = 1$.*

The probability that this strategy succeeds is $V(n) := \left(\prod_{j=s}^n q_j\right)\left(\sum_{j=s}^n r_j\right)$.

Proof. We have to consider two cases.

1. In the first case we assume that $p_j < 1$ for all $j \in [1,n]$. For a fixed $k \in [0, n-1]$, we will refer to the set of $\mathcal{I}_{k+1}, \mathcal{I}_{k+2}, \dots, \mathcal{I}_n$ as the tail from k onwards and the number of success in it will be modeled by the r.v. $\mathcal{S}_k = \mathcal{I}_{k+1} + \mathcal{I}_{k+2} + \cdots + \mathcal{I}_n$. We want to compute the index k^* that maximizes the probability $\mathbb{P}(S_{k^*} = 1)$ of having only one success in the tail, which will be the one to stop at.

   To compute the probability $\mathbb{P}(\mathcal{S}_k = 1)$, we first notice that it is possibile to rewrite that event as the union of disjoint events E_j as follows:

$$\{\mathcal{S}_k = 1\} = \bigcup_{j=k+1}^{n} E_j, \quad E_j = \text{“only } \mathcal{I}_j = 1 \text{ in the tail } \mathcal{I}_{k+1}, \mathcal{I}_{k+2}, \dots, \mathcal{I}_n\text{”}. \tag{2}$$

   This implies that $\mathbb{P}(\mathcal{S}_k = 1) = \sum_{j=k+1}^n \mathbb{P}(E_j)$, with $\mathbb{P}(E_j) = p_j n_j$ where n_j represents the probability that all variables in the tail other than $\mathcal{I}_j$ are equal to 0 and it can be computed as $\prod_{i=k+1, i\neq j}^n q_i$. For example, $\mathbb{P}(E_{k+3}) = q_{k+1}q_{k+2}p_{k+3}\cdots q_{n-1}q_n$. In the following we use the fact that $p_j < 1$ for all $j \in [1,n]$ hence $q_j > 0$. Putting it all together we have:

$$\begin{aligned}\mathbb{P}(\mathcal{S}_k = 1) &= \sum_{j=k+1}^{n} \mathbb{P}(E_j) = \sum_{j=k+1}^{n} p_j n_j = \sum_{j=k+1}^{n} p_j \prod_{\substack{i=k+1\\ i\neq j}}^{n} q_i \\ &= \sum_{j=k+1}^{n} \frac{p_j}{q_j} \cdot q_j \prod_{\substack{i=k+1\\ i\neq j}}^{n} q_i = \sum_{j=k+1}^{n} r_j \prod_{i=k+1}^{n} q_i.\end{aligned} \tag{3}$$

   Now we focus on finding the optimal rule for the stopping problem. Let $\mathcal{T}$ be the set of all rules t such that $\{t = k\} \in \sigma(\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_k)$. Consider $\mathcal{C} \subset \mathcal{T}$ which comprises all the rules that wait for a fixed time k and then stop on the first success. We can notice that, if we apply a rule $t \in \mathcal{C}$, we stop on the last success (i.e. t succeeds) iff $\mathcal{S}_t = 1$. Hence, maximizing the probability of success of a rule $t \in \mathcal{C}$ is equivalent to maximizing the probability that $\mathcal{S}_t = 1$. The optimal rule in $\mathcal{C}$ is therefore the index k^* that maximizes $\mathbb{P}(\mathcal{S}_{k^*} = 1)$. It's easy to see that $\mathbb{P}(\mathcal{S}_k = 1)$ is always a unimodal function for $k = 0, 1, \dots, n-1$ which means it only has one maximum.

   (a) To show this, we assume that $r_2 + r_3 + \cdots + r_n \geq 1$ and we relate the fact that $\mathbb{P}(\mathcal{S}_k = 1)$ is increasing to the sum of the odds r_j:

$$
\begin{aligned}
\mathbb{P}(\mathcal{S}_k = 1) \text{ is increasing in } k &\iff \mathbb{P}(\mathcal{S}_k = 1) < \mathbb{P}(\mathcal{S}_{k+1} = 1) \\
&\iff \left(r_{k+1} + \sum_{j=k+2}^{n} r_j\right)\left(q_{k+1} \prod_{i=k+2}^{n} q_i\right) < \sum_{j=k+2}^{n} r_j \prod_{i=k+2}^{n} q_i \\
&\iff \left(r_{k+1} + \sum_{j=k+2}^{n} r_j\right) q_{k+1} < \sum_{j=k+2}^{n} r_j \\
&\iff \frac{1-q_{k+1}}{q_{k+1}} q_{k+1} < (1-q_{k+1}) \sum_{j=k+2}^{n} r_j \\
&\iff 1 < \sum_{j=k+2}^{n} r_j =: f(k) \qquad (4)
\end{aligned}
$$

Notice that $f(k)$ is a monotonically decreasing function of k.

This means that there will be a point $k^* = \min\{k \in [0, n-2] : f(k) \leq 1\}$ such that $f(k) \leq 1 \; \forall k \geq k^*$. By using 4 we conclude that from that point onwards $p^* := \mathbb{P}(\mathcal{S}_{k^*} = 1) \geq \mathbb{P}(\mathcal{S}_{k^*+1} = 1) \geq \cdots \geq \mathbb{P}(\mathcal{S}_n = 1)$. In other words, k^* is the point of maximum for $\mathbb{P}(\mathcal{S}_k = 1)$ and it is reached when the sum of the odds becomes for the first time less or equal than 1.

We now know how to compute the best waiting time k^* related to the class $\mathcal{C}$ but we need to show that this rule is the best in general. To do so, we extend $\mathcal{C}$ to a class $\mathcal{C}'$ which comprises rules that wait for a random time W, with $\{W = k\} \in \sigma(\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_k)$, and then stop on the first success. Since $\mathcal{S}_k$ is by definition independent from $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_k$, this implies that the optimal rule in $\mathcal{C}'$ has the same optimal probability of success p^* of the one in $\mathcal{C}$ as $\mathbb{P}(\mathcal{S}_W = 1 | \mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_k, W = k) = \mathbb{P}(\mathcal{S}_k = 1)$ for every k. Immediately we have that $\mathbb{P}(\mathcal{S}_{k^*} = 1) \geq \mathbb{P}(\mathcal{S}_W = 1 | W \geq k^*)$. To conclude this part, we just need to link the condition $f(k) \leq 1$ with the one stated in the theorem 1. Due to the definition of k^*, we have that $k' := k^* - 1 = \max\{k \in [0, n-2] : f(k) > 1\}$. We can now notice that $f(k') = \sum_{j=k^*+1}^{n} r_j = \sum_{j=s}^{n} r_j$ which is the sum in theorem and therefore we conclude that $s = \sup\left\{k \in [1, n] : \sum_{j=k}^{n} r_j \geq 1\right\}$.

(b) The case where $r_2 + r_3 + \cdots + r_n < 1$ is trivial, since this means by 4 that $\mathbb{P}(\mathcal{S}_k = 1)$ is a decreasing function for every k. Hence its point of maximum is reached at $k^* = 0$ and therefore the rule imposes to stop on the very first success encountered.

To sum up, the optimal rule imposes to stop at the first success at index $k \geq s$, with $s = k^* + 1$, and has a probability of stopping at the very last success of $V(n) := \mathbb{P}(\mathcal{S}_{k^*} = 1) = \left(\prod_{j=s}^{n} q_j\right)\left(\sum_{j=s}^{n} r_j\right)$.

2. Finally, assume that $p_j = 1$ for some $j \in [1, n]$. We have to consider two cases:

   (a) If $p_n = 1$ then the optimal stopping has to happen on n. In fact, since $r_n = \infty$, by following the definition of s in the statement we have $s = n$.

   (b) If $p_n < 1$, let j^* be the last index $j \in [1, n-1]$ such that $p_{j^*} = 1$. This implies that $p_{j^*+1} < 1, \ldots, p_n < 1$.

      i. If $s \geq j^* + 1$, i.e. the sum of the odds exceeds or equals 1 before j^*, due to the probabilities of the events being less than 1 we can use the same reasoning proposed in the first case. This can be done by setting $\mathcal{I}'_1 = \mathcal{I}_{j^*+1}, \mathcal{I}'_2 = \mathcal{I}_{j^*+2}, \ldots, \mathcal{I}'_h = \mathcal{I}_n$ with $h = n - j^*$.

      ii. If $s \leq j^*$, we cannot stop on any $j < j^*$ or we would lose with probability 1. This means that the first valid index to stop at is j^*. This implies $s = j^*$ and concludes the proof.

□

1.1 The Odds-algorithm

The following algorithm is based on the previous result.

Algorithm 1 Odds-algorithm $(p_1, \ldots, p_n)$

$Q \leftarrow (1 - p_n)$
$R \leftarrow p_n / Q$
if $R \geq 1$ **then**
  return $s = n, V(n) = Q \cdot R$
for $i \leftarrow n - 1$ **down to** 1 **do**
  $q_i \leftarrow 1 - p_i$
  $R \leftarrow R + p_i / q_i$
  $Q \leftarrow Q \cdot q_i$
  if $R \geq 1$ **then**
    return $s = i, V(n) = Q \cdot R$

The returned index s is the time from which onwards it is optimal to stop on the first encountered success. The output $V(n)$ is the value of the optimal probability. The algorithm is correct because it follows directly from the theorem 1.1.

1.2 A tipical lower bound

In the following we present the proof in Bruss (2000) of the typical lowerbound of $1/e$ to the probability of success of the strategy of the Odds-theorem $V(n)$.

Theorem 1.2 (A tipical lower bound for $V(n)$, Bruss (2000))**.** *Let $s(n)$ be the index given by the Odds-theorem. If $R_{s(n)} = \sum_{j=s(n)}^{n} r_j = R$ then*

1. $V(n) = \sum_{j=s(n)}^{n} r_j \prod_{j=s(n)}^{n} q_j > Re^{-R}$
2. *If $R_{s(n)} \to 1$ with $R_{s(n)}^{(2)} = \sum_{k=s(n)}^{n} r_k^2 \to 0$ as $n \to \infty$, then $V(n) \to 1/e$.*

Proof. 1. By rewriting $R_{s(n)} = \sum_{j=s(n)}^{n} r_j = R$ as $\sum_{j=s(n)}^{n}(1/q_j - 1)$ by some simple algebraic steps we can obtain the arithmetic mean of the $1/q_j$ as $\frac{1}{m}\sum_{j=s(n)}^{n} 1/q_j = \frac{R+m}{m}$, where $m = n - s(n) + 1$. We can upper bound the geometric mean of the q_j with their arithmetic mean using the AM-GM inequality as follows:

$$\sqrt[m]{\prod_{j=s(n)}^{n} 1/q_j} = \sqrt[m]{1/\prod_{j=s(n)}^{n} q_j} \leq \frac{R+m}{m} = 1 + \frac{R}{m} \tag{5}$$

This implies that $\prod_{j=s(n)}^{n} q_j \geq (1 + R/m)^{-m}$. By the fundamental limit $(1 + R/m)^{-m} \to e^{-R}$ as $m \to \infty$, we can conclude that:

$$V(n) = R \prod_{j=s(n)}^{n} q_j \geq R(1 + R/m)^{-m} > Re^{-R}. \tag{6}$$

2. Let $Q_{s(n)}^{-1} = \prod_{k=s(n)}^{n} \frac{1}{q_k} = \prod_{k=s(n)}^{n}(1 + r_k)$. Taking the logarithm and using the inequality $\ln(1 + r_k) \geq r_k - r_k^2$, we can write $-\ln(Q_{s(n)}) = \sum_{k=s(n)}^{n} \ln(1 + r_k) \geq \sum_{k=s(n)}^{n}(r_k - r_k^2) = R_{s(n)} - R_{s(n)}^{(2)}$. By exponentiating both sides we have $Q_{s(n)} \leq \exp\{-R_{s(n)} + R_{s(n)}^{(2)}\} \to 1/e$, thanks to the two limits stated in the hypothesis. Using this and the equation (6), we obtain that $Q_{s(n)} \to 1/e$ and thus $V(n) = Q_{s(n)} R_{s(n)} \to 1/e$, as $n \to +\infty$.

□

In a subsequent paper (Bruss (2003)), Bruss proved a much more general result related to the success probability $V(n)$.

Theorem 1.3. *If* $\sum_{j=1}^{n} r_j \geq 1$ *then* $V(n) > 1/e$.

This results is much more general since it gives a lower bound on the probability and not an asymptotic value. Moreover, the hypothesis are much simpler to verify than in 1.2.

1.3 The secretary problem

We want, as described in the beginning, to maximize the probability of stopping on the best candidate in a random permutation of n candidates, where all the $n!$ permutations are equally likely.

The random variable associated with the k-th candidate in a fixed permutation is $\mathcal{I}_j = 1$, i.e. a record, with probability $1/j$, hence $r_j = 1/(j-1)$. Basically, the event $\mathcal{A}_j$ means "the j-th candidate is better than the previous ones" and the indicator function is $\mathcal{I}_j = 1$ if the related event happens. All the $\mathcal{I}_j$'s are independent, so all the hypotheses of the Odds-theorem are verified. The Odds-algorithm returns $V(n) = Q_{s(n)} \cdot R_{s(n)}$ where $R_{s(n)} = \frac{1}{n-1} + \cdots + \frac{1}{s(n)-1}$ and $Q_{s(n)} = \frac{s-1}{n}$. Notice that the $\sum_{j=1}^{n} r_j \geq 1$ (r_1 diverges) for every n and therefore we can use 1.3 to assert that the probability of picking the best candidate, $V(n)$, is always at least 37% independently from the number of candidates. This is a very important results that highlights the power of the odds theorem.

Another very interesting property of the secretary problem is that $s(n)/n \to 1/e$ as $n \to +\infty$. This implies that, even without computing s, for large n the optimal stopping rule can be approximated by simply rejecting the first $\sim n/e$ interviewed applicants, and then to stop at the first applicant who is better then the previous ones (choosing the last applicant only if a better candidate is not found before).

2 The hiring problem

We now consider the hiring problem Broder et al. (2010), which extends the secretary problem. The setting is the same, but now we are not interested in stopping at the best candidate: we hire continuously, with no fixed limit. The first thing that comes to mind when thinking of an extension of the secretary problem is the criterion used to hire a new applicant. Should we hire a candidate whose quality score is better than the score of the previously hired one, or when is it above a given threshold? This will lead us to define and study different strategies. The most interesting aspect about each of the strategies is the implied trade-off between the rate at which employees are hired and their quality.

2.1 The model

Let's suppose that each applicant i has a quality score $Q_i \sim \mathcal{U}(0,1)$ and the applicants are continosly interviewed. The Q_i values are i.i.d, except for the first employee, who has a fixed $q \in (0,1)$. This is a necessity because if it had $U \sim \mathcal{U}(0,1)$ then the probability that the i-th applicant has an higher quality value would be $\mathbb{P}(Q_i \geq U) = 1 - U$ and the number of needed interviews to hire the next employee would be a geometric random variable $\mathcal{G}$ with parameter $1-U$ and expected value $\frac{1}{1-U}$ so

$$\mathbb{E}_{U \sim \mathcal{U}(0,1)}\left(\frac{1}{1-U}\right) = \int_0^1 \frac{1}{1-x} dx = \infty. \tag{7}$$

Here we used the *law of iterated expectations*, that is

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X) \tag{8}$$

for random variables X and Y defined on the same probability space. In this case the expected number of interviews is not even defined because the integral diverges.

Our interest in this work is to study two strategies for which the applicant i is accepted if and only if his/her Q_i is at least the *average* of the quality scores of the hired applicants, with a proper definition of average. These two strategies are called *hiring above the mean* and *hiring above the median* and they are reasonable approaches which could also be applied in real word scenarios. Since we are interested in the tradeoff between the quality of the hired applicants and the frequency of the hires, there is not a proper definition of *optimal* strategy.

2.2 Baseline strategies

In this section we consider some simple strategies to introduce concepts that will help us to analyze more complicated ones in 2.3 and 2.4.

2.2.1 t-threshold strategy

We may decide to hire applicants with a quality score that is at least a fixed quality threshold t: the probability to hire the applicant i is $\mathbb{P}(Q_i \geq t) = 1 - t \; \forall i$, and the quality values of the employees are $\sim \mathcal{U}(t, 1)$ from the beginning. This strategy gives us some guarantees about the quality $\geq t$ and the rate of growth of the company $1/(1-t)$ (that is a deterministic value), but it would be more reasonable to prefer a faster growing at the beginning, when there are a few employees, and a more selective hiring process as the company grows.

2.2.2 *Max* strategy

Consider the scenario where we start with only one employee with quality score $q \in (0, 1)$ and then hire an applicant if and only if his/her quality score is greater than the max quality score of the current employees.

For simplicity, let us consider the *gap* $G_i := 1 - Q_i$ rather than the quality value Q_i. It holds $G_i \sim \mathcal{U}(0, G_{i-1})$, where, for $i = 0$, $G_0 = g = 1 - q$ and $\mathbb{E}_{G_i \sim \mathcal{U}(0, G_{i-1})}(G_i | G_{i-1}) = \frac{1}{G_{i-1}} \int_0^{G_{i-1}} x dx = \frac{G_{i-1}}{2}$ so $\mathbb{E}(G_n) = g/2^n$ given all the gaps. Using this strategy, the gaps exponentially decrease as the company grows.

Also, we can represent the gaps as the recurrence $G_i = G_{i-1} U_i$ if $i \geq 1$, and $G_0 = g$, where $U_i \sim \mathcal{U}(0, 1)$ are independent uniform random variables. Solving the recurrence, we obtain $G_n = g \prod_{i=1}^{n} U_i$, so the gap of the n-th employee is affected by the gap g of the first one, an effect also present in the following strategies. Unfortunately, for similar reasons as in equation (7), the expected number of interviews between two hirings is actually not defined, which means large waits between hires.

2.3 Hiring above the mean

Let A_i be the average quality of the first $i + 1$ employees, with $A_0 = q$ such that $A_i = \frac{1}{i+1} \sum_{j=0}^{i} Q_j$. If the adopted strategy is to hire only candidates whose quality score is above the mean of the hired ones, as the number of employees grows, the average quality score approaches one.

Given a strategy, it is interesting to analyze the expectations, the hiring rate and the rate of convergence of the A_i. We redefine the *gap* as $G_i := 1 - A_i$. It follows that $\lim_{i \to +\infty} G_i = 0$ and $g = 1 - q$.

Lemma 2.1. *For any $t \geq 0$, the conditional distribution of G_{i+t} given G_i is the same as that of $G_i \prod_{j=1}^{t} (1 - \frac{U_j}{i+j+1})$, where the U_j's are independent random variables $\sim \mathcal{U}(0, 1)$.*

Proof. Let us proceed by induction on $t \geq 0$. For $t = 0$ we obtain that G_{i+0} is distributed like G_i, which is trivial. Now, for $t > 0$, suppose that the statement holds for $t - 1$. Then, conditioned on

$G_i, \ldots, G_{i+t-1}$, the quality score of the $(i+t)$-th employee $Q_{i+t} \sim 1 - G_{i+t-1}U_t$. So, conditioned on $G_1, \ldots, G_{i+t-1}$, we have

$$\begin{aligned}
G_{i+t} = 1 - A_{i+t} = 1 - \frac{1}{i+t+1}\sum_{j=0}^{i+t} Q_j &= 1 - \frac{\sum_{j=0}^{i+t-1} Q_j + Q_{i+t}}{i+t+1} \\
&= 1 - \frac{A_{i+t-1}(i+t) + Q_{i+t}}{i+t+1} \sim 1 - \frac{(1-G_{i+t-1})(i+t) + (1 - G_{i+t-1}U_t)}{i+t+1} \\
&= G_{i+t-1}\left(\frac{i+t+U_t+1-1}{i+t+1}\right) \sim G_{i+t-1}\left(1 - \frac{U_t}{i+t+1}\right), \qquad (9)
\end{aligned}$$

where all the steps consist of simple algebraic manipulations. Now, iterating this result, we can obtain the lemma $G_{i+t} \sim G_{i+t-1}(1 - \frac{U_t}{i+t+1}) \sim \cdots \sim G_i \prod_{j=1}^{t}\left(1 - \frac{U_j}{i+j+1}\right)$. □

Note that from the previous result we can obtain $G_n \sim g\prod_{j=1}^{n}\left(1 - \frac{U_j}{j+1}\right)$, and from the independence of $\{U_j\}_{j=1}^{n}$'s its expected value is $\mathbb{E}(G_n) = \mathbb{E}\left(g\prod_{j=1}^{n}\left(1 - \frac{U_j}{j+1}\right)\right) = g\prod_{j=1}^{n}\left(1 - \frac{1}{2(j+1)}\right)$.

Now we are interested about the asymptotic values of the expected gap and the expected number of interviews after hiring n applicants.

For the next proofs we will use the Taylor series of the natural logarithm: let x be a real number such that $|x| < 1$. Then, stopping at the third term,

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + o(x^3) = x - \frac{x^2}{2} + \Theta(x^3). \qquad (10)$$

Other relations that will be useful are

$$H_n = \sum_{j=1}^{n} \frac{1}{j} = \ln n + \Theta(1) \qquad (11)$$

where H_n is the n-th harmonic number, and

$$\sum_{j=1}^{n} (1/j)^2 \in \Theta(1). \qquad (12)$$

Proposition 2.2. $\mathbb{E}(G_n) \in \Theta(1/\sqrt{n})$.

Proof.

$$\begin{aligned}
\mathbb{E}(G_n) = g\prod_{j=1}^{n}\left(1 - \frac{1}{2(j+1)}\right) &= g\exp\left\{\sum_{j=1}^{n}\ln\left(1 - \frac{1}{2(j+1)}\right)\right\} \\
&\overset{(a)}{=} g\exp\left\{\sum_{j=1}^{n}\left(-\frac{1}{2(j+1)} + \Theta\left(\frac{1}{j^2}\right)\right)\right\} \\
&\overset{(b)}{=} g\exp\left\{-\frac{1}{2}\sum_{t=2}^{n+1}\left(\frac{1}{t}\right) + \Theta(1)\right\} = g\exp\left\{-\frac{1}{2}\left(H_n + \frac{1}{n+1} - 1\right) + \Theta(1)\right\} \\
&\overset{(c)}{=} g\exp\left\{-\frac{1}{2}\ln n + \Theta(1)\right\} \in \Theta\left(\frac{1}{\sqrt{n}}\right) \qquad (13)
\end{aligned}$$

where in (a) we used the Taylor expansion of the natural logarithm (10) stopped at the second term; in (b) we changed the index for the summation and we used the relation (12). In (c) we made use of (11) and $1/(n+1) \in \Theta(1)$. □

Lemma 2.3. $\mathbb{E}(1/G_n) = \frac{1}{g}\prod_{j=1}^{n}(j+1)\ln(1+1/j) \in \Theta(\sqrt{n})$.

Proof.

$$\begin{aligned}\mathbb{E}\left(\frac{1}{G_n}\right) &\overset{(a)}{=} \frac{1}{g}\prod_{j=1}^{n}\mathbb{E}\left(\frac{1}{1-U_j/(j+1)}\right) = \frac{1}{g}\prod_{j=1}^{n}\int_0^1 \frac{j+1}{j+1-u}du = \frac{1}{g}\prod_{j=1}^{n}(j+1)\ln\left(1+\frac{1}{j}\right)\\ &\overset{(b)}{=} \frac{(n+1)!}{g}\prod_{j=1}^{n}\left(\frac{1}{j}-\frac{1}{2j^2}+\Theta\left(\frac{1}{j^3}\right)\right) \overset{(c)}{=} \frac{n+1}{g}\prod_{j=1}^{n}\left(1-\frac{1}{2j}+\Theta\left(\frac{1}{j^2}\right)\right)\\ &= \frac{n+1}{g}\exp\left\{\sum_{j=1}^{n}\ln\left(1-\frac{1}{2j}+\Theta\left(\frac{1}{j^2}\right)\right)\right\} \overset{(d)}{=} \frac{n+1}{g}\exp\left\{\sum_{j=1}^{n}\left(-\frac{1}{2j}+\Theta\left(\frac{1}{j^2}\right)\right)\right\}\\ &\overset{(e)}{=} \frac{n+1}{g}\exp\left\{-\frac{1}{2}\ln n+\Theta(1)\right\} \in \Theta(\sqrt{n}) \end{aligned} \tag{14}$$

where in (a) we used the independence of the $\{U_j\}_{j=1}^{n}$'s, the linearity of the expected value and then its definition. In (b) we used (10) and the fact that $\prod_{j=1}^{n}(j+1) = (n+1)!$, while in (c) the same for the denominator. In (d) we used (10) stopped at the second term and in (e) analogous considerations as we have seen in the previous proof. □

Now, let T_n be the number of candidates that are interviewed before n are hired.

Proposition 2.4. $\mathbb{E}(T_n) = \frac{1}{g}\sum_{i=1}^{n}\prod_{j=1}^{i-1}(j+1)\ln(1+1/j) \in \Theta(n^{3/2})$.

Proof. Let T'_j be the number of candidates interviewed between the $(j-1)$-st hire and the j-th one, such that $T_n = \sum_{j=1}^{n} T'_j$. The conditional distribution of T'_j given G_{j-1} is $\mathcal{G}(G_{j-1})$ so its expected value is $1/G_{j-1}$ and, for the law of iterated expectations (8),

$$\mathbb{E}(T'_j) = \mathbb{E}\left(\mathbb{E}(T'_j|G_{j-1})\right) = \mathbb{E}(1/G_{j-1}). \tag{15}$$

Therefore

$$\mathbb{E}(T_n) = \sum_{j=1}^{n}\mathbb{E}(1/G_{j-1}) = \sum_{j=1}^{n}\Theta\left(\sqrt{j}\right) = n\Theta\left(\sqrt{n}\right) \in \Theta\left(n^{3/2}\right). \tag{16}$$

□

2.4 Hiring above the median

In this context we still assume that the probability distribution of the applicants' quality score is a uniform random variable $\mathcal{U}(0,1)$. The first employee has a fixed quality score $q \in (0,1)$ and we always mantain the invariant of having an odd number of employees. This implies that we always hire two new applicants with a quality score above the median M_k of the current $2k+1$ employees.

We hire an infinite number of candidates and $\lim_{k\to\infty} M_k = 1$. We will proceed in the following with the analysis of the *gap* $G'_k := 1 - M_k$ and we let $g = G'_0 = 1-q$. Both M_k and G'_k refer to the setting with $2k+1$ employees. Before proceeding, it is necessary to recall the definition of the *beta distribution* of parameters α and β denoted by the simbol $\mathcal{B}\mathrm{e}(\alpha,\beta)$ and defined using the *beta function* $B(\alpha,\beta)$.

$$p_{\mathcal{B}\mathrm{e}(\alpha,\beta)}(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1} \qquad B(\alpha,\beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$\mathop{\mathbb{E}}_{X\sim\mathcal{B}\mathrm{e}(\alpha,\beta)}(X) = \frac{\alpha}{\alpha+\beta} \qquad \mathop{\mathbf{Var}}_{X\sim\mathcal{B}\mathrm{e}(\alpha,\beta)}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

A fundamental property of the beta distribution is the following:

Proposition 2.5. *Given a sample of size n from a continuous uniform distribution, its k-th order statistic (i.e. its k-th smallest value) is distributed as a* $\mathcal{B}\mathrm{e}(k, n+1-k)$.

The first result that we prove from Broder et al. (2010) is the derivation of the conditional distribution of G'_{t+k} given G'_k.

Lemma 2.6. *For any $t, k \geq 0$ the conditional distribution of G'_{t+k} given G'_k is the same as $G'_k \prod_{j=1}^{t} B_j$, where the B_j's are independent and $B_j \sim \mathcal{B}\mathrm{e}(k+j+1, 1)$.*

Proof. The general idea is the following: in the setting with $2k+1$ employees, the quality scores of the employees above the median M_k (which are k in number) and that of the next two hires is distributed as $\mathcal{U}(M_k, 1)$. By adding the two new hires it is easy to see that the new median M_{k+1} will be the minimum of $k+1$ $\mathcal{U}(M_k, 1)$ r.v.s. by displaying the $2(k+1)+1$ scores in non-decreasing order. Formally, we proceed by induction on $t \geq 0$. For $t = 0$ the claim is trivially proven. We let $t > 0$ and we suppose that the claim holds up to $t-1$. In this case we have $2(k+t-1)+1$ employees. With the next two hires, the amount of employees above the median is $(k+t-1)+2 = k+t+1$. Let:

- $U_1, \dots, U_{k+t+1}$ be indipendent $\mathcal{U}(0,1)$ random variables;
- $U'_1, \dots, U'_{k+t+1}$ be a family of independent r.v.s where $U'_i \sim 1 - G'_{k+t+1} U_i$.

Now condition on $G' := (G_0, \dots, G'_{k+t-1})$. We now have that M_{t+k}, by the reasoning described at the beginning, is distributed as:

$$\begin{aligned} \min\{U'_1, \dots, U'_{k+t+1}\} &\sim \min\{1 - G'_{k+t-1}U_1, \dots, 1 - G'_{k+t-1}U_{k+t+1}\} \\ &\overset{\text{(a)}}{=} 1 - G'_{k+t-1} \max\{U_1, \dots, U_{k+t+1}\} \\ &\overset{\text{(b)}}{\sim} 1 - G'_{k+t-1} B_t \end{aligned} \tag{17}$$

In (a) we used the relationship between max and min and $M_k = 1 - G'_k$ and in (b) we used the proposition 2.5 since the max is the $(k+t+1)$-th order statistic. From this it immediately follows that the conditional distirbution of G'_{k+t} given G' is given by $G'_{k+t-1}\,\mathcal{B}\mathrm{e}(k+t+1, 1)$ and the thesis follows by induction. □

We can now use this result to obtain the expectation of the gap and also the number of interviews to reach $2k+1$ employees, in a similar fashion to that of propositions 2.2 and 2.4.

Proposition 2.7. $\mathbb{E}(G'_k) = g \prod_{j=1}^{k} (1 - \frac{1}{j+2}) \in \Theta(1/k)$.

We now shift the focus to T'_k, the number of interviews until there are $2k+1$ employees.

Proposition 2.8. $\mathbb{E}(T'_k) = k(k+1)/g$.

The following proposition helps us in tracking the mean quality of the first n employees.

Proposition 2.9. *Let A'_n be the mean quality score of the first n emplyees when hiring above the median. Then $\mathbb{E}(A'_n) = 1 - \Theta(\log n/n)$.*

2.5 Discussion and comparison

Given the results about hiring above the mean and hiring above the median, it is interesting to make a comparison to understand the similarities and the differences.

Firstly, note that the expected gap and the expected number of interviews depend, in the same way for both the strategies, on the gap of the first employee g. This quantity has a multiplicative effect on the expected gap, and appears as $1/g$ in the expected number of interviews. The multiplicative effect holds also for the gap of the n-th employee in the max strategy. It is interesting to note how much the initial conditions affect the entire trend of the company, independently from the selected strategy.

Now we compare the asymptotic behaviour of the expected mean quality of the employees and the expected number of interviews.

Given the number of employees n for both the strategies, the expected mean quality is:

- $1 - \Theta(1/\sqrt{n})$ for hiring above the mean;
- $1 - \Theta(\log n/n)$ for hiring above the median

and, since $\log n/n \in o(1/\sqrt{n})$, i.e. $\log n/n$ is asymptotically dominated by $1/\sqrt{n}$ as $n \to +\infty$, hiring above the median assures a greater expected mean quality of the employees as hiring above the mean.

On the other hand, the expected number of interviews is:

- $\Theta(n^{3/2})$ for hiring above the mean;
- $\Theta(n^2)$ for hiring above the median

and $n^{3/2} \in o(n^2)$, so hiring above the mean leads to a faster growth of the company as hiring above the median.

This difference between the behaviour of the strategies is not intuitive and defines a trade-off between the quality of the employees and the growth of the company.

3 Conclusions

In this work we studied two interesting problems. The first one, the secretary problem, allowed us to study the Odds-theorem and the related algorithm, which are simple and can be applied to a wide class of problems. The second one, the hiring problem, is a modification of the secretary problem that admits continuous hirings. Due to this novel aspect, the problem is enriched with interesting properties, which induce the discussions of the previous section. Further extensions of these settings can be considered: for example, in the secretary problem the odds strategy can assume an integral form when the number of candidates is not a fixed constant but a Poisson random process. No optimal strategy has been developed for the case where the r_i's are not known, even though dynamic programming can be used for small instances (Dendievel, 2012). For the hiring problem, interesting extensions presented in Broder et al. (2010) are those of practical interest such as the one where the quality score of each candidate is only estimated and not known exactly or the "hiring and firing" setting which also allows to remove underperforming employees.

References

A. Z. Broder, A. Kirsch, R. Kumar, M. Mitzenmacher, E. Upfal, and S. Vassilvitskii. The hiring problem and lake wobegon strategies. *SIAM Journal on Computing*, 39(4):1233–1255, 2010.

F. T. Bruss. Sum the odds to one and stop. *Ann. Probab.*, 28(3):1384–1391, 06 2000. doi: 10.1214/aop/1019160340. URL `https://doi.org/10.1214/aop/1019160340`.

F. T. Bruss. A note on bounds for the odds theorem of optimal stopping. *Annals of Probability*, 31(4): 1859–1861, 2003.

R. Dendievel. New developments of the odds theorem, 2012.