F. THOMAS BRUSS (Brussels)

Odds -theorem and monotonicity

Abstract Given a finite sequence of events and a well-defined notion of events being interesting, the Odds-theorem (Bruss (2000)) gives an online strategy to stop on the last interesting event. This strategy is optimal for independent events, and it is obtained in a straightforward way by an algorithm which is optimal itself (odds-algorithm). Here we study questions in how far the optimal value mirrors monotonicity properties of the underlying sequence of probabilities of events. We make these questions precise, motivate them, and then give complete answers. The motivation is enhanced by certain problems where it seems desirable to apply the odds-algorithm but where a lack of information does not allow to do so without incorporating sequential estimation. In view of this goal, the notion of a plug-in odds-algorithm is introduced. Several applications are included.

2010 Mathematics Subject Classification: 60G40.

Key words and phrases: Odds-algorithm; Plug-in odds-algorithm; Secretary problem; Multiple stopping; Group interviews; Games; Clinical trial; Prophet inequality.

1. Content, related work, and motivation. The Odds-theorem is an easy-to-apply theorem in the domain of optimal stopping. It gives an online strategy to stop on the very last interesting event of a given sequence of events. Its interest lies in the flexibility of the notion interesting, in its optimality for independent events, and in the odds-algorithm which flows from it. The Odds-theorem can also be useful for a similar setting with conditionally independent events. If not stated otherwise, we assume independence.

The present article is the second addendum to the Odds-Theorem, after the short note of B. (2003) proving the general lower bound $1/e$ for the win probability. ("B." in references stands for the author’s name.) In this article we mainly examine questions of monotonicity. Monotonicity questions arise, among other instances, when the decision maker has some influence on the order of events within the sequence, or when the length of the sequence may vary after observations have begun.

1.1. Related work. We will focus for the major part of the present article on the original Odds-theorem, including its continuous-time version
for processes with independent increments presented in the same paper (B. (2000)).

In the mean-time, there exist several interesting variations of the underlying model and/or its payoff-function. These include multiple stopping problems such as e.g. in Matsui and Ano (2014,2016), or modified payoffs (see e.g. Tamaki (2010,2011)), or again models modified in such a way that they may be helpful for related game problems in continuous-time, such as in Sza-jowski (2007). For a best-choice problem with dependent criteria see e.g. the earlier work of Gnedin(1994).

Moreover, the Odds-Theorem can be adapted for conditional independent events, as in Ferguson(2016) and other papers. For a review of results on developments of the Odds Theorem see Dendievel(2013).

An interesting alternative to the Odds-theorem has been recently given by Goldenshluger et al. (2019). The proof is complete under the hypothesis that the optimal strategy is monotone. Also, some recent results are related with the context of Odds-Theorem. We mention the work of Grau Ribas (private communication (2019)), who proved the dynamic programming approach to be efficient for a new interesting payoff function.

1.2. Motivation. Questions in how far optimal win probabilities mirror monotonicity properties of the underlying sequence of probabilities of events are shown to be relevant in several contexts. As we will see in the Section Applications (Section 5), they can help us to decide without computations which game we want to play if we have a choice, or may give us advice in scheduling interviews of candidates in order to make a better choice. They may help us to find answers for new questions concerning well-known selection problems, but also be important in different contexts, as for instance in planning or reorganizing clinical trials or sequential medical treatments.

The last example is one of those cases, quite typical in practice, where not all ingredients to apply the odds-algorithm are readily available. This may have several reasons. For example, the odds, which are needed, may not be known in advance since they depend on one or several unknown parameters. Then these parameters should be estimated, of course, and the idea is to insert these estimates sequentially into a suitably adapted version of the odds algorithm. Or, the horizon $n$ (number of observations) may not be fixed because one would like to model $n$ as the outcome of a random variable $N$. This is exemplified in Section 4.1 of Bruss(2000) which introduces the continuous-time version of the odds algorithm, and where $N$ is the outcome of the number of arrivals of a counting process with independent increments such as e.g. a Poisson process. If its rate is unknown, it must be sequentially estimated.

2. Odds-algorithm and plug-in odds-algorithms. In this section we recall the Odds-algorithm, and then present versions of the odds-algorithm
which incorporate sequential estimation. We call the latter ones plug-in odds algorithms. The author has used this terminology before in oral presentations, but Subsections 2.1 and 2.2 below are the first explicit reference in the literature.

We first recall the Odds-Theorem and the odds-algorithm. Let \((E_k)_{k=1,2,\ldots,n}\) be a sequence of \(n\) events defined on some probability space \((\Omega, A, P)\). Suppose we have a well-defined criterion according to which \(E_k\) is an interesting event. Let \(I_k = 1\{E_k \text{ is interesting}\}\). Correspondingly, we call \(P(I_k = 1)/P(I_k = 0)\) the odds of \(E_k\) being interesting. Further, let \(p_k = P(I_k = 1), q_k = 1 - p_k, r_k = p_k/q_k, k = 1, 2, \ldots, n,\) and let

\[
Q_k = \prod_{j=k}^{n} q_j, \quad R_k = \sum_{j=k}^{n} r_j. \tag{1}
\]

Then, if the \(I_k\) are independent, the index \(k\) maximizing \(Q_k R_k\), denoted by \(s\), yields the optimal value \(V(n)\), namely \(V(n) = Q_s R_s\). This index \(s\) is called the optimal threshold. The optimal strategy is to stop at the first index \(t \geq s\) such that \(I_t = 1\). Here \(s = 1\) if \(R_1 < 1\), otherwise \(s = \max\{1 \leq k \leq n : R_k \geq 1\}\).

This is the Odds-theorem of B. (2000) from which follows the odds-algorithm (B. (2000), sect. 2.1).

2.1. Plug-in versions of the odds-algorithm. Note that we may use two equivalent definitions of the optimal threshold \(s\), namely

**Definition 2.1** \(s = 1\) if \(R_1 < 1\), otherwise \(s = \max\{1 \leq k \leq n : R_k \geq 1\}\),

**Definition 2.2** \(s = 1\) if \(R_1 < 1\), otherwise \(s = \min\{1 \leq k \leq n : R_{k+1} < 1\}\).

Let us briefly comment on the reason for giving these equivalent definitions. **Advantage of Def. 2.1**: If the odds \(r_n, r_{n-1}, \ldots\) are known beforehand, then we can sum successively backwards \(r_n + r_{n-1} + \cdots\) and stop as soon as the value 1 is reached or exceeded. In other words, \(s\) is a deterministic stopping time, and the value \(V(n) = Q_s R_s\) is immediately available. Thus, the optimal strategy and the optimal value are obtained at the same time, and not a single computation is redundant. With these three properties the odds-algorithm qualifies as a so-called triple-ace algorithm. This is why Def. 2.1 of \(s\) is used in B. (2000).

**Advantage of Def. 2.2**: If the \(p_j, j = 1, 2, \ldots, n\) and thus \(r_j, j = 1, 2, \ldots, n\) are not known, and so Def. 2.1 is useless since, starting with \(n\), we cannot decide beforehand whether \(R_n = r_n \geq 1, R_{n-1} \geq 1, \ldots\). To enable sequential estimation, time must now run in its natural order for the algorithm. Suppose we have fixed an estimator for the odds. Then Def. 2.2 can in principle always
be applied if we ignore the decision for the first observation $I_1$. Indeed, if all $p_j$’s are a function of the same unknown parameters then we get from $I_1$ (for example via a Bayesian update) a first estimate of the following $p_j$’s, that is for $j \geq 2$. For the case of one unknown parameter this method is exemplified in B. (2005, p. 82), and for extended versions described in B. (2018).

**Definition 2.3** A plug-in odds-algorithm for unknown odds is any form of the odds-algorithm which, after having determined a sequential estimator for the odds, uses Def. 2.2 to obtain the threshold $s := \hat{s}$ by replacing the unknown odds $r_1, r_2, \cdots$ stepwise by their sequential estimates $\hat{r}_1, \hat{r}_2, \cdots$. The unknown $p_j$’s are estimated sequentially with the chosen estimator, and their estimates $\hat{p}_1, \hat{p}_2, \cdots$ are used to compute the corresponding $\hat{q}_k = 1 - \hat{p}_k$ and $\hat{r}_k = \hat{p}_k/\hat{q}_k$.

Since the odds in the plug-in algorithm are random variables, the threshold $\hat{s}$ becomes also a random variable, and we do not know whether the "optimal rule" would be a monotone rule. No optimality is claimed. Of course, the first estimates are hardly reliable. As $n$ increases one can obtain reasonable results by using the first observations for the sequential estimation only (no stopping). For larger $n$ this method seems to perform well, as seen in Bruss and Louchard (2009).

**2.2. Continuous time versions.** In the same spirit, the continuous-time version of the odds-algorithm on a horizon $[0, T]$ can be adapted for sequential updates. Recall the definitions of $\lambda(t)$ and $h(t)$ in B.(2000), subsection 4.1. Let $t$ be a jump time of the process counting the events (arrivals). Note that stopping makes only sense in jump times producing interesting events. Look at (7) of B.(2000) and replace

$$\lambda(u) := \hat{\lambda}_t(u); \quad h(u) := \hat{h}_t(u), \quad t \leq u \leq T,$$

where $\hat{\lambda}_t(u)$ and $\hat{h}_t(u)$ are the corresponding updates for the chosen estimators, as seen at time $t$. Writing (informally) $I_t = 1$ if the event observed at time $t$ is interesting, the algorithm now prescribes to stop at time $\tau$ (recall Def. 2.2) defined by

$$\tau = \inf\{0 \leq t \leq T : I_t = 1 \text{ and } \int_t^T \hat{\lambda}_t(u)\hat{h}_t(u) \leq 1\}.$$

Note that this version can be applied for both unknown odds (re-interpreted as the conditional probability $h(u)$ of an event seen in $u$ of being interesting), and also for unknown $n$ (interpreted as the random number of arrivals in a counting process with unknown intensity $\lambda(u)$).

The idea is clear and we do not formulate a corresponding Definition 2 (would be more technical). Again, no claim of optimality is made. However, simulations for larger $n$ gave again rather encouraging results.
In Section 5 we will see an example showing that understanding monotonicity in values obtained through the odds-algorithm, on which we will focus in Section 3, also helps to predict the random range of values obtained by plug-in versions.

3. Monotonicity. Before addressing questions of monotonicity it is useful to point out that the answer for the value \( V(n) = Q_s R_s \) (see (1)) is complete. It also covers the case \( Q_s = 0 \) and \( R_s = \infty \) since \( V(n) \) stays well-defined, as shown below.

Corollary 3.1 For \( n \) fixed with optimal threshold \( s \): If \( Q_s = 0 \), then we have \( V(n) = Q_{s+1} \).

Proof If \( s \) is optimal with \( p_s = 1 \), then clearly \( s \) must be the last \( j \) such that \( p_j = 1 \). With \( 1 = p_s = r_s q_s \) the product \( q_s r_s \) is no undetermined form. Therefore, from (1) and (2), using \( R_s = R_{s+1} + r_s \), \( Q_s = q_s Q_{s+1} \),

\[
Q_s R_s = Q_{s+1} (q_s R_{s+1} + q_s r_s) = Q_{s+1} (q_s R_{s+1} + 1) = Q_{s+1} = \prod_{j=s+1}^{n} q_j, \tag{2}
\]

because \( R_{s+1} < 1 \) by definition of \( s \), and \( q_s = 1 - p_s = 0 \). The answer stays also correct for \( s = n \) with the standard definition that an empty product equals 1.

Remark 3.2 We conclude from Corollary 3.1 that, whenever we deal with \( Q_s R_s \), we can always assume \( Q_s > 0 \); otherwise \( Q_s R_s \) reduces to \( Q_{s+1} > 0 \). This simple result will be used repeatedly in what follows.

3.1. Extended terminology In the following we define the setting of the Odds-theorem for \( n \) varying. It seems evident what is meant by \( n \) varying, but a clear terminology will keep arguments simple.

We speak of a \( n \)-problem for an underlying sequence of probabilities \( p_1, p_2, \ldots \), if the problem of stopping on the last interesting event applies to the stopped sequence \( I_1, \ldots, I_n \) with \( E(I_k) = p_k, k = 1, 2, \ldots, n \). More precisely:

(i) We say that we win in the \( n \)-problem if we succeed to stop on the last index \( k \in \{1, 2, \ldots, n\} \) with \( I_k = 1 \).

(ii) A stopping time \( \sigma \) is said to be optimal for the \( n \)-problem if \( \sigma \) maximizes the win probability for the \( n \)-problem. The corresponding value is denoted by \( V(n) \).
(iii) We say that \( s(n) \) is the optimal threshold for the \( n \)-problem, if the stopping time

\[
\sigma_n = \inf\{s(n) \leq k \leq n : I_k = 1\} \land n
\]  

(3)

solves the \( n \)-problem. Here, as for \( n \) fixed, it is understood that one cannot return to an \( I_j = 1 \) passed over before, and that the stopping time is also non-anticipative, i.e. \( \{\sigma_n = k\} \in \mathcal{F}_k \) where \( \mathcal{F}_k \) denotes the \( \sigma \)-field generated by \( I_1, I_2, \ldots, I_k \).

From the Odds-theorem (B. (2000, p. 1385)) we have correspondingly

\[
V(n) = Q(s,n)R(s,n),
\]  

(4)

where, for \( 1 \leq k \leq n \),

\[
Q(k,n) = \prod_{j=k}^{n} q_j; \quad R(k,n) = \sum_{j=k}^{n} r_j,
\]  

(5)

and

\[
s := s(n) = \begin{cases} 
1 & \text{, if } R(1,n) \leq 1 \\
\sup\{1 \leq k \leq n : R(k,n) \geq 1\} & \text{, otherwise.}
\end{cases}
\]  

(6)

According to (6), we will use the simplified notation \( Q(s,n) := Q(s(n),n) \) and \( R(s,n) := R(s(n),n) \) whenever this is not ambiguous.

We are now ready to tackle questions of interest concerning the monotonicity of \( V(n) \) in \( n \). Since we have an explicit and simple formula for \( V(n) \), such questions, including when \( V(n) = V(n+j) \) for fixed \( j \in \mathbb{N} \) will hold, are not deep, of course. Our focus will be on trying to see certain facts quickly, and what their implications are. Also, our objective is to increase intuition of what will happen if the underlying sequence changes in a certain way.

The following Lemma 3.3 and Theorem 3.4 are the basic results.

**Lemma 3.3** For an underlying sequence \( p_1, p_2, \ldots \) with \( 0 \leq p_j \leq 1 \) for all \( j \), let

\[
N^* = \sup\{n \in \mathbb{N} : R(1,n) \leq 1\}.
\]  

(7)

Then the optimal win probability \( V(n) \) is non-decreasing in \( n \) for \( 1 \leq n \leq N^* \).

**Proof** Since \( s(n) = 1 \) for \( n \in \{1, 2, \ldots, N^*\} \) by definition of \( N^* \), we have from the optimality of the threshold \( s(n) \) the value \( V(n) = Q(1,n)R(1,n) \)
for \( n \leq N^* \). Thus, by definition of \( Q(k,n) \) and \( R(k,n) \) in (1) and (2), we get for \( 1 \leq n < N^* \):

\[
V(n + 1) = Q(1, n + 1) R(1, n + 1) = Q(1, n) q_{n+1} (R(1, n) + r_{n+1})
\]

\[
= Q(1, n) (q_{n+1} R(1, n) + p_{n+1}) \geq Q(1, n) R(1, n) (q_{n+1} + p_{n+1})
\]

\[
= Q(1, n) R(1, n) = V(n)
\]

where the inequality follows from \( R(1, n) \leq 1 \). Thus \( V(n + 1) \geq V(n) \) for \( n < N^* \).

The index \( N^* \) defined in Lemma 3.3 is a benchmark in the sense that assumptions for \( p_1, p_2, \ldots, p_{N^*} \) are irrelevant for the monotone behaviour of \( V(n) \). As we shall see in the following, from \( N^* \) onward, \( V(n) \) mimicks (simple) monotonicity assumptions of the \( p_n \) on \( \{N^*, N^* + 1, \ldots\} \). We will see later on that the latter is not necessarily true for strict monotonicity.

**Theorem 3.4** The sequence of optimal values \( V(n)_{n \geq N^*} \) for the \( n \)-problems reflects monotonicity properties of an underlying sequence \( (p_n)_{n \geq N^*} \) as follows:

(A) If the sequence \( (p_n)_{n \geq N^*} \) is non-increasing, then the optimal values are non-increasing for \( n \geq N^* \).

(B) If \( (p_n)_{n \geq N^*} \) is non-decreasing, then \( V(n) \) is non-decreasing for all \( n \in \mathbb{N} \).

**Proof**

(A) We first show that if the success probabilities \( p_j \) are non-increasing, then the optimal threshold \( s(n) \) for the \( n \)-problem defined in (6) satisfies

\[
\forall n \in \{1, 2, \cdots\} : s(n) = s \implies s(n + 1) \in \{s, s + 1\}. \quad (8)
\]

Indeed, we first note that by definition \( 1 \leq s(j) \leq j \) and \( s(j) \leq s(j + 1) \) since all odds are non-negative. Also, \( R(s, n) - 1 < r_s \) since, from (6), \( R(s + 1, n) = R(s, n) - r_s < 1 \) and \( R(s, n) \geq 1 \). Moreover, with non-increasing \( p_j \) we see that we have non-increasing odds \( r_j = p_j/q_j \). Consequently, there are only two possibilities by passing from \( n \) to \( n + 1 \):

- if \( R(s + 1, n + 1) = R(s, n) - r_s + r_{n+1} \geq 1 \) we get \( s(n + 1) = s(n) + 1 = s + 1 \), otherwise \( s(n + 1) = s(n) + 1 = s(n + 1) = s \). This proves statement (8).

Hence we have to consider for the proof of (A) only two cases, namely (i) \( s(n + 1) = s(n) \), and (ii) \( s(n + 1) = s(n) + 1 \). Clearly, we can limit our interest to \( n \geq N^* \).
(i) Let \( s(n + 1) = s \). Then (4) and (5) imply that the inequality \( V(n + 1) \leq V(n) \) is equivalent to the inequality

\[ Q(s, n + 1)R(s, n + 1) \leq Q(s, n)R(s, n), \]

which, according to (5), is again equivalent to

\[ q_{n+1}Q(s, n) (R(s, n) + r_{n+1}) \leq Q(s, n)R(s, n). \]

Recall Remark 3.2 and divide by \( Q(s, n) > 0 \). Using \( r_{n+1} = p_{n+1}/q_{n+1} \) we see that inequality (10) becomes

\[ q_{n+1}R(s, n) + p_{n+1} \leq R(s, n). \]

This inequality is always true for \( n \geq N^* \), since \( p_k + q_k = 1 \) for all \( k \), and since for \( n \geq N^* \) we have \( R(s, n) \geq 1 \) by definition of \( s := s(n) \).

(ii) If \( s(n + 1) = s + 1 \), then we must prove that the condition

\[ Q(s + 1, n + 1)R(s + 1, n + 1) \leq Q(s, n)R(s, n) \]

will hold for \( n \geq N^* \). This is slightly more involved.

By definition of \( Q(s, n) \) and \( R(n, s) \) the inequality (12) is now equivalent to

\[ \frac{q_{n+1}}{q_s} Q(s, n) (R(s, n) + r_{n+1} - r_s) \leq Q(s, n)R(s, n). \]

We first note that the case \( q_s = q_{n+1} \) is trivial because then we have also \( r_{n+1} = r_s \) so that both sides of (13) become \( Q(s, n)R(s, n) \), and thus the statement is true.

Hence we can confine our interest to \( q_s \neq q_{n+1} \). Since we assumed the \( p_j \) non-increasing, this means \( p_s > p_{n+1} \) and \( r_s > r_{n+1} \).

Independent, we have seen already that we can focus our interest on \( Q(n, s) > 0 \), implying \( 0 < p_s < 1 \) and \( 0 < q_{n+1} < 1 \). Therefore, dividing inequality (13) by \( Q(s, n) > 0 \) and multiplying it by \( q_s > 0 \), it becomes

\[ R(s, n) (q_s - q_{n+1}) \geq q_{n+1} (r_{n+1} - r_s). \]

Since the rhs of (14) can be written \( p_{n+1} - q_{n+1}(p_s/q_s) \) we obtain, using \( q_s - q_{n+1} = p_{n+1} - p_s < 0 \),

\[ R(s, n) \leq \frac{-p_{n+1} + r_s q_{n+1}}{-p_{n+1} + p_s}. \]
With non-increasing $p_j$'s we have non-decreasing $q_j$'s so that

$$r_s q_{n+1} \geq r_s q_s = p_s.$$ 

Therefore the rhs of (15) is greater or equal to 1 as it should be in the non-trivial case $n \geq N^*$ by definition of $R(s, n)$.

However, here we have to observe an additional combined constraint. By passing from $n$ to $n + 1$, the optimal threshold index $s(n + 1)$ for the $(n + 1)$-problem becomes $s(n) + 1$ if and only if

$$R(s(n), n) \geq 1,$$

and

$$R(s(n) + 1, n) < 1, \text{ and } R(s(n) + 1, n + 1) \geq 1.$$ 

Since $R(s(n), n) - R(s(n)+1, n) = r_s(n)$ and $R(s(n)+1, n+1) - R(s(n)+1, n) = r_{n+1} \geq r_s(n)$ the constraints in (16) are satisfied if the rhs of inequality (15) does not exceed or reach the value $1 + r_s(n)$. Indeed, we see that inequality (15) is sharp, namely, with $s := s(n)$,

$$\frac{r_s q_{n+1} - p_{n+1}}{p_s - p_{n+1}} = 1 + r_s. \quad (17)$$

To see this, note that $1 + r_s$ can be written as $q_s^{-1}$. Since $q_s > 0$, the equation (17) is equivalent to $q_{n+1} p_s - p_{n+1} q_s = p_s - p_{n+1}$, and this is straightforward to verify.

This completes the proof of part (A).

(B) Suppose now that the sequence $(p_n)_{n \geq N_0}$ is non-decreasing. Note that, although this means that the $q_n$ are non-increasing, we cannot use here a duality argument based on re-interpreting the $I_k$ as $1 - I_k$. Therefore, the proof of (B) does not follow directly from the proof of (A), but we can use several parts of it.

We first note that the $(n + 1)$-optimal threshold $s(n + 1)$ can now no longer coincide with $s(n)$. In fact, with $p_{n+1} \geq p_{s(n)}$, and thus $r_{n+1} \geq r_{s(n)}$, we see from (6) that the sum of odds $R(s(n + 1), n + 1)$ would otherwise not be the minimal tail sum of odds to reach or exceed 1. This implies that the part (i) of the proof of (A) is now irrelevant, and that (ii) should now read $s(n + 1) \geq s(n) + 1$.

To begin with $s(n+1) \geq s(n) + 1$, suppose first that $s(n+1) = s(n) + 1$. Then we can use the proof of the part (A) literally by reversing all
inequality signs in the equivalence (9)-(13). Also, the equality (17) stays valid. Hence
\[ V(n + 1) = Q(s(n) + 1, n + 1)R(s(n) + 1, n + 1) \]
\[ \geq Q(s(n), n)R(s(n), n) = V(n), \]
so that the statement (B) is proved for \( s(n + 1) = s(n) + 1 \).

Furthermore, we can now use an important part of the proof of the Odds-theorem B. (2000). It is the part dealing with uni-modality (see p. 1386, line 3 up to equation (4)). It was shown there that the function \( f(k, n) := Q(k, n)R(k, n) \) is, for fixed \( n \), unimodal in \( k \). The unimodality holds in the sense that \( f(k, n) \) is either non-increasing for all \( 1 \leq k \leq n \), or else non-decreasing up to its maximum, and non-increasing thereafter.

Now, replacing \( n \) by \( n + 1 \) we know then from the inequality figuring in (18) that the index \( k = s(n) + 1 \) must belong to the non-decreasing wing of the graph of \( f(k, n + 1) \). Since \( s(n + 1) = \arg \max_{1 \leq k \leq s(n + 1)} \{f(k, n + 1)\} \), this uni-modality property implies
\[ f(s(n) + 1, n + 1) \leq f(s(n) + 2, n + 1) \]
\[ \leq \cdots \leq f(s(n + 1), n + 1). \]

Note that the rhs of (19) corresponds to \( f(s(n + 1), n + 1) = V(n + 1) \) (by definition of \( s(n + 1) \)), the inequality (18) affirms that \( V(n) \leq f(s(n) + 1, n + 1) \). The latter is true since the index \( s(n) + 1 \) with \( s(n) \leq s(n) + 1 \leq s(n + 1) \) lies in the non-decreasing part of \( f(\cdot, n + 1) \). Hence, taking both together, we have
\[ V(n) \leq f(s(n) + 1, n + 1) \leq V(n + 1). \]

This proves part (B) and completes the proof of Theorem 3.4. \[\blacksquare\]

The following easy observation is worth pointing out.

**COROLLARY 3.5** \( \forall n \in \{1, 2, \cdots\} : s(n + 1) = s(n) \implies V(n + 1) \leq V(n) \)

**PROOF** In the part A (i) of Theorem 3.4 we only used \( R(s(n), n) \geq 1 \) for \( n \geq N^* \). However, the latter holds by definition and is independent of monotonicity assumptions (although the hypothesis \( s(n + 1) = s(n) \) itself is not, as just seen before). \[\blacksquare\]

### 4. Uniqueness of optimal thresholds and values.

Corollary 1 of B. (2000, p. 1387) says for fixed \( n \): \( V(n) = Q_s R_s = Q_{s-1} R_{s-1} \) if and only if \( R_s = 1 \). This translates for the \( n \)-problem based on an underlying sequence \( (p_j) \) directly into:
Corollary 4.1

\[ V(n) = Q(s(n) - 1, n)R(s(n) - 1, n) \iff R(s(n), n) = 1, \quad (20) \]

that is, two consecutive indices \( s - 1 \) and \( s \) are both optimal thresholds for the \( n \)-problem if and only if the sum of the relevant odds in the \( n \)-problem equals 1.

The following results are complementary to the preceding one. We give criteria for values of different \( n \)-problems to coincide. Since we have an explicit formula for \( V(n) \) in terms of \( p_1, p_2, \ldots, p_n \) and \( s(n) \), we have a straightforward equivalence, namely \( V(n+1) = V(n) \) if and only if \( Q(s(n+1), n+1) = Q(s(n), n)R(s(n), n) \). Since from (5), (by putting for \( Q(a, b) = 1 \) and \( R(a, b) = 0 \) for \( b < a \)),

\[ Q(s(n+1), n+1) = q_{n+1}Q(s(n), n) / \prod_{j=s(n)}^{s(n+1)-1} q_j \quad (21) \]

\[ R(s(n+1), n+1) = R(s(n), n) - R(s(n), s(n+1) - 1) + r_{n+1}, \quad (22) \]

we can solve the equation \( V(n) = V(n+1) \) explicitly for \( R(s(n), n) \). This requires to compute \( s(n+1) \), which is no problem, of course, but means additional work. In the same way we could derive for an arbitrarily fixed \( j \in \{1, 2, \ldots\} \) a criterion for \( V(n+j) = V(n) \) to hold. It clearly suffices to adapt the above formulae in (21) and (22) and then to solve the equivalence equation again for \( R(s(n), n) \). As our primary goal is to increase the ease of application of the Odds-Theorem and to see implications as quickly as possible, (21) and (22) are slightly too complicated for this purpose. In the case of monotonicity things are simpler:

Theorem 4.2 If the underlying sequence \( (p_j) \) with \( 0 < p_j < 1 \) is non-increasing, then

\[ V(n+1) = V(n) \]

if and only if one of the two following conditions are satisfied

(a) \( R(s(n), n) = 1 \) or \( R(s(n+1), n) = 1 \)

(b) \( p_{s(n)} = p_{n+1} \).

Proof We recall from the first part of the proof of Theorem 3.4 (see (8)) that, with \( s := s(n) \), we have \( s(n+1) \in \{s, s+1\} \).
Let first \( s(n+1) = s \). Then \( V(n+1) = V(n) \) means \( Q(s,n+1)R(s,n+1) = Q(s,n)R(s,n) \). Replacing in the proof of A (i) all inequality signs by "=",
this means

\[
V(n+1) = V(n) \iff q_{n+1}R(s,n) + p_{n+1} = R(s,n).
\]

Since the rhs equation holds if and only \( R(s,n) = 1 \), we have proved (a) for the case \( s(n+1) = s(n) \).

Second, if \( s(n+1) = s+1 \), then (13) and (14) in the proof of A (ii) say, when replacing again all inequality signs by "=",
that \( V(n+1) = V(n) \) if and only if one of the following conditions hold

\((\alpha)\) \( q_{s(n)} = q_{n+1} \), that is \( p_{s(n)} = p_{n+1} \) and thus \( r_{s(n)} = r_{n+1} \),

or else, from (15) and (17),

\((\beta)\) \( R(s,n) = 1 + r_s \).

The condition \((\alpha)\) is what we called in (13) the "trivial" case. Since the sequence \( (p_j) \) is non-increasing it implies \( p_{s(n)} = p_{s(n)+1} = \cdots = p_n = p_{n+1} \).

Concerning condition \((\beta)\) in the case \( p_{s(n)} \neq p_{s(n)+1} \), with \( R(s,n) = R(s+1,n) + r_s \) we see that it can only hold if \( R(s+1,n) = 1 \). Since we are in the case \( s(n+1) = s + 1 \), the condition reads \( R(s(n+1),n) = 1 \), and hence Theorem 4.2 is proved.

We note that the Condition (ii) in the theorem is very transparent. If the monotone sequence \( (p_j) \) is piecewise constant on a stretch beginning at \( k \) say, it suffices to check whether the length of the stretch has the length at least \( 1/p_k \). Condition (i) is in general harder to see (or to exclude) but monotonicity makes it again easier.

Essentially the same holds for a corresponding Theorem for monotone non-decreasing \( p_j \)'s except that \( s(n+1) \) may now be larger that \( s(n) + 1 \).

From the preceding two results we get immediately:

**Corollary 4.3** If \( (p_j) \) is non-increasing, then optimal thresholds and optimal values of different \( n \)-problems are all unique if \( R(s(n),n) \neq 1 \) and \( R(s(n+1),n) \neq 1 \) and \( p_{s(n)} \neq p_{n+1} \) for all \( n \).

### 4.1. Strict monotonicity

For a given underlying sequence \( (p_j) \) we say that we have a *coincidence* in \( n \) and \( n + k \), if and only if \( V(n) = V(n + k) \). Using Theorem 4.2 and Corollary 4.3 we can prove many results about coincidences and their frequencies in an underlying sequence. It turns out that most questions which may come to our mind are seemingly easy to answer. Here is a small collection:
For example, are there monotone sequences with repeated coincidences \( V(n) = V(n+1) = V(n+2) = \ldots \)? Clearly yes. Any sequence which is constant from a certain index \( j \) onward, will do.

Or, can we find a sequence where all \( V(n) \) are different from a certain index onward? Yes, an easy choice is one where the sequence of partial sums \( S_n := p_1 + p_2 + \cdots + p_n \) converges.

Or a third one, can we find for each monotone sequence \( (p_j) \) with diverging partial sums a minoried sequence with diverging partial sums such \( \{V(n) = V(n+1) \text{ infinitely often}\} \)? Again the answer is yes, and first-year analysis suffices for the proof. Questions of this kind may not be of general interest, however, and thus this should do.

Nevertheless, one fact is clearly of interest, namely, monotone sequences without coincidences yield strictly monotone values \( V(n) \). This will be used in Subsection 5.4.

5. Applications

5.1. Our first example shows that understanding monotonicity may override wrong intuitions. Suppose we are offered two games: a 4-game with probabilities \( .1, .2, .24, .25 \) say, or the 5-game \( .1, .2, .24, .25, .251 \). One may feel that it should be harder to succeed in the 5-game since the expected number of ones is small in both games, and the last 1 can be hidden in more places to in the 5-problem than in the 4-problem, with \( p_5 \) being only slightly larger than \( p_4 \).

Theorem 3.4 tells us without computation that we should choose the 5-problem, however. We win in the latter with probability \( .4215 \), which is around 1.68 percent better than playing the 4-game optimally.

5.2. To give a different kind of example, we note that an interesting event need not be associated with a value as such. For instance, in so-called compassionate-use treatments, stopping at the last successful treatment in a sequential clinical trial does not distinguish the last fortunate patient in any way. Stopping on the last success means stopping with the first patient covering all successes, and thus preventing unnecessary treatments thereafter. The latter is the true objective.

The odds of a successful treatment must typically be estimated sequentially (see B. (2018), 3.2 and 3.3). Here the idea is to plug in the estimates into the odds-algorithm, that is, to use the plug-in version given in Section 2.1. The independence of the \( I_k \) is now a priori lost by definition and no optimality can be claimed. However, since different patients are independent of each other in their reaction to treatments, the chosen plug-in odds-algorithm may still yield a good approximation of the optimal strategy. Also, since new patients tend to join the line if treatments seem more and more successful, or
former patients may withdraw from the line if the contrary seems the case, it is good to see, for both patients and doctors, that this is fully in line with Theorem 3.4.

A similar reasoning holds if the physician has a fixed time-horizon $[0, T]$ but cannot estimate beforehand the arrival rate of demands for treatments. In such case the plug-in version of Section 2.2 can be used. If the success rate must be estimated at the same time, then this makes a double task which is difficult if the number of patients is small. However, it seems hard to find a convincing alternative approach, and thus choosing a plug-in algorithm may still be the best one can do.

5.3. May a reasoning based on piecewise monotonicity also be helpful? Yes, but this depends on where monotonicity begins and on the corresponding $s(n)$. For example, look at the interesting group interview problem of Hsiau and Yang(2000). (Example 3.3 of B. (2000) shows the concise solution with the odds-algorithm). This is a give-and-take problem in the sense that we can (formally) win by interviewing all candidates together but we can hardly expect to make good interviews. Suppose we reserve 5 days for seeing 15 candidates, say, and begin for external reasons with group sizes 3 each on the 1st and 2nd day. Since $I_1 = p_1 = 1$ and $p_k < 1$ for $k \geq 2$ by definition, we cannot arrange for having increasing $p_k$’s. If we want groups of sizes 2, 3, 4 in any order on days 3, 4, 5 it turns out best we choose from the 6 remaining possibilities the schedule $(3, 3, 4, 3, 2)$ and get the best with probability $0.448\cdots$.

5.4. The last example, concerning the classical secretary problem will be given in detail:

5.4.1. Classical secretary problem. Grau Ribas (private communication (2018)) instigated by Bayón et al.(2018), asked whether the optimal win probability in the classical secretary problem (CSP) with $n$ candidates is strictly decreasing for $n \geq 3$. Neither Grau Ribas, nor the author, nor peers we asked found the statement in the literature, and a reader knowing a source is kindly requested to inform the author.

The answer, given by Theorem 5.1 given below is affirmative, and says more.

Let $I_k = 1\{k$th candidate has relative rank 1$\}, k = 1, 2, \cdots$. It is well known that the $I_k$ are independent with $p_k = P(I_k = 1) = 1/k$. The CSP for the $n$-problem is the problem to stop online with maximum probability on rank 1 of $n$ uniquely rankable candidates, i.e. on the last indicator $I_k = 1$ for $k \leq n$. Note that $V(2) = 1/2 = V(3)$ so that we confine our interest to $n \geq 3$.

THEOREM 5.1 In the classical secretary problem with $n$ candidates:

(i) The optimal win probability $V(n)$ is strictly decreasing in $n$ for $n \geq 3$. 
(ii) The optimal thresholds $s(n)$ are all unique for $n \geq 3$.

**Proof** Let $V(n)$ be the optimal win probability for $n$ candidates. We have $p_n = 1/n, q_n = 1 - p_n = (n - 1)/n$ and $r_n = p_n/q_n = 1/(n - 1)$. Since $(p_n)$ is decreasing, Theorem 3.4 (A) implies that $V(n)$ is non-increasing. It follows that $V(n)$ is strictly decreasing for $n \geq 3$ if and only if $V(n) \neq V(n + 1)$ for $n \geq 3$. With $(p_j)$ being strictly decreasing, Corollary 4.3 says then that (i) and (ii) hold both at the same time if $R(s(n), n) \neq 1$ and $R(s(n + 1), n) \neq 1$ for all $n \geq 3$.

Using from above $p_n, q_n,$ and $r_n$ in $R(k, n)$ we get

$$R(k, n) = \sum_{j=k}^{n} \frac{1}{j-1} = \sum_{j=k}^{n-1} j^{-1} = H(n - 1) - H(k - 2),$$

(23)

where $H(n)$ denotes the $n$th partial sum of the harmonic series $H(n) = 1 + 2^{-1} + 3^{-1} + \cdots + n^{-1}$. It is well-known from number theory that $H(n)$ is non-integer for $n \geq 2$. If we can show

$$\forall n \geq 3 \text{ and } 1 \leq k < n - 1 : H := H(n) - H(k) \notin \mathbb{N} \text{ for } n \geq 3,$$

(24)

then clearly $R(k, n) \neq 1$ for $n \geq 3$.

Indeed, (24) is easy to show. (The author, as well as a referee, believe this should be in the literature, but did not find it. The author keeps his proof on request for the reader.)

This proves then (i) and (ii) of Theorem 5.1 at the same time. ■

5.5. Prophet comparisons. It is not hard to find for the CSP alternative proofs for the first part showing that $V(n)$ is non-increasing. The following prophet argument in order to compare $V(n)$ and $V(n + 1)$ leads to a proof which may be the among the most elegant ones.

Suppose a prophet knows the position of rank $n + 1$, the worst candidate. Let $V_P(n + 1)$ be the prophet’s value. Clearly $V_P(n + 1) \geq V(n + 1)$. Knowing the position of rank $(n + 1)$ the prophet will ignore it and thus face an equivalent $n$-problem, since the relative ranks on the other $n$ positions do not change. Hence $V(n) := V_P(n + 1) \geq V(n + 1)$.

A similar prophet argument was already used to show that the value $v_n$ of the celebrated Robbins’ Problem for $n$ observations is non-decreasing in $n$ (see Bruss and Ferguson(1993), or, independently, Assaf and Samuel-Cahn(1996)). For a more general formulation see the notion of a half-prophet in Bruss and Ferguson(1996).

The preceding argument was easy due to the nice structure of the CSP model where all positions of ranks are equally likely. In more general settings such prophet tricks are usually more involved. Note however that Theorem
3.4 stays always useful for this, since, whatever index $k$, fixed or random, is singled out for a prophet’s value comparison, cancelling $p_k$ in the underlying sequence does not affect any monotonicity property.

**Remark 5.2** If the number of candidates $n$ in the secretary problem is a random variable $N$, then the optimal strategy is in general no longer a simple threshold strategy. As shown in Presman and Sonin(1972) (the fathers of the secretary problem with unknown $n$) the stopping region may, depending on the law $P(N = n)_{n=1,2,...}$ split into stopping islands. They found the corresponding monotonicity criteria. (The unified approach for unknown $N$ in continuous time (B. (1984)) may be seen as an interesting alternative model from the point of view of applicability in everyday problems.) Note that the phenomenon “stopping islands” may appear also in plug-in algorithms when the updated threshold $\hat{s}$ makes larger jumps.

**Acknowledgement.** The authors thanks the two referees for their fruitful comments, in particular for the advice to re-write Section 2 in a more explicit form. This also improved Section 5.

6. References

[1] D. Assaf and E. Samuel-Cahn. The secretary problem: minimizing the expected rank with i.i.d. random variables. *Adv. in Appl. Probab.*, 28(3): 828–852, 1996. ISSN 0001-8678. doi: 10.2307/1428183. Cited on p. 39.

[2] L. Bayón, P. Fortuny Ayuso, J. M. Grau, A. M. Oller-Marcén, and M. M. Ruiz. The Best-or-Worst and the Postdoc problems. *J. Comb. Optim.*, 35(3):703–723, 2018. ISSN 1382-6905. doi: 10.1007/s10878-017-0203-4. Cited on p. 38.

[3] F. T. Bruss. Sum the odds to one and stop. *Ann. Probab.*, 28(3):1384–1391, 2000. ISSN 0091-7798. doi: 10.1214/aop/1019160340. Cited on pp. 25, 26, 27, 28, 30, 34, 38, and 43.

[4] F. T. Bruss. Die Kunst der richtigen Entscheidung. *Spektrum der Wissenschaft*, June:78–82, 2005. Cited on p. 28.

[5] F. T. Bruss. A mathematical approach to comply with ethical constraints in compassionate use treatments. *Math. Sci.*, 43(1):10–22, 2018. ISSN 0312-3685. Cited on pp. 28 and 37.

[6] F. T. Bruss and T. S. Ferguson. Minimizing the expected rank with full information. *J. Appl. Probab.*, 30(3):616–626, 1993. ISSN 0021-9002. doi: 10.2307/3214770. Cited on p. 39.
[7] F. T. Bruss and T. S. Ferguson. Half-prophets and Robbins’ problem of minimizing the expected rank. In *Athens Conference on Applied Probability and Time Series Analysis, Vol. I* (1995), volume 114 of Lect. Notes Stat., pages 1–17. Springer, New York, 1996. doi: 10.1007/978-1-4612-0749-8_1. Cited on p. 39.

[8] F. T. Bruss and G. Louchard. The odds algorithm based on sequential updating and its performance. *Adv. in Appl. Probab.*, 41(1):131–153, 2009. ISSN 0001-8678. doi: 10.1239/aap/1240319579. Cited on p. 28.

[9] Bruss, F. Thomas. A unified approach to a class of best choice problems with an unknown number of options. *Ann. Probab.*, 12(3): 882–889, 1984. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(198408)12:3<882:AUATAC>2.0.CO;2-M&origin=MSN. Cited on p. 40.

[10] Bruss, F. Thomas. A note on bounds for the odds theorem of optimal stopping. *Ann. Probab.*, 31(4):1859–1861, 2003. ISSN 0091-1798. doi: 10.1214/aop/1068646368. Cited on p. 25.

[11] R. Dendievel. New developments of the odds-theorem. *Math. Sci.*, 38 (2):111–123, 2013. ISSN 0312-3685. Cited on p. 26.

[12] T. S. Ferguson. The sum-the-odds theorem with application to a stopping game of Sakaguchi. *Math. Appl.*, 44(1):45–61, 2016. ISSN 1730-2668. doi: 10.14708/ma.v44i1.1192. Cited on p. 26.

[13] A. V. Gnedin. On a best-choice problem with dependent criteria. *J. Appl. Probab.*, 31(1):221–234, 1994. ISSN 0021-9002. doi: 10.2307/3215248. Cited on p. 26.

[14] A. Goldenshluger, Y. Malinovsky, and A. Zeevi. A Unified Approach for Solving Sequential Selection Problems. *arXiv e-prints*, art. arXiv:1901.04183, Jan 2019. Cited on p. 26.

[15] S.-R. Hsiau and J.-R. Yang. A natural variation of the standard secretary problem. *Statist. Sinica*, 10(2):639–646, 2000. ISSN 1017-0405. Cited on p. 38.

[16] T. Matsui and K. Ano. A note on a lower bound for the multiplicative odds theorem of optimal stopping. *J. Appl. Probab.*, 51(3):885–889, 2014. ISSN 0021-9002. doi: 10.1239/jap/1409932681. Cited on p. 26.

[17] T. Matsui and K. Ano. Lower bounds for Bruss’ odds problem with multiple stoppings. *Math. Oper. Res.*, 41(2):700–714, 2016. ISSN 0364-765X. doi: 10.1287/moor.2015.0748. Cited on p. 26.
[18] E. L. Presman and I. M. Sonin. The problem of best choice in the case of a random number of objects. *Teor. Veroiatnost. i Primenen.*, 17:695–706, 1972. ISSN 0040-361x. Cited on p. 40.

[19] K. Szajowski. A game version of the Cowan-Zabczyk-Bruss’ problem. *Statist. Probab. Lett.*, 77(17):1683–1689, 2007. ISSN 0167-7152. doi: 10.1016/j.spl.2007.04.008. Cited on p. 26.

[20] M. Tamaki. Sum the multiplicative odds to one and stop. *J. Appl. Probab.*, 47(3):761–777, 2010. ISSN 0021-9002. doi: 10.1239/jap/1285335408. Cited on p. 26.

[21] M. Tamaki. Maximizing the probability of stopping on any of the last $m$ successes in independent Bernoulli trials with random horizon. *Adv. in Appl. Probab.*, 43(3):760–781, 2011. ISSN 0001-8678. doi: 10.1239/aap/1316792669. Cited on p. 26.
Optymalne zatrzymywanie w oparciu o algorytm ilorazu szans
a monotoniczność wartości problemu.

F. Thomas Bruss

Streszczenie Rozważamy optymalne zatrzymywanie na wyróżnionym zdarzeniu w sekwencyjnym eksperymentie ze skończona liczbą opcji. Twierdzenie o ilorazie szans (Bruss(2000)) wyznacza taką strategię, która maksymalizuje szansę na właściwy wybór. Strategia ta jest optymalna, gdy mamy sekwencję niezależnych eksperymentów, a jej wyznaczanie jest proste. Służy do tego wspomniany algorytm oparty o sumowanie ilorazu szans. W pracy analizowane są szczególne własności takich zadań. Badana jest monotoniczność wartości optymalnej w powiązaniu z monotonicznością podstawowej sekwencji prawdopodobieństw zdarzeń. Podana jest motywacja do takich badań, a następnie udzielono pełnych odpowiedzi. Motywację wzmagają problemy, w których pożądana jest zastosowanie algorytmu szans, ale w których brak informacji nie pozwala na to bez zastosowania sekwencyjnej estymacji nieznanym parametrów. W związku z takimi zagadnieniami wprowadzono pojęcie adaptacyjnego algorytmu ilorazu szans. Rozważania są ilustrowane przykładami.

2010 Klasyfikacja tematyczna AMS (2010): 60G40; 62L15.

Słowa kluczowe: oscylacje harmoniczne, rozmaitości konfiguracji, rozwiązania przybliżone.

F. Thomas Bruss studied Mathematics in Saarbrücken (Germany), and, with a delegation grant from Saarbrücken, in Cambridge (UK) and in Sheffield (UK). He holds the Diplom-Matematiker as well as his doctorate Dr. rer. nat. in Mathematics of the University of Saarbrücken. His scientific career began in 1977 at the University of Namur. In 1978 he received the docteur légal Dr. en sc. and obtained tenure as First assistant a year later. His time in Namur also included visiting positions at the University of Zaire (1981) and at Strathclyde University Glasgow (1984). He then moved to the United States, first as Visiting Associate Professor at UC Santa Barbara, then as Adjunct Professor (Feodor-Lynen) fellow at the University of Arizona, and then as Visiting Associate Professor at UCLA.

In 1990 Thomas returned to Europe as Professor of the Vesalius College of the Vrije Universiteit Brussel. Independently, he obtained the inscription sur liste des professeurs of France. In 1993 he was appointed chair of Mathématiques Générales and Probabilités of the Université Libre de Bruxelles (ULB) where he has stayed ever since. Since then he also held visiting positions at the University of Antwerpen, the University of Namur and the Université Catholique de Louvain. Now, retired from the chair of Mathématiques Générales, Thomas continues as Professeur de l’université of ULB, and as Invited Professor of the Université Catholique de Louvain.

F. Thomas Bruss
Université Libre de Bruxelles
Département de Mathématique
CP 210, B-1050 Brussels, Belgium
E-mail: tbruss@ulb.ac.be

(Received: 17th of May 2019; revised: 17th of July 2019)