Optimal Stopping on Patterns in Strings Generated by Independent Random Variables

F.Thomas Bruss* and Guy Louchard †

January 9, 2003

Abstract

Strings are generated by sequences of independent draws from a given alphabet. For a given pattern of length $l, H = H_1H_2 \dots H_l$ say, and an integer $n \geq l$, our goal is to maximize the probability of stopping on the last appearance of the pattern H in such a string of length n (if any), given that we must not return on a preceding pattern H. This contrasts the goals of several other investigations on patterns in strings such as computing the expected occurence time and the probability of finding exactly r patterns in a string of fixed length. Several motivations are given for our problem ranging from relatively simple best choice problems to more difficult stopping problems allowing for a variety of interesting applications. We solve this problem completely in the homogeneous case for uncorrelated patterns. However, several of these results extend immediately to the inhomogeneous case. In the homogeneous case, optimal success probabilities are shown to vary, dependent on characteristics of the pattern, essentially between the value $1/e$ and a new asymptotic constant $0.619\dots$ These results mark a considerable difference of the true optimal success probability compared with what a first approximation by heuristic arguments would yield. It is interesting to see that the so-called odds-algorithm which gives the optimal rule for $l = 1$ yields an excellent approximation of the optimal rule for any l. This is important for applications because the odds-algorithm is very convenient. We finally give a detailed asymptotic analysis for the homogeneous case.

Keywords: Odds-algorithm, simple odds-algorithm, algorithms on words, investment problems, secretary problems, apartment problem, single-island stopping rule, online decisions, pattern correlation, Poisson clumping heuristic, asymptotic analysis .

AMS Subject classification: $60G40$

1 Introduction

Considerable work has been done on strings generated by sequences of independent draws from a given alphabet. To give an introductory review we first agree on the following terminology.

We say that we are in the *independent* case (or Bernoulli case in the stringology terminology) if the letter produced by the k-th draw, X_k say, is independent of $X_1, \dots, X_{k-1}, X_{k+1}, \dots$. Moreover we call the case *homogeneous*, if the law of X_k does not depend on k, whereas we call it *inhomogeneous* otherwise. If we are interested in the occurence of a certain fixed pattern of consecutive letters $H = H_1H_2, \dots, H_l$ we distinguish two cases, namely

$i)$: patterns may not overlap,
$ii)$: patterns may overlap.

*Université Libre de Bruxelles, Département de Mathématiques et ISRO, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, email:tbruss@ulb.ac.be

†Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, email:louchard@ulb.ac.be

Case i) is conceptually easier. As an example, let $H = AABCAA$, where the letters are drawn from the Latin alphabet $A, B, C, \ldots, Z$. Then the string

$$DEFAABCAABCAAG$$

contains the pattern H only once in case i), but twice in case ii). To avoid ambiguities we define the first occurence of H as the occurence time in case i) that is, H occurs at time 9 in our example. Clearly this distinction is only important if some RHS-subpattern (suffix) coincides exactly with some LHS-subpattern (prefix) of H as it is the case in our example for the subpattern AA.

The *autocorrelation pattern* of a pattern (see [14]) is a suitable way to characterize a pattern by its subpattern structure. It is a string of $0's$ and $1's$ obtained as follows. We copy the pattern successively below itself, each time moving one letter to the right. We return 1 in the kth position if the kth position of H and all succeeding letters coincide (that is $H_k H_{k+1} \ldots H_l$ coincides with $H_1 H_2 \ldots H_{l-k+1}$) . Otherwise we return 0. In our example, the autocorrelation pattern of $H = AABCAA$ is 100011. The autocorrelation pattern of H is denoted by $H * H$. Note that it begins, for any H, with a 1. To stay consistent with the idea of (positive) correlation in Statistics, we say that a pattern is *autocorrelated* if and only if the sum of digits in the autocorrelation pattern is *strictly greater* than one. Otherwise we say that the pattern is non-autocorrelated, or simply *uncorrelated*. Clearly, a pattern of length one is always uncorrelated.

Important results on pattern occurences are given in Guibas and Odlyzko, [14], Szpankowski [26] and Knuth, Morris, Pratt [17]. Known results on two-persons games defined on strings can be found in Conway; see [14] and Li [19].

The objective of this paper is to solve stopping problems for specified patterns on strings. The paper is organized as follows: in Sec.2, we present the Stopping Problems and give examples of possible applications. Sec.3 is devoted to the homogeneous case with uncorrelated pattern. The optimal stopping problem is solved in detail and we show the importance of monotonicity (unimodality) of occurrence probabilities. We also indicate the extension to the inhomogeneous case, showing, in particular, that the form of the optimal strategy stays the same. In Sec. 4 we analyze the asymptotic properties of our algorithms. We also discuss the Poisson clumping heuristic as an alternative approach and show where these methods give similar results and where they diverge. Finally we give an explicit numerical example.

2 The Stopping Problem

The following problem attracts our particular interest.

Problem 2.1 Given an alphabet and a fixed pattern $H = H_1 H_2, \ldots, H_l$, we are allowed to observe sequentially the outcomes $X_1 X_2 \ldots X_n$ of $n > l$ draws. Whenever H is achieved we are allowed either to stop or else to continue. What strategy maximizes the probability of stopping on the very last appearance of H up to X_n?

It is interesting to see that the general solution of Problem 2.1 is harder than one would expect but before we go into more details, we recall the solution for the special case $l = 1$.

Solution of Problem 2.1 for $l = 1$. Let $l = 1$, that is H consists of a single letter from the given alphabet. Further let $p_k := P(X_k = H)$ be the probability that the k-th draw produces H, and let the odds of H be denoted by $r_k := p_k/(1 - p_k)$. Define

$$s = \begin{cases} 1, & \text{if } \sum_{k=1}^{n} r_k < 1 \\ \sup\{k : \sum_{j=k}^{n} r_j \geq 1\}, & \text{otherwise.} \end{cases} \tag{1}$$

Then the optimal strategy for Problem 2.1 is to stop on the first appearance of H, if any, from draw number s onwards. The corresponding optimal successs probability is given by

$$\pi^* = \pi^*(n, \{p_k\}_{k=1,2,\ldots,n}) = \sum_{k=s}^{n} r_k \prod_{k=s}^{n} q_k, \tag{2}$$

where $q_k := 1 - p_k$.

Proof. This follows immediately from the Odds-theorem of Bruss [7] (See Theorem 1, p.1385). ■

Related Problems and Applications.

The solution allows for applications, as e.g. the solution of some dice problems, the secretary problem (see [7]), the best choice problem with group interviews (see Hsiau and Yang [15] and [7]), the "last peak problem" (Tamaki [27]), and others. An integral-version of (1) and (2) can also be applied to some random$-n$ problems in continuous time such as the apartment problem (see Cowan and Zabczyk [10] and Bruss [6]). The latter has recently attracted new interest (see Ano and Ando [4], Kurishima and Ano [18], Suchwalko and Szajowski [24] and Szajowski [25]), though partially in a different context. These and other applications motivate at the same time our interest in the more general Problem 2.1, the solution of which would allow to tackle more general problems. One interesting step into this direction is the recent model of Hsiau and Yang [16] solving the Markovian version of Bruss [7], which means the Markovian version for $l = 1$ and fixed horizon n.

As a motivation for our fixed horizon problem with general pattern length l, we give just one example of a problem taken from investment problems on the stock market.

Investment problem. This problem can be seen as a generalisation of Tamaki's last-peak problem. It may appeal to short-term traders and chart technicians: we codify ups and downs and/or jump ranges of a certain stock price over consecutive days by letters from a fixed alphabet. Here we suppose independence of increments of stock prices in either absolute or relative terms. Given an horizon of n days and a pattern H how do we maximize the probability of buying or selling the *last* time that our stock-price pattern H occurs in the horizon?

We shall return to this problem at several instances of Section 4.

We see many other instances of applications. However, here is a limitation to the obvious one. Although it is tempting to advertize our results also for the use in search engines, this would not always be convincing. For example, if patterns are combinations of real-world words then modelling the occurrence of letters by independent draws is naive: "Surfing on the ..." is very likely to end up in "web", less likely to end up in "water", and anything else has negligible probabilities. Hence some Markovian set-up like a longer-time directional reinforcement (see e.g. Mauldin et al. [20] and Allard and Monticino [3]) would be needed. But our model does not cope with such an extension.

The following problem is a natural generalisation of Problem 2.1

Problem 2.2 Under the conditions of Problem 2.1, what strategy maximizes the probability of stopping on the m-th last occurence of H?

Again, for the special case $l = 1$, the solution of Problem 2.2 is known (see Bruss and Paindaveine, [8]). For $m > 1$ the optimal solution is of the same form as in the case $m = 1$ in the sense that there exists a threshold draw number s_m from which onwards it is optimal to stop on the first occurence of H. Here we imply $s_1 = s$, of course, by notation. The algorithm to compute s_m for $m > 1$ is more complicated than the odds-algorithm (1) where we simply add up the odds for H, from the end of the string (that is from the draw number n) until we hit or exceed the threshold 1. However, it is interesting to observe (see [8]) that the following simple algorithm analogous to the odds-algorithm defined by (1) yields an almost optimal performance under rather mild conditions: this is the algorithm defined by summing up the odds from the end to determine the threshold index

$$\tilde{s}_m = \begin{cases} 1, & \text{if } \sum_{k=1}^{n} r_k < m \\ \sup\{k : \sum_{j=k}^{n} r_j \geq m\}, & \text{otherwise.} \end{cases} \tag{3}$$

This algorithm will be called the *simple odds-algorithm for m stops.* Hence $\tilde{s}_1 = s_1 = s$, and this threshold is optimal for $m = 1$. In general, $\tilde{s}_m \neq s_m$, however.

Now let, for $k = 0, 1, \dots, n - l + 1$ and $l > 1$,

$$p_{k+l} := P(X_{k+1}X_{k+2}, \dots, X_{k+l} = H), \tag{4}$$

and, as before, $r_k := p_k/(1 - p_k)$. A central question of this paper is: how good is the strategy to stop on the first occurence of $H = H_1H_2, \dots, H_l$ after the threshold draw number defined in (1), or more generally, in (3)?
In order to be able to answer this question, we must first know the optimal strategy for general patterns, of course, and what it achieves. This will require a step-by-step approach presented in the following sections.

3 The homogeneous case with uncorrelated pattern

This is the case where p_{k+l} defined in (4) does not depend on k, so that

$$\begin{aligned} p &:= \prod_{j=1}^{l} P(H_j) = P(H) \\ &= p_{k+l}, k = 0, 1, \dots, n - l + 1, \end{aligned} \tag{5}$$

and where the autocorrelation pattern of H is given by

$$H * H = \underbrace{100\dots0}_{l}. \tag{6}$$

We first look at the case of $m = 1$, that is we want to find the strategy which maximizes the probability of stopping on the last occurence of $H = H_1H_2 \dots H_l$ within a string of fixed length n.
Before we state and prove the results of this section, we shortly motivate our method of attack.
We define

Definition 3.1 *The string $X_kX_{k+1} \dots X_n$ is called (k, n)-string, where $1 \leq k \leq n$.*

For a (k, n)-string to contain the pattern H it is necessary that $n - k + 1 \geq l$ or, equivalently, $1 \leq k \leq n - l + 1$. Among all such (k, n)-strings there must be one with maximum probability of containing H exactly once. Suppose that this maximum probability is attained for $k = t$ (not necessarily unique). To avoid ambiguities we take t to be the smallest such k. It is then intuitive that we should stop on the first occurence of H in this (t, n)-string , which is, with maximum probability, at the same time the very last pattern H in the whole $(1, n)$-string . However, to prove this intuition to be correct we must show that the stopping region is *simple*, or, in other words, that the optimal rule is a single-island stopping rule. (see Presman and Sonin [21] and Samuel-Cahn [22]). This means that, once it is optimal to stop on a (t, n)-string then it is optimal to stop until the end. This can be formulated in the following way. Following the notation of Szpankowski [26] we let

$$O_n := \#\text{ occurences of } H \text{ in a string of length } n. \tag{7}$$

Note that we need not distinguish here between the cases where we count overlapping occurrences of H separately or not, since H is uncorrelated. To garantee the single-island property of the optimal rule, we must show that if $P(O_{n-t+1} = 1)$ is maximal in t, then $P(O_{n-k+1} = 1)$ is decreasing for $k = t, t + 1, \dots n$. This property is implied by the following theorem, which we will prove through several steps.

Theorem 3.2 *The sequence $(P(O_n = 1))_n$ is unimodal in n on $\{l, l + 1, l + 2, \dots\}$.*

Remark 3.3 *Generating function*

The generating function of the $P(O_n = 1)$ can be derived from Szpankowski [26]. It is given by

$$g(z) = \sum_{n=l}^{\infty} P(O_n = 1) z^n = \frac{pz^l}{(1 - z + pz^l)^2}. \tag{8}$$

However, it seems difficult to derive a proof of Theorem 3.2 directly by (8). (An asymptotic analysis of (8) is given in Sec. 4). We choose a detour by expressing $P(O_n = 1)$ in terms of the times of the first and the second completion of H.

Lemma 3.4 *Let T_m be the time (draw number) of the m-th completion of the pattern H in an infinite string, let $p_{j,m} := P(T_m = j)$, and let p_j the probability of j being a completion time. Then*

$$p_j = \sum_{m=1}^{\infty} p_{j,m}, \tag{9}$$

and

$$P(O_n = 1) = \sum_{k=l}^{n} p_{k,1} \sum_{j=n-k+1}^{\infty} p_{j,1}. \tag{10}$$

Proof. Equation (9) is clear since $p_{j,1} + p_{j,2} + \ldots$ is the total probability of H being completed at time j. To see (10) we use

$$\{O_n = 1\} = \{T_1 \le n\} \cap \{T_2 > n\} = (\cup_{k=l}^{n} \{T_1 = k\}) \cap (\cup_{k=n+1}^{\infty} \{T_2 = k\}).$$

Since H is uncorrelated and the $X_1, X_2, \ldots$ are independent, $T_2 - T_1$ follows the same distribution as T_1. Hence, using the independence of T_1 and $T_2 - T_1$,

$$\begin{aligned} P(O_n = 1) &= \sum_{k=l}^{n} P(T_1 = k, T_2 - T_1 > n - k) \\ &= \sum_{k=l}^{n} p_{k,1} P(T_2 - T_1 > n - k) = \sum_{k=l}^{n} p_{k,1} \sum_{j=n-k+1}^{\infty} p_{j,1}. \end{aligned}$$

■

Linking the generating functions of T_1 and $P(O_n = 1)$.
Note that, from Lemma 3.4, we can recover the probability generating function $f(z)$ of $P(T_1 = j)$. Using (10) we obtain from the generating function $g(z) = \sum_{n=l}^{\infty} P(O_n = 1) z^n$ for $|z| \le 1$,

$$\begin{aligned} g(z) &= \sum_{n=l}^{\infty} \sum_{k=l}^{n} p_{k,1} z^k \sum_{j=n-k+1}^{\infty} p_{j,1} z^{n-k} \\ &= \sum_{k=l}^{\infty} p_{k,1} z^k \sum_{n=k}^{\infty} z^{n-k} \sum_{j=n-k+1}^{\infty} p_{j,1}, \end{aligned}$$

where we used Fubini and $|z| \le 1$. This can be written as

$$f(z) \sum_{u=0}^{\infty} z^u \sum_{j=u+1}^{\infty} p_{j,1} = f(z) \sum_{j=l}^{\infty} p_{j,1} \sum_{u=0}^{j-1} z^u,$$

where we used again Fubini, and also $p_{j,1} = 0$, for $j < l$. The latter can be written as

$$g(z) = \frac{f(z)(f(z) - 1)}{z - 1}.$$

Using (8) and solving for f yields

$$f(z) = \frac{pz^l}{1 - z + pz^l}. \tag{11}$$

The latter is also given in Szpankowski [26]. From (11), we can also derive the probability generating function of $P[T_1 > j]$. This is given by

$$\frac{f(z) - 1}{z - 1} = \frac{1}{1 - z + pz^l}. \tag{12}$$

The following lemma gives a recurrence relation for the $P(T_1 = j)$, which will be used later on. It is also of interest on its own.

Lemma 3.5 *$p_{j,1} = P(T_1 = j)$ satisfies, for $j \geq l$, the recurrence relation*

$$\begin{aligned} p_{l,1} &= p, \\ p_{j+1,1} &= p_{j,1} - [\![j \geq 2l - 1]\!] p p_{j+1-l,1}, \end{aligned}$$

where $p = P(H)$ and $[\![A]\!] = \mathbb{1}_{\{A\}}$. [1]

Proof. The trick to prove this is to use the independence of variables combined with a reflection argument. Let Ω be the sample space of all possible sequences with generic element ω. Let $\omega \in \{T_1 = j\}$. Hence ω is of the form

$$\omega = \omega_j X_{j+1} X_{j+2} \ldots$$

where

$$\omega_j = X_1 X_2 \ldots X_{j-l+1} \ldots X_j,$$

and $X_{j-l+1} \ldots X_j = H$. Now let (ω_j, X_{j+1}) denote the concatenation of ω_j with X_{j+1}. The string (ω_j, X_{j+1}) has the same probability as the string (X_{j+1}, ω_j) . The latter is a string of the form

$$X_{j+1} X_1 X_2 \ldots X_{j-l+1} \ldots X_j$$

that is a string of length $j + 1$ which completes the pattern H at location $j + 1$. Now note that (X_{j+1}, ω) lies in the set $\{T_1 = j + 1\}$ unless $X_{j+1} X_1 X_2 \ldots X_{l-1} = H$. In the latter case it is an element of $\{T_1 = l, T_2 = j + 1\}$ because $X_1 X_2 \ldots X_j$ achieved the first pattern H at location j. Hence, by homogeneity,

$$\begin{aligned} P(T_1 = j + 1) &= P(T_1 = j) - P(T_1 = l, T_2 = j + 1) \\ &= P(T_1 = j) - P(T_1 = l) P(T_2 = j + 1 | T_1 = l). \end{aligned} \tag{13}$$

Since H is uncorrelated, $T_2 - T_1$ is independent of T_1. Hence we have

$$P(T_2 = j + 1 | T_1 = l) = P(T_2 - T_1 = j + 1 - l),$$

and so, as $T_2 - T_1$ and T_1 follow the same distribution,

$$P(T_2 = j + 1 | T_1 = l) = P(T_1 = j + 1 - l) = p_{j+1-l,1}. \tag{14}$$

From (13) we obtain using $P(T_1 = l) = P(H) = p$ and (14)

$$p_{j+1,1} = p_{j,1} - p p_{j+1-l,1}.$$

Finally we note that the negative term vanishes unless $j \geq 2l - 1$. Hence the Lemma is proved. ■

We now need an upper bound for $P(H)$ if H is uncorrelated. This is provided by the following lemma

[1] Here we use the indicator function notation proposed by Knuth et al. [13].

Lemma 3.6 *Let* $\mu(l) := \sup\{P(H) : H = H_1H_2\dots H_l, \quad H \quad uncorrelated\}$. *Then*

$$\mu(l) \leq \frac{1}{l}(1-1/l)^{l-1} \sim e^{-1}/l, l \text{ large.}$$

Proof. The inequality is trivial for $l = 1$ with $\mu(1) \leq 1$. Hence let $l \geq 2$. We first note that for each uncorrelated pattern H we have $H_1 \neq H_l$, because otherwise the correlation string would be $(1\dots1)$, where the dots stand for either 0's or 1's. Hence, since $P(H_1) + P(H_l) \leq 1$, using independence,

$$P(H) \leq P(H_1)\prod_{k=2}^{l-1} P(H_k)(1 - P(H_1)), \tag{15}$$

where the product term is defined to be 1 for $l = 2$. Now

$$\prod_{k=2}^{l-1} P(H_k) \leq [\max\{P(H_1), 1 - P(H_1)\}]^{l-2}, \tag{16}$$

so that from (15) and (16)

$$\begin{aligned} P(H) &\leq \max\{(P(H_1))^{l-1}(1 - P(H_1)), (1 - P(H_1))^{l-1}P(H_1)\} \\ &\leq \sup_{a\in[0,1]}\{a^{l-1}(1-a)\} \end{aligned}$$

It is straightforward to check that $\varphi(a) := a^{l-1}(1 - a)$ is maximized on $[0, 1]$ by $a = 1 - 1/l$ with $\varphi_{max} = \frac{1}{l}(1 - 1/l)^{l-1}$. The asymptotic relationship for growing l is evident. This proves the Lemma. ■

The following lemma prepares the asymptotic analysis of Section 4.

Lemma 3.7
(i) The sequence $(p_{j,1})_{j\geq l}$ *is monotone decreasing to* 0 *in* j.
(ii) The sequence $(y_j)_{j\geq l}$ *defined by* $y_j := p_{j+1,1}/p_{j,1}$ *is monotone decreasing to some limit* λ^* *satisfying* $\lambda^* = 1 - p/\lambda^{*(l-1)}$.

Proof. From Lemma 3.5 we have $p_{j+1,1} \leq p_{j,1}$ for $j \geq l$ so that the sequence $(p_{j,1})_{j\geq l}$ is decreasing. Since it is bounded from below by 0 it must converge to 0 because $\sum_{k=l}^{\infty} p_{k,1} = 1$. This proves (i).
To prove (ii) we first note that, for $l = 1$, the y_j are constant because T_1 is then a geometric random variable. Hence let $l \geq 2$. According to Lemma 3.5 $p_{j+1,1} = p_{j,1}$ for $l \leq j \leq 2l - 2$ so that the y_j are constant on $l \leq j \leq 2l - 2$. Hence we suppose now that $j \geq 2l - 1$. From Lemma 3.5 we have then

$$p_{j+1,1} = p_{j,1} - pp_{j+1-l,1},$$

or equivalentlty, since $p_{j,1} > 0$, using the definition of y_j,

$$y_j = 1 - p\frac{p_{j+1-l,1}}{p_{j,1}}.$$

We can write this in terms of telescoping product, namely

$$\begin{aligned} y_j &= 1 - p\frac{p_{j+1-l,1}}{p_{j+2-l,1}}\frac{p_{j+2-l,1}}{p_{j+3-l,1}}\cdots\frac{p_{j-1,1}}{p_{j,1}} \\ &= 1 - p\prod_{k=j+1-l}^{j-1}\frac{1}{y_k}. \end{aligned} \tag{17}$$

Suppose now that $\lim_{j\to\infty} y_j$ exists and equals λ^*. Then λ^* must satisfy the corresponding limit equation of (17), that is

$$\lambda^* = 1 - p/\lambda^{*(l-1)}. \tag{18}$$

To show that (y_j) does converge, we use induction to show that $(y_j)_{j\geq l}$ decreases. This is true for $j = l, l+1, \dots, 2l-2$, where $y_j = 1$.

Suppose now it holds for $j = l, l+1, \dots, n$. Then from (17)

$$y_{n+1} = 1 - p \prod_{k=n+2-l}^{n} \frac{1}{y_k} = 1 - p \frac{1}{y_{n-1}} \frac{1}{y_{n-2}} \cdots \frac{1}{y_{n+1-l}} \left(\frac{y_{n+1-l}}{y_n} \right) \leq y_n, \tag{19}$$

since the bracketed term of (19) is, by induction hypothesis, greater or equal one.
Hence $(y_j)_{j\geq l}$ is decreasing and the proof is complete. ■

The following result specifies the limit λ^* described in Lemma 3.7.(ii), which will be needed in Section 4.

Lemma 3.8 $(y_j) \downarrow \lambda^*$ *where* λ^* *is the* largest *root of the equation*

$$\lambda = 1 - p/\lambda^{(l-1)}.$$

Proof. Define

$$g(\lambda) = 1 - p/\lambda^{(l-1)} - \lambda.$$

Then

$$g(\lambda) = 0 \Leftrightarrow G(\lambda) = \lambda^{l-1} - \lambda^l - p = 0.$$

Note that $G(0) = G(1) = -p$. Also,

$$G'(\lambda) = (l-1)\lambda^{l-2} - l\lambda^{l-1} = \lambda^{l-2}(l - 1 - l\lambda) = 0$$

iff $\lambda = (l-1)/l$. Since

$$G((l-1)/l) = ((l-1)/l)^{l-1}(1 - (l-1)/l) - p = ((l-1)/l)^{l-1}/l - p,$$

we have

$$\max_{\lambda\in[0,1]} g(\lambda) \geq 0 \Leftrightarrow \max_{\lambda\in[0,1]} G(\lambda) \geq 0$$

iff $p \leq ((l-1)/l)^{l-1}/l$. But this is always true for an uncorrelated H as we have shown in Lemma 3.6. This means that $g(\lambda) = 0$ has at least one solution as $g(0+) = -\infty$ and $g(1) = -p$. Now it is easy to check that $g''(\lambda) < 0$ on $]0,1[$ and hence g is concave on $]0,1[$. This reduces the number of roots of $g(\lambda) = 0$ to exactly one or two. Since $g''(\lambda) < 0$ we have that $h(\lambda) := g(\lambda) + \lambda$ is concave on $]0,1[$. Also, $h'(\lambda) < 1$ on $](l-1)/l, 1]$. Thus, in particular, $h'(\lambda) < 1$ on $]\lambda^*, 1]$, where λ^* is the largest root of $g(\lambda) = 0$.
Let us now analyze the iteration

$$y_j = 1, l \leq j \leq 2l-2,$$

and, for $j \geq 2l-2$,

$$y_j = 1 - p \prod_{k=j+1-l}^{j-1} \frac{1}{y_k}. \tag{20}$$

Set

$$z_k := y_k/\lambda^*, z_k > 1, l \leq k \leq 2l-2.$$

We obtain, with (18),

$$\frac{\lambda^{*l}}{p}(z_j - 1) = -\left[\prod_{k=j+1-l}^{j-1} \frac{1}{z_k} - 1 \right].$$

Set now

$$\varepsilon_k := 1 - z_k, \varepsilon_k < 0, l \le k \le 2l-2.$$

This leads to

$$\log\left[1 + \frac{\lambda^{*l}}{p}\varepsilon_j\right] = -\sum_{k=j+1-l}^{j-1} \log(1-\varepsilon_k), \tag{21}$$

which shows that $\varepsilon_k \le 0$, for all k. As (y_j) is monotone decreasing, it must converge to λ^*.

To analyze (21), we note that, in a neighbourhood of λ^*, we have, by definition of ε_k,

$$\frac{\lambda^{*l}}{p}\varepsilon_j \sim -\sum_{k=j+1-l}^{j-1} \varepsilon_k.$$

The recurrence defined by

$$\frac{\lambda^{*l}}{p}\varepsilon_j = -\sum_{k=j+1-l}^{j-1} \varepsilon_k \tag{22}$$

has a characteristic equation

$$\frac{\xi^{l-1}-1}{\xi-1} = \frac{\lambda^{*l}}{p}\xi^{l-1}. \tag{23}$$

The dominant root ξ^* is real and given, with $\tilde{p} := p^{1/(l-1)}$, by

$$\xi^* = \tilde{p} + \tilde{p}^2/(l-1) + \mathcal{O}(\tilde{p}^3).$$

A more detailed analysis is given in the next section. We obtain $\varepsilon_j \sim C\xi^{*j}$, where C is fixed by some initial conditions in the neighbourhood of λ^*. ■

Note that λ^* is equal to $1/z^*$, where z^* is the dominant singularity (of smallest module) of $F(z) := 1 - z + pz^l$. The justification of this equality is given in Sec.4 .

Let $H = (H_1, H_2, \ldots, H_l)$ be as before a pattern of length l of letters from A. Define I_k to be the random indicator variable of the event that the pattern appears ending at k, i.e.

$$I_k = [\![X_{k-l+1} = H_1, \ldots, X_k = H_l]\!] \tag{24}$$

for $k = l, l+1, \ldots$. For completeness, we define I_k to be identically 0 if $k < l$. We assume that the pattern, H, is uncorrelated. In terms of the I_k, this assumption is (see Section 1)

$$I_k = 1 \text{ implies } I_j = 0 \text{ for } j = k-1, \ldots, k-l+1. \tag{25}$$

Recall that O_n denotes the number of times the pattern has occured by time n (see (7), that is

$$O_n = \sum_{k=1}^{n} I_k, \tag{26}$$

for $n = 1, 2, \ldots$. We now show that the sequence $P(O_n = 1)$ is unimodal in n, in the sense that this probability will either decrease for $n = 1, 2, \ldots$ (this is possible for $l = 1$) or else increase and then decrease. Our original proof consisted of several quite technical parts using the recurrence relation given in Lemma 3.5. However T.S. Ferguson (2002) instigated by our result found a nicer approach. The interesting part of this is that Ferguson saw that a *stronger* statement (saying that a certain ratio of probabilities is increasing) was easier to prove than a weaker statement about differences we actually needed, and what we had proved. The essence of it is the following Lemma (proved by Ferguson). The referee conjectured this Lemma too and proposed an elegant alternative proof of the single-island property of the optimal rule.

Lemma 3.9 . *(Ferguson 2002, private communication).* $P(O_n = 1)/P(O_n = 0)$ *is nondecreasing in* n *on the set* $\{n > 0 : P(O_n = 0 > 0)\}$.

Proof. Suppose $P(O_{n+1} = 0) > 0$. We are to show

$$\frac{P(O_{n+1} = 1)}{P(O_{n+1} = 0)} \geq \frac{P(O_n = 1)}{P(O_n = 0)} \tag{27}$$

If $P(I_{n+1} = 1) = 0$, then (27) holds with equality. So suppose $P(I_{n+1} = 1) > 0$. Since,

$$P(O_{n+1} = 0) = P(O_n = 0) - P(O_n = 0, I_{n+1} = 1)$$

and

$$P(O_{n+1} = 1) = P(O_n = 1) + P(O_n = 0, I_{n+1} = 1) - P(O_n = 1, I_{n+1} = 1),$$

we have

$$\frac{P(O_{n+1} = 1)}{P(O_{n+1} = 0)} = \frac{P(O_n = 1) + P(O_n = 0, I_{n+1} = 1) - P(O_n = 1, I_{n+1} = 1)}{P(O_n = 0) - P(O_n = 0, I_{n+1} = 1)}$$

Eq. (27) holds if, and only if

$$[P(O_n = 0) + P(O_n = 1)]P(O_n = 0, I_{n+1} = 1) \geq P(O_n = 0)P(O_n = 1, I_{n+1} = 1).$$

Using (25), this holds if, and only if

$$\begin{aligned} &[P(O_n = 0) + P(O_n = 1)]P(O_{n-l+1} = 0, I_{n+1} = 1) \\ &\geq \; P(O_n = 0)P(O_{n-l+1} = 1, I_{n+1} = 1). \end{aligned} \tag{28}$$

Since the I_i are independent of I_k's with $|k - i| \geq l$, we have that O_{n-l} is independent of I_n. Since $P(O_{n+1} = 1) > 0$, (28) holds if, and only if

$$[P(O_n = 0) + P(O_n = 1)]P(O_{n-l+1} = 0) \geq P(O_n = 0)P(O_{n-l+1} = 1).$$

Thus (27) holds if, and only if

$$1 + \frac{P(O_n = 1)}{P(O_n = 0)} \geq \frac{P(O_{n-l+1} = 1)}{P(O_{n-l+1} = 0)}.$$

Therefore the result is true by induction, provided it is true for the first l integers, which is obvious.
■

Proof of Theorem 3.2
We want to show

$$P(O_n = 1) > P(O_{n-1} = 1) \Rightarrow P(O_{n-1} = 1) > P(O_{n-2} = 1) \tag{29}$$

Proof. We first note that unless $P(I_n = 1) > 0$ we cannot have $P(O_n = 1) > P(O_{n-1} = 1)$ and so we suppose $1 > P(I_n = 1) > 0$. Further, the implication is then trivial if $P(O_{n-1} = 1) = 0$ because $\{O_{n-1} = 0\} \subseteq \{O_{n-2} = 0\}$. Hence we suppose $P(O_{n-1} = 1) > 0$. Further, (29) is trivially satisfied if $n < l + 2$, hence we also suppose $n \geq l + 2$. Now we use again

$$\begin{aligned} P(O_n = 1) &= P(O_{n-1} = 1, I_n = 0) + P(O_{n-1} = 0, I_n = 1) \\ &= P(O_{n-1} = 1) - P(O_{n-1} = 1, I_n = 1) + P(O_{n-1} = 0, I_n = 1) \\ &= P(O_{n-1} = 1) + P(I_n = 1)[P(O_{n-l} = 0) - P(O_{n-l} = 1)] \end{aligned}$$

Thus, with the conditions $P(I_n = 1) > 0$ and $P(O_{n-1} = 1) > 0$ being satified, the implication (29) is equivalent to

$$P(O_{n-l} = 0) > P(O_{n-l} = 1) \Rightarrow P(O_{n-l-1} = 0) > P(O_{n-l-1} = 1).$$

This again is equivalent to

$$P(O_{n-l-1}=1)\geq P(O_{n-l-1}=0)\Rightarrow P(O_{n-l}=1)\geq P(O_{n-l}=0) \tag{30}$$

But since $n\geq l+2$, we can rewrite (30) in the form

$$P(O_k=1)\geq P(O_k=0)\Rightarrow P(O_{k+1}=1)\geq P(O_{k+1}=0)$$

with $k=1,2,\ldots$ Therefore the result follows from Lemma 3.9 saying that $P(O_k=1)/P(O_k=0)$ is non-decreasing, because the condition $P(O_k=0)>0$ is now granted. ■

We are now ready to describe the optimal strategy.

Theorem 3.10 *Let*

$$s^* = argmax_{1\leq s\leq n} P(O_{n-s+1}=1).$$

The optimal strategy for stopping on the last pattern H in $X_1X_2\ldots X_n$ is to stop at the first complete pattern H (if any) within the string $X_{s^*}X_{s^*+1}\ldots X_n$. The corresponding value is given by

$$V=V(n,l,p)=P(O_{n-s^*+1}=1).$$

Proof. We first note that the above defined strategy is successful if and only if

a) $T_1\leq n$
b) $\exists$ exactly one H in the (s^*+l-1,n)-string.

We lose on $\{T_1>n\}$ and put the stopping time $\tau=n+1$ on $\{T_1>n\}$, with payoff 0. Let $\mathcal{F}_k$ be the σ-field generated by $\{X_1,X_2,\ldots,X_k\}$, and let $S_{\mathcal{F}}$ be the set of all random variables with values in $\{1,2,\ldots,n\}$ satisfying $\{\tau=k\}\in\mathcal{F}_k$. Our goal is to find the value V defined by

$$V=\sup_{\tau\in S_{\mathcal{F}}} P(J_\tau=1), \tag{31}$$

where

$$J_\tau = [\![\tau= \text{ is the last time of completion in } X_1,X_2,\ldots,X_n]\!].$$

Since the J_k are in general not $\mathcal{F}_k$-measurable we do the classical trick (see e.g. Chow et al [9], p.46) to replace J_k by $L_k=\mathbb{E}(J_k|\mathcal{F}_k)$, which are all $\mathcal{F}_k$-measurable and solve the equivalent stopping problem $\{L_k,\mathcal{F}_k\}_{k=1,\ldots,n}$.

We shall now prove

$$V=P(O_{n-s^*+1}=1), \tag{32}$$

but we give here our referee's argument which is more elegant than our original one. Indeed, let v_n be the optimal value (success probability) when we have just observed pattern H, but still have a string of length n. If we stop immediately, the reward is the probability that no further pattern appears in the future, i.e. $P(O_n=0)$, where n denotes now the number of remaining draws. If we do not stop and continue observations the optimal value is given by

$$\sum_{j=l}^{n}\tilde{p}_{j,1}v_{n-j},$$

where $\tilde{p}_{j,1}:=\tilde{p}_{j,1}(n)$ denotes the probability that the first pattern H within the *last* n observations is completed at time j.

Thus the principle of optimality yields

$$v_n = max\left\{P(O_n=0),\sum_{j=l}^{n}\tilde{p}_{j,1}v_{n-j}\right\}.$$

This stopping problem is then in the monotone case if the following condition holds:

$$P(O_n = 0) \geq \sum_{j=l}^{n} \tilde{p}_{j,1} P(O_{n-j} = 0) \Rightarrow P(O_{n-1} = 0) \geq \sum_{j=l}^{n} \tilde{p}_{j,1} P(O_{n-1-j} = 0).$$

Considering that $\sum_{j=l}^{n} \tilde{p}_{j,1} P(O_{n-j} = 0) = P(O_n = 1)$, this condition is equivalent to the one given in Lemma 3.9. This completes the proof. ■

Remark 3.11 *More on Problem 2.2.*

It will turn out that Problem 2.2 is quite similar to Problem 2.1. This is why we give only the essential steps.

Theorem 3.12 *Let*

$$s_m^* = argmax_{1 \leq s \leq n} P(O_{n-s+1} = m).$$

The optimal strategy for stopping on the mth last pattern H in $X_1 X_2 \dots X_n$ is to stop at the first complete pattern H (if any) within the string $X_{s_m^*} X_{s_m^*+1} \dots X_n$.

Proof. This follows from the proof of theorem 3.10 by substituting s^* by s_m^*, T_1 by T_m and then using the following Lemma . ■

Lemma 3.13 *: $P(O_n = m+1)/P(O_n = m)$ is non-decreasing in n on the set $\{n > 0 : P(O_n = m) > 0\}$.*

Proof. Similarly as in the proof of Ferguson's Lemma, we can decompose

$$\begin{aligned} P(O_{n+1} = m+1) &= P(O_n = m+1) - P(O_n = m+1, I_{n+1} = 1) + P(O_n = m, I_{n+1} = 1) \\ &= P(O_n = m+1) + P(I_{n+1} = 1)[P(O_{n-l+1} = m) - P(O_{n-l+1} = m+1)] \end{aligned}$$

because H is uncorrelated. Hence we can obtain the result by simply following all subsequent steps of proof of Lemma 3.9. ■

This implies also the corollary:

Corollary 3.14 $\forall m = 1, 2, \dots : P(O_n = m)_{n \geq ml}$ *is unimodal in n.*

Remark 3.15

Recall that T_k is the time of completion of the kth pattern in an infinite string. Since the pattern H is uncorrelated, each completion is, in the homogeneous case, the birth time of an identically distributed time interval up to the next completion. Hence the process $(T_k)_{k \geq 1}$ is a renewal process where all $T_k - T_{k-1}$ are i.i.d. with the same law, which is just the law of T_1. This is the reason why the following extension of Lemma 3.5 is possible. Recall the notation

$$p_{j,k} := P(T_k = j).$$

Clearly $p_{j,k} = 0$ for $j = 1, 2, \dots, kl - 1$ and

$$p_{kl,k} = (P(H))^k = p_l^k = p^k. \tag{33}$$

From the renewal process property we find now by the same reflection argument as in the proof of Lemma 3.5 the analogous recurrence relation

$$p_{j+1,m} = p_{j,m} - [\![j \geq (m+1)l - 1]\!] p p_{j+1-l,m}, \text{ for } j \geq ml. \tag{34}$$

This recurrence relation is of independent interest.

Remark 3.16 *Inhomogeneous case:*

Note that neither in the preceeding Lemma 3.13 nor in Lemma 3.9 any homogeneity condition was used so that Ferguson's proof assures the single-island property of the optimal rule also in the *inhomogeneous case.* However other questions concerning the inhomogeneous case will, as well as the correlated-pattern case, require more research so that the remainder of this paper is still centered on the homogeneous case.

4 An Asymptotic Analysis

4.1 Analysis of $P(O_n = 1)$

Fix l and let us analyze the case where p is small. We study the homogeneous case only and start from (8), i.e.

$$g(z) = \frac{pz^l}{F(z)^2},$$

with $F(z) := 1 - z + pz^l$. We will derive several asymptotic equivalents for $p \downarrow 0$. $F(z)$ possesses a dominant singularity (of smallest module) given by

$$z^* = 1 + c_1 p + c_2 p^2 + c_3 p^3 + c_4 p^4 + \mathcal{O}(p^5), p \downarrow 0, \tag{35}$$

where, by inverting $F(z^*) = 0$ (using Lagrange's inversion theorem), we obtain the coefficients

$$\begin{aligned} c_1 &= 1, \\ c_2 &= l, \\ c_3 &= l(3l-1)/2, \\ c_4 &= l(4l-1)(2l-1)/3. \end{aligned}$$

Let us now write down the asymptotic expansion of $1/F^2(z)$ in the neighbourhood of z^*. We derive

$$\frac{1}{F^2(z)} = \alpha_2/(z-z^*)^2 + \alpha_1/(z-z^*) + \mathcal{O}(1),$$

where

$$\alpha_1 = \frac{-z^*(z^*-1)l(l-1)}{(z^*(l-1)-l)^3},$$

and

$$\alpha_2 = \frac{z^{*2}}{(z^*(l-1)-l)^2}.$$

By classical singularity analysis, (see Flajolet and Sedgewick [11]), this gives

$$\varphi_1(j) := P(O_j = 1) = [z^j]g(z) \sim \frac{p\alpha_2(j-l+1)z^{*l}}{z^{*(j+2)}} - \frac{p\alpha_1 z^{*l}}{z^{*(j+1)}} + \mathcal{O}(R^{-j}), j \to \infty \tag{36}$$

for $R > z^*$, where $[z^n]g(z)$ denotes the coefficient of z^n in the Taylor expansion of $g(z)$. We can, of course, simplify (36) by using $z^{*l} = -(1-z^*)/p$.

Also, the generating function of $\Delta_n := P(O_n = 1) - P(O_{n-1} = 1)$ is given by $h(z) = (1-z)g(z)$, and we obtain

$$\varphi_2(j) := \Delta_j = [z^j]h(z) \sim (1-z^*)\varphi_1(j) + p\alpha_2 z^{*l}/z^{*(j+1)}. \tag{37}$$

Finally we see that

$$\frac{1}{F(z)} = \alpha_3/(z-z^*) + \mathcal{O}(1), \tag{38}$$

where

$$\alpha_3 = \frac{z^*}{z^*(l-1)-l}.$$

Hence, from (11),

$$\varphi_3(j) := P(T_1 = j) = [z^j]f(z) \sim -\frac{p\alpha_3 z^{*l}}{z^{*(j+1)}}, \tag{39}$$

which shows that the limit λ^* in Lemma 3.7 is indeed equal to $1/z^*$.
We must now evaluate

$$j^* := \inf_j \{P(O_j = 1) \text{ is maximal on } [l..\infty]\}.$$

A first solution would be to compute $\varphi_1'(j)$ and to define j_1^* such that $\varphi_1'(j_1^*) = 0$. This leads to

$$j_1^* \sim \frac{z^*(l-1) - z^*\ln(z^*)(l-1) - l}{\ln(z^*)(z^*(l-1)-l)}. \tag{40}$$

Plugging (35) into (40) gives

$$j_1^* = 1/p - 1/2 + p(l^2 - l - 1/6)/2 + p^2(2l-1)(16l^2 - 16l - 1)/24 + \mathcal{O}(p^3).$$

A more precise way is to start directly from (37) and to define j_2^* such that $\varphi_2(j_2^*) = 0$. This leads to

$$j_2^* \sim \frac{-z^*}{(z^*-1)(z^*(l-1)-l))}. \tag{41}$$

Again, plugging (35) into (41) gives

$$j_2^* = 1/p + pl(l-1)/2 + p^2 2l(2l-1)(l-1)/3 + \mathcal{O}(p^3).$$

We note that j_1^* has only the term $1/p$ in common with j_2^*.

The optimal probability is asymptotically given, from (36), by

$$\begin{aligned} P_2^* = \varphi_1(j_2^*) &= e^{-1}[1 + p(2l-1)/2 + p^2(36l^2 - 36l + 7)/24] + \mathcal{O}(p^3). \\ &= e^{-1}[1 + g(p,l)], \text{ say.} \end{aligned} \tag{42}$$

If we use j_1^*, we have two terms in common with (42).

The interpretation of P_2^* is as follows: as we see from (42), the optimal winning probability is close to $1/e$ provided $g(p,l)$ is small. Indeed, it is interesting to see that this fact is independent of the sequence of occurrences probabilities, provided that the string is just long enough for the odds of occurrences to add up to 1 (see Bruss [7]). Now it is true that this Theorem assumes independence of occurrences in time and that this very hypothesis is not satisfied here (because H depends on l states). On the other hand, we have seen through the analysis of the convergence of $p_{j+1,1}/p_{j,1}$ to λ^*, that the approximation $p_{j+k,1} \sim \lambda^{*k} p_{j,1}$ is excellent (even for smaller j) if $p = P(H)$ is small. By the renewal process property the same holds for the $p_{j,m}$, that is $p_{j+k,m} \sim \lambda^{*k} p_{j,m}$. Inter-occurrence times are therefore approximately geometrically distributed as if they were perfectly independent RV's in the homogeneous case. Therefore, numerically, this has the same effect as if all occurrences were independent.

Finally, we can analyze, in the neighbourhood of j_2^*, the ratio $P(O_j = 1)/P(O_j = 0)$. Indeed, $P(O_j = 0) \equiv P(T_1 > j)$ and, from (12), (38), we readily derive

$$P(T_1 > j) \sim -\frac{\alpha_3}{z^{*(j+1)}}.$$

Now, with (36), (41) and $j = j_2^* - d$, this gives

$$R(d) := P(O_{j_2^*-d} = 1)/P(O_{j_2^*-d} = 0) \sim \frac{z^*(d-1)-d}{z^*(l-1)-l}$$

We note that $R(l) \sim 1$, which is intuitively obvious. Also, with (35),

$$R(d) \sim 1 + (l-d)p + (-l + d + 2l^2 - 2dl)p^2 + (l - d + 9l^3/2 - 9l^2/2 - 9dl^2/2 + 9dl/2)p^3 + \mathcal{O}(p^4);$$

4.2 Analysis of λ^* and ξ^*

$\lambda^* = 1/z^*$ can be analyzed as in the previous subsection. Starting from (35) or directly from (18), we obtain

$$\lambda^* = 1 - p + (-l+1)p^2 - 1/2(3l-2)(l-1)p^3 - 1/3(l-1)(4l-3)(2l-1)p^4 + \mathcal{O}(p^5), p \downarrow 0. \quad (43)$$

Let us now turn to (23). We see that, for $j \leq l-2$, ξ^* is given by a series $\sum_1^j c_k \tilde{p}^k + \mathcal{O}(\tilde{p}^{j+1})$, the first terms of which are given by

$$\xi^* = \tilde{p} + \frac{1}{l-1}\tilde{p}^2 + \frac{l+2}{2(l-1)^2}\tilde{p}^3 + \frac{(l+3)(l+1)}{3(l-1)^3}\tilde{p}^4 + \mathcal{O}(\tilde{p}^5). \quad (44)$$

For $j = l-1 \ldots 2(l-2)$, λ^{*l} contributes with the first two terms of its expansion: $\lambda^* = 1 - p + \ldots$,etc.

4.3 The optimal probability for $p = p_{max}$

Recall that, from Lemma 3.6, $p \leq \gamma(l)$, with $\gamma(l) := \frac{1}{l}(1-1/l)^{l-1}$. What is the asymptotic expression for P_2^* for $p = \gamma(l)$ and l large? First of all, $F(z)$ becomes, with $\varepsilon := 1/l$

$$F(z) = 1 - z + (1-\varepsilon)^{1/\varepsilon - 1}\varepsilon z^{1/\varepsilon}.$$

This leads to a very simple singularity: $z^* = 1/(1-\varepsilon)$. We obtain

$$\begin{aligned} F(z) &\sim (\varepsilon-1)^2/(2\varepsilon)(z-z^*)^2 + (2\varepsilon-1)(\varepsilon-1)^3/(6\varepsilon^2)(z-z^*)^3 \\ &+ (3\varepsilon-1)(2\varepsilon-1)(\varepsilon-1)^4/(24\varepsilon^3)(z-z^*)^4 + (4\varepsilon-1)(3\varepsilon-1)(2\varepsilon-1)(\varepsilon-1)^5/(120\varepsilon^4)(z-z^*)^5 \\ &+ \mathcal{O}((z-z^*)^6). \end{aligned}$$

Of course, there is a general expression for all terms, but we need only these ones here. Note that z^* is now a fourth-order singularity of $g(z)$. Omitting the details, this leads to

$$\begin{aligned} \varphi_1(j) &\sim \frac{2}{135}(1-\varepsilon)^{j-1}[4 - 6\varepsilon + (-45j^2 - 45j - 6)\varepsilon^2 + (45j^3 + 90j^2 + 45j + 4)\varepsilon^3], \quad (45) \\ \varphi_2(j) &\sim \frac{-2}{135}\varepsilon(1-\varepsilon)^{j-2}[4 + (-6+90j)\varepsilon + (-180j^2 - 90j - 6)\varepsilon^2 + (45j^3 + 90j^2 + 45j + 4)\varepsilon^3]. \end{aligned}$$

The solution of

$$\varphi_2(j_2^*) = 0 \quad (46)$$

is given by

$$j_2^* \sim \beta/\varepsilon + \gamma + \mathcal{O}(\varepsilon). \quad (47)$$

We give only the asymptotically relevant terms in the expansion. β is the suitable solution of

$$-4 - 90\beta + 180\beta^2 - 45\beta^3 = 0.$$

This is given (see e.g. Abramowitz and Stegun [1]) by

$$\beta = \frac{2}{3}10^{1/2}\cos[\arctan(1761^{1/2}/67)/3] + 4/3 = 3.404953466\ldots$$

γ is the solution of the linear equation

$$-135\beta^2\gamma + 6 + 360\beta\gamma - 90\beta^2 + 90\beta - 90\gamma = 0,$$

which gives $\gamma = -1.70247673318\ldots$

The optimal probability is finally given by

$$\varphi_1(j_2^*) \sim \frac{2}{135}e^{-\beta}(4 - 45\beta^2 + 45\beta^3) = 0.6192522270\ldots \quad (48)$$

Note that the optimal win probability is now much larger than $1/e$.

4.4 Bounds and "typical" bounds

There are no non-trivial uniform bounds for optimal success probability. The lower bound is 0, assumed for $p = 0$ *or* $n < l$, the upper bound is 1 for $p = 1$ *and* $n \geq l$. However there are two "typical" bounds worth to remember.

The lower asymptotic bound for $p > 0$ and fixed l is $1/e$, well-known from the secretary problem and the like. (See e.g. Samuels [23] for a review on secretary problems; see also the recent paper of Gnedin [12] on related problems). The upper bound is new. It is obtained by a specific choice of a pattern of the form $ABBB \ldots B$ or $BAAA \ldots A$ with maximal pattern probability $p = \frac{1}{l}(1 - 1/l)^{l-1}$ (see Lemma 3.6). It is about $0.619\ldots$ and is realized for any string of length n not less than about $3.4\ldots \times l$. Most interesting cases have optimal success probabilities somewhere in this range. This range is interesting.

As an illustration let us refer to our investment problem of Section 2. There we speak of two complementary events "up" and "down". Suppose equality of prices on two consecutive days have a negligible probability: it is comforting for an investor to know that, for a pattern down-up-up-up or up-down-down-down, an horizon of just $3.4\ldots \times 4 \sim 14$ trading days can, depending on the up-or-down probability, be enough for having a success probability of more than 60%.

4.5 Aldous' Poisson Clumping Heuristic

If n becomes large compared to l then the local dependence created by the pattern (with memory $l-1$) becomes negligible and hence Poisson clumping heuristic (Aldous [2]) becomes the prime candidate for approximation. (For coumpound Poisson approximation see also the recent user's guide by Barbour and Chryssaphinou [5]).

Here is what Poisson Clumping Heuristic can do. We follow the notation of Aldous (see subsection B5). For fixed $H = H_1 H_2, \ldots, H_l$ let S_H be the set of random times k such that $X_{k-(l-1)} X_{k-(l-2)} \ldots X_{k-1} X_k = H$, and let C_H be the number of times pattern H appears in a position overlapping this pattern H including H itself. Knowing $P(H_1), P(H_2), \ldots, P(H_l)$ we can directly compute $\mathbb{E}(C_H)$ and use what Aldous ([2]) calls the fundamental identity to get the H-clump rate, that is

$$\lambda_H = P(H)/\mathbb{E}(C_H). \tag{49}$$

This yields the approximation

$$P(T_1 \geq l + j) \sim e^{-j\lambda_H}, \quad j = 0, 1, 2, \ldots$$

and hence, by the memoryless property, the process $T_1, T_2, \ldots$ is approximately a homogeneous Poisson process with rate λ_H.

For uncorrelated patterns H we have $C_H = 1$, of course, so that from(49) $\lambda_H = P(H) = p$. Maximizing the probability of $\{O_n = 1\}$ over the tail-string of length n is trivial in terms of a Poisson process, namely, allowing for real n,

$$\text{argmax}_n \{pne^{-pn}\} = \frac{1}{p}. \tag{50}$$

Also, the unimodality question, which took some effort to prove in our problem, is here no problem at all: $\varphi(x) := xe^{-x}$ satisfies $\varphi'(x) \geq 0$ on $[0, 1]$ and $\varphi(1) = e^{-1}$. The odds-algorithm (see Bruss [7]) gives for $l = 1$ the optimal threshold time

$$s^* = \text{argmax}_{1 \leq s \leq n} P(O_{n-s+1} = 1) = \left\lceil \frac{1-p}{p} \right\rceil. \tag{51}$$

We see that the approximation (50) for (51) is good for small p, in particular since the function xe^{-x} is quite flat in the neighbourhood of its maximum $x = 1$. Another advantage of the Poisson clumping heuristic is that it works in principle for general patterns with clump size λ_H, provided λ_H is small,

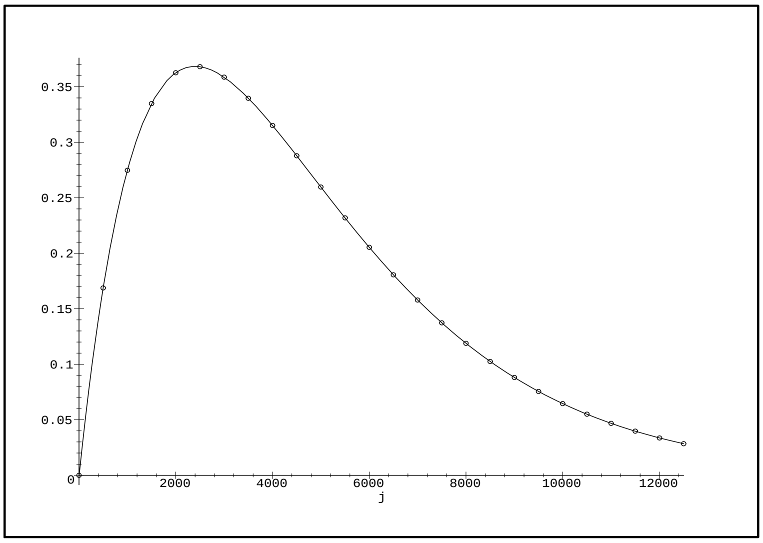

Figure 1: Observed and asymptotic $P(O_j = 1)$ values (asymptotic=line, exact=circle) for an alphabet of 7 letters with probability 1/7 each and an uncorrelated pattern of length 4. The maximal value is close to $1/e$.

again with the question of unimodality being no problem at all. With some smoothness condition for pattern probabilities varying in time it can also be extended to the inhomogeneous case. All these advantages make the Poisson Clumping Heuristic an impressive tool.

For our more specific problem, it has of course its limits. It is a tool desiged to obtain long-range approximations, and not designed for cases where, in Aldous' terminology, the "clumps" can be large compared to the "area". If l increases it does therefore little to explain optimal win probabilities which can be impressively large compared to $1/e$. Recall for example our investment problem for the uncorrelated pattern $UDDD$ respectively $DUUU$, with $P(U) = 1 - P(D)$ and $P(U) = 1/l$ respectively $P(U) = 1 - 1/l$, where the optimal win probability attains values in the region of $0.619\ldots$

Remark 4.1

We have proved no result of sufficiently general interest in the inhomogeneous case, but simulations indicate that the odds-algorithm will yield good approximations in this case as well. Given the simplicity of the odds-algorithm, this fact attracts our particular interest.

4.6 A numerical example

Let us now test our asymptotics on two examples. First we have chosen an alphabet of 7 letters, each drawn with a probability 1/7. We look for an uncorrelated pattern of 4 letters. Figure 1 shows the exact probability $P(O_n = 1)$ given by $[z^j]g(z)$ and the asymptotic equivalent $\varphi_1(j)$ given by (36) . Figure 2 shows the same values near the maximum, with $\varphi_1(j)$ shown above with only the α_2 term in (36) and below with α_2 and α_1 terms. The fit with the two terms is excellent. Figure 3 gives the exact values of Δ_j, near the maximum, given by $[z^j]h(z)$ together with $\varphi_2(j)$. z^* is given by $1.0004171885876380379\ldots$. We give all numbers with precision to enable the reader to see the variation of our asymptotics. The asymptotics (35) with respectively $[c_1, c_2]$, $[c_1, c_2, c_3]$, and $[c_1, c_2, c_3, c_4]$, are given by $1.00041718699396562\ldots$, $1.0004171885834131678\ldots$, $1.0004171885876258568\ldots$. The extremum position j^* is given by $j^* = 2401$ and $P_{opt} := [z^{j^*}]g(z) = .36841687858219491068\ldots$, $[z^{j^*-1}]g(z) = .36841687842139940037\ldots$. These two values are very close. Finally we compute

$$j_1^* = 2400.5024739145375530\ldots$$

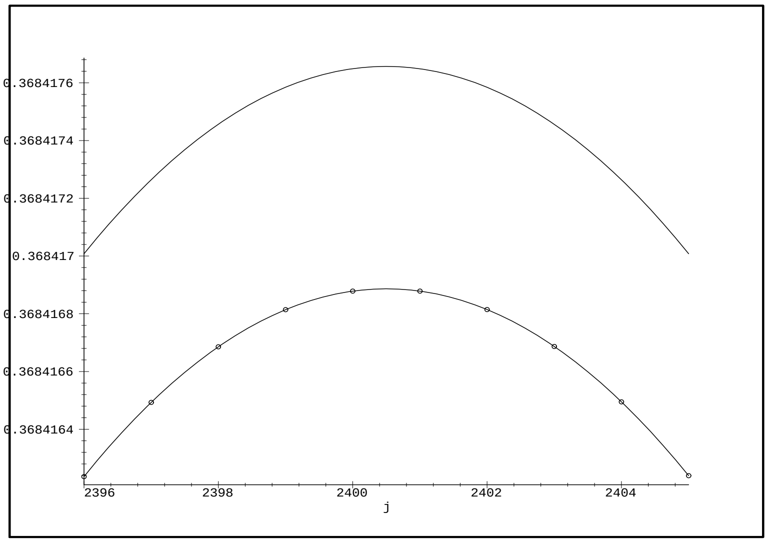

Figure 2: Relates to Figure 1. Observed and asymptotic $P(O_j = 1)$ values near the maximum (asymptotic=line, exact=circle)

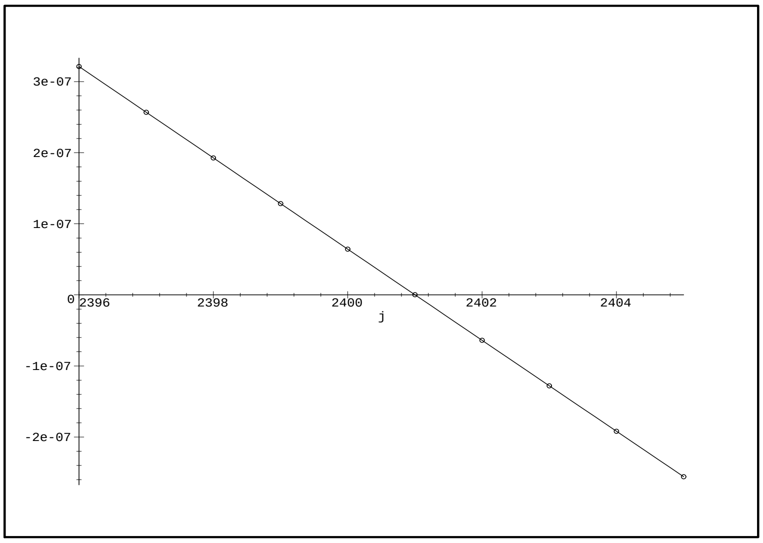

Figure 3: Relates to Figure 1. Observed and asymptotic Δ_j values near the maximum (asymptotic=line, exact=circle)

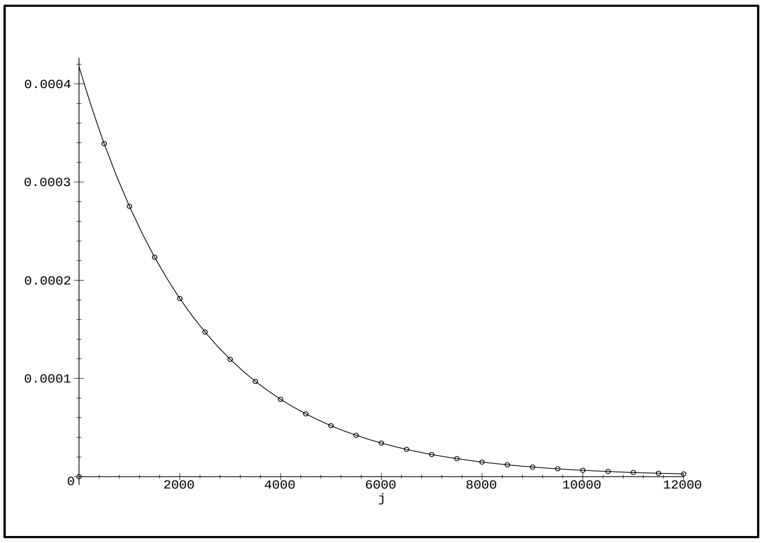

Figure 4: Relates to Figure 1. Observed and asymptotic $P(T_1 = j)$ values (asymptotic=line, exact=circle)

$$\begin{aligned} j_2^* &= 2401.0025086728926115\ldots \\ P_2^* &= .36841687585687129705\ldots \end{aligned}$$

Figure 4 shows the exact values of $P(T_1 = j)$ given by $[z^j]f(z)$ and the asymptotic equivalent $\varphi_3(j)$ given by (39) .

Let us now turn to λ^* and ξ^*. We obtain $\lambda^* = 0.99958298538609975959\ldots$ and $\xi_3^* = 0.076724664308750244034\ldots$ (dominant root). The equivalents (43) and (44) (with λ^* given up to p^6 precision) lead respectively to $.99958298538609975966 4\ldots$ and $.076724664308750239230\ldots$. Eq. (23) possesses two other conjugate roots $\xi_1 = -.038153737860556103082\ldots - .063101261148889674174\ldots I$, $\xi_2 = \overline{\xi_1}$, with $\rho := |\xi_1| = .073739249191406916608\ldots, \theta := \arg(\xi_1) = -2.1146228290417202536\ldots$. The solution of (22) is given by

$$\varepsilon(j) = A_1\xi_1^j + A_2\xi_2^j + A_3\xi^{*j}$$

Denote by $\tilde{\varepsilon}(j)$ the observed values given by our iteration (20). Using the initial values for $j = 10, 11, 12$, we solve the system

$$\begin{aligned} \tilde{\varepsilon}(10) &= \varepsilon(10), \\ \tilde{\varepsilon}(11) &= \varepsilon(11), \\ \tilde{\varepsilon}(12) &= \varepsilon(12), \end{aligned}$$

for $\{A_1, A_2, A_3\}$. This leads to $A_1 = \beta_1 + \beta_2 I$ where $\beta_1 = -818.12206420325739169\ldots, \beta_2 = -33.996600796082949275\ldots, A_3 = -757.75336792717202639\ldots$. Of course, we must have $A_1 = \overline{A_2}$, so we use the asymptotic equivalent

$$\varepsilon(j) = 2\rho^j[\beta_1 cos(j\theta) - \beta_2 sin(j\theta)] + A_3\xi^{*j}. \quad (52)$$

A comparison between $\tilde{\varepsilon}(j)$ and $\varepsilon(j)$ as given by (52) is shown in Figure 5. The ratio $\tilde{\varepsilon}(j)/\varepsilon(j)$ is given in Figure 6 which reflects the periodic component of (52).

Next we have chosen an alphabet of two letters and look for an optimal uncorrelated pattern of 7 letters. Figure 7 shows the exact probability $P(O_n = 1)$ given by $[z^j]g(z)$ and the asymptotic equivalent $\varphi_1(j)$ given by (45). Figure 8 shows the same values near the maximum. The fit is excellent.

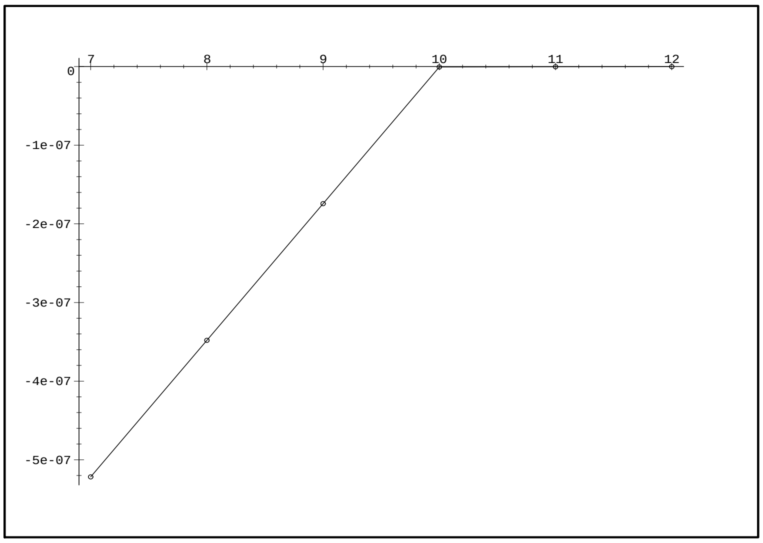

Figure 5: Relates to Figure 1. $\tilde{\varepsilon}(j)$ and $\varepsilon(j)$ (asymptotic=line, observed=circle)

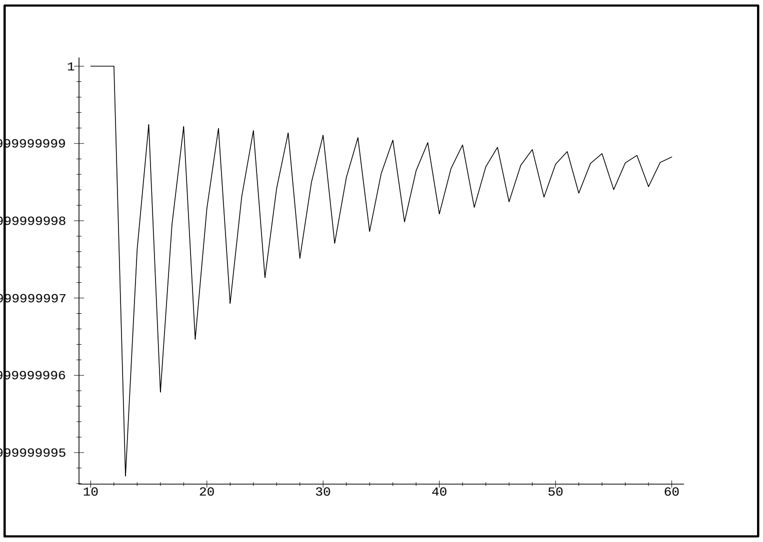

Figure 6: Relates to Figure 1. The ratio $\tilde{\varepsilon}(j)/\varepsilon(j)$

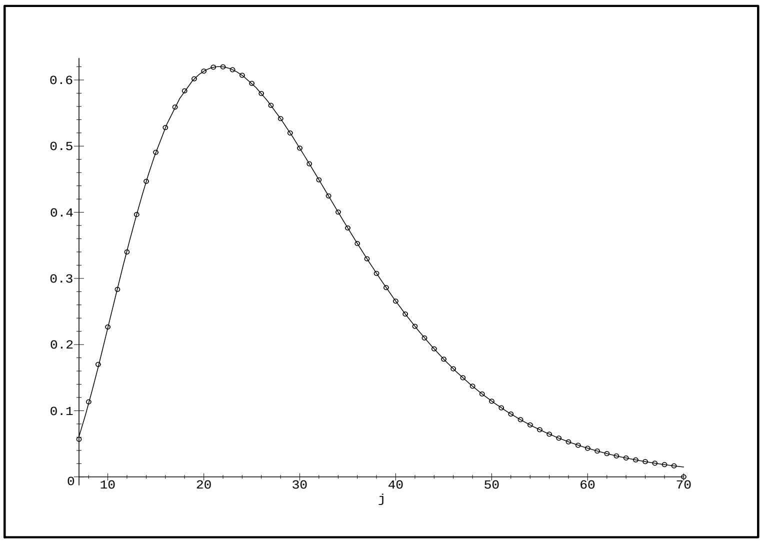

Figure 7: Observed and asymptotic $P(O_j = 1)$ values (asymptotic=line, exact=circle) for an alphabet of 2 letters and an optimal uncorrelated pattern of length 7. In contrast to the situation of Figure 1 the maximum value is now close to the upper asymptotic bound of $0.619\ldots$

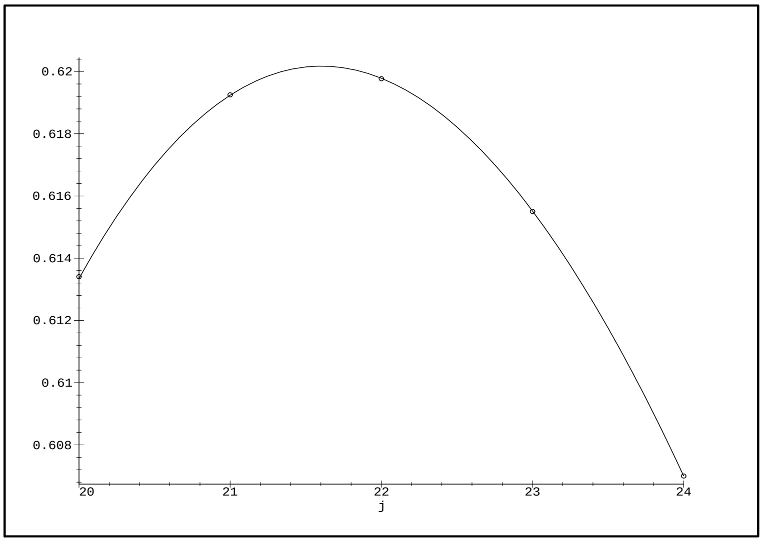

Figure 8: Relates to Figure 7. Observed and asymptotic $P(O_j = 1)$ values (asymptotic=line, exact=circle) near the maximum for an alphabet of 2 letters and an optimal uncorrelated pattern of length 7.

Eq.(46) leads to $j_2^* = 22.1079\ldots$ while (47) gives $j_2^* \sim 22.132\ldots$ (recall that the error is $\mathcal{O}(\varepsilon)$), to be compared with the observed maximum location at $j^* = 22$. The observed optimal probability is given by $P_{opt} := [z^{j^*}]g(z) = .619768\ldots$, and $\varphi_1(j^*) = .619784\ldots$. These values should be compared with the asymptotic optimal win probability given by (48): $0.6192522270\ldots$ The fit is still very good for all practical purposes.

Acknowledgement
We are grateful to Prof. T.S. Ferguson for his proof of Lemma 3.9 and to a referee for a more elegant argument in the proof of Theorem 3.2 as well as other helpful comments.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of mathematical Functions.* Dover, 1965.

[2] D. Aldous. *Probability Approximations via the Poisson Clumping Heuristic.* Springer-Verlag, 1989.

[3] P. Allard and M. Monticino. Optimal stopping rules for directionally reinforced processes. *Advances in Applied Probability*, 33:483–504, 2001.

[4] K. Ano and M. Ando. A note on Bruss' stopping problem with random availability. In *Papers in honor of Thomas S. Ferguson, IMS Lectures Notes- Monograph Series*, volume 35, pages 71–82. 2000.

[5] A. D. Barbour and O. Chryssaphinou. Compound Poisson approximations: a users guide. *Advances in Applied Probability*, 2(3):964–1002, 2001.

[6] F.T. Bruss. On an optimal selection problem of Cowan and Zabczyk. *Journal of Applied Probability*, 24:918–928, 1987.

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

[8] F.T. Bruss and D. Paindaveine. Selecting a sequence of last successes in independent trials. *Journal of Applied Probability*, 37:389–399, 2000.

[9] Y. S. Chow, H. Robbins, and D. Siegmund. *The Theory of optimal stopping.* Dover Publications, 1991.

[10] R. Cowan and J. Zabczyk. On optimal selection problem associated with the Poisson process. *Theory of Probability and Applications*, 23:584–592, 1978.

[11] P. Flajolet and B. Sedgewick. Analytic combinatorics: complex asymptotics and generating functions. Technical Report 2026, INRIA, 1993.

[12] A. Gnedin. Best choice problem from the planar Poisson process. Technical report, 2002. http://front.math.ucdavis.edu/math.PR/0209050.

[13] R.L. Graham, D.E. Knuth, and O. Patashnik. *Concrete mathematics, A foundation for Computer Science.* Addison-Wesley, 1988.

[14] L.J. Guibas and A.M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory*, 30:183–208, 1981.

[15] S.R. Hsiau and J.R. Yang. A natural variation of the standard secretary problem. *Statistica Sinica*, 10:639–646, 2000.

[16] S.R. Hsiau and J.R. Yang. Selecting the last success in Markov-dependent trials. *Journal of Applied Probability*, 39(2):271–281, 2002.

[17] D.E. Knuth, J. Morris, and V. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6:189–195, 1977.

[18] A. Kurishima and K. Ano. A Poisson arrival selection problem for Gamma prior intensity with natural number parameter. *Scientiae Mathematicae Japonicae*, 7:209–223, 2002.

[19] S.Y.R Li. A martingale approach to the study of the occurrence of patterns in repeated experiments. *Annals of Probability*, 8(6):1171–1175, 1980.

[20] R.D. Mauldin, M. Monticino, and H. von Weizsäcker. Directionally reinforced random walks. *Advances in Mathematics*, 117(2):239–252, 1996.

[21] E.L. Presman and I.M. Sonin. The best choice problem for a random number of objects. *Theory of Probability and Applications*, 17:657–668, 1972.

[22] E. Samuel-Cahn. The best choice problem with random freeze on jobs. *Stochastic Processes and Their Applications*, 55:315–327, 1995.

[23] S.M. Samuels. Secretary problems. In *Handbook of Sequential Analysis (B.K Gosh and P.K. Sen, ed.)*, pages 381–405. Marcel Dekker, NY, 1991.

[24] A. Suchwalko and K. Szajowski. On the Bruss stopping problem with general gain function. In *Optimal Stopping and Stochastic Games, Bedlewo, Poland*, 2002. http://www.krc.karelia.ru/events/isdg/text/isdpois.pdf.

[25] K. Szajowski. Game version of the Bruss problem. In *Optimal Stopping and Stochastic Games, Bedlewo, Poland*, 2002. http://www.neyman.im.pwr.pl/oss.2002/abstract/SZAJOW.pdf.

[26] W. Szpankowski. *Average case Analysis of Algorithms on sequences.* Wiley, 2001.

[27] M. Tamaki. Optimal stopping on trajectories and the ballot theorem. *Journal of Applied Probability*, 38(4):946–959, 2001.