J. Appl. Prob. **47,** 761–777 (2010)
Printed in England
© *Applied Probability Trust* 2010

SUM THE MULTIPLICATIVE ODDS TO ONE AND STOP

MITSUSHI TAMAKI,* *Aichi University*

Abstract

We consider the optimal stopping problem of maximizing the probability of stopping on any of the last m successes of a sequence of independent Bernoulli trials of length n, where m and n are predetermined integers such that $1 \le m < n$. The optimal stopping rule of this problem has a nice interpretation, that is, it stops on the first success for which the sum of the m-fold multiplicative odds of success for the future trials is less than or equal to 1. This result can be viewed as a generalization of Bruss' (2000) odds theorem. Application will be made to the secretary problem. For more generality, we extend the problem in several directions in the same manner that Ferguson (2008) used to extend the odds theorem. We apply this extended result to the full-information analogue of the secretary problem, and derive the optimal stopping rule and the probability of win explicitly. The asymptotic results, as n tends to ∞, are also obtained via the planar Poisson process approach.

Keywords: Optimal stopping; secretary problem; planar Poisson process; odds theorem; nonhomogeneous Poisson process; monotone problem; one-stage look-ahead rule

2010 Mathematics Subject Classification: Primary 60G40

1. Introduction

Let n be a given positive integer, and suppose that n independent Bernoulli trials are performed one at a time, each of which results in a *success* or a *failure*. That is, if we let X_j equal 1 if the jth trial is a success and 0 if it is a failure, then $X_1, X_2, \dots, X_n$ are independent Bernoulli random variables that are observed sequentially. When we seek an optimal stopping rule of this sequential observation problem with the objective of maximizing the probability of stopping on the last success, Bruss (2000) gave an elegant solution, given below, where $p_j = \mathrm{P}\{X_j = 1\}$, $q_j = 1 - p_j$, and $r_j = p_j/q_j$ represents the odds of success on the jth trial (if $p_j = 1$ and r_j is taken to be ∞).

Theorem 1.1. (Bruss' sum-the-odds theorem.) *For the above stopping problem, the optimal rule stops on the first success $X_k = 1$ with $k \ge s$, if any, where*

$$s = \min\left\{k \ge 1 : \sum_{j=k+1}^{n} r_j \le 1\right\}. \tag{1.1}$$

Moreover, the maximal win (i.e. achieving the objective) probability is

$$v = \left(\prod_{j=s}^{n} q_j\right)\left(\sum_{j=s}^{n} r_j\right).$$

Received 2 March 2010; revision received 29 April 2010.
* Postal address: Department of Business Administration, Aichi University, Nagoya Campus, 370 Shimizu, Kurozasa, Miyoshi, Aichi 470-0296, Japan. Email address: tamaki@vega.aichi-u.ac.jp

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

The optimal rule (1.1) has a nice interpretation, i.e. it stops on the first success for which the sum of the odds of success for the future trials is less than or equal to 1 (we are indifferent between stopping and continuing if the sum of the odds is equal to 1). We refer to this result as the *sum-the-odds theorem* (called simply STOT) according to Ferguson (2008). Bruss (2003) later showed that the best possible lower bound on the maximal probability is e^{-1} under some condition, more precisely, $v > \mathrm{e}^{-1}$ if $\sum_{j=1}^{n} r_j \geq 1$. See also Hill and Krengel (1992) and Hsiau and Yang (2000) for related works.

Bruss (2000) started his argument by raising the question of correctly guessing the last '6' when a fair die is tossed a fixed number, n, of times. The STOT answers this question immediately. But what about the question of correctly guessing any one of the last two '6', i.e. guessing either the last '6' or the second to last '6'. This question seems a natural extension of the Bruss question, but the STOT cannot answer this question because the optimality criterion is different. In this paper we attempt to answer problems of this kind. The criterion we adopt here is more generally described as maximizing the probability of stopping on any of the last m successes for a predetermined m (we assume that $n > m$ unless otherwise specified, because, for $n \leq m$, the optimal rule evidently stops on the first success). The optimal rule of this problem also has a nice interpretation. That is, it can be shown that the optimal rule stops on the first success for which the sum of the m-fold multiplicative odds of success for the future trials is less than or equal to 1, if we define the sum of the *j-fold multiplicative odds* of successes on the kth trial by

$$R_{k,j} = \sum_{k \leq i_1 < i_2 < \cdots < i_j \leq n} r_{i_1} r_{i_2} \cdots r_{i_j}$$

for $1 \leq j \leq n-k+1$ (if $\max_{k \leq i \leq n} p_i = 1$, $R_{k,j}$ is taken to be ∞) and $R_{k,j} = 0$ for $j > n-k+1$. More explicitly, we have the following result.

Theorem 1.2. (Sum-the-multiplicative-odds theorem (STMOT).) *For the stopping problem of maximizing the probability of stopping on any of the last m successes in n independent Bernoulli trials, the optimal rule stops on the first success $X_k = 1$ with $k \geq s_m$, if any, where*

$$s_m = \min\{k \geq 1 \colon R_{k+1,m} \leq 1\}. \tag{1.2}$$

Moreover, the maximal win probability is

$$v_m = \left(\prod_{j=s_m}^{n} q_j\right)\left(\sum_{j=1}^{m} R_{s_m,j}\right). \tag{1.3}$$

Since $R_{k+1,m} = 0$ for $k \geq n-(m-1)$, we find that $s_m \leq n-(m-1)$ from (1.2), which implies that, as is intuitively clear, the optimal rule stops on the success if the remaining number of trials is less than m. Obviously, the optimal rule may be described as

$$N_m^* = \min\{k \geq s_m \colon X_k = 1\}. \tag{1.4}$$

The proof of Theorem 1.2 (the STMOT) will be given in more generality in Section 2. The STMOT with $m = 1$ is just the STOT. Recall the question of correctly guessing any of the last two '6' in the dice game. This problem occurs if $m = 2$ and $q_j = q = \frac{5}{6}$, $r_j = r = \frac{1}{5}$ for each j in the STMOT, and it is easy to see that $s_2 = n-7$ if $n \geq 8$ from (1.2), because $R_{k+1,2} = \binom{n-k}{2} r^2 \leq 1$ is equivalent to $n-k \leq 7$. In other words, it is optimal to announce

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

the first '6', if any, in the last eight tosses as being either one of the last two '6'. Moreover, the probability of a correct guess is

$$q^8\left[\binom{8}{1}r+\binom{8}{2}r^2\right]\approx 0.6326$$

from (1.3).

The STOT has been extended into other directions in Ferguson (2008). We show in Section 2 that Ferguson's extension can be also made to the STMOT. In Section 3 we apply the STMOT to the celebrated secretary problem which corresponds to the special case $p_i = 1/i,\ 1 \le i \le n$, and examine the asymptotic behavior as $n \to \infty$. In Section 4 we consider the full-information analogue of the secretary problem, and derive the optimal rule and the corresponding probability of win explicitly. We also obtain the asymptotic results using the planar Poisson process model. See Bruss and Paindaveine (2000), Bruss and Louchard (2009), and Hsiau and Yang (2002) for other generalizations of the STOT.

2. The general model

In the STMOT, Bernoulli random variables $X_1, X_2, \dots, X_n$ are assumed to be independent for a finite n and the payoff for not stopping is assumed to be 0. In this section we attempt to extend the STMOT into the following directions. First, an infinite number of Bernoulli trials is allowed. Second, the payoff for not stopping is ω, which may be different from 0. Third, the Bernoulli random variables are allowed to be dependent. Fourth, at stage i, in addition to observing X_i, other dependent random variables are allowed to be observed that may influence the assessment of the probability of success at future stages. The method we use here is to modify the original problem into a monotone stopping problem by not allowing stopping on a failure, and then apply a simple result that gives conditions for the one-stage look-ahead rule (the 1-sla rule) to be optimal in a monotone problem (see, e.g. Ferguson (2006, Chapter 5) or Chow *et al.* (1971)). This method is exactly the same as that which Ferguson (2008, Section 2) used to extend the STOT, so we mimic his argument throughout this section.

Since stopping on a failure is forbidden, the notion of a 'stage' must be changed accordingly, that is, a stage is defined to contain all the observations up to and including the next success if any. Let $Z_i,\ i = 1, 2, \dots$, denote the set of random variables observed after success $i-1$ up to and including success i. If there are less than i successes, we let $Z_i = 0$, where '0' represents a special absorbing state. Thus, we treat the following general model. Let $Z_1, Z_2, \dots$ be a stochastic process on an arbitrary space with an absorbing state called 0. We make the assumption that, with probability 1, the process will eventually be absorbed at 0. We observe the process sequentially and must predict within m stages in advance when the state 0 will first be hit. If we predict correctly, we win 1, if we predict incorrectly, we win nothing, and if the process hits 0 before we predict, we win ω (it is assumed here that $\omega < 1$, because, if $\omega \ge 1$, it is clearly optimal never to stop). This is a stopping rule problem in which stopping at stage k yields the payoff

$$\begin{aligned} Y_k &= \omega\,\mathbf{1}(Z_k = 0) + \mathbf{1}(Z_k \neq 0)\,\mathrm{P}\{Z_{k+m} = 0 \mid \mathcal{G}_k\}, \qquad k = 1, 2, \dots, \\ Y_\infty &= \omega, \end{aligned} \tag{2.1}$$

where $\mathcal{G}_k = \sigma(Z_1, \dots, Z_k)$ is the σ-field generated by $Z_1, \dots, Z_k$ and $\mathbf{1}(E)$ represents the indicator function of an event E. The assignment $Y_\infty = \omega$ means that if we never stop, we

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

win ω. From (2.1) we evaluate

$$\mathrm{E}[Y_{k+1} \mid \mathcal{G}_k] = \omega\,\mathrm{P}\{Z_{k+1} = 0 \mid \mathcal{G}_k\} + \mathrm{P}\{Z_{k+1} \neq 0,\ Z_{k+m+1} = 0 \mid \mathcal{G}_k\}.$$

The 1-sla rule calls for stopping at stage k if $Y_k \geq \mathrm{E}[Y_{k+1} \mid \mathcal{G}_k]$. On the set $\{Z_k = 0\}$, this reduces to $\omega \geq \omega$, which is always true. On the set $\{Z_k \neq 0\}$, this reduces to

$$\mathrm{P}\{Z_{k+m} = 0 \mid \mathcal{G}_k\} \geq \omega\,\mathrm{P}\{Z_{k+1} = 0 \mid \mathcal{G}_k\} + \mathrm{P}\{Z_{k+1} \neq 0,\ Z_{k+m+1} = 0 \mid \mathcal{G}_k\}. \tag{2.2}$$

However, an absorbing state 0 implies that $\{Z_j = 0\} \subset \{Z_{j+1} = 0\}$ for each j, so we easily see that

$$\mathrm{P}\{Z_{k+m} = 0 \mid \mathcal{G}_k\} = \mathrm{P}\{Z_{k+1} = 0 \mid \mathcal{G}_k\} + \mathrm{P}\{Z_{k+1} \neq 0,\ Z_{k+m} = 0 \mid \mathcal{G}_k\}$$

and

$$\begin{aligned}\mathrm{P}\{Z_{k+1} \neq 0,\ Z_{k+m+1} = 0 \mid \mathcal{G}_k\} &= \mathrm{P}\{Z_{k+1} \neq 0,\ Z_{k+m} = 0 \mid \mathcal{G}_k\} \\ &\quad + \mathrm{P}\{Z_{k+m} \neq 0,\ Z_{k+m+1} = 0 \mid \mathcal{G}_k\}.\end{aligned}$$

Applying these two expressions to (2.2) yields

$$(1-\omega)\,\mathrm{P}\{Z_{k+1} = 0 \mid \mathcal{G}_k\} \geq \mathrm{P}\{Z_{k+m} \neq 0,\ Z_{k+m+1} = 0 \mid \mathcal{G}_k\}.$$

Therefore, the 1-sla rule is

$$N_m = \min\left\{k\colon Z_k = 0 \text{ or } \left(Z_k \neq 0 \text{ and } \frac{W_k}{V_k} \leq 1-\omega\right)\right\}, \tag{2.3}$$

where

$$V_k = \mathrm{P}\{Z_{k+1} = 0 \mid \mathcal{G}_k\}, \qquad W_k = \mathrm{P}\{Z_{k+m} \neq 0,\ Z_{k+m+1} = 0 \mid \mathcal{G}_k\}.$$

If $V_k = 0$ on $\{Z_k \neq 0\}$ then it is a mistake to stop at k since we can do at least as well by continuing one more step. Therefore, in this and subsequent formulae, we take the ratio in (2.3) to be $+\infty$ when $V_k = 0$, even if the numerator is 0 as well.

The problem is said to be *monotone* if when the 1-sla rule calls for stopping at any stage, then it will continue to call for stopping at all future stages no matter what the future observations turn out to be. More specifically, the problem is monotone if

$$A_1 \subset A_2 \subset A_3 \subset \cdots \quad \text{almost surely (a.s.)},$$

where $A_k = \{Y_k \geq \mathrm{E}[Y_{k+1} \mid \mathcal{G}_k]\}$ for all k. From this we see that a sufficient condition for the problem to be monotone is

$$W_k/V_k \text{ is a.s. nonincreasing in } k. \tag{2.4}$$

We have the following result.

Theorem 2.1. *Suppose that the process $Z_1, Z_2, \ldots$ has an absorbing state 0 such that*

$$\mathrm{P}\{Z_k \textit{ is absorbed at } 0\} = 1$$

and that the stopping problem with reward sequence (2.1) satisfies condition (2.4). Then the 1-sla rule (2.3) is optimal.

Proof. In the present context, it is obvious that $|Y_n|$ is bounded by $1 + |\omega|$ and

$$\lim_{n\to\infty} Y_n = Y_\infty = \omega.$$

These are sufficient conditions under which the 1-sla rule is optimal (see Theorem 1 of Ferguson (2008)).

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

2.1. Application to the STMOT

We return to the problem of stopping on any of the last m successes of a finite sequence of possibly dependent Bernoulli trials, $X_1, X_2, \dots, X_n$. We model the information given to the decision maker through an increasing sequence of σ-fields, $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$, and allow him/her to use a stopping rule adapted to this sequence. We assume that, for every j, the event $\{X_j = 1\}$ is in $\mathcal{F}_j$. In the general model formulation, if the kth success occurs at stage j then $\mathcal{G}_k = \mathcal{F}_j$. Assume that $\omega = 0$. Then the following corollary is immediate from Theorem 2.1.

Corollary 2.1. *Suppose that n Bernoulli random variables $X_1, X_2, \dots, X_n$ are observed sequentially. Let $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ be an increasing sequence of σ-fields such that $\{X_j = 1\}$ is in $\mathcal{F}_j$ for all $1 \le j \le n$. Let*

$$V_k = \mathrm{P}\{X_{k+1} + \cdots + X_n = 0 \mid \mathcal{F}_k\}, \tag{2.5}$$

$$W_k = \mathrm{P}\{X_{k+1} + \cdots + X_n = m \mid \mathcal{F}_k\}. \tag{2.6}$$

Then the optimal rule is described as

$$N_m = \min\left\{k \ge 1 \colon X_k = 1 \text{ and } \frac{W_k}{V_k} \le 1\right\}, \tag{2.7}$$

provided that the following condition is satisfied:

$$W_k / V_k \text{ is a.s. nonincreasing in } k. \tag{2.8}$$

This may be considered as a generalized STMOT in the sense that the ratio W_k/V_k in (2.7) can be written as, from (2.5) and (2.6),

$$\sum_{k+1 \le i_1 < i_2 < \cdots < i_m \le n} \left(\prod_{j=1}^{m} \frac{P_k(i_j \mid i_1, i_2, \dots, i_{j-1})}{1 - P_k(i_j \mid i_1, i_2, \dots, i_{j-1})} \right), \tag{2.9}$$

if we define

$$P_k(i_j \mid i_1, i_2, \dots, i_{j-1}) = \mathrm{P}\{X_{i_j} = 1 \mid \mathcal{F}_k, K(i_1, i_2, \dots, i_j)\},$$

where $K(i_1, i_2, \dots, i_j)$ represents, on the index set $S = \{k+1, k+2, \dots, n\}$, an information pattern described as

$$K(i_1, i_2, \dots, i_j) = \{X_i = 1 \text{ for } i \in \{i_1, i_2, \dots, i_{j-1}\},\ X_i = 0 \text{ for } i \in S - \{i_1, i_2, \dots, i_j\}\}.$$

It is easy to see that Corollary 2.1 implies the STMOT. In the STMOT, the X_j are independent and $\mathcal{F}_j = \sigma(X_1, \dots, X_j)$. So the conditioning in the definition of $P_k(i_j \mid i_1, i_2, \dots, i_{j-1})$ may be ignored, and $P_k(i_j \mid i_1, i_2, \dots, i_{j-1}) = p_{i_j}$. Then from (2.9) we have

$$\frac{W_k}{V_k} = \sum_{k+1 \le i_1 < i_2 < \cdots < i_m \le n} \left(\prod_{j=1}^{m} \frac{p_{i_j}}{q_{i_j}} \right), \tag{2.10}$$

so that in this case, $N_m = N_m^*$ of (1.4). We see from this that the W_k/V_k are nonrandom and nonincreasing in k because increasing k decreases the possible number of products on the

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

right-hand side of (2.10), so that (2.8) is satisfied. Thus, the problem is monotone and the 1-sla rule, N_m^*, is optimal. Since v_m is expressed as

$$v_m = \sum_{j=1}^{m} \mathrm{P}\left\{\sum_{i\in S^*} X_i = j\right\}, \tag{2.11}$$

where $S^* = \{s_m, s_m+1, \dots, n\}$ and the X_i are independent, we can write

$$\begin{aligned}\mathrm{P}\left\{\sum_{i\in S^*} X_i = j\right\} &= \sum_{s_m\le i_1<\cdots<i_j\le n}\left(\prod_{i\in\{i_1,\dots,i_j\}} \mathrm{P}\{X_i=1\} \prod_{i\in S^*-\{i_1,\dots,i_j\}} \mathrm{P}\{X_i=0\}\right)\\ &= \left(\prod_{i\in S^*}\mathrm{P}\{X_i=0\}\right)\sum_{s_m\le i_1<\cdots<i_j\le n}\left(\prod_{i\in\{i_1,\dots,i_j\}}\frac{\mathrm{P}\{X_i=1\}}{\mathrm{P}\{X_i=0\}}\right)\\ &= \left(\prod_{i=s_m}^{n} q_i\right)R_{s_m,j},\end{aligned}$$

which, combined with (2.11), proves (1.3) of the STMOT.

3. Application to the secretary problem

The secretary problem can be described as follows. A known number, n, of rankable applicants (1 being the best and n the worst) appear one at a time in a random order, with all $n!$ permutations equally likely. That is, each of the successive ranks of n applicants constitutes a random permutation. Suppose that all that can be observed are the relative ranks of the applicants as they appear. If Y_j denotes the relative rank of the jth applicant among the first j applicants, the sequentially observed random variables are $Y_1, Y_2, \dots, Y_n$. It is well known that

(a) $Y_1, Y_2, \dots, Y_n$ are independent random variables;

(b) $\mathrm{P}\{Y_j = i\} = 1/j,\ 1\le i\le j, 1\le j\le n$.

The jth applicant is called a *candidate* if he/she is relatively best, i.e. $Y_j = 1$. The problem we consider here is to stop on any of the last m successes, that is, any of the last m candidates (stopping is identified with selection of an applicant in the secretary problem). The independent random variables of Section 1 are therefore $X_1, X_2, \dots, X_n$, where $X_j = \mathbf{1}(Y_j = 1)$ from (a). Since $p_j = \mathrm{P}\{X_j = 1\} = 1/j$ and so $r_j = 1/(j-1)$ from (b), we immediately have the following result from the STMOT.

Lemma 3.1. *For the secretary problem, the optimal rule passes up the first $s_m - 1$ applicants and then selects the first candidate, if any, where*

$$s_m = \min\left\{k\ge 1: \sum_{k+1\le i_1<i_2<\cdots<i_m\le n}\prod_{j=1}^{m}\frac{1}{i_j-1}\le 1\right\}. \tag{3.1}$$

The maximal win probability is

$$v_m = \frac{s_m-1}{n}\sum_{k=1}^{m}\left(\sum_{s_m\le i_1<i_2<\cdots<i_k\le n}\prod_{j=1}^{k}\frac{1}{i_j-1}\right). \tag{3.2}$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

For small values of n, s_m and v_m can easily be computed from (3.1) and (3.2), respectively. For example, we have, for $m = 2$,

$$s_2 = 2 \quad \text{and} \quad v_2 = \tfrac{17}{24} \approx 0.708 \quad \text{for } n = 5,$$
$$s_2 = 3 \quad \text{and} \quad v_2 = \tfrac{1303}{2016} \approx 0.646 \quad \text{for } n = 10.$$

Of interest are the approximate values of s_m and v_m for large n.

Lemma 3.2. *Let n tend to* ∞. *Then, asymptotically,*

$$\text{(i)} \quad s_m^* = \lim_{n\to\infty} \frac{s_m}{n} = \exp\{-(m!)^{1/m}\},$$
$$\text{(ii)} \quad v_m^* = \lim_{n\to\infty} v_m = \exp\{-(m!)^{1/m}\} \sum_{j=1}^{m} \frac{(m!)^{j/m}}{j!}.$$

Proof. We give an intuitive proof. If we let n tend to ∞ and write s as the limit of k/n, then, using t_j for i_j/n and $\mathrm{d}t_j$ for $1/n$, $1 \le j \le m$, the sum in (3.1) becomes a Riemann approximation to a multiple integral,

$$\begin{aligned}
&\sum_{k+1\le i_1<i_2<\cdots<i_m\le n} \left(\frac{1}{i_1-1}\right)\left(\frac{1}{i_2-1}\right)\cdots\left(\frac{1}{i_m-1}\right) \\
&\quad \to \int_{s\le t_1<t_2<\cdots<t_m\le 1} \left(\frac{\mathrm{d}t_1}{t_1}\right)\left(\frac{\mathrm{d}t_2}{t_2}\right)\cdots\left(\frac{\mathrm{d}t_m}{t_m}\right) \\
&\quad = \int_s^1 \frac{\mathrm{d}t_1}{t_1} \int_{t_1}^1 \frac{\mathrm{d}t_2}{t_2} \cdots \int_{t_{m-1}}^1 \frac{\mathrm{d}t_m}{t_m} \\
&\quad = \frac{(-\log s)^m}{m!}.
\end{aligned}$$

Equating this to 1 yields $s = s_m^*$ and proves (i) from (3.1). In a similar way, we have, from (3.2),

$$v_m \to v_m^* = (s_m^*) \sum_{j=1}^{m} \int_{s_m^*}^1 \frac{\mathrm{d}t_1}{t_1} \int_{t_1}^1 \frac{\mathrm{d}t_2}{t_2} \cdots \int_{t_{j-1}}^1 \frac{\mathrm{d}t_j}{t_j} = (s_m^*) \sum_{j=1}^{m} \frac{(-\log s_m^*)^j}{j!},$$

which gives (ii) via (i).

Table 1 presents some numerical values of s_m^* and v_m^* for given m. The case in which $m = 1$ is the classical secretary problem studied in Lindley (1961), Dynkin (1963), and Gilbert and Mosteller (1966). Our problem seems similar to that of choosing one of the m best considered in Gusein-Zade (1966), Gilbert and Mosteller (1966), and Frank and Samuels (1980). However, these two problems are different because the kth last candidate is not necessarily the kth best

TABLE 1: Values of s_m^* and v_m^* for several m.

m	1	2	3	4	5	10
s_m^*	0.3679	0.2431	0.1625	0.1093	0.0739	0.0108
v_m^*	0.3679	0.5869	0.7260	0.8167	0.8767	0.9822

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

among all candidates for $k \geq 2$, and the only information available at each stage in our problem is record (candidate) information and not (relative) rank information. The reader is referred to Ferguson (1989) and Samuels (1991) for reviews of the secretary problem.

We can easily obtain further asymptotic quantities of interest by using a nonhomogeneous Poisson process (NPP) model. Presman and Sonin (1972, p. 772) found the NPP to be an appropriate setting in which to define the infinite version of the secretary problem as the limit of the corresponding finite problems. This model facilitates the derivation of the asymptotic values. Suppose that n applicants appear at fractional times $1/n, 2/n, \ldots, n/n$ instead of at times $1, 2, \ldots, n$. As $k/n \to x$ with $n \to \infty$, we can find from properties (a) and (b) that the occurrence of candidates in the limiting problem constitutes an NPP with intensity function $\lambda(x) = 1/x,\ 0 < x \leq 1$. That is, if we denote by $M(s,t)$ the number of candidates appearing in the time interval $(s,t),\ 0 < s < t \leq 1$, then

$$\mathrm{P}\{M(s,t) = k\} = \mathrm{e}^{-\Lambda(s,t)} \frac{\{\Lambda(s,t)\}^k}{k!}, \qquad k \geq 0, \tag{3.3}$$

where

$$\Lambda(s,t) = \int_s^t \lambda(x)\,\mathrm{d}x = \log\left(\frac{t}{s}\right).$$

This can be seen as a refinement of Theorem 1 of Gilbert and Mosteller (1966). Let $f(x)$ be the win probability when we use a (limiting) threshold rule with value x, which stops on the first candidate that appears after time x, if any. Then, from (3.3),

$$f(x) = \sum_{k=1}^{m} \mathrm{P}\{M(x,1) = k\} = \sum_{k=1}^{m} \frac{x(-\log x)^k}{k!}.$$

It is easy to see that $f(x)$ is maximized at $x = s_m^*$ and yields $f(s_m^*) = v_m^*$, as desired from Lemma 3.2. Note that the value s_m^* represents not only the optimal threshold value but also the probability of no choice (no stop), i.e. the probability that no candidate appears until the very end, which follows from $\mathrm{P}\{M(s_m^*, 1) = 0\} = s_m^*$. Since $M(s_m^*, 1)$ is a Poisson random variable with mean $\Lambda(s_m^*, 1) = -\log s_m^* = (m!)^{1/m}$, we have the following result.

Lemma 3.3. (i) *The expected number of candidates that appear after s_m^* is a geometric mean of m integers $1, 2, \ldots, m$, that is,*

$$\mathrm{E}[M(s_m^*, 1)] = (m!)^{1/m}.$$

(ii) *The probability of stopping on the kth last candidate conditional on the win being achieved is*

$$\frac{(m!)^{k/m}}{k!} \bigg/ \sum_{j=1}^{m} \frac{(m!)^{j/m}}{j!}, \qquad 1 \leq k \leq m.$$

(iii) *Let K_m be the expected rank of the applicant chosen conditional on stopping. Then*

$$K_m = \tfrac{1}{2}[1 + \exp\{(m!)^{1/m}\}].$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

Proof. Part (i) is immediate from $\mathrm{E}[M(s_m^*, 1)] = \Lambda(s_m^*, 1)$. Part (ii) follows because this conditional probability is just

$$\frac{\mathrm{P}\{M(s_m^*, 1) = k\}}{\sum_{j=1}^{m} \mathrm{P}\{M(s_m^*, 1) = j\}}.$$

To prove part (iii), let T_m be the stopping time of the optimal rule, i.e. the arrival time of the first candidate after s_m^*, if any. If no candidate appears, T_m is defined to be 1. Since $T_m > t$ is equivalent to $M(s_m^*, t) = 0$ for $t < 1$, the density of T_m is

$$f_{T_m}(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{P}\{T_m > t\} = -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{s_m^*}{t}\right) = \frac{s_m^*}{t^2}.$$

Obviously, $\mathrm{P}\{T_m = 1\} = s_m^*$. We only stop on a candidate, i.e. relatively best applicant, so the expected rank of the applicant chosen at time t is $1/t$ (see, e.g. Mucci (1973)). Hence,

$$K_m = \int_{s_m^*}^{1} \frac{1}{t} \frac{f_{T_m}(t)}{1 - \mathrm{P}\{T_m = 1\}}\,\mathrm{d}t,$$

which yields (iii).

For example, $K_1 = 1.859$, $K_2 = 2.557$, $K_3 = 3.577$, $K_4 = 5.073$, and $K_5 = 7.267$. These show that, for small m, the optimal rule chooses a pretty good applicant though it has probability s_m^* of no choice as a risk. As for the *rank minimization problem* whose objective is to minimize the expected rank of the applicant chosen with the severe condition that one applicant must be chosen, Chow *et al.* (1964) showed that the minimal asymptotic expected rank is approximately 3.869. See also Krieger and Samuel-Cahn (2009). We can give the arrival time distribution of the kth last candidate.

Lemma 3.4. *Let Z_k denote the arrival time of the kth last candidate. Then the density function of Z_k, $k \geq 1$, is*

$$f_k(z) = \frac{(-\log z)^{k-1}}{(k-1)!}, \qquad 0 < z < 1.$$

Proof. Observe that the event $Z_k \leq z$ occurs if and only if $M(z, 1) \leq k - 1$ for each k. Hence,

$$f_k(z) = \frac{\mathrm{d}}{\mathrm{d}z}\mathrm{P}\{Z_k \leq z\} = \frac{\mathrm{d}}{\mathrm{d}z}\left(\sum_{j=0}^{k-1} \mathrm{P}\{M(z, 1) = j\}\right) = \frac{(-\log z)^{k-1}}{(k-1)!},$$

which completes the proof.

For later use, we return to the finite problem and review the distribution of the number of candidates. Define

$$M_n = X_1 + X_2 + \cdots + X_n,$$

where $X_j = \mathbf{1}(Y_j = 1)$ is as defined at the beginning of this section. Then M_n denotes the total number of candidates. It is well known (see, e.g. Equation (2.5.9) of Arnold *et al.* (1998) or Sections 6.2, 6.3, and 9.5 of Blom *et al.* (1994)) that the probability mass function of M_n is expressed as

$$p_n(k) = \mathrm{P}\{M_n = k\} = \frac{1}{n!}\begin{bmatrix} n \\ k \end{bmatrix}, \qquad 1 \leq k \leq n,$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

where the notation $\left[{n \atop k}\right]$, $1 \le k \le n$, $1 \le n$, is a real number called the *Stirling number of the first kind* (see Knuth (1992) for this notation). This number can be simply calculated from the following recursive relation:

$$\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} + (n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix}, \qquad 1 \le k \le n,\ 2 \le n,$$

with $\left[{1 \atop 1}\right] = 1$ and $\left[{n \atop k}\right] = 0$ for $k = 0$ or $k > n$. Therefore, $p_n(k)$ satisfies the recursion

$$p_n(k) = \frac{1}{n} p_{n-1}(k-1) + \left(1 - \frac{1}{n}\right) p_{n-1}(k), \qquad 1 \le k \le n,\ 2 \le n, \tag{3.4}$$

with $p_1(1) = 1$ and $p_n(k) = 0$ for $k = 0$ or $k > n$.

4. Full-information analogue

In contrast to the no-information problem considered in Section 3, the full-information analogue is the problem in which the observations are the true values of n applicants $Y_1, Y_2, \dots, Y_n$, assumed to be independent and identically distributed random variables from a known continuous distribution, taken without loss of generality to be the uniform distribution on the interval $[0, 1]$. Let $L_k = \max\{Y_1, Y_2, \dots, Y_k\}$ be the maximum of the first k observations and call the kth observation or the kth applicant a *record* if $L_k = Y_k$. It is desired to obtain a stopping rule that maximizes the probability of stopping on any of the last m successes, that is, the last m records. The case in which $m = 1$ is the full-information best-choice problem solved in Gilbert and Mosteller (1966). For ease of description, let $a_k = \mathrm{P}\{M_k < m\}$, $k \ge 0$, be the probability that the number of candidates is less than m when the total number of applicants is k in the secretary problem, namely, $a_k = \sum_{i=1}^{m-1} p_k(i)$ for $k \ge m$ and $a_k = 1$ for $k < m$ ($a_0 = 1$ for convenience). The main results can be summarized as follows.

Theorem 4.1. (a) *Optimal stopping rule: for a given positive integer m, there exists a nondecreasing sequence of the thresholds $\{b_j(m),\ 1 \le j\}$ defined as $b_j(m) = 0$ for $1 \le j < m$ and as a unique solution $x \in (0, 1)$ to the equation*

$$\sum_{i=m}^{j} p_i(m) \binom{j}{i} \left(\frac{1-x}{x}\right)^i = 1 \tag{4.1}$$

for $j \ge m$, such that the optimal rule is to choose the first record $Y_k (= L_k)$ that exceeds the threshold $b_{n-k}(m)$. Henceforth, we simply write b_j for $b_j(m)$ unless otherwise specified.

(b) *Optimal probability: let $P^*_{n,m}$ denote the optimal win probability as a function of n and m; then*

$$P^*_{n,m} = \sum_{r=1}^{n} P(r),$$

where

$$P(1) = \frac{1}{n} \sum_{k=0}^{n-1} a_k \binom{n-1}{k} \sum_{j=k+1}^{n} \binom{n}{j} (1 - b_{n-1})^j b_{n-1}^{n-j} \tag{4.2}$$

and, for $2 \le r \le n$,

$$P(r) = \frac{1}{r-1} \sum_{k=0}^{n-r} a_k \binom{n-r}{k} \sum_{i=1}^{r-1} [P_1(i,k) + P_2(i,k)], \tag{4.3}$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

with

$$P_1(i,k) = \sum_{j=k+1}^{n} \frac{\binom{n}{j}}{\binom{n}{k}} \frac{(1-b_{n-r})^j b_{n-r}^{n-j} - (1-b_{n-i})^j b_{n-i}^{n-j}}{n-k}$$

and

$$P_2(i,k) = \sum_{j=k+1}^{n-r+1} \frac{\binom{n-r+1}{j}}{\binom{n-r+1}{k}} \frac{(1-b_{n-i})^j b_{n-i}^{n-j}}{n-r-k+1}.$$

(c) *Asymptotics: let c_m be the unique root t to the equation*

$$\sum_{i=m}^{\infty} p_i(m) \frac{t^i}{i!} = 1; \tag{4.4}$$

then, as $n \to \infty$,

$$P_{n,m}^* \to P_m^* = \mathrm{e}^{-c_m} J_m(c_m) + \{K_m(c_m) - c_m J_m(c_m)\} I(c_m), \tag{4.5}$$

where

$$I(t) = \int_1^{\infty} \frac{\mathrm{e}^{-tx}}{x}\,\mathrm{d}x,$$

$$J_m(t) = \sum_{j=0}^{\infty} a_j \frac{t^j}{j!} = \sum_{j=0}^{m-1} \frac{t^j}{j!} + \sum_{j=m}^{\infty} a_j \frac{t^j}{j!},$$

$$K_m(t) = \sum_{i=1}^{\infty} \min(i,m) \frac{t^i}{i!} + \sum_{i=m+1}^{\infty} \left(\sum_{j=m}^{i-1} a_j \right) \frac{t^i}{i!}.$$

Proof. (a) In Corollary 2.1, the Bernoulli random variables are $X_1, X_2, \ldots, X_n$, where $X_k = \mathbf{1}(Y_k = L_k)$, and the σ-field $\mathcal{F}_k$ is the σ-field generated by the variables $Y_1, Y_2, \ldots, Y_k$. We easily see that

$$V_k = (L_k)^{n-k}.$$

On the other hand, we may compute W_k by conditioning on the number, J, of future observations whose values are greater than L_k. Here J is a binomial random variable with parameters $(n-k, 1-L_k)$, i.e.

$$\mathrm{P}\{J = j\} = \binom{n-k}{j} (1-L_k)^j L_k^{n-k-j},$$

and, moreover, given $J = j$, the probability that the number of future records (candidates) is m is $p_j(m)$ by exchangeability. Thus,

$$W_k = \sum_{j=m}^{n-k} p_j(m) \binom{n-k}{j} (1-L_k)^j L_k^{n-k-j},$$

and, hence,

$$\frac{W_k}{V_k} = \sum_{j=m}^{n-k} p_j(m) \binom{n-k}{j} \left(\frac{1-L_k}{L_k} \right)^j \tag{4.6}$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

for $k \le n - m$ and 0 otherwise. Since the nondecreasing nature of L_k implies that

$$\frac{1 - L_k}{L_k} \ge \frac{1 - L_{k+1}}{L_{k+1}} \quad \text{a.s.},$$

we have, for $k \le n - m - 1$,

$$\frac{W_k}{V_k} - \frac{W_{k+1}}{V_{k+1}} \ge \sum_{j=m}^{n-k-1} p_j(m) \binom{n-k-1}{j-1} \left(\frac{1 - L_k}{L_k}\right)^j \ge 0,$$

which shows that condition (2.8) is satisfied. Hence, the 1-sla rule, N_m of (2.7), is optimal and written as, from (4.6) and the definition of the threshold, (4.1),

$$N_m = \min\{k \ge 1 \colon Y_k = L_k \ge b_{n-k}\}.$$

(b) Let $P(r)$ be the win probability when stopping on the rth observation. If we denote by W the win event then, conditioning on the rth observation being a record with value x, i.e. $L_r = Y_r = x$, we can express

$$P(r) = \int_{b_{n-r}}^{1} \mathrm{P}\{W \mid L_r = Y_r = x\} \frac{1}{r} f_r(x)\, \mathrm{d}x, \tag{4.7}$$

where $f_r(x) = rx^{r-1}$ is a density of L_r. We first consider the case in which $r \ge 2$. Note that, since the threshold b_j is nondecreasing, stopping on the rth observation leads to W if and only if

(i) the previous maximum is less than the corresponding threshold, i.e. $L_{r-1} < b_{n-i}$ if $L_{r-1} = Y_i$, $1 \le i \le r - 1$;

(ii) the number of future records is less than m.

Let A and B denote the events in (i) and (ii), respectively. Then it is easy to see that A and B are conditionally independent given $L_r = Y_r = x$, so

$$\mathrm{P}\{W \mid L_r = Y_r = x\} = \mathrm{P}\{A \mid L_r = Y_r = x\}\, \mathrm{P}\{B \mid L_r = Y_r = x\}.$$

Given $L_r = Y_r = x$, L_{r-1} can be viewed as the maximum of $r - 1$ independent random variables each uniformly distributed on $(0, x)$, and the arrival time of the previous maximum is uniform on $\{1, 2, \dots, r - 1\}$ independently of L_{r-1}. Hence,

$$\mathrm{P}\{A \mid L_r = Y_r = x\} = \frac{1}{r-1} \sum_{i=1}^{r-1} \left[\frac{\min(x, b_{n-i})}{x}\right]^{r-1}. \tag{4.8}$$

On the other hand, we have

$$\mathrm{P}\{B \mid L_r = Y_r = x\} = \sum_{k=0}^{n-r} a_k \binom{n-r}{k} (1 - x)^k x^{n-r-k}, \tag{4.9}$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

using a similar reasoning to that used to derive W_k in (a). Therefore, we have, from (4.7)–(4.9),

$$P(r) = \frac{1}{r-1}\sum_{k=0}^{n-r} a_k \binom{n-r}{k} \times \sum_{i=1}^{r-1}\left\{\int_{b_{n-r}}^{b_{n-i}} x^{n-k-1}(1-x)^k\,\mathrm{d}x + b_{n-i}^{r-1}\int_{b_{n-i}}^{1} x^{n-k-r}(1-x)^k\,\mathrm{d}x\right\},$$

which gives (4.3) via the well-known formula

$$\int_c^1 x^{m-1}(1-x)^{n-1}\,\mathrm{d}x = \sum_{j=n}^{m+n-1} \frac{\binom{m+n-1}{j}}{\binom{m+n-1}{n-1}} \frac{(1-c)^j c^{m+n-1-j}}{m}$$

for $0 < c \le 1$ and positive integers m and n. The case in which $r = 1$ is special. Stopping on the first observation leads to W if and only if the number of future records is less than m. Hence, substituting

$$\mathrm{P}\{W \mid L_1 = Y_1 = x\} = \sum_{k=0}^{n-1} a_k \binom{n-1}{k}(1-x)^k x^{n-1-k}$$

into (4.7) yields (4.2).

(c) To derive the asymptotic result (4.5), we can use a planar Poisson process (PPP) model which is known to facilitate the derivation of the asymptotic values for some full-information problems (see, e.g. Gnedin (1996), (2004), Samuels (2004), Bruss and Swan (2009), and Tamaki (2009)). According to Samuels (2004, Sections 9 and 10), we use a Poisson process with unit rate on the semi-infinite strip $[0, 1] \times [0, \infty)$. This turns the problem upside down, making the 'best' become the 'smallest'. The process is scanned from left to right by shifting a vertical detector and the scanning can be stopped each time a point in the PPP, referred to as an *atom* henceforth, is detected. A link to the finite problems can be established by suitably embedding the finite independent and identically distributed sequences in the PPP in a similar manner as given for the Gilbert and Mosteller full-information best-choice problem in Gnedin (1996, Section 3).

Suppose that an atom is identified as a point (t, y) if the atom appears at time t as a record (the relatively best atom as in the finite problem) having value y in the PPP. Let $Q_k(t, y)$ denote the probability that k records appear in the future after skipping the point (t, y). Then if we denote by $\mathrm{P}_{\mathrm{Poisson}}(j, \mu)$ the Poisson probability of j events for a given mean μ, i.e. $\mathrm{P}_{\mathrm{Poisson}}(j, \mu) = \mathrm{e}^{-\mu}(\mu)^j/j!$, we have

$$Q_k(t, y) = \sum_{j=k}^{\infty} p_j(k)\,\mathrm{P}_{\mathrm{Poisson}}(j, y(1-t)), \qquad k \ge 0, \tag{4.10}$$

because $\mathrm{P}_{\mathrm{Poisson}}(j, y(1-t))$ is the probability that there exist j atoms in the box domain $[t, 1] \times [0, y]$ whose area is $y(1-t)$ and, conditional on there being j atoms, the probability that k of them are records is $p_j(k)$ by exchangeability.

Let, in particular,

$$V(t, y) = Q_0(t, y), \qquad W(t, y) = Q_m(t, y).$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

Solving for the locus of point (t, y) at which the win probability, if we stop immediately, is equal to the win probability when stopping on the next record, if any, or, equivalently,

$$\sum_{k=0}^{m-1} Q_k(t, y) = \sum_{k=1}^{m} Q_k(t, y),$$

yields $W(t, y)/V(t, y) = 1$, which implies that $y(1-t) = c_m$, where c_m, is defined by (4.4) (for simplicity, we write c for c_m throughout the proof). Since $W(t, y)/V(t, y) \le 1$ implies that $W(t', y')/V(t', y') \le 1$ for $t' > t$, $y' < y$, we are in the monotone case of optimal stopping (continuous analogue of Corollary 2.1) and we can conclude that the optimal rule stops with the first record, if any, that lies below the threshold curve $y = c/(1-t)$. Let T be the arrival time of the first (leftmost) atom that lies below the threshold curve $y = c/(1-t)$, and let S be the time when the value of the best (lowest) atom above the threshold is equal to the threshold. Then T and S are independent and their distributions are given by

$$\mathrm{P}\{T > t\} = \mathrm{P}_{\mathrm{Poisson}}\left(0, \int_0^t g(r)\,\mathrm{d}r\right), \tag{4.11}$$

$$\mathrm{P}\{S > s\} = \mathrm{P}_{\mathrm{Poisson}}\left(0, \int_0^s (g(s) - g(r))\,\mathrm{d}r\right), \tag{4.12}$$

where $g(r) = c/(1-r)$, $0 < r < 1$.

Considering that the win probability when we stop at point (t, y) is

$$u(t, y) = \sum_{k=0}^{m-1} Q_k(t, y), \tag{4.13}$$

we have, by exploiting the virtual stopping time $\min(S, T)$ which makes the calculations simpler,

$$\begin{aligned} P_m^* = &\int_0^1 \int_0^t u\left(s, \frac{c}{1-s}\right) f_S(s) f_T(t)\,\mathrm{d}s\,\mathrm{d}t \\ &+ \int_0^1 \int_0^s \left(\frac{1-t}{c} \int_0^{c/(1-t)} u(t, y)\,\mathrm{d}y\right) f_T(t) f_S(s)\,\mathrm{d}t\,\mathrm{d}s, \end{aligned} \tag{4.14}$$

where $f_T(t)$ and $f_S(s)$ are the densities of T and S, respectively.

We have, from (4.13) and (4.10),

$$\begin{aligned} u\left(s, \frac{c}{1-s}\right) &= \sum_{k=0}^{m-1} \sum_{j=k}^{\infty} p_j(k)\,\mathrm{P}_{\mathrm{Poisson}}(j, c) \\ &= \sum_{j=0}^{m-1} \left(\sum_{k=0}^{j} p_j(k)\right) \mathrm{P}_{\mathrm{Poisson}}(j, c) \\ &\quad + \sum_{j=m}^{\infty} \left(\sum_{k=0}^{m-1} p_j(k)\right) \mathrm{P}_{\mathrm{Poisson}}(j, c) \\ &= \sum_{j=0}^{m-1} \mathrm{P}_{\mathrm{Poisson}}(j, c) + \sum_{j=m}^{\infty} a_j\,\mathrm{P}_{\mathrm{Poisson}}(j, c) \\ &= \mathrm{e}^{-c} J_m(c). \end{aligned} \tag{4.15}$$

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

Similarly,

$$\begin{aligned}\int_0^{c/(1-t)} u(t,y)\,\mathrm{d}y &= \sum_{k=0}^{m-1}\sum_{j=k}^{\infty} p_j(k)\int_0^{c/(1-t)} \mathrm{e}^{-y(1-t)}\frac{[y(1-t)]^j}{j!}\,\mathrm{d}y\\ &= (1-t)^{-1}\sum_{k=0}^{m-1}\sum_{j=k}^{\infty} p_j(k)\int_0^{c} \mathrm{e}^{-x}\frac{x^j}{j!}\,\mathrm{d}x\\ &= (1-t)^{-1}\sum_{k=0}^{m-1}\sum_{j=k}^{\infty} p_j(k)\biggl(\sum_{i=j+1}^{\infty} \mathrm{P}_{\mathrm{Poisson}}(i,c)\biggr), \end{aligned} \tag{4.16}$$

where the last equality follows from the well-known identity which states that, for any $c > 0$, the event 'there are more than j arrivals by time c' is the same as the event 'the waiting time for the $(j+1)$th arrival is less than c'. However, interchanging the order of summation successively, we have

$$\begin{aligned}&\sum_{k=0}^{m-1}\sum_{j=k}^{\infty} p_j(k)\biggl(\sum_{i=j+1}^{\infty} \mathrm{P}_{\mathrm{Poisson}}(i,c)\biggr)\\ &\quad= \sum_{k=0}^{m-1}\sum_{i=k+1}^{\infty}\biggl(\sum_{j=k}^{i-1} p_j(k)\biggr)\mathrm{P}_{\mathrm{Poisson}}(i,c)\\ &\quad= \sum_{i=1}^{m}\biggl[\sum_{k=0}^{i-1}\biggl(\sum_{j=k}^{i-1} p_j(k)\biggr)\biggr]\mathrm{P}_{\mathrm{Poisson}}(i,c) + \sum_{i=m+1}^{\infty}\biggl[\sum_{k=0}^{m-1}\biggl(\sum_{j=k}^{i-1} p_j(k)\biggr)\biggr]\mathrm{P}_{\mathrm{Poisson}}(i,c)\\ &\quad= \sum_{i=1}^{m}\biggl[\sum_{j=0}^{i-1}\biggl(\sum_{k=0}^{j} p_j(k)\biggr)\biggr]\mathrm{P}_{\mathrm{Poisson}}(i,c)\\ &\qquad+ \sum_{i=m+1}^{\infty}\biggl[\sum_{j=0}^{m-1}\biggl(\sum_{k=0}^{j} p_j(k)\biggr) + \sum_{j=m}^{i-1}\biggl(\sum_{k=0}^{m-1} p_j(k)\biggr)\biggr]\mathrm{P}_{\mathrm{Poisson}}(i,c)\\ &\quad= \sum_{i=1}^{m} i\,\mathrm{P}_{\mathrm{Poisson}}(i,c) + \sum_{i=m+1}^{\infty}\biggl(m + \sum_{j=m}^{i-1} a_j\biggr)\mathrm{P}_{\mathrm{Poisson}}(i,c)\\ &\quad= \mathrm{e}^{-c}K_m(c). \end{aligned} \tag{4.17}$$

Substituting the three expressions (4.15)–(4.17) into (4.14), we obtain

$$\begin{aligned}P_m^* &= \mathrm{e}^{-c}J_m(c)\,\mathrm{P}\{S \le T\} + \mathrm{e}^{-c}c^{-1}K_m(c)\,\mathrm{P}\{S > T\}\\ &= \mathrm{e}^{-c}J_m(c) + (K_m(c) - cJ_m(c))\mathrm{e}^{-c}c^{-1}\,\mathrm{P}\{S > T\}. \end{aligned} \tag{4.18}$$

Note that we have, from (4.11) and (4.12),

$$\begin{aligned}\mathrm{P}\{S > T\} &= \int_0^1 \mathrm{P}\{S > t\} f_T(t)\,\mathrm{d}t\\ &= \int_0^1 (1-t)^{-c}\mathrm{e}^{-ct/(1-t)}c(1-t)^{c-1}\,\mathrm{d}t\\ &= c\mathrm{e}^{c}I(c), \end{aligned}$$

which, combined with (4.18), gives the desired result (4.5). Thus, the proof is complete.

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

TABLE 2: Values of c_m and P_m^* for several m.

m	1	2	3	4	5	10
c_m	0.8044	1.5151	2.3731	3.3573	4.4523	11.2433
P_m^*	0.5802	0.8424	0.9465	0.9834	0.9953	0.9999

Table 2 presents some numerical values of c_m and P_m^*.

Remark 4.1. Since $p_n(n) = 1/n!$, (4.1) for $j = m$ can be written as

$$\frac{1}{m!}\left(\frac{1-x}{x}\right)^m = 1,$$

which yields

$$b_m(m) = \frac{1}{1 + (m!)^{1/m}}.$$

Similarly, considering that

$$p_n(n-1) = \frac{1}{2 \cdot (n-2)!}, \qquad p_n(n-2) = \frac{3n-1}{24 \cdot (n-3)!},$$

which can be obtained from the recursive formula (3.4), we have, as another expression of (4.1) for $j = m+1$ and $j = m+2$,

$$\frac{m+1}{m!}\left(\frac{1-x}{x}\right)^m + \frac{1}{2 \cdot (m-1)!}\left(\frac{1-x}{x}\right)^{m+1} = 1 \tag{4.19}$$

and

$$\frac{(m+1)(m+2)}{2 \cdot m!}\left(\frac{1-x}{x}\right)^m + \frac{m+2}{2 \cdot (m-1)!}\left(\frac{1-x}{x}\right)^{m+1} + \frac{3m+5}{24 \cdot (m-1)!}\left(\frac{1-x}{x}\right)^{m+2} = 1,$$

respectively. In particular, from (4.19), $b_2(1) = (1+\sqrt{6})/5$ and $b_3(2) = \sqrt{3}/3$.

Acknowledgement

The author is grateful to an anonymous referee for his/her careful reading.

References

ARNOLD, B. C., BARLAKRISHNAN, N. AND NAGARAJA, H. N. (1998). *Records*. John Wiley, New York.

BLOM, G., HOLST, L. AND SANDEL, D. (1994). *Problems and Snapshots from the World of Probability*. Springer, New York.

BRUSS, F. T. (2000). Sum the odds to one and stop. *Ann. Prob.* **28,** 1384–1391.

BRUSS, F. T. (2003). A note on bounds for the odds theorem of optimal stopping. *Ann. Prob.* **31,** 1859–1861.

BRUSS, F. T. AND LOUCHARD, G. (2009). The odds algorithm based on sequential updating and its performance. *Adv. Appl. Prob.* **41,** 131–153.

BRUSS, F. T. AND PAINDAVEINE, D. (2000). Selecting a sequence of last successes in independent trials. *J. Appl. Prob.* **37,** 389–399.

BRUSS, F. T. AND SWAN, Y. C. (2009). A continuous-time approach to Robbins' problem of minimizing the expected rank. *J. Appl. Prob.* **46,** 1–18.

CHOW, Y. S., ROBBINS, H. AND SIEGMUND, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston, MA.

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press

Chow, Y. S., Moriguti, S., Robbins, H. and Samuels, S. M. (1964). Optimal selection based on relative rank (the 'secretary problem'). *Israel J. Math.* **2,** 81–90.

Dynkin, E. B. (1963). The optimal choice of the instant for stopping a Markov process. *Soviet Math. Dokl.* **4,** 627–629.

Ferguson, T. S. (1989). Who solved the secretary problem? *Statist. Sci.* **4,** 282–289.

Ferguson, T. S. (2006). *Optimal Stopping and Applications*. Available at http://www.math.ucla.edu/∼tom/Stopping/Contents.html.

Ferguson, T. S. (2008). The sum-the-odds theorem with application to a stopping game of Sakaguchi. Preprint.

Frank, A. Q. and Samuels, S. M. (1980). On an optimal stopping problem of Gusein-Zade. *Stoch. Process. Appl.* **10,** 299–311.

Gilbert, J. P. and Mosteller, F. (1966). Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.* **61,** 35–73.

Gnedin, A. V. (1996). On the full information best-choice problem. *J. Appl. Prob.* **33,** 678–687.

Gnedin, A. V. (2004). Best choice from the planar Poisson process. *Stoch. Process. Appl.* **111,** 317–354.

Gusein-Zade, S. M. (1966). The problem of choice and the optimal stopping rule for a sequence of independent trials. *Theory Prob. Appl.* **11,** 472–476.

Hill, T. P. and Krengel, U. (1992). A prophet inequality related to the secretary problem. In *Strategies for Sequential Search and Selection in Real Time* (Amherst, MA, 1990; Contemp. Math. **125**), American Mathematical Society, Providence, RI, pp. 209–215.

Hsiau, S.-R. and Yang, J.-R. (2000). A natural variation of the standard secretary problem. *Statistica Sinica* **10,** 639–646.

Hsiau, S.-R. and Yang, J.-R. (2002). Selecting the last success in Markov-dependent trials. *J. Appl. Prob.* **39,** 271–281.

Knuth, D. E. (1992). Two notes on notation. *Amer. Math. Monthly* **99,** 403–422.

Krieger, A. M. and Samuel-Cahn, H. (2009). The secretary problem of minimizing the expected rank: a simple suboptimal approach with generalizations. *Adv. Appl. Prob.* **41,** 1041–1058.

Lindley, D. V. (1961). Dynamic programming and decision theory. *Appl. Statist.* **10,** 39–51.

Mucci, A. G. (1973). On a class of secretary problems. *Ann. Prob.* **1,** 417–427.

Presman, È. L. and Sonin, I. M. (1972). The best choice problem for a random number of objects. *Theory Prob. Appl.* **17,** 657–668.

Samuels, S. M. (1991). Secretary problem. In *Handbook of Sequential Analysis*, eds B. K. Ghosh and P. K. Sen, Dekker, New York, pp. 381–405.

Samuels, S. M. (2004). Why do these quite different best-choice problems have the same solutions? *Adv. Appl. Prob.* **36,** 398–416.

Tamaki, M. (2009). Optimal choice of the best available applicant in full-information models. *J. Appl. Prob.* **46,** 1086–1099.

https://doi.org/10.1239/jap/1285335408 Published online by Cambridge University Press