Adv. Appl. Prob. **23**, 823–834 (1991)
Printed in N. Ireland

SOME REMARKS ON NEVZOROV'S RECORD MODEL

DIETMAR PFEIFER,* *University of Oldenburg*

Abstract

We give a simple proof for the independence of record indices in Nevzorov's record model which is based on ranks. An application of these results to a probabilistic analysis of a particular searching algorithm with non-equiprobable orderings is also discussed.

RECORD INDEX; RANKS; POISSON BINOMIAL DISTRIBUTION; NON-EQUIPROBABLE ORDERINGS; SEARCHING ALGORITHM

1. Introduction

For a sequence of random variables $\{X_n\}$ let the record indices $\{I_n\}$ be defined by

$$I_n = \begin{cases} 1, & \text{if } X_n > \max(X_1, \cdots, X_{n-1}), \\ 0, & \text{otherwise}, \end{cases} \qquad n \geqq 2. \tag{1.1}$$

By convention, $I_1 = 1$. If $I_n = 1$ for some n, we say that X_n is a *record value* of the sequence. In case that the sequence $\{X_n\}$ is i.i.d. with a continuous c.d.f. Rényi (1962) has shown that the $\{I_n\}$ are independent random variables with success probabilities

$$P(I_n = 1) = \frac{1}{n}, \qquad n \in \mathbb{N}. \tag{1.2}$$

For alternative proofs and applications of this result, see for example the recent monographs by Resnick (1987) or Pfeifer (1989a). A more general setup in which the $\{X_n\}$ are independent but not identically distributed has been introduced by Nevzorov (1986); cf. also his recent survey article (1988). The assumptions here are that the c.d.f. F_n of X_n is of the specific form

$$F_n(x) = P(X_n \leqq x) = F^{\alpha_n}(x), \qquad x \in \mathbb{R} \tag{1.3}$$

with parameters $\alpha_n > 0$ and a fixed continuous c.d.f. F. His basic result states that under (1.3), the $\{I_n\}$ are still independent, but this time with success probabilities

$$P(I_n = 1) = \frac{\alpha_n}{\sum_{i=1}^{n} \alpha_i} \tag{1.4}$$

Received 25 September 1989; revision received 5 September 1990.
* Postal address: Fachbereich 6, Mathematik, Universität Oldenburg, Postfach 25 03, D-W2900 Oldenburg, Germany.

which covers Rényi's result for $\alpha_n \equiv \text{const}$. For an application of this model to the so-called *secretary problem*, see Pfeifer (1989b). In the latter paper, Nevzorov's result was proved using embeddings into suitable non-homogeneous extremal processes. Here we shall give a more simple proof of this fundamental result using the ranks of the underlying random variables which at the same time allows for extensions to even more general situations. As an application, we study the probabilistic behaviour of a particular searching algorithm in computer science in a more general setting of non-equiprobable orderings of the elements.

2. Main results

For (not necessarily i.i.d.) random variables $X_1, \cdots, X_n$ let $\Sigma_n = \Sigma_n(X_1, \cdots, X_n)$ denote the order of $X_1, \cdots, X_n$ with respect to increasing ranks, i.e. $\Sigma_n = \sigma \Leftrightarrow X_{\sigma(1)} \leqq \cdots \leqq X_{\sigma(n)}$. To avoid ambiguities in case of ties, i.e. $X_i = X_j$ for some $i < j$, choose σ such that $\sigma^{-1}(i) > \sigma^{-1}(j)$. Then

$$\{\Sigma_n = \sigma\} = \{X_{\sigma(n)} \geqq X_{\sigma(n-1)} \geqq \cdots \geqq X_{\sigma(1)}\}.$$

The following lemma is a key result.

Lemma 1. *Let* $X_1, \cdots, X_{n+1}$ *be arbitrary random variables and* I_{n+1} *be defined as in* (1.1). *Then*

$$\begin{aligned} P(I_{n+1} = 1 \mid \Sigma_n = \sigma) &= \frac{P(X_{n+1} > \max\{X_1, \cdots, X_n\}, \Sigma_n = \sigma)}{P(\Sigma_n = \sigma)} \\ &= \frac{P(X_{n+1} > X_{\sigma(n)} \geqq \cdots \geqq X_{\sigma(1)})}{P(X_{\sigma(n)} \geqq \cdots \geqq X_{\sigma(1)})} \end{aligned} \tag{2.1}$$

whenever $P(\Sigma_n = \sigma) > 0$.

Proof. Straightforward.

In view of relation (2.1) it is clear that I_{n+1} and Σ_n are independent if the right-hand side of (2.1) is independent of σ's with $P(\Sigma_n = \sigma) > 0$ which is the case, for instance, for $X_1, \cdots, X_{n+1}$ having *exchangeable* identically distributed components with zero probability of ties. In this case,

$$P(X_{\sigma(n)} > \cdots > X_{\sigma(1)}) = \frac{1}{n!}, \qquad P(X_{n+1} > X_{\sigma(n)} > \cdots > X_{\sigma(1)}) = \frac{1}{(n+1)!}$$

and hence

$$P(I_{n+1} = 1 \mid \Sigma_n = \sigma) = \frac{\dfrac{1}{(n+1)!}}{\dfrac{1}{n!}} = \frac{1}{n+1},$$

independent of σ which also means that I_{n+1} and $(I_1, \cdots, I_n)$ are independent (cf. Rényi (1962), p. 109, and Pfeifer (1987) where mixtures of i.i.d. random variables were considered).

To give an example, consider random sampling of n elements out of a set $\{x_1, \cdots, x_m\}$ of $m \geqq n$ distinct, ordered elements without replacement. If X_k denotes the result of the kth drawing, then $(X_1, \cdots, X_n)$ has exchangeable (but not independent) components with the required properties, each X_k being uniformly distributed over $\{x_1, \cdots, x_m\}$. Hence $I_1, \cdots, I_n$ are independent with success probabilities given by (1.2).

For $m = n$ the random vector $(X_1, \cdots, X_n)$ is just a *random permutation* of $\{x_1, \cdots, x_n\}$, each ordering of the n elements being equiprobable. Here Σ_n^{-1} denotes the order in which the elements $x_1, \cdots, x_n$ are successively drawn.

We shall now show how Nevzorov's result follows readily from Lemma 1.

Theorem 1 (Nevzorov). *Suppose that* $X_1, \cdots, X_n$, $n \in \mathbb{N}$, *are independent random variables with* (*continuous*) *c.d.f.'s* $F_1, \cdots, F_n$ *given by* (1.3). *Then the record indices* $I_1, \cdots, I_n$ *are independent with success probabilities given by* (1.4).

Proof. Due to Lemma 1 it suffices to show that the right-hand side of (2.1) is independent of any permutation σ of $\{1, \cdots, n\}$, for all $n \in \mathbb{N}$. By a standard argument in extreme value theory, we may assume that F is an exponential d.f., hence by Result 2 in Dansie (1983),

$$P(X_{\sigma(n)} > \cdots > X_{\sigma(1)}) = \prod_{i=1}^{n} \frac{\alpha_{\sigma(i)}}{\sum_{j=1}^{i} \alpha_{\sigma(j)}}$$

and likewise

$$P(X_{n+1} > X_{\sigma(n)} > \cdots > X_{\sigma(1)}) = \prod_{i=1}^{n} \frac{\alpha_{\sigma(i)}}{\sum_{j=1}^{i} \alpha_{\sigma(j)}} \cdot \frac{\alpha_{n+1}}{\sum_{j=1}^{n} \alpha_{\sigma(j)} + \alpha_{n+1}}.$$

The right-hand side of (2.1) thus gives

$$P(I_{n+1} = 1 \mid \Sigma_n = \sigma) = \frac{\alpha_{n+1}}{\sum_{j=1}^{n} \alpha_{\sigma(j)} + \alpha_{n+1}} = \frac{\alpha_{n+1}}{\sum_{j=1}^{n+1} \alpha_j}$$

which is independent of σ and hence proves the theorem.

Similar to the case of random sampling without replacement, Nevzorov's result extends to the case that $(X_1, \cdots, X_n)$ is a random permutation of n distinct, ordered elements $\{x_1, \cdots, x_n\}$, with

$$
\begin{aligned}
P(\Sigma_n = \sigma) &= P(\Sigma_n^{-1} = \sigma^{-1}) = P(X_{\sigma(n)} > \cdots > X_{\sigma(1)}) \\
&= P(X_1 = x_{\sigma^{-1}(1)}, \cdots, X_n = x_{\sigma^{-1}(n)}) = \prod_{i=1}^{n} \frac{\alpha_{\sigma(i)}}{\sum_{j=1}^{i} \alpha_{\sigma(j)}},
\end{aligned} \tag{2.2}
$$

Σ_n^{-1} denoting again the order of drawing of the individual elements which corresponds to Plackett's (1975) first-order model.

Note that in the (simple) model of sampling without replacement, the distribution of the maximum element, N, say, is uniform over the set $\{x_1, \cdots, x_n\}$ since

$$
\begin{aligned}
P(N = x_k) &= P(I_k = 1, I_{k+1} = \cdots = I_n = 0) \\
&= \frac{1}{k} \prod_{j=k+1}^{n} \left(1 - \frac{1}{j}\right) = \frac{1}{n}
\end{aligned} \tag{2.3}
$$

for $1 \leqq k \leqq n$; likewise for $k = n$.

In what follows we want to show that *any* distribution of N can be realized by a (uniquely determined) Nevzorov model (i.e., with *independent* record indices $I_1, \cdots, I_n$) with property $\sum_{i=1}^{n} \alpha_i = 1$. In view of relations (1.4) and (2.2), this is only a norming condition and will be assumed to hold henceforth.

Lemma 3. *Let* Σ_n^{-1}, *the distribution of the possible orderings of the elements* $x_1, \cdots, x_n$, *be chosen according to* (2.2). *Then*

$$
P(N = x_k) = \alpha_k, \qquad 1 \leqq k \leqq n.
$$

Proof. According to (1.4) we have, similarly to (2.3),

$$
\begin{aligned}
P(N = x_k) &= P(I_k = 1, I_{k+1} = \cdots = I_n = 0) \\
&= \frac{\alpha_k}{\sum_{i=1}^{k} \alpha_i} \prod_{j=k+1}^{n} \left(1 - \frac{\alpha_j}{\sum_{i=1}^{j} \alpha_i}\right) \\
&= \frac{\alpha_k}{\sum_{i=1}^{k} \alpha_i} \prod_{j=k+1}^{n} \left(\frac{\sum_{i=1}^{j-1} \alpha_i}{\sum_{i=1}^{j} \alpha_i}\right) = \frac{\alpha_k}{\sum_{i=1}^{n} \alpha_i} = \alpha_k
\end{aligned}
$$

for $1 \leqq k < n$; similarly for $k = n$ (cf. also Plackett (1975)).

The following result shows that a Nevzorov model in general describes, in some sense, independent sampling from a non-uniform distribution.

Theorem 2. *Let the random variable* Z *have a distribution given by*

$$
P(Z = k) = \alpha_k, \qquad 1 \leqq k \leqq n.
$$

Construct a realization of a random variable recursively as follows:

1. *Choose* $\sigma(n) \in \{1, 2, \cdots, n\}$ *according to the distribution* P^Z.
2. *Suppose* $\sigma(n), \sigma(n-1), \cdots, \sigma(n-k+1), 1 \leqq k < n$ *have been selected. Choose* $\sigma(n-k) \in \{1, 2, \cdots, n\}$ *according to the conditional distribution* $P^z(\cdot \mid Z \notin \{\sigma(n), \sigma(n-1), \cdots, \sigma(n-k+1)\})$, *independent of the choices before.*

Then σ *is a realization of* Σ_n.

Proof. Follows from Plackett (1968), p. 533, Harville (1973) or Hartigan (1968).

The construction of Theorem 2 is equivalent to independent sampling (in reverse order) *without* replacement from $\{1, 2, \cdots, n\}$ according to the distribution P^Z; therefore a realization σ of Σ_n could also be obtained by independent sampling *with* replacement until all numbers $1, \cdots, n$ have been drawn, omitting multiple occurrences.

3. An application to a searching algorithm

A fundamental algorithm in computer science is the linear search for the maximum element in a field of $n \in \mathbb{N}$ ordered elements $x_1, \cdots, x_n$ by comparisons (cf. Knuth (1973), Section 1.2.10 and Kemp (1984), Section 3.1). The quantity of specific interest here is the number of storages of the current maximum until all elements of the field have been examined. In our terminology, this is precisely the number $S_n = \sum_{i=2}^n I_i$ of positive *record indices* (except I_1) in the sequence $x_1, \cdots, x_n$ (note that in the above references the storage of the first element x_1 is not counted). If one assumes that the $n!$ possible orderings of the elements follow a specific distribution then S_n is a random variable with a distribution induced by the ordering distribution. Usually the probabilistic analysis of the algorithm is based on the assumption of equiprobable orderings, i.e. the probability of $1/n!$ for each particular ordering (see the references above). Since the approaches there are combinatorial in nature using generating functions and recursions for the relevant probabilities, the main interest is concerned with the calculation of $P(S_n = k)$, $0 \leqq k \leqq n-1$ (involving Stirling numbers of the first kind; cf. also Rényi (1962)) or $E(S_n)$ and $\operatorname{Var}(S_n)$. However, somewhat more can be said here due to the fact that in the case of equiprobable orderings, we have precisely the (exchangeable) case of sampling *without* replacement discussed in Section 2, hence $I_1, \cdots, I_n$ are *independent* random variables with success probabilities given by (1.2) and thus

$$(3.1)\qquad \begin{aligned} E(S_n) &= \sum_{i=2}^n \frac{1}{i} = \log n + C - 1 + O\left(\frac{1}{n}\right) \\ \operatorname{Var}(S_n) &= \sum_{i=2}^n \frac{1}{i}\left(1 - \frac{1}{i}\right) = \log n + C - \frac{\pi^2}{6} + O\left(\frac{1}{n}\right) \end{aligned} \qquad (n \to \infty)$$

where $C = 0{\cdot}577216$ denotes *Euler's constant* (cf. also Kemp (1984), p. 25). Similarly, the distribution of S_n (which is *Poisson binomial*) can be obtained.

Moreover, as a sum of independent $\{0, 1\}$-valued random variable the distribution of S_n can be approximated by a Poisson distribution for large n; namely, if T_n denotes a Poisson random variable with mean $E(S_n) = \sum_{i=2}^n 1/i$ (or, likewise, $\log n + C - 1$), we have

$$\sup_{x \in \mathbb{R}} |P(S_n \leqq x) - P(T_n \leqq x)| = \frac{\frac{1}{6}\pi^2 - 1}{2\sqrt{2\pi e}} \frac{1}{\log n + C - 1} + O(\log^{-\frac{3}{2}} n)$$

$$= \frac{0{\cdot}078}{\log n + C - 1} + O(\log^{-\frac{3}{2}} n) \qquad (n \to \infty)$$

whereas a corresponding normal approximation would yield an approximation error of $O(\log^{-\frac{1}{2}} n)$ which is worse (see Deheuvels and Pfeifer (1988)).

Since

$$\sup_{k \geqq 0} |P(S_n = k) - P(T_n = k)| \leqq \sup_{x \in \mathbb{R}} |P(S_n \leqq x) - P(T_n \leqq x)|$$

we also obtain

$$P(S_n = k) = e^{-\lambda_n} \frac{\lambda_n^k}{k!} + O\left(\frac{1}{\log n}\right) \qquad (n \to \infty),$$

uniformly in $k \geqq 0$, where $\lambda_n = \log n + C - 1$, $n \geqq 2$, which improves Kemp's (1984) estimate.

In the remainder of the paper we want to analyse the probabilistic behaviour of the searching algorithm in a Nevzorov model, i.e. in a model with non-equiprobable orderings of the elements. In view of Lemma 3, such a model is fairly general since arbitrary distributions for the maximum N can be treated. However, the independence of record indices also shows that for a probabilistic analysis of the algorithm, it is *not* sufficient to know the distribution of N alone since there are of course also models with *dependent* record indices.

Since in a Nevzorov model, the number of intermediate storages, S_n, is still given by $S_n = \sum_{i=2}^n I_i$, a sum of independent $\{0, 1\}$-valued random variables most of the 'classical' results immediately carry over to the more general situation. For instance, putting

$$p_i = P(I_i = 1) = \frac{\alpha_i}{\sum_{j=1}^{i} \alpha_j}, \quad 2 \leqq i \leqq n,$$

we obtain

$$E(S_n) = \sum_{i=2}^{n} p_i, \qquad \operatorname{Var}(S_n) = \sum_{i=2}^{n} p_i(1 - p_i)$$

and

$$P(S_n=0)=\prod_{i=2}^{n}P(I_i=0)=\prod_{i=2}^{n}(1-p_i)$$

$$P(S_n=1)=P\Big(\bigcup_{i=2}^{n}\{I_i=1\}\cap\bigcap_{2\leqq j\leqq n,j\neq i}\{I_j=0\}\Big)$$

$$=\sum_{i=2}^{n}p_i\prod_{2\leqq j\leqq n,j\neq i}(1-p_j)=\sum_{i=2}^{n}\frac{p_i}{1-p_i}\prod_{j=2}^{n}(1-p_j)$$

$$P(S_n=n-2)=P\Big(\bigcup_{i=2}^{n}\{I_i=0\}\cap\bigcap_{2\leqq j\leqq n,j\neq i}\{I_j=1\}\Big)$$

$$=\sum_{i=2}^{n}(1-p_i)\prod_{2\leqq j\leqq n,j\neq i}p_j=\sum_{i=2}^{n}\frac{1-p_i}{p_i}\prod_{j=2}^{n}p_j$$

$$P(S_n=n-1)=\prod_{i=2}^{n}P(I_i=1)=\prod_{i=2}^{n}p_i$$

as well as the recursive relation

$$\begin{aligned}P(S_{n+1}=k)&=\sum_{i=0}^{k}P(S_n=k-i)P(I_{n+1}=i)\\&=P(S_n=k-1)P(I_{n+1}=1)+P(S_n=k)P(I_{n+1}=0)\\&=p_{n+1}P(S_n=k-1)+(1-p_{n+1})P(S_n=k),\qquad 0\leqq k\leqq n,\end{aligned}$$

which generalizes known results for the i.i.d. case (Knuth (1973), p. 97, Kemp (1984), p. 23).

Also, a Poisson approximation for the distribution of S_n is possible when $\sum_{i=2}^{n}p_i^2/\sum_{i=2}^{n}p_i$ is small, since for a Poisson random variable T_n with mean $E(S_n)=\sum_{i=2}^{n}p_i$ we have

$$\sup_{x\in\mathbb{R}}|P(S_n\leqq x)-P(T_n\leqq x)|=\frac{1}{2\sqrt{2\pi e}}\frac{\sum_{i=2}^{n}p_i^2}{\sum_{i=2}^{n}p_i}+O\left(\max\left\{\frac{\sum_{i=2}^{n}p_i^2}{\left\{\sum_{i=2}^{n}p_i\right\}^{\frac{3}{2}}},\left\{\frac{\sum_{i=2}^{n}p_i^2}{\sum_{i=2}^{n}p_i}\right\}^2\right\}\right)\qquad(n\to\infty)$$

(cf. Deheuvels and Pfeifer (1988)) which gives

$$P(S_n=k)=e^{-\lambda_n}\frac{\lambda_n^k}{k!}+O\left(\frac{\sum_{i=2}^{n}p_i^2}{\sum_{i=2}^{n}p_i}\right)\qquad(n\to\infty),$$

uniformly in $k\geqq 0$, where now $\lambda_n=\sum_{i=2}^{n}p_i$, $n\geqq 2$.

There is, however, a significant difference between the equiprobable and the non-equiprobable distribution case since in the first situation, the linear search through the set $\{x_1, \cdots, x_n\}$ is as good as a search in arbitrary order. This is no longer the case in a more general situation such as a Nevzorov model, and it is worth while to ask for the 'best' searching strategy here. The subsequent theorem gives an answer to this problem, based on the following two auxiliary results.

Lemma 4. *Let $a, x, y > 0$ be real numbers with $x < y$. Then*

$$\frac{x}{a+x} \cdot \frac{y}{a+x+y} > \frac{y}{a+y} \cdot \frac{x}{a+x+y}$$

$$\left(1 - \frac{x}{a+x}\right)\left(1 - \frac{y}{a+x+y}\right) = \left(1 - \frac{y}{a+y}\right)\left(1 - \frac{x}{a+x+y}\right)$$

$$\frac{x}{a+x} + \frac{y}{a+x+y} > \frac{y}{a+y} + \frac{x}{a+x+y}.$$

Proof. Straightforward.

Lemma 5. *Let X, Y, Z be independent random variables such that Y is stochastically smaller than Z, i.e. $P(Y \leqq x) \geqq P(Z \leqq x)$ for all $x \in \mathbb{R}$. Then $X + Y$ is stochastically smaller than $X + Z$.*

Proof. Straightforward.

Theorem 3. *Let $\beta_1 \geqq \beta_2 \geqq \cdots \geqq \beta_n > 0$, $n \in \mathbb{N}$ be real numbers and $(\alpha_1, \cdots, \alpha_n)$ be a permutation of $(\beta_1, \cdots, \beta_n)$. Then under the assumptions of Lemma 3, S_n is stochastically smallest if $\alpha_i = \beta_i$, $1 \leqq i \leqq n$, simultaneously minimizing $E(S_n)$.*

Proof. Again let $p_i = P(I_i = 1) = \alpha_i / \sum_{j=1}^{i} \alpha_j$, $2 \leqq i \leqq n$. Suppose for a permutation $(\alpha_1, \cdots, \alpha_n)$ we have $\alpha_i < \alpha_{i+1}$ for some $1 \leqq i < n$. Put $a = \sum_{j=1}^{i-1} \alpha_j$, $x = \alpha_i$, $y = \alpha_{i+1}$ and

$$p_i' = \frac{y}{a+y}, \qquad p_{i+1}' = \frac{x}{a+x+y}.$$

Then by Lemma 4

$$\begin{aligned} &p_i' p_{i+1}' < p_i p_{i+1}, \\ (3.2) \qquad &(1 - p_i')(1 - p_{i+1}') = (1 - p_i)(1 - p_{i+1}),\ p_i' + p_{i+1}' < p_i + p_{i+1}. \end{aligned}$$

By the independence of record indices and Lemma 5, it suffices to show that $I_i' + I_{i+1}'$ is stochastically smaller than $I_i + I_{i+1}$ where I_i', I_{i+1}' are independent $\{0, 1\}$-valued random variables with success probabilities p_i' and p_{i+1}', respectively.

But by (3.2), it suffices to show

$$\begin{aligned} P(I_i' + I_{i+1}' \leqq 0) &= (1-p_i')(1-p_{i+1}') \\ &= (1-p_i)(1-p_{i+1}) = P(I_i + I_{i+1} \leqq 0) \\ P(I_i' + I_{i+1}' \leqq 1) &= 1 - p_i' p_{i+1}' \\ &> 1 - p_i p_{i+1} = P(I_i + I_{i+1} \leqq 1). \end{aligned}$$

Hence S_n is (even strictly) stochastically smaller under the permutation $(\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \alpha_i, \alpha_{i+2}, \cdots, \alpha_n)$ than under $(\alpha_1, \cdots, \alpha_n)$. Since by (3.10), $p_i' + p_{i+1}' < p_i + p_{i+1}$, also $E(S_n)$ is strictly smaller under the permutation $(\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \alpha_i, \alpha_{i+2}, \cdots, \alpha_n)$. This proves the theorem.

Theorem 3 thus says that in a Nevzorov model the number S_n of intermediate storages is stochastically smallest (with minimal mean) if the set $\{x_1, \cdots, x_n\}$ is searched through in the (optimal) order $x_{\tau(1)}, \cdots, x_{\tau(n)}$ where τ is a permutation such that $\alpha_{\tau(1)} \geqq \cdots \geqq \alpha_{\tau(n)}$. This is also intuitively obvious since

$$\alpha_{\tau(1)} = P(N = x_{\tau(1)}) = \max_{1 \leqq j \leqq n} \alpha_j = \max_{1 \leqq j \leqq n} P(N = x_j)$$

and

$$\begin{aligned} \frac{\alpha_{\tau(k)}}{1 - \sum_{j=1}^{k-1} \alpha_{\tau(j)}} &= P(N = x_{\tau(k)} \mid N \notin \{x_{\tau(1)}, \cdots, x_{\tau(k-1)}\}) \\ &= \max_{k \leqq i \leqq n} \frac{\alpha_{\tau(i)}}{1 - \sum_{j=1}^{k-1} \alpha_{\tau(j)}} \\ &= \max_{1 \leqq j \leqq n} P(N = x_j \mid N \notin \{x_{\tau(1)}, \cdots, x_{\tau(k-1)}\}). \end{aligned}$$

Another conclusion is that if the set $\{x_1, \cdots, x_n\}$ is searched through in the optimal order described above then S_n is stochastically *largest* in the equidistribution case since here

$$p_i = \frac{\alpha_{\tau(i)}}{\sum_{j=1}^{i} \alpha_{\tau(j)}} \leqq \frac{\alpha_{\tau(i)}}{i\alpha_{\tau(i)}} = \frac{1}{i} = p_i^*, \qquad 1 \leqq i \leqq n,$$

where the p_i^* are the success probabilities for the record indices in the i.i.d. case. This means that if the distribution of the possible orderings of the elements, Σ_n^{-1}, is known, the average number $E(S_n)$ of intermediate storages grows at most at a logarithmic rate if the proper strategy is chosen. However, if the distribution of Σ_n^{-1} is *not* known, the probabilistic behaviour of S_n may be arbitrarily bad with $S_n \equiv n - 1$ in the worst case.

We conclude with some examples.

Example 1 (*triangular distribution*). *Let* $\alpha_i = 2i/n(n+1)$, $1 \leqq i \leqq n$. *Then* $p_i = 2/(i+1)$, $2 \leqq i \leqq n$ *and hence*

$$(3.3)\quad \begin{aligned} E(S_n) &= 2\sum_{j=3}^{n+1}\frac{1}{j} = 2\log n + 2C - 3 + O\Big(\frac{1}{n}\Big) \\ \operatorname{Var}(S_n) &= 2\sum_{j=3}^{n+1}\Big(\frac{1}{j}-\frac{2}{j^2}\Big) = 2\log n + 2C + 2 - \frac{2\pi^2}{3} + O\Big(\frac{1}{n}\Big) \end{aligned} \qquad (n \to \infty)$$

as well as

$$P(S_n = k) = e^{-\lambda_n}\frac{\lambda_n^k}{k!} + O\Big(\frac{1}{\log n}\Big) \qquad (n \to \infty)$$

with $\lambda_n = 2\log n + 2C - 3$, *uniformly in* $k \geqq 0$. *Here,* $\tau = (n\, n-1 \cdots 2\, 1)$ *would be optimal* (*i.e. searching in reverse order*) *with a resulting average of*

$$(3.4)\quad \begin{aligned} E(S_n) &= \frac{2n+2}{2n+1}\sum_{j=2}^{n}\frac{1}{j} - \frac{2n}{2n+1}\sum_{j=n+1}^{2n-1}\frac{1}{j} \\ &= \log n + C - 1 - \log 2 + O\Big(\frac{\log n}{n}\Big) \qquad (n \to \infty) \end{aligned}$$

which can be obtained after some straightforward calculations.

A comparison between (3.3) and (3.4) shows that the optimal searching strategy requires only about one half of the number of storages (on average) as in the linear search case. However, the saving is only about log 2 steps (on average) compared with the equidistribution case (cf. (3.1)).

Example 2 (*truncated geometric distribution*). *Let* $\alpha_i = \alpha^{i-1}(1-\alpha)/(1-\alpha^n)$, $1 \leqq i \leqq n$ *with some positive real number* $\alpha \neq 1$. *Then* $p_i = \alpha^{i-1}(1-\alpha)/(1-\alpha^i)$, $2 \leqq i \leqq n$.

Case I. $\alpha < 1$. *Some numerical analysis shows that*

$$\max\Big\{\alpha - \alpha^n, -\frac{1-\alpha}{\alpha\log\alpha}\log\Big(\frac{1-\alpha^{n+1}}{1-\alpha^2}\Big)\Big\} \leqq E(S_n) = \frac{1-\alpha}{\alpha}\sum_{j=2}^{n}\frac{\alpha^j}{1-\alpha^j}$$
$$\leqq \min\Big\{\frac{\alpha-\alpha^n}{1-\alpha^2}, -\frac{1-\alpha}{\alpha\log\alpha}\log\Big(\frac{1-\alpha^n}{1-\alpha}\Big)\Big\}.$$

Moreover, with $n \to \infty$, $E(S_n)$ *converges to some finite limit* s *with*

$$\max\Big\{\alpha, \frac{(1-\alpha)\log(1-\alpha^2)}{\alpha\log\alpha}\Big\} \leqq s \leqq \min\Big\{\frac{\alpha}{1-\alpha^2}, \frac{(1-\alpha)\log(1-\alpha)}{\alpha\log\alpha}\Big\},$$

and, by the monotone convergence theorem, S_n *converges in distribution to some integrable random variable* S *with mean* $E(S) = s$ *which says that for large* n, *the distribution of the number of intermediate storages is almost independent of* n.

Note that in Case I, $\alpha_1 > \cdots > \alpha_n$, *hence linear search is the optimal strategy here.*

Case II. $\alpha > 1$. *Here*

$$\max\left\{\frac{\alpha-1}{\alpha}(n-1), \frac{\alpha-1}{\alpha\log\alpha}\log\left(\frac{\alpha^{n+1}-1}{\alpha^2-1}\right)\right\}$$
$$\leqq E(S_n)=\frac{\alpha-1}{\alpha}\sum_{j=2}^{n}\frac{\alpha^j}{\alpha^j-1}$$
$$\leqq \min\left\{\frac{\alpha}{\alpha+1}(n-1), \frac{\alpha-1}{\alpha\log\alpha}\log\left(\frac{\alpha^{n}-1}{\alpha-1}\right)\right\},$$

i.e.

$$E(S_n)=\frac{\alpha-1}{\alpha}n+O(1) \qquad (n\to\infty)$$

which is significantly worse than in the equidistribution case and comes close to the worst case for large α.

Note that the optimal strategy here would again be a search in reverse order; a corresponding probabilistic analysis can easily be reduced to Case I, *with* α *being replaced by* $1/\alpha$.

Acknowledgements

I wish to thank the referee for pointing out some references which simplified the presentation of the paper.

References

Dansie, B. R. (1983) A note on permutation probabilities. *J. R. Statist. Soc.* B **45,** 22–24.

Deheuvels, P. and Pfeifer, D. (1988) On a relationship between Uspensky's theorem and Poisson approximation. *Ann. Inst. Statist. Math.* **40,** 671–681.

Hartigan, J. A. (1968) Note on discordant observations. *J. R. Statist. Soc.* B **30,** 545–550.

Harville, D. A. (1973) Assigning probabilities to the outcomes of multi-entry competitions. *J. Amer. Statist. Assoc.* **68,** 312–316.

Kemp, R. (1984) *Fundamentals of the Average Case Analysis of Particular Algorithms.* Wiley–Teubner, New York and Stuttgart.

Knuth, D. W. (1973) *The Art of Computer Programming.* Vol. 1: *Fundamental Algorithms.* Addision-Wesley, Reading, MA.

Nevzorov, V. B. (1986) Two characterizations using records. In *Stability Problems for Stochastic Models.* Lecture Notes in Mathematics **1233,** Springer-Verlag, 79–85.

Nevzorov, V. B. (1988) Records. *Theory Prob. Appl.* **32,** 201–228.

Pfeifer, D. (1987) On a joint strong approximation theorem for record and inter-record times. *Prob. Theory Rel. Fields* **75,** 213–221.

Pfeifer, D. (1989a) *Einführung in die Extremwertstatistik.* Teubner-Verlag, Stuttgart.

Pfeifer, D. (1989b) Extremal processes, secretary problems and the $1/e$ law. *J. Appl. Prob.* **26,** 722–733.

PLACKETT, R. L. (1968) Random permutations. *J. R. Statist. Soc.* B **30,** 517–534.

PLACKETT, R. L. (1975) The analysis of permutations. *Appl. Statist.* **24,** 193–202.

RÉNYI, A. (1962) Théorie des éléments saillants d'une suite d'observations. *Colloquium on Combinatorial Methods in Probability Theory,* 104–115, Mathematisk Institut, Aarhus Universitet, Denmark.

RESNICK, S. I. (1987) *Extreme Values, Regular Variations and Point Processes.* Springer, New York.