Ann Inst Stat Math (2018) 70:951–968
https://doi.org/10.1007/s10463-017-0621-1

Asymptotic results for jump probabilities associated to the multiple scan statistic

Markos V. Koutras[1] · **Demetrios P. Lyberopoulos**[1]

Received: 14 October 2015 / Revised: 2 June 2017 / Published online: 16 November 2017

Abstract The concept of pattern arises in many applications comprising experimental trials with two or more possible outcomes in each trial. A binary scan of type r/k is a special pattern referring to success–failure strings of fixed length k that contain at least r-successes, where r, k are positive integers with $r \leq k$. The multiple scan statistic $W_{t,k,r}$ is defined as the enumerating random variable for the overlapping moving windows occurring until trial t which include a scan of type r/k. In the present work, we consider a sequence of independent binary trials with not necessarily equal probabilities of success and develop upper bounds for the probability of the event that the multiple scan statistic will perform a jump from ℓ to $\ell + 1$ (where ℓ is a nonnegative integer) in a finite time horizon.

Keywords Multiple scan statistic · Upper bound · Demisubmartingale · N-demisupermartingale · Demimartingale

1 Introduction

One of the most popular topics of the theory of patterns has certainly been the study of scan statistics, that is of random variables enumerating the moving windows in a sequence of binary outcomes trials which contain a prescribed number of successes, see, e.g. Balakrishnan and Koutras (2002) or Glaz et al. (2001). Moreover, relative

✉ Demetrios P. Lyberopoulos
dilyber@webmail.unipi.gr

Markos V. Koutras
mkoutras@unipi.gr

1 Department of Statistics and Insurance Science, University of Piraeus, 80 Karaoli and Dimitriou street, 185 34 Piraeus, Greece

waiting times problems have been addressed under several frameworks, referring, for instance, to the assumptions made for the sequence of binary trials, the scans' enumerating schemes and patterns more general than scans (cf., e.g. Inoue and Aki 2009, Hsieh and Wu 2013 and Koutras and Alexandrou 1995). Note that a useful tool for studying "compound patterns" (scans belong to this general family of patterns) has been the theory of martingales (cf., e.g. Pozdnyakov et al. 2005 and Pozdnyakov and Steele 2009).

Despite this and the fact that the classes of demi(sub)martingales and N-demi(super)martingales have really drawn the attention of many researchers during the last decades [see, e.g. the pioneering papers of Newman and Wright (1982) and Christofides (2003), respectively, where the latter two concepts were introduced], there seems to be a gap in taking advantage of the results developed for the aforementioned generalizations of martingales to arrive at useful outcomes for scan statistics problems.

A first attempt to fill in this gap for the case of the multiple scan statistic is made here. Our interest focuses on it mainly for two reasons: the first one is that the multiple scan statistic has been studied in a quite smaller extent than its simple counterparts, while the second one refers to the fact that it appears in several problems of the statistical literature, which refer to Scans Theory and can be fruitfully exploited in a wide range of research areas, such as quality control, actuarial science, reliability theory and molecular biology (see, e.g. Balakrishnan and Koutras 2002).

The main aim of the present work is to develop some asymptotic results for jump probabilities associated to a multiple scan statistic. After introducing the necessary definitions and notations in Sect. 2, an upper bound is developed in Sect. 3 for an upcrossing probability in a nondecreasing sequence of random variables. In the same section, we exploit the previous result to establish an asymptotic upper bound for the probability of the event that the multiple scan statistic will perform a jump from ℓ to $\ell + 1$ until the t-th trial (ℓ is a nonnegative integer). Some further asymptotic results are also discussed.

In Sect. 4, we extend the aforementioned upcrossing inequality for the case of sequences belonging to the class of demisubmartingales and prove that the multiple scan statistic is member of this class. In addition, we investigate some interesting theoretical results for the multiple scan statistic process concerning its membership to the classes of demimartingales and N-demisupermartingales. Finally, a numerical study of the developed bounds and a brief discussion on potential applications are carried out in Sects. 5 and 6, respectively.

2 Preliminaries

In the sequel, $\mathbb{N} := \{1, 2, \ldots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. The set of all real numbers will be denoted by $\mathbb{R}$, while if $d \in \mathbb{N}$ then $\mathbb{R}^d$ will stand for the Euclidean space of dimension d. Moreover, $x \wedge y := \min\{x, y\}$ and $x^+ := \max\{x, 0\}$ for $x, y \in \mathbb{R}$. For $n \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$ the i-canonical projection from $\mathbb{R}^n$ onto $\mathbb{R}$ is denoted by π_i.

Throughout this paper, we consider an arbitrary but fixed probability space $(\varOmega, \varSigma, P)$. A set $N \in \varSigma$ with $P(N) = 0$ is called a P-null set. A sequence $\{Z_j\}_{j \in \mathbb{N}}$ of $\varSigma$-measurable functions satisfies a property P-a.s. if there exists a P-null set, say

O, such that the property is satisfied by $\{Z_j\}_{j\in\mathbb{N}}$ for all $\omega \notin O$. Furthermore, χ_E stands for the indicator (or characteristic) function of the set $E \in \Sigma$.

On defining $T_Z := \{B \subseteq \mathbb{R}^d : Z^{-1}(B) \in \Sigma\}$ for any $\Sigma - \mathfrak{B}_d$-measurable function Z, it is clear that $\mathfrak{B}_d \subseteq T_Z$, where $\mathfrak{B}_d := \mathfrak{B}(\mathbb{R}^d)$ stands for the Borel σ-algebra of subsets of $\mathbb{R}^d$ ($d \in \mathbb{N}$). We shall be denoting by $P_Z : T_Z \longrightarrow \mathbb{R}$ the image measure of P under Z. The restriction of P_Z to $\mathfrak{B}_d$ is denoted again by P_Z, while R_Z stands for the range of Z. The notation $\mathbf{PB}(n, p_1, \ldots, p_n)$, where $n \in \mathbb{N}$ and $p_j \in (0, 1)$ for $j \in \{1, \ldots, n\}$, stands for the law of Poisson's binomial distribution. Note that $\mathbf{PB}(n, p_1, \ldots, p_n)$ is in fact the distribution of a sum of n-Bernoulli random variables with probabilities of success $p_1, \ldots, p_n$ (cf., e.g. Wang 1993, Section 3 for more details).

The family of all real-valued P-integrable functions on Ω will be denoted by $\mathcal{L}^1(P)$. Functions that are P-a.s. equal are not identified. The expectation of the random variable Z is denoted by $\mathbb{E}_P[Z]$.

Recall also that any family $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ of σ-subalgebras of Σ, such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for each $n \in \mathbb{N}$, is called a filtration for the measurable space (Ω, Σ). Moreover, a sequence $\{Z_n\}_{n\in\mathbb{N}}$ of random variables on Ω is said to be adapted to a filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ if each Z_n is $\mathcal{F}_n$-measurable. If $\mathcal{F}_n$ is the σ-algebra generated by the random vector $(Z_1, \ldots, Z_n)$, i.e. $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$ for each $n \in \mathbb{N}$, then $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ is said to be the canonical filtration for $\{Z_n\}_{n\in\mathbb{N}}$, and it will be denoted by $\{\mathcal{F}_n^{(Z)}\}_{n\in\mathbb{N}}$.

Definition 1 Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^1(P)$. Then $\{Z_n\}_{n\in\mathbb{N}}$ is said to be:
(a) a P-**martingale** (with respect to $\{\mathcal{F}_n^{(Z)}\}_{n\in\mathbb{N}}$), if

$$\mathbb{E}_P[(Z_{n+1} - Z_n) f(Z_1, \ldots, Z_n)] = 0 \quad \text{for each } n \in \mathbb{N} \tag{1}$$

and for every measurable function f on $\mathbb{R}^n$ such that the above expectations exist. If condition (1) but with "$\geq$" in the place of the equality is satisfied for every f as above but with $f \geq 0$, then $\{Z_n\}_{n\in\mathbb{N}}$ is said to be a P-submartingale (with respect to $\{\mathcal{F}_n^{(Z)}\}_{n\in\mathbb{N}}$).
(b) a ***P*-demimartingale**, if condition (1) but with "$\geq$" in the place of the equality is satisfied for every f coordinatewise nondecreasing function on $\mathbb{R}^n$ such that the above expectations exist.
(c) a ***P*-demisubmartingale**, if condition (1) but with "$\geq$" in the place of the equality is satisfied for every f as in (b) but with $f \geq 0$.
(d) a ***N*-demimartingale** under P, if condition (1) but with "$\leq$" in the place of the equality is satisfied for every f as in (b). In particular, if $f \geq 0$ then $\{Z_n\}_{n\in\mathbb{N}}$ is said to be a ***N*-demisupermartingale** under P.

From the definitions given above it is clear that the class of all P-martingales is a subset of the class of all demimartingales, which in its own turn is a subclass of the demisubmartingales' one. Moreover, it is obvious that any P-submartingale is also a demisubmartingale as well as that any N-demimartingale is also a N-demisupermartingale. For more on Definition 1 and the way that the notions given in there are related to each other, the interested reader may refer to the excellent monograph by Prakasa Rao (2012).

Next we insert, for completeness sake, a short proof of the equivalence between the usual definition of a P-submartingale and its alternative one, given in the second part of Definition 1, (a): Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of random variables which satisfies the usual definition of a P-submartingale. It then follows that $\{Z_n\}_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^1(P)$ such that

$$\mathbb{E}_P[\chi_A(Z_{n+1}-Z_n)] \geq 0 \quad \text{for each } n \in \mathbb{N} \text{ and } A \in \mathcal{F}_n^{(Z)};$$

hence $\mathbb{E}_P[(Z_{n+1}-Z_n)\chi_B(Z_1,\ldots,Z_n)] \geq 0$ for each $B \in \mathfrak{B}_n$, which implies successively that condition

$$\mathbb{E}_P[(Z_{n+1}-Z_n)f(Z_1,\ldots,Z_n)] \geq 0 \quad \text{for each } n \in \mathbb{N}$$

holds true for every f characteristic, f simple and f nonnegative measurable function. Thus, $\{Z_n\}_{n\in\mathbb{N}}$ satisfies the second part of Definition 1, (a). Since the inverse implication is obvious, the equivalence of the two definitions follows.

3 Some new asymptotic results for the multiple scan statistic

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of binary trials on Ω, each resulting in either a *success* (that is $\{X_n = 1\}$) or a *failure* (that is $\{X_n = 0\}$) with probabilities of success p_n ($0 < p_n < 1$). Then, for any fixed $k \in \mathbb{N}$ and for each $m \in \mathbb{N}$ such that $m \leq k$, the sequence $X_n, X_{n+1}, \ldots, X_{n+m-1}$ of random variables on Ω is called a **moving window** (for $\{X_n\}_{n\in\mathbb{N}}$) of length m. In particular, if $P(\sum_{j=n}^{n+m-1} X_j \geq r) > 0$ then the above subsequence of $\{X_n\}_{n\in\mathbb{N}}$ is said to be a ***(P-)*scan** or ***(P-)*generalized run** of type r/k, that is, the term "scan of type r/k" refers to subsequences $X_n, X_{n+1}, \ldots, X_{n+m-1}$ of length $m \leq k$ such that the number of successes contained therein is at least r with positive probability.

For each $n \in \mathbb{N}$ and $k \in \mathbb{N}$ consider the random variable $Y_{n,k}$ on Ω defined by

$$Y_{n,k} := \sum_{j=\max\{n-k+1,1\}}^{n} X_j.$$

In what follows, we set $X_0 := 0$ and consider every sum over an empty index set conventionally equal to zero.

The multiple scan statistic declares the total number of the overlapping moving windows including a scan of type r/k until trial t, where $r, k \in \mathbb{N}$ with $r \leq k$ and $t \in \mathbb{N}$. To state it more formally, if we let r, k, t be as above, then the random variable $W_{t,k,r}$ defined on Ω by means of

$$W_{t,k,r} := \sum_{n=k}^{t} \chi_{[r,\infty)}(Y_{n,k})$$

will be called the (t,k,r)-multiple scan statistic associated with the sequence $\{Y_{n,k}\}_{n\in\mathbb{N}}$ or briefly the **multiple scan statistic** if no confusion arises. Moreover, the sequence $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ will be called the multiple scan statistic process. Note that $W_{t,k,r} = 0$ for every $t < k$ and that $R_{W_{t,k,r}} = \{0, \ldots, t-k+1\}$ for every $t \geq k$.

The above definition of the multiple scan statistic has been adopted by Boutsikas et al. (2009) as well as by Chen and Glaz (1996, 2005) for the two-dimensional case. However, it should be mentioned that the nomenclature "multiple scan statistic" can also be met in the literature under concepts different than the one defined above, see, e.g. Cucala (2008) and Naus and Wallenstein (2004).

Some useful notions and facts are also recalled next. Let $\{Z_n\}_{n\in\{1,\ldots,m\}}$ be a sequence of random variables on Ω, and recall (see, e.g. Prakasa Rao 2012) that, for any fixed $a, b \in \mathbb{R}$ with $a < b$, by $U_{a,b}$ is denoted the random variable defined on Ω by means of

$$U_{a,b} := U_{a,b}(Z_1, \ldots, Z_m) := \max\{k \in \mathbb{N} : J_{2k} < m+1\},$$

where

$$J_{2k-1} := \begin{cases} m+1, & \text{if } \{n \in \mathbb{N} : J_{2k-2} < n \leq m,\ Z_n \leq a\} = \emptyset \\ \min\{n \in \mathbb{N} : J_{2k-2} < n \leq m,\ Z_n \leq a\}, & \text{otherwise} \end{cases}$$

and

$$J_{2k} := \begin{cases} m+1, & \text{if } \{n \in \mathbb{N} : J_{2k-1} < n \leq m,\ Z_n \geq b\} = \emptyset \\ \min\{n \in \mathbb{N} : J_{2k-1} < n \leq m,\ Z_n \geq b\}, & \text{otherwise} \end{cases}$$

for each $k \in \mathbb{N}$ (convention $J_0 := 0$). The random variable $U_{a,b}$ counts the number of (a, b)-upcrossings for $\{Z_n\}_{n\in\{1,\ldots,m\}}$. For more, see Prakasa Rao (2012), Sections 2.4 and 3.4.

We shall now prove the next result, which will be exploited in the sequel and is quite useful in its own as well.

Lemma 1 *Let $m_1, m_2 \in \mathbb{N}$ with $m_1 < m_2$. If $\{Z_n\}_{n\in\mathbb{N}}$ is a P-a.s. nondecreasing sequence of random variables on Ω then the inequality*

$$P(U_{a,b}(Z_{m_1}, \ldots, Z_{m_2}) = 1) \leq \frac{1}{b-a}\mathbb{E}_P[(Z_{m_2} - a)^+ - (Z_{m_1} - a)^+]$$

holds true for all $a, b \in \mathbb{R}$ with $a < b$.

Proof First fix on arbitrary $m_1, m_2 \in \mathbb{N}$ with $m_1 < m_2$. Next consider the sequence $\{Z_n^{(m_1,m_2)}\}_{n\in\mathbb{N}}$ of random variables on Ω defined as follows

$$Z_n^{(m_1,m_2)} := \begin{cases} Z_{m_1}, & \text{if } n \in \{1, \ldots, m_1 - 1\} \\ Z_n, & \text{if } n \in \{m_1, \ldots, m_2\} \\ Z_{m_2}, & \text{if } n \in \{m_2 + 1, \ldots\}. \end{cases} \tag{2}$$

Evidently $U_{a,b}(Z_1^{(m_1,m_2)}, \ldots, Z_m^{(m_1,m_2)}) = U_{a,b}(Z_{m_1}, \ldots, Z_{m_2})$ for each $m \in \mathbb{N}$ with $m \geq m_2$ and for all $a, b \in \mathbb{R}$ with $a < b$.

It follows by the monotonicity of $\{Z_n\}_{n\in\mathbb{N}}$ that $U_{a,b}(Z_{m_1}, \ldots, Z_{m_2})$ takes on two values (i.e. 0, 1) as well as that

$$(b-a)\chi_{\{Z_{m_1}\leq a, Z_{m_2}\geq b\}} \leq Z_{m_2} - Z_{m_1}.$$

Therefore, we may state that

$$P(Z_{m_1} \le a, Z_{m_2} \ge b) \le \frac{1}{b-a}\mathbb{E}_P[Z_{m_2} - Z_{m_1}]$$

for all $a, b \in \mathbb{R}$ with $a < b$.

Applying now the latter inequality for all a, b as above and for the sequence $\{\widetilde{Z}_n(a)\}_{n\in\mathbb{N}}$ of random variables on Ω defined as $\widetilde{Z}_n(a) := (Z_n - a)^+$, we deduce that

$$\begin{aligned}\frac{1}{b-a}\mathbb{E}_P[\widetilde{Z}_{m_2}(a) - \widetilde{Z}_{m_1}(a)] &\ge P\big((Z_{m_1} - a)^+ = 0, (Z_{m_2} - a)^+ \ge b - a\big)\\ &\ge P\big(Z_{m_1} - a < 0, Z_{m_2} - a > b - a, Z_{m_2} > a\big)\\ &= P\big(Z_{m_1} < a, Z_{m_2} > b\big),\end{aligned}$$

thereof we obtain an equivalent expression for the requested inequality. □

For the remainder of this section, let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent binary trials such that $p_n := P(X_n = 1) \in (0, 1)$ and $q_n := 1 - p_n$ for each $n \in \mathbb{N}$.

Fix now arbitrary $r, k \in \mathbb{N}$ with $r \le k$ and consider arbitrary $\varepsilon \in (0, 1/2)$ and $\ell \in \mathbb{N}_0$. Also consider the sequence $\{U_t(\ell, \varepsilon)\}_{t\in\mathbb{N}}$ of random variables on Ω defined by

$$U_t(\ell,\varepsilon) := \begin{cases} U_{\ell+\varepsilon,\ell+1-\varepsilon}(W_{k,k,r}, \ldots, W_{t,k,r}), & \text{if } t \in \{k, k+1, \ldots\} \\ 0, & \text{otherwise.}\end{cases}$$

Note that the random variable $U_t(\ell, \varepsilon)$ denotes the number of upcrossings of the multiple scan statistic process from below $\ell + \varepsilon$ to above $\ell + 1 - \varepsilon$, that is the number of its jumps from ℓ to $\ell + 1$, until trial t. Clearly, $U_t(\ell, \varepsilon)$ does not depend on the specific choice of ε (provided that $0 < \varepsilon < 1/2$); hence, for each $t \in \mathbb{N}$ the random variable

$$U_t(\ell) := \lim_{\varepsilon\to 0^+} U_t(\ell, \varepsilon)$$

declares the number of jumps that the multiple scan statistic process makes from ℓ to $\ell + 1$ until trial t, as well.

The next proposition provides the main asymptotic result of this paper concerning the multiple scan statistic.

Proposition 1 *Let $\theta \in (0, \infty)$ and r, k be arbitrary but fixed positive integers such that $r \le k$.*

If $p_t \to 0^+$ as $t \to \infty$, so that the asymptotic condition

$$\lim_{t\to\infty}(t - k + 1)\binom{k-1}{r-1}p_t^r q_t^{k-r+1} = \theta \tag{3}$$

is satisfied, then for each $\ell \in \mathbb{N}_0$ the following holds true:

$$P\left(\bigcup_{s=k}^{\infty}\{W_{s+1,k,r} = \ell + 1,\ W_{s,k,r} = \ell\}\right) \le u(r, k, \ell, \theta), \tag{4}$$

where

$$u(r,k,\ell,\theta) := \sum_{n=k+\ell}^{\infty} \Big[P(Y_{n,k} \geq r) \wedge [1 - F_*(\ell;\theta,r,k)]\Big] - (1-\ell)^+ P(Y_{k,k} \geq r).$$

The function F_* appearing above is the cumulative distribution function of a compound Poisson distribution with structural parameter θ and compounding cumulative distribution function

$$G(x) := G(x;r,k) := \begin{cases} 0, & \text{if } x \leq 0 \\ 1 - \frac{\binom{k-x-1}{r-1}}{\binom{k-1}{r-1}}, & \text{if } x \in \{1,\ldots,k-r\} \\ 1, & \text{if } x \geq k-r+1. \end{cases}$$

Proof First fix on arbitrary $\ell \in \mathbb{N}_0$ and $r, k \in \mathbb{N}$ with $r \leq k$. Then note that by the definition of the random variable $U_t(\ell)$ the range of $U_t(\ell)$, $t \geq k$, equals $\{0, 1\}$, and therefore

$$\mathbb{E}_P[U_t(\ell)] = P(U_t(\ell) = 1) = P\left(\bigcup_{s=k}^{t} \{W_{s+1,k,r} = \ell+1,\, W_{s,k,r} = \ell\}\right).$$

As a consequence, we may write

$$\lim_{t\to\infty} \mathbb{E}_P[U_t(\ell)] = P\left(\bigcup_{s=k}^{\infty} \{Y_{s+1,k} > r,\, W_{s,k,r} = \ell\}\right) \tag{5}$$

and we may state that the limit $\lim_{t\to\infty} \mathbb{E}_P[U_t(\ell)]$ does exist.

Furthermore, Lemma 1 yields for every $0 < \varepsilon < 1/2$, and for each $t \in \mathbb{N}$ with $t \geq k$ that

$$\begin{aligned} \mathbb{E}_P[U_t(\ell,\varepsilon)] \ &\leq \frac{\mathbb{E}_P\big[\big(W_{t,k,r} - (\ell+\varepsilon)\big)^+ - \big(W_{k,k,r} - (\ell+\varepsilon)\big)^+\big]}{\big(\ell+1-\varepsilon\big) - \big(\ell+\varepsilon\big)} \qquad (6) \\ &= \frac{1}{1-2\varepsilon}\mathbb{E}_P\Big[\Big(W_{t,k,r} - \ell - \varepsilon\Big)\chi_{\{W_{t,k,r} > \ell+\varepsilon\}}\Big] \\ &\quad - \frac{1}{1-2\varepsilon}\mathbb{E}_P\Big[\Big(W_{k,k,r} - \ell - \varepsilon\Big)\chi_{\{W_{k,k,r} > \ell+\varepsilon\}}\Big] \\ &= \frac{1}{1-2\varepsilon}\mathbb{E}_P\Big[\Big(W_{t,k,r} - \ell - \varepsilon\Big)\chi_{\{W_{t,k,r} \geq \ell+1\}}\Big] \\ &\quad - \frac{1}{1-2\varepsilon}\mathbb{E}_P\Big[\Big(W_{k,k,r} - \ell - \varepsilon\Big)\chi_{\{W_{k,k,r} \geq \ell+1\}}\Big] \\ &\leq \frac{1}{1-2\varepsilon}\mathbb{E}_P\Big[\Big(W_{t,k,r} - \ell - \varepsilon\Big)\chi_{\{W_{t,k,r} \geq \ell+1\}}\Big] \\ &\quad - \frac{1-\varepsilon}{1-2\varepsilon} P(W_{k,k,r} \geq \ell+1); \end{aligned}$$

letting now $\varepsilon \to 0^+$ and making use of the Monotone Convergence Theorem we obtain

$$\begin{aligned}
\mathbb{E}_P[U_t(\ell)] &\le \mathbb{E}_P\big[(W_{t,k,r}-\ell)\chi_{\{W_{t,k,r}\ge \ell+1\}}\big] - P(W_{k,k,r}\ge \ell+1) \qquad (7)\\
&\le \mathbb{E}_P\big[(W_{t,k,r}-W_{k+\ell-1,k,r})\chi_{\{W_{t,k,r}\ge \ell+1\}}\big] - (1-\ell)^+ P(Y_{k,k}\ge r)\\
&= \mathbb{E}_P\left[\sum_{n=k+\ell}^{t} \chi_{[r,\infty)}(Y_{n,k})\chi_{\{W_{t,k,r}\ge \ell+1\}}\right] - (1-\ell)^+ P(Y_{k,k}\ge r)\\
&\le \sum_{n=k+\ell}^{t} [P(Y_{n,k}\ge r)\wedge P(W_{t,k,r}\ge \ell+1)] - (1-\ell)^+ P(Y_{k,k}\ge r). \qquad (8)
\end{aligned}$$

Noting that for each $t \in \mathbb{N}$ with $t \ge k+\ell$ and for each $n \in \{k+\ell,\dots,t\}$ we have

$$P(Y_{n,k}\ge r) \le P(Y_{n,k}\ge 1) = 1 - P(Y_{n,k}=0) = 1 - \prod_{j=n-k+1}^{n} q_j < 1$$

with the last equality being justified by the P-independence of $\{X_n\}_{n\in\mathbb{N}}$, we conclude by the Cauchy criterion for series convergence that

$$\sum_{n=k+\ell}^{\infty} [P(Y_{n,k}\ge r)\wedge P(W_{t,k,r}\ge \ell+1)] < \infty.$$

Inequality (8) yields now, for each $t \in \mathbb{N}$ with $t \ge k$

$$\mathbb{E}_P[U_t(\ell)] \le \sum_{n=k+\ell}^{\infty} [P(Y_{n,k}\ge r)\wedge P(W_{t,k,r}\ge \ell+1)] - (1-\ell)^+ P(Y_{k,k}\ge r)$$

and letting $t \to \infty$ we obtain, by virtue of the Fubini Theorem

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{E}_P[U_t(\ell)] &\le \sum_{n=k+\ell}^{\infty} \lim_{t\to\infty}[P(Y_{n,k}\ge r)\wedge P(W_{t,k,r}\ge \ell+1)]\\
&\quad - (1-\ell)^+ P(Y_{k,k}\ge r)\\
&\le \sum_{n=k+\ell}^{\infty} [P(Y_{n,k}\ge r)\wedge \lim_{t\to\infty} P(W_{t,k,r}\ge \ell+1)]\\
&\quad - (1-\ell)^+ P(Y_{k,k}\ge r).
\end{aligned}$$

This result, condition (5), Corollary 3.3.3 of Boutsikas et al. (2009) and assumption (3) prove the asymptotic result stated in (4). □

According to the proof of Proposition 1, if we consider a $t \in \mathbb{N}$ with $t \ge k$ the probability that the multiple scan statistic process will perform a jump from ℓ to $\ell+1$ until trial t will be given by

$$\begin{aligned}\varpi_{t,k,r}(\ell) &:= P\left(\bigcup_{s=k}^{t}\{W_{s+1,k,r}=\ell+1,\, W_{s,k,r}=\ell\}\right)\\ &= P\left(\bigcup_{s=k+\ell-1}^{t}\{W_{s+1,k,r}=\ell+1,\, W_{s,k,r}=\ell\}\right)\\ &= \sum_{s=k+\ell-1}^{t} P(Y_{s+1,k}\geq r,\, W_{s,k,r}=\ell);\end{aligned}$$

the last equality holds true because the events $\{W_{s+1,k,r}=\ell+1,\, W_{s,k,r}=\ell\}$, $s\in\mathbb{N}$, are disjoint. Therefore, we get that

$$\varpi_{t,k,r}(\ell)\leq \sum_{n=k+\ell}^{t+1} P(Y_{n,k}\geq r) =: u_0(r,k,\ell,t).$$

Clearly, $u_0(r,k,\ell,t)$ is one of the most non-sophisticated or more "naive" upper bounds (if not the most one) for the probability $\varpi_{t,k,r}(\ell)$. It is also evident that it is outperformed by the upper bound provided in condition (8); hence $u_0(r,k,\ell) := \lim_{t\to\infty} u_0(r,k,\ell,t) = \sum_{n=k+\ell}^{\infty} P(Y_{n,k}\geq r) < \infty$ is outperformed by the upper bound given in the above proposition.

It is also worth noticing that the expectation in the right side of inequality (7) can be also bounded as follows:

$$\begin{aligned}\mathbb{E}_P\big[(W_{t,k,r}-\ell)\chi_{\{W_{t,k,r}\geq\ell+1\}}\big] &= \mathbb{E}_P[W_{t,k,r}]-\ell-\mathbb{E}_P[(W_{t,k,r}-\ell)\chi_{\{W_{t,k,r}\leq\ell\}}]\\ &\leq \sum_{n=k}^{\infty} P(Y_{n,k}\geq r)-\ell+\ell P(W_{t,k,r}\leq\ell)\\ &= \sum_{n=k}^{\infty} P(Y_{n,k}\geq r)-\ell[1-P(W_{t,k,r}\leq\ell)].\end{aligned}$$

Thus, for fixed $\ell\in\mathbb{N}_0$ and $r,k\in\mathbb{N}$ with $r\leq k$ and for each $t\in\mathbb{N}$ with $t\geq k$

$$u_1(r,k,\ell,t) := \sum_{n=k}^{\infty} P(Y_{n,k}\geq r)-\ell[1-P(W_{t,k,r}\leq\ell)]-(1-\ell)^+P(Y_{k,k}\geq r)$$

is an upper bound for the probability $\varpi_{t,k,r}(\ell)$. Then as in the proof of Proposition 1, we apply the Monotone Convergence Theorem and take into account Corollary 3.3.3 of Boutsikas et al. (2009) to infer that if condition (3) is satisfied then

$$\begin{aligned}\lim_{t\to\infty} u_1(r,k,\ell,t) &= \sum_{n=k}^{\infty} P(Y_{n,k}\geq r)-\ell[1-F_*(\ell;\theta,r,k)]-(1-\ell)^+P(Y_{k,k}\geq r)\\ &= u_1(r,k,\ell,\theta)\end{aligned}$$

(where F_* is as in Proposition 1) is an upper bound for the probability $\varpi_{k,r}(\ell) := P(\bigcup_{s=k}^{\infty}\{W_{s+1,k,r} = \ell+1,\ W_{s,k,r} = \ell\})$.

Obviously, $u_1(r,k,\ell,\theta)$ is tighter than $u_0(r,k,\ell)$. It cannot be claimed, though, that it is **admissible**, in the sense that it takes values in (0, 1), for all r, k, ℓ as above. Furthermore, the corresponding finite time horizon bound is more efficient than its naive counterpart, i.e. $u_1(r,k,\ell,t) < u_0(r,k,\ell,t)$, if and only if

$$P(W_{t,k,r} \leq \ell) < 1 - \frac{1}{\ell}\left[\sum_{n\in\Psi} P(Y_{n,k} \geq r) + (1-\ell)^+ P(Y_{k,k} \geq r)\right],$$

where $\Psi := \{k, \ldots, k+\ell-1\} \cup \{t+2, t+3, \ldots\}$.

Corollary 1 *Under the assumptions of Proposition 1 the upper bound*

$$u(r,k,\ell,\theta) = \sum_{n=k+\ell}^{\infty}\Big[P(Y_{n,k} \geq r) \wedge [1 - F_*(\ell;\theta,r,k)]\Big] - (1-\ell)^+ P(Y_{k,k} \geq r)$$

can be evaluated, for fixed $\ell \in \mathbb{N}_0$ and $r, k \in \mathbb{N}$ with $r \leq k$, and for each $t \in \mathbb{N}$ with $t \geq k$, by the aid of the following recursive formulae:

$$P(Y_{n,k} = y) = \begin{cases} \prod_{j=1}^{k} q_{n-k+j}, & \text{if } \ y = 0 \\ (1/y)\sum_{j=1}^{y}(-1)^{j-1}P(Y_{n,k} = y-j)T(j), & \text{if } \ y > 0, \end{cases} \tag{9}$$

where

$$T(j) := \sum_{i=1}^{k}\left(\frac{p_{n-k+i}}{q_{n-k+i}}\right)^j \quad \text{for each } j, \tag{10}$$

and

$$F_*(\ell;\theta,r,k) := \sum_{y=0}^{\ell} f_*(y;\theta,r,k),$$

with

$$f_*(y;\theta,r,k) = \begin{cases} e^{-\theta}, & \text{if } \ y = 0 \\ (\theta/y)\sum_{j=1}^{(k-r+1)\wedge y} j[G(j) - G(j-1)]f_*(y-j;\theta,r,k), & \text{if } \ y > 0, \end{cases} \tag{11}$$

where G is as in Proposition 1.

Proof The P-independence of the binary trials $\{X_n\}_{n\in\mathbb{N}}$ yields for all r, k, t as above and for each $n \in \{k, \ldots, t\}$ that $P_{Y_{n,k}} = \mathbf{PB}(k;\ p_{n-k+1}, \ldots, p_n)$; hence, conditions (9) and (10) hold true (cf., e.g. Shah 1994). For the computation of the compound Poisson part of $u(r,k,\ell,\theta)$ one may use either the recursive scheme of equation (12.4.6) of Bowers et al. (1997) or Panjer's recursions (cf., e.g. Bowers et al. 1997, Theorem 12.4.3), which lead to condition (11). Thus, our corollary follows by Proposition 1. $\square$

In a similar way, the upper bounds $u_0(r, k, \ell)$ and $u_1(r, k, \ell, \theta)$ can be explicitly computed as well.

It follows by (9) and (10) that if $p_t \to 0^+$ as $t \to \infty$, then the probability $P(Y_{t,k} = y)$ will do so for any $y \in \{1, \ldots, k\}$; hence, there exists a positive integer $t_0 \geq k$ such that $\sum_{n=k+\ell}^{\infty} P(Y_{n,k} \geq r) \simeq \sum_{n=k+\ell}^{t_0} P(Y_{n,k} \geq r)$, where r, k, ℓ are as in Corollary 1. The latter suggests that it is most likely that there are values of r, k, ℓ, t for which all or some of the above upper bounds prove to be admissible.

Since it is immediate by the definition of $U_t(\ell)$ that for any fixed $\ell \in \mathbb{N}_0$ the function $t \longmapsto U_t(\ell)$ is nondecreasing, it follows by the Monotone Convergence Theorem that $\lim_{t\to\infty} \mathbb{E}_P[U_t(\ell)] = \mathbb{E}_P[\lim_{t\to\infty} U_t(\ell)] = \mathbb{E}_P[U_\infty(\ell)]$, where the random variable $U_\infty(\ell) := \lim_{t\to\infty} U_t(\ell)$ denotes the number of jumps from ℓ to $\ell+1$ for the multiple scan statistic process.

Consider now the random variables $U_\infty := \sum_{\ell=0}^{\infty} U_\infty(\ell)$ and $U_t := \sum_{\ell=0}^{\infty} U_t(\ell)$, which obviously stand for the number of unit jumps that the multiple scan statistic process presents in an infinite time horizon and until trial t, respectively. Then the following result is a consequence of the definition of U_t, condition (8) and the Fubini Theorem.

Corollary 2 *Under the assumptions of Proposition 1 and for each $t \in \mathbb{N}$ with $t \geq k$ the probability $P\left(\bigcup_{\ell=0}^{t-k+1} \bigcup_{s=k}^{\infty} \{W_{s+1,k,r} = \ell + 1,\ W_{s,k,r} = \ell\}\right)$ is upper bounded by*

$$\sum_{\ell=0}^{t-k+1} \sum_{n=k+\ell}^{t} [P(Y_{n,k} \geq r) \wedge P(W_{t,k,r} \geq \ell + 1)] - (1 - \ell)^+ P(Y_{k,k} \geq r).$$

Motivated by Corollary 2, we raise the question whether upper bounds similar to the one of Proposition 1 can be obtained for the expectation $\mathbb{E}_P[U_\infty]$ as well? In fact, this is not a trivial task and will not concern us more here.

4 An alternative approach via demisubmartingales

It is made clear by the preceding analysis that Lemma 1 is vital for establishing all the asymptotic results given in Sect. 3 as it possesses a key role in the proof of Proposition 1. The inequality provided by Lemma 1 was established under the assumption that $\{Z_n\}_{n\in\mathbb{N}}$ is a P-a.s. nondecreasing sequence of random variables, and was subsequently applied for a scan process. We shall now extend the result of Lemma 1 for the case where the same sequence is assumed to be a P-demisubmartingale, thereof providing a useful tool to obtain upcrossing inequalities for a wider range of sequence of random variables. As an illustration we indicate how one can use the more general result for the case of the multiple scan statistic. This approach, in addition to an alternative proof of Lemma 1, leads to some interesting theoretical results for the multiple scan statistic, which possess their own merit.

To start with let us first point out that if a sequence $\{Z_n\}_{n\in\mathbb{N}}$ of random variables on Ω is either a P-demisubmartingale or a N-demisupermartingale under P, then the same applies for the sequence $\{Z_n^{(m_1,m_2)}\}_{n\in\mathbb{N}}$ defined by means of (2).

To verify that, note that for each $n \in \{m_1, \ldots, m_2 - 1\}$ and for every f nonnegative coordinatewise nondecreasing function on $\mathbb{R}^n$ we may write

$$\begin{aligned} I_{n,m_1,m_2}(f) &:= \mathbb{E}_P\left[\left(Z_{n+1}^{(m_1,m_2)} - Z_n^{(m_1,m_2)}\right) f\left(Z_1^{(m_1,m_2)}, \ldots, Z_{m_1}^{(m_1,m_2)}, \ldots, Z_n^{(m_1,m_2)}\right)\right] \\ &= \mathbb{E}_P[(Z_{n+1} - Z_n) f(Z_{m_1}, \ldots, Z_{m_1}, \ldots, Z_n)] \\ &= \mathbb{E}_P[(Z_{n+1} - Z_n) h(Z_1, \ldots, Z_{m_1}, \ldots, Z_n)] \end{aligned}$$

with the function $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ being defined as follows:

$$h(z_1, \ldots, z_{m_1}, \ldots, z_n) := f(z_{m_1}, \ldots, z_{m_1}, \ldots, z_n) \text{ for each } (z_1, \ldots, z_n) \in \mathbb{R}^n.$$

The proof is completed by noting that since f is a nonnegative coordinatewise nondecreasing function on $\mathbb{R}^n$, the same will hold true for h as well.

We are now ready to present the extension of the upcrossing inequality given in Lemma 1, for the case of demisubmartingales.

Proposition 2 *Let $m_1, m_2 \in \mathbb{N}$ with $m_1 < m_2$.*

If the sequence $\{Z_n\}_{n\in\mathbb{N}}$ is a P-demisubmartingale, then the following inequality holds true:

$$\mathbb{E}_P[U_{a,b}(Z_{m_1}, \ldots, Z_{m_2})] \leq \frac{\mathbb{E}_P[(Z_{m_2} - a)^+ - (Z_{m_1} - a)^+]}{b - a}$$

for all $a, b \in \mathbb{R}$ such that $a < b$.

Proof First fix on arbitrary $m_1, m_2 \in \mathbb{N}$ with $m_1 < m_2$. Next define the sequence $\{\widetilde{Z}_n\}_{n\in\mathbb{N}}$ of random variables on Ω as follows:

$$\widetilde{Z}_n := \begin{cases} Z_{n+m_1-1}, & \text{if } n \in \{1, \ldots, m_2 - m_1\} \\ Z_{m_2}, & \text{otherwise.} \end{cases}$$

We shall now proceed by carrying out the next two steps.
(a) We shall prove that the sequence $\{\widetilde{Z}_n\}_{n\in\mathbb{N}}$ is a P-demisubmartingale.

First note that for $n \in \{1, \ldots, m_2 - m_1\}$ and for every f nonnegative coordinatewise nondecreasing function on $\mathbb{R}^n$ we get

$$\begin{aligned} \widetilde{H}_n(f) &:= \mathbb{E}_P\left[\left(\widetilde{Z}_{n+1} - \widetilde{Z}_n\right) f\left(\widetilde{Z}_1, \ldots, \widetilde{Z}_n\right)\right] \\ &= \mathbb{E}_P[(Z_{n+m_1} - Z_{n+m_1-1}) f(Z_{m_1}, Z_{m_1+1}, \ldots, Z_{n+m_1-1})]; \end{aligned}$$

hence, letting $j = n + m_1 - 1 \in \{m_1, \ldots, m_2 - 1\}$ we deduce for every f nonnegative coordinatewise nondecreasing function on $\mathbb{R}^{j-m_1+1}$ that

$$\widetilde{H}_n(f) = \mathbb{E}_P\left[\left(Z_{j+1}^{(m_1,m_2)} - Z_j^{(m_1,m_2)}\right) f\left(Z_{m_1}^{(m_1,m_2)}, Z_{m_1+1}^{(m_1,m_2)}, \ldots, Z_j^{(m_1,m_2)}\right)\right]. \tag{12}$$

Since, as stated before, $\{Z_n^{(m_1,m_2)}\}_{n\in\mathbb{N}}$ is a P-demisubmartingale, it follows that for each $j \in \{m_1, \ldots, m_2-1\}$ and for every h nonnegative coordinatewise nondecreasing function on $\mathbb{R}^j$ the inequality

$$\mathbb{E}_P\left[\left(Z_{j+1}^{(m_1,m_2)} - Z_j^{(m_1,m_2)}\right) h\left(Z_1^{(m_1,m_2)}, \ldots, Z_j^{(m_1,m_2)}\right)\right] \geq 0$$

holds true. Consider now the function $h_1 : \mathbb{R}^j \longrightarrow \mathbb{R}$ given by

$$h_1(z_1, \ldots, z_{m_1-1}, z_{m_1}, \ldots, z_j) := \mathbf{1}(z_1, \ldots, z_{m_1-1}) f(z_{m_1}, \ldots, z_j)$$

for each $(z_1, \ldots, z_j) \in \mathbb{R}^j$, where $\mathbf{1}$ is the unit function on $\mathbb{R}^{m_1-1}$. Clearly, h_1 is a nonnegative coordinatewise nondecreasing function on $\mathbb{R}^j$ since f is so on $\mathbb{R}^{j-m_1+1}$; hence, applying the last inequality for $h = h_1$ and taking into account (12) we infer that $\widetilde{H}_n(f) \geq 0$. The latter along with our assumption that $\{Z_n\}_{n\in\mathbb{N}}$ is a P-demisubmartingale proves the statement given in the beginning of step (a).
(b) Consider, now, for any fixed $a, b \in \mathbb{R}$ with $a < b$ the random variable

$$\widetilde{U}_{a,b} := U_{a,b}(\widetilde{Z}_1, \ldots, \widetilde{Z}_{m_2-m_1+1}) = U_{a,b}(Z_{m_1}, \ldots, Z_{m_2}).$$

It follows by (a) that we may apply a well-known upcrossing inequality for demisubmartingales (cf. Newman and Wright 1982 or better see, e.g. Prakasa Rao 2012, Theorem 2.4.1) to deduce that

$$\mathbb{E}_P[\widetilde{U}_{a,b}] \leq \frac{\mathbb{E}_P[(\widetilde{Z}_{m_2-m_1+1} - a)^+ - (\widetilde{Z}_1 - a)^+]}{b-a},$$

which proves the proposition. □

Lemma 2 *Let $k \in \mathbb{N}$ be arbitrary but fixed, and let $r \in \mathbb{N}$ such that $r \leq k$. Then the following hold true:*

(i) *The multiple scan statistic process $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ is a P-submartingale.*

If the sequence $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ is in addition either a P-demimartingale or a N - demisupermartingale under P, then

(ii) *there exists a P-null set $O_W \in \Sigma$ such that for any $\omega \notin O_W$ condition $W_{t,k,r}(\omega) = 0$ holds for each $t \in \mathbb{N}$;*
(iii) $P(\bigcap_{j=1}^r \{X_j = 1\}) = 0.$

Proof First fix on an arbitrary $k, r \in \mathbb{N}$ such that $r \leq k$.
(i) It then follows for each $t \in \mathbb{N}$ that $\mathbb{E}_P[W_{t,k,r}] = \sum_{n=r}^t P(Y_{n,k} \geq r) < \infty$ as well as

$$W_{t+1,k,r} - W_{t,k,r} = \chi_{[r,\infty)}(Y_{t+1,k});$$

hence, for every f nonnegative measurable function on $\mathbb{R}^t$ such that the expectation

$$H_{t,k,r}(f) := H_{t,k,r}(W_{1,k,r},\ldots,W_{t,k,r};f)$$
$$:= \mathbb{E}_P[(W_{t+1,k,r}-W_{t,k,r})f(W_{1,k,r},\ldots,W_{t,k,r})]$$

exists, we get

$$H_{t,k,r}(f) = \mathbb{E}_P[\chi_{[r,\infty)}(Y_{t+1,k})f(W_{1,k,r},\ldots,W_{t,k,r})] \geq 0, \tag{13}$$

which completes the proof of (i).
(ii) Assume that $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ is a P-demimartingale. It then follows that condition (13) is satisfied for each $t \in \mathbb{N}$ and for every f coordinatewise nondecreasing function on $\mathbb{R}^t$ such that the expectation $H_{t,k,r}(f)$ exists; hence, for $f = f_1 := -1$ we get $P(Y_{t+1,k} \geq r) = 0$ for each $t \in \mathbb{N}$, implying that $P(\bigcup_{t\in\mathbb{N}}\{Y_{t+1,k} \geq r\}) = 0$ or equivalently that $P(\bigcap_{t\in\mathbb{N}}\{Y_{t+1,k} < r\}) = 1$, and so there is a universal P-null set $O_W := O_{Y;r,k} \in \Sigma$ such that for any $\omega \notin O_W$ condition $Y_{n+1,k}(\omega) < r$ holds true for each $n \in \mathbb{N}$.

Assuming now that $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ is a N-demisupermartingale under P, we get by (i) that for each $t \in \mathbb{N}$ and for every f nonnegative coordinatewise nondecreasing function on $\mathbb{R}^t$, such that the expectation $H_{t,k,r}(f)$ exists, condition (13) holds true but with "=" in the place of "≥"; hence, by an application of the latter for $f = f_2 := 1$ we infer again that $P(Y_{t+1,k} \geq r) = 0$ for each $t \in \mathbb{N}$, which proves assertion (ii).
(iii) If the sequence $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ is in addition either a P-demimartingale or a N-demisupermartingale under P, then by the proof of (ii) we get that $P(Y_{r,k} = r) = 0$ or equivalently $P(\bigcap_{j=1}^{r}\{X_j = 1\}) = 0$. □

By virtue of assertion (i) of Lemma 2 we may apply Proposition 2 in order to rediscover the inequality that plays a key role in the proof of Proposition 1, that is (6). So, the latter proposition can be derived once again by repeating the part of its proof beyond the aforementioned key inequality. Therefore, all asymptotic results of Sect. 3 remain valid.

Thus, it can be stated that the alternative approach for proving the asymptotic results of Sect. 3 differs from the first (and more direct) one, presented in that section, in the following: herein the submartingale property of the multiple scan statistic process together with the new upcrossing inequality, given in Proposition 2, was exploited for deriving inequality (6), while in Sect. 3 the same inequality was proved by an application of Lemma 1 in conjunction with the fact that the multiple scan statistic process is a P-a.s. nondecreasing sequence of random variables on Ω.

Remarks. In view of Lemma 2, we may state the next interesting facts for the multiple scan statistic process $\{W_{t,k,r}\}_{t\in\mathbb{N}}$:
(a) If the sequence $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ is either a P-demimartingale or a N-demisupermartingale under P, then by assertion (ii) of the above lemma we get that $P(T_r^{(k)} \leq t) = 1 - P(W_{t,k,r} = 0) = 0$ for each $t \in \mathbb{N}$, where $T_r^{(k)}$ is the waiting time for the first occurrence of a scan of type r/k, that is $T_r^{(k)} := \min\{n \in \mathbb{N} : Y_{n,k} \geq r\}$. Consequently, $P(T_r^{(k)} = \infty) = 1$.
(b) If the sequence $\{X_n\}_{n\in\mathbb{N}}$ is P-independent, then $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ cannot be either a P-demimartingale or a N-demisupermartingale under P; this is so because otherwise

Table 1 Bounds of $\varpi_{t,k,r}(\ell)$ for four different choices of parameters (I: $p = 0.1,\ \theta = 0.1,\ (r,k) = (4,8),\ \ell = 4$, II: $p = 0.025,\ \theta = 0.01,\ (r,k) = (3,4),\ \ell = 3$, III: $p = 0.005,\ \theta = 0.0005,\ (r,k) = (3,9),\ \ell = 5$, IV: $p = 0.005,\ \theta = 0.0005,\ (r,k) = (3,9),\ \ell = 0$)

I	t	55	56	57	58	59
	Upper bound	0.187234	0.191489	0.195744	0.2	0.204255
II	t	227	228	229	230	231
	Upper bound	0.00124131	0.00124693	0.00125255	0.00125816	0.00126378
III	t	155	156	157	158	159
	Upper bound	0.00145779	0.00146805	0.00147832	0.00148858	0.00149885
IV	t	155	156	157	158	159
	Upper bound	0.00149885	0.00150912	0.00151938	0.00152965	0.00153991

assertion (iii) of Lemma 2 would hold true, which would imply that $p_j = 0$ for some $j \in \{1, \ldots, r+1\}$.
(c) In a similar fashion with (b), we infer that if $\{X_n\}_{n\in\mathbb{N}}$ is a P-homogeneous Markov chain of first order with transition probabilities $\{p_{s,t}\}_{s,t\in\{0,1\}}$ (cf., e.g. Shiryaev 1984), such that

$$p_{1,1} = P(X_{n+1} = 1 \mid X_n = 1) > 0 \quad \text{for each } n \in \mathbb{N}$$

then $\{W_{t,k,r}\}_{t\in\mathbb{N}}$ cannot be either a P-demimartingale or a N-demisupermartingale under P, since the Markov property yields $P(\bigcap_{j=1}^{r}\{X_j = 1\}) = p_{1,1}^{r-1}p_1 > 0$ for $r \in \mathbb{N}$.

5 Numerical results

In this section, Corollary 1 is exploited in order to compute upper bounds of the probability $\varpi_{t,k,r}(\ell)$ for various values of the parameters r, k, ℓ, θ as well as of the success probabilities p_t.

A simple approach for obtaining these bounds so as condition (3) and condition $\lim_{t\to 0^+} p_t = 0$ are fulfilled is the following one: first fix on an arbitrary $\theta \in (0,\infty)$. Next assign to $r, k, \ell \in \mathbb{N}$, where $r \leq k$, the specific values you wish (e.g. $r = 12,\ k = 24,\ \ell = 10$) as well as a very small value (e.g. $p_0 = 10^{-i}$ for $i \in \{1,2,3\}$) to all success probabilities p_t. Finally, set $t_* := \left\lfloor \frac{\theta}{\binom{k-1}{r-1}p_0^r(1-p_0)^{k-r+1}} \right\rfloor + k - 1$, where $\lfloor x \rfloor$ denotes the integer part of the real number x. Then Corollary 1 can be applied for computing the requested upper bounds for all $t \geq t_*$. Table 1 gives such bounds for five successive values of t with the first entry referring to t_*.

In Table 2 we present for specific choices of the parameters r, k, ℓ, θ, p the bound stated in Proposition 1 along with the corresponding simulated values for the probability $\varpi_{t,k,r}(\ell)$ and the *relative error*

$$\text{r.e.} := \frac{\text{upper bound-simulated value}}{\text{simulated value}}.$$

Table 2 Computed bounds versus simulated $\varpi_{t,k,r}(\ell)$ for $(r, k, \ell, \theta, p) = (4, 10, 7, 0.1, 0.1)$ and $s = 10{,}000$

t	Sim	Upper bound	r.e.
60	0.0173	0.0293	0.6936
80	0.0253	0.0426	0.6838
100	0.0350	0.0559	0.5971
120	0.0462	0.0692	0.4978
140	0.0571	0.0825	0.4448
160	0.0699	0.0958	0.3705
180	0.0831	0.1091	0.3129
200	0.0983	0.1225	0.2462
220	0.1119	0.1358	0.2136
240	0.1260	0.1491	0.1833
260	0.1412	0.1624	0.1501
280	0.1595	0.1757	0.1016
300	0.1749	0.1890	0.0806

The simulated values have been obtained at each t as follows: First a sequence $x_1, \ldots, x_t$ of t-random numbers from a Bernoulli distribution with success probability p is generated. Then the sums $y_{n,k} = \sum_{j=n-k+1}^{n} x_j$ and $w_{n,k,r} = \sum_{n=k}^{t} \chi_{[r,\infty)}(y_{n,k})$ are recorded in separate lists. If the inequality $\max\{w_{k,k,r}, \ldots, w_{t,k,r}\} \geq \ell + 1$ holds true then 1 is assigned to the generated list of random numbers else 0 is done so. This routine is performed for each of the s-simulated sequences (usually $s = 10{,}000$; a larger value of s, namely $s = 100{,}000$, was used when greater accuracy was necessary); the requested simulated probability is then given as the ratio of the sum of the assigned units divided by s.

Extensive numerical experimentation revealed that, in general, the upper bound of Proposition 1 appears to be quite conservative. The behaviour of approximation errors with respect to the values of the parameters r, k, ℓ, θ, p does not seem to follow a specific pattern. However, under the condition $\theta \simeq p$ and for θ, p taking values close to 0, the bound produces tight values for the probability $\varpi_{t,k,r}(\ell)$ and in addition the corresponding relative errors behave as a nonincreasing function of the number of trials t (see Table 2); therefore, one might state that our bound improves as time evolves.

It is also worth mentioning that for all sets of parameter values that were examined the computing time for evaluating our bound was significantly smaller than the computing time of the simulated values for $\varpi_{t,k,r}(\ell)$. The evaluation of the simulated values proved to be extremely time-consuming, and in some cases (e.g. for parametric vectors with θ, p close to 0 and large l) computationally intractable. Thus, it can be stated that there are sets of parametric values for which our bound provides a rough, nevertheless the only attainable estimate (in terms of computing time) for the probabilities $\varpi_{t,k,r}(\ell)$.

In closing we mention that the upper bound provided by Proposition 1 is not a probability, and therefore, it may produce for certain values of the parameters (e.g. $p = 0.075,\ \theta = 1.2,\ r = 2,\ k = 4,\ \ell = 7$) inadmissible bounds (greater than one). Once again our numerical experimentation revealed that in general the bound generates admissible values for

- *small values of* θ, namely $\theta \in (0, 1)$;
- *small or moderate values of* r;
- *small success probabilities* p_t, i.e. between 0.001 and 0.1.

6 A discussion on possible applications

In this section we shall discuss a couple of real-life situations where an interest for either the event $\bigcup_{s=k}^{t}\{W_{n,k,r} = \ell, W_{n+1,k,r} = \ell + 1\}$ or the event $\bigcup_{s=k}^{\infty}\{W_{n,k,r} = \ell, W_{n+1,k,r} = \ell + 1\}$; hence, for the probabilities $\varpi_{t,k,r}(\ell)$ or $\varpi_{k,r}(\ell)$ (where $\ell \in \mathbb{N}_0$ and $r, k \in \mathbb{N}$ with $t \geq k \geq r$), respectively, may emerge. The first one was motivated by Section 5 of Boutsikas and Koutras (2002), while the second was inspired from sports practice.

Example 1 If each random variable X_n is equal to the indicator function of the event that the reserve process of an insurance company drops below a critical level (say $\alpha > 0$) at time n, then the multiple scan statistic $W_{t,k,r}$ will count the number of times that the (discrete-time) reserve process presents r-such falls in moving time windows of length k. Providing that values of $W_{t,k,r}$ exceeding ℓ raise a warning about the evolution of the company's reserve, $\varpi_{k,r}(\ell)$ will give the probability of a "solvency alarm" for the insurance company.

Example 2 A scouter of a basketball team wishes to make a (short) transfer list of three-point shooters currently possessing moderate percentages, potentially improvable though. A first step towards this direction is to have a look at the shooters' stats and decide whether they will be included in his initial list basing on the following criterion: *detect those players scoring for more than* $\ell = 7$ *times* $r = 3$ *out of the last* $k = 5$ *three-point shots every* $t = 100$ *shots*. Assume now that the outcomes of the three-point attempts of each player do not depend with each other as well as that the success probability for each attempt deteriorates to zero as time passes, and that the expected number of successful attempts will eventually settle to a positive number θ (i.e. the requirements of Proposition 1). In fact, the latter assumptions may easily be regarded as very realistic ones, since they can be interpreted as the "fatigue effect" ($\lim_{t\to\infty} p_t = 0$) in the performance of a three-point shooter (it is reasonable to assume that $\theta > 0$, since otherwise the player should definitely avoid calling himself a shooter!). So, in this framework $\varpi_{100,3,5}(7)$ denotes the probability of detecting a member of the scouter's (initial) list.

Implications for more areas of application (e.g. molecular biology) of our results can be found in Section 5 of Boutsikas and Koutras (2002).

Acknowledgements D.P.L. would like to dedicate this work in memory of his father Panagiotis (Takis). His support during this research endeavour of D.P.L. is one of the many moving memories which the co-author will always recall, full of love and gratitude! The authors would like to thank the anonymous reviewers for their helpful and constructive comments that contributed to improving the original version of this paper. This research has been co-financed by the European Union (European Social Fund—ESF) and

Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: Aristeia II. Investing in knowledge society through the European Social Fund.

References

Balakrishnan, N., Koutras, M. V. (2002). *Runs and scans with applications*. New York: Wiley.

Boutsikas, M. V., Koutras, M. V. (2002). Modeling claim exceedances over thresholds. *Insurance*: *Mathematics & Economics*, *30*, 67–83.

Boutsikas, M. V., Koutras, M. V., Milienos, F. S. (2009). Extreme value results for scan statistics. In J. Glaz, V. Pozdnyakov, S. Wallenstein (Eds.), *Scan statistics*: *Methods and applications* (pp. 55–85). Boston: Birkhäuser.

Bowers, N. L., Gerber, H. U., Hickman, J., Jones, D. A., Nesbitt, C. J. (1997). *Actuarial mathematics* (2nd ed.). Illinois: The Society of Actuaries.

Chen, J., Glaz, J. (1996). Two-dimensional discrete scan statistics. *Statistics & Probability Letters*, *31*, 59–68.

Chen, J., Glaz, J. (2005). Approximations for multiple scan statistics. In R. Baeza-Yates, J. Glaz, H. Gzyl, J. Hüsler, J. L. Palacios (Eds.), *Recent advances in applied probability* (pp. 97–114). Boston: Springer.

Christofides, T. C. (2003). Maximal inequalities for N-demimartingales. *Archives of Inequalities and Applications*, *1*, 387–397.

Cucala, L. (2008). A hypothesis-free multiple scan statistic with variable window. *Biometrical Journal*, *50*, 299–310.

Glaz, J., Naus, J. I., Wallenstein, S. (2001). *Scan statistics*. New York: Springer.

Hsieh, Y.-F., Wu, T.-L. (2013). Recursive equations in finite Markov chain imbedding. *Annals of the Institute of Statistical Mathematics*, *65*, 513–527.

Inoue, K., Aki, S. (2009). On waiting time distributions associated with compound patterns in a sequence of multi-state trials. *Annals of the Institute of Statistical Mathematics*, *61*, 499–516.

Koutras, M. V., Alexandrou, V. A. (1995). Runs, scans and urn model distributions: A unified Markov chain approach. *Annals of the Institute of Statistical Mathematics*, *47*, 743–766.

Naus, J. I., Wallenstein, S. (2004). Multiple window and cluster size scan procedures. *Methodology and Computing in Applied Probability*, *6*, 389–400.

Newman, C. M., Wright, A. L. (1982). Associated random variables and martingale inequalities. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *59*, 361–371.

Pozdnyakov, V., Steele, J. M. (2009). Martingale methods for patterns and scan statistics. In J. Glaz, V. Pozdnyakov, S. Wallenstein (Eds.), *Scan statistics*: *Methods and applications* (pp. 289–317). Boston: Birkhäuser.

Pozdnyakov, V., Glaz, J., Kulldorff, M., Steele, J. M. (2005). A martingale approach to scan statistics. *Annals of the Institute of Statistical Mathematics*, *57*, 21–37.

Prakasa Rao, B. L. S. (2012). *Associated sequences, demimartingales and nonparametric inference. Probability and its applications*. Basel: Springer.

Shah, B. K. (1994). On the distribution of the sum of independent integer valued random variables. *The American Statistician*, *27*, 123–124.

Shiryaev, A. N. (1984). *Probability, volume 95 of Graduate Texts in Mathematics* (2nd ed.). Berlin: Springer.

Wang, Y. H. (1993). On the number of successes in independent trials. *Statistica Sinica*, *3*, 295–312.