J. Appl. Prob. **54,** 363–378 (2017)
*doi:*10.1017/jpr.2017.5

NEW MAXIMAL INEQUALITIES FOR *N*-DEMIMARTINGALES WITH SCAN STATISTIC APPLICATIONS

MARKOS V. KOUTRAS * ** AND
DEMETRIOS P. LYBEROPOULOS,* *** *University of Piraeus*

Abstract

In the present work, some new maximal inequalities for nonnegative N-demi(super)-martingales are first developed. As an application, new bounds for the cumulative distribution function of the waiting time for the first occurrence of a scan statistic in a sequence of independent and identically distributed (i.i.d.) binary trials are obtained. A numerical study is also carried out for investigating the behavior of the new bounds.

Keywords: N-demimartingale; N-demisupermartingale; demimartingale; scan statistic; demisubmartingale; bound; i.i.d. binary trials

2010 Mathematics Subject Classification: Primary 60E15
Secondary 60G48

1. Introduction

Demimartingales have drawn the attention of many researchers during the last decades (see e.g. the pioneering papers of Newman and Wright [14] and Christofides [7], respectively, where the latter two concepts were introduced), because of the fact that they can be exploited to develop useful tools for dealing with stochastical dependence. The martingale theory possesses an important feature: many of the techniques and tools derived by it remain valid or can be easily extended under very general assumptions on the underlying structure. For more details, we refer the interested reader to, e.g. Prakasa Rao [18].

This is not always straightforward or guaranteed, though. On the contrary, it may prove to be quite tricky as it is suggested by a recent contribution of Dai *et al.* [8], in which a counterexample for the validity of some Chow type maximal inequalities for N-demimartingales was provided. This counterexample also applies to another well-known maximal inequality for N-demimartingales (see [7, Theorem 2.1] or [18, Theorem 3.2.1]), which could have been of special interest for the purposes of this paper, should it be valid.

The main objective of this work consists in developing some alternatives to the latter inequality. More precisely, after recalling the necessary preliminary notions and notation (see Section 2), in Section 3 some new maximal inequalities for nonnegative N-demi(super)martingales are derived. These inequalities are implemented, in Section 4, for obtaining bounds for the cumulative distribution function (CDF) of the waiting time for the first occurrence of a certain type of scan. In Section 4 we conclude with a numerical investigation of the behavior of the bounds.

Received 3 November 2015; revision received 31 October 2016.
* Postal address: Department of Statistics and Insurance Science, University of Piraeus, 80 Karaoli and Dimitriou Street, 185 34 Piraeus, Greece.
** Email address: mkoutras@unipi.gr
*** Email address: dilyber@webmail.unipi.gr

Note that scan statistics are defined as random variables enumerating the moving windows in a sequence of binary outcomes trials that contain a prescribed number of successes. The associated waiting time problems have proven to be of particular importance due to the widespread applicability of scan statistics in a substantial number of scientific areas such as actuarial science, reliability theory, and molecular biology; see, e.g. [3], [15], and [6], respectively. For a survey of the area, we refer the interested reader to Balakrishnan and Koutras [1] and to the edited volume of Glaz *et al.* [12]. In Pozdnyakov *et al.* [17] and Pozdnyakov and Steele [16], a martingale approach (for other approaches, see, e.g. [9], [10], and [19]) has been exploited for scan and pattern related problems. Despite this, there seems to be a gap in taking advantage of results related to the aforementioned generalizations of martingales to arrive at useful outcomes for scan statistics problems. This gap was the motivation for the last section of our work, and hopefully for a small contribution in addressing it.

2. Preliminaries

The notation $\mathbb{N}$ stands for the set of all positive integers, and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. The set of all real numbers is denoted by $\mathbb{R}$, while $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$. If $d \in \mathbb{N}$ then $\mathbb{R}^d$ denotes the Euclidean space of dimension d. Moreover, $x \wedge y := \min\{x, y\}$, $x \vee y := \max\{x, y\}$, and $x^+ := x \vee 0$ for $x, y \in \mathbb{R}$. For $n \in \mathbb{N}$ and $i \in \{1, \dots, n\}$, the i-canonical projection from $\mathbb{R}^n$ onto $\mathbb{R}$ is denoted by π_i.

Throughout what follows, we consider an arbitrary but fixed probability space $(\Omega, \Sigma, \mathbb{P})$.

By $\sigma(Z) := \{Z^{-1}(B) : B \in \mathfrak{B}\}$ we denote the σ-algebra generated by the Σ-measurable function Z, where $\mathfrak{B} := \mathfrak{B}(\mathbb{R})$ stands for the Borel σ-algebra of subsets of $\mathbb{R}$. On defining $T_Z := \{B \subseteq \mathbb{R} : Z^{-1}(B) \in \Sigma\}$ for any Σ-measurable function Z, it is clear that $\mathfrak{B} \subseteq T_Z$. We denote by $\mathbb{P}_Z : T_Z \to \mathbb{R}$ the *image measure of* $\mathbb{P}$ *under* Z. The restriction of $\mathbb{P}_Z$ to $\mathfrak{B}$ is denoted again by $\mathbb{P}_Z$, while R_Z stands for the range of Z. The notation $\text{bin}(n, p)$, where $n \in \mathbb{N}$ and $p \in (0, 1)$, stands for the law of the binomial distribution. Moreover, its probability mass and cumulative distribution function at point $x \in \mathbb{R}$ will be denoted by $b(x; n, p)$ and $F_b(x; n, p)$, respectively, while $M_b(\alpha, n, p) := \sum_{x=\alpha}^{n} x b(x; n, p)$ for each $\alpha \in \{1, \dots, n\}$.

A set $N \in \Sigma$ with $\mathbb{P}(N) = 0$ is called a $\mathbb{P}$-*null set*. A sequence $\{Z_j\}_{j \in \mathbb{N}}$ of Σ-measurable functions *satisfies a property* $\mathbb{P}$-*a.s.* ($\mathbb{P}$-almost surely) if there exists a $\mathbb{P}$-null set, say O, such that the property is satisfied by $\{Z_j\}_{j \in \mathbb{N}}$ for all $\omega \notin O$.

The family of all real-valued $\mathbb{P}$-integrable functions on Ω will be denoted by $\mathcal{L}^1(\mathbb{P})$. Functions that are $\mathbb{P}$-a.s. equal are not identified. The (unconditional) expectation of the random variable Z is denoted by $\mathbb{E}_{\mathbb{P}}[Z]$. If $Z \in \mathcal{L}^1(\mathbb{P})$ and $\mathcal{F}$ is a σ-subalgebra of Σ, then each function $\widetilde{Z} \in \mathcal{L}^1(\mathbb{P} \mid \mathcal{F})$ satisfying, for each $F \in \mathcal{F}$, the equality $\int_F Z \, \mathrm{d}\mathbb{P} = \int_F \widetilde{Z} \, \mathrm{d}\mathbb{P}$ is said to be a *version of the conditional expectation of* Z *given* $\mathcal{F}$, and is denoted by $\mathbb{E}_{\mathbb{P}}[Z \mid \mathcal{F}]$. Furthermore, for any $E \in \Sigma$ we set $\mathbb{P}(E \mid F) := \mathbb{E}_{\mathbb{P}}[\mathbf{1}_E \mid \mathcal{F}]$, where $\mathbf{1}_E$ stands for the indicator (or characteristic) function of E.

A family $\{\mathcal{F}_j\}_{j \in \mathbb{N}}$ of σ-subalgebras of Σ, such that $\mathcal{F}_j \subseteq \mathcal{F}_{j+1}$ for each $j \in \mathbb{N}$, is called a *filtration* for the measurable space (Ω, Σ). Moreover, a sequence $\{Z_j\}_{j \in \mathbb{N}}$ of random variables on Ω is said to be *adapted to a filtration* $\{\mathcal{F}_j\}_{j \in \mathbb{N}}$ if each Z_j is $\mathcal{F}_j$-measurable. If $\mathcal{F}_j = \sigma(\bigcup_{i=1}^{j} \sigma(Z_i))$ for each $j \in \mathbb{N}$ then $\{\mathcal{F}_j\}_{j \in \mathbb{N}}$ is said to be the *canonical filtration* for $\{Z_j\}_{j \in \mathbb{N}}$, and is denoted by $\{\mathcal{F}_j^{(Z)}\}_{j \in \mathbb{N}}$.

3. Maximal inequalities for nonnegative N-demimartingales

Some notions that are fundamental for the purposes of this section are first recalled.

https://doi.org/10.1017/jpr.2017.5 Published online by Cambridge University Press

Definition 1. Let $\{Z_j\}_{j\in\mathbb{N}}$ be a sequence in $\mathcal{L}^1(\mathbb{P})$. Then $\{Z_j\}_{j\in\mathbb{N}}$ is said to be

(i) a $\mathbb{P}$-martingale (with respect to $\{\mathcal{F}_j^{(Z)}\}_{j\in\mathbb{N}}$), if

$$\mathbb{E}_{\mathbb{P}}[(Z_{j+1}-Z_j)f(Z_1,\dots,Z_j)]=0 \quad \text{for each } j\in\mathbb{N} \tag{1}$$

and for every measurable function f on $\mathbb{R}^j$ such that the above expectations exist;

(ii) a $\mathbb{P}$-demimartingale, if condition (1) but with '$\geq$' in the place of the equality is satisfied for every coordinatewise nondecreasing function f on $\mathbb{R}^j$ such that the above expectations exist;

(iii) a $\mathbb{P}$-demisubmartingale, if condition (1) but with '$\geq$' in the place of the equality is satisfied for every f as in (ii) but with $f\geq 0$;

(iv) an N-demimartingale under $\mathbb{P}$, if condition (1) but with '$\leq$' in the place of the equality is satisfied for every f as in (ii). In particular, if $f\geq 0$ then $\{Z_n\}_{n\in\mathbb{N}}$ is said to be an N-demisupermartingale under $\mathbb{P}$.

From the definitions given above it is clear that the class of all $\mathbb{P}$-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 N-demimartingale is also an N-demisupermartingale. For more on Definitions 1 and the way that the notions given there are related to each other, we refer the interested reader to Prakasa Rao [18].

The next result is provided as an alternative to [7, Theorem 2.1].

Proposition 1. *If $\{Z_j\}_{j\in\mathbb{N}}$ is an N-demimartingale under $\mathbb{P}$ such that $Z_j\geq 0$ $\mathbb{P}$-a.s. for each $j\in\mathbb{N}$, then for any fixed $t\in\mathbb{N}$ and for each $\varepsilon>0$ the following inequality holds:*

$$\mathbb{P}\Big(\max_{1\leq j\leq t} Z_j>\varepsilon\Big)\leq 1-\frac{1}{\varepsilon}\mathbb{E}_{\mathbb{P}}[Z_1]+\frac{1}{\varepsilon}\sum_{m=1}^{t}\mathbb{E}_{\mathbb{P}}[Z_m\,\mathbf{1}_{\{Z_m>\varepsilon\}}]. \tag{2}$$

If, in addition, $\{Z_j\}_{j\in\mathbb{N}}$ is upper bounded by a positive number ζ then

$$\zeta\mathbb{P}\Big(\max_{1\leq j\leq t} Z_j\geq\varepsilon\Big)\geq \mathbb{E}_{\mathbb{P}}[Z_1]-\mathbb{E}_{\mathbb{P}}[Z_t\,\mathbf{1}_{\{\max_{1\leq j\leq t} Z_j<\varepsilon\}}].$$

Proof. First fix on arbitrary $t\in\mathbb{N}$ and $\varepsilon>0$. Define next the random variable $\tau_{t,\varepsilon}\colon\Omega\longrightarrow\mathbb{R}$ by means of

$$\begin{aligned}\tau_{t,\varepsilon}(\omega)&:=\tau_{t,\varepsilon}(Z_1,\dots,Z_t)(\omega)\\&=\begin{cases}\inf\{j\in\{1,\dots,t\}\colon Z_j(\omega)>\varepsilon\} & \text{if } \omega\in\bigcup_{j=1}^{t}\{Z_j>\varepsilon\},\\ t & \text{if } \omega\in\bigcap_{j=1}^{t}\{Z_j\leq\varepsilon\},\end{cases}\end{aligned} \tag{3}$$

for each $\omega\in\Omega$. Clearly, $\tau_{t,\varepsilon}$ is an $\mathcal{F}_t^{(Z)}$-measurable function and $R_{\tau_{t,\varepsilon}}=\{1,\dots,t\}$.

Moreover, it can be proven that $\tau_{t,\varepsilon}$ is a coordinatewise nonincreasing function of $Z_1,\dots,Z_t$ by distinguishing the following cases.

(i) Let $\omega\in\bigcup_{j=1}^{t}\{Z_j>\varepsilon\}$. Then we may claim that there exists a $j_1\in\{1,\dots,t\}$ such that $\inf\{j\in\{1,\dots,t\}\colon Z_j(\omega)>\varepsilon\}=j_1$, and so $\tau_{t,\varepsilon}(Z_1,\dots,Z_t)(\omega)=j_1$. Consider now

https://doi.org/10.1017/jpr.2017.5 Published online by Cambridge University Press

a sequence $\{\widehat{Z}_j\}_{j\in\mathbb{N}}$ of random variables on Ω such that

(a) $\widehat{Z}_{j_2}(\omega) > Z_{j_2}(\omega)$ for some $j_2 \in \{j_1, \dots, t\}$ and $\widehat{Z}_j(\omega) = Z_j(\omega)$ for all $j \neq j_2$. Then, we have

$$\tau_{t,\varepsilon}(\widehat{Z}_1, \dots, \widehat{Z}_{j_2}, \dots, \widehat{Z}_t)(\omega) = \tau_{t,\varepsilon}(Z_1, \dots, \widehat{Z}_{j_2}, \dots, Z_t)(\omega) = j_1; \tag{4}$$

(b) there exists a $j_3 \in \{1, \dots, j_1 - 1\}$ such that $\widehat{Z}_{j_3}(\omega) > Z_{j_3}(\omega)$ and $\widehat{Z}_j(\omega) = Z_j(\omega)$ for all $j \neq j_3$. If $\widehat{Z}_{j_3}(\omega) \le \varepsilon$ then condition (4), with j_3 in the place of j_2, holds, while if $\widehat{Z}_{j_3}(\omega) > \varepsilon$, we obtain

$$\tau_{t,\varepsilon}(\widehat{Z}_1, \dots, \widehat{Z}_{j_3}, \dots, \widehat{Z}_t)(\omega) = \tau_{t,\varepsilon}(Z_1, \dots, \widehat{Z}_{j_3}, \dots, Z_t)(\omega) = j_3 < j_1.$$

(ii) Let $\omega \in \bigcap_{j=1}^{t}\{Z_j \le \varepsilon\}$. Consider a sequence $\{\breve{Z}_j\}_{j\in\mathbb{N}}$ of random variables on Ω such that $\breve{Z}_{j_4}(\omega) > Z_{j_4}(\omega)$ for some $j_4 \in \{1, \dots, t\}$ and $\breve{Z}_j(\omega) = Z_j(\omega)$ for all $j \neq j_4$. If $\breve{Z}_{j_4}(\omega) > \varepsilon$ we obtain $\tau_{t,\varepsilon}(\breve{Z}_1, \dots, \breve{Z}_{j_4}, \dots, \breve{Z}_t)(\omega) = j_4 < t$, otherwise the equality $\tau_{t,\varepsilon}(\breve{Z}_1, \dots, \breve{Z}_{j_4}, \dots, \breve{Z}_t)(\omega) = t$ holds.

Consequently, $\mathbf{1}_{[0,t]}(\tau_{t,\varepsilon})$ is a coordinatewise nondecreasing function of $Z_1, \dots, Z_t$.

The latter, along with our assumption that $\{Z_j\}_{j\in\mathbb{N}}$ is an N-demimartingale under $\mathbb{P}$, yields $\mathbb{E}_{\mathbb{P}}[Z_1] \le \mathbb{E}_{\mathbb{P}}[Z_{\tau_{t,\varepsilon}\wedge t}]$ (see [13] or better, see, e.g. [18, Theorem 3.1.7]); therefore,

$$\mathbb{E}_{\mathbb{P}}[Z_1] \le \mathbb{E}_{\mathbb{P}}[Z_{\tau_{t,\varepsilon}}\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j > \varepsilon\}}] + \mathbb{E}_{\mathbb{P}}[Z_t\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j \le \varepsilon\}}]$$

and the following inequality ensues:

$$\varepsilon\mathbb{P}\Big(\max_{1\le j\le t} Z_j \le \varepsilon\Big) \ge \mathbb{E}_{\mathbb{P}}[Z_1] - \mathbb{E}_{\mathbb{P}}[Z_{\tau_{t,\varepsilon}}\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j > \varepsilon\}}]. \tag{5}$$

But since the last expectation is equal to

$$\sum_{m=1}^{t} \mathbb{E}_{\mathbb{P}}[Z_{\tau_{t,\varepsilon}}\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j > \varepsilon,\ \tau_{t,\varepsilon}=m\}}] = \sum_{m=1}^{t} \mathbb{E}_{\mathbb{P}}[Z_m\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j > \varepsilon,\ \tau_{t,\varepsilon}=m\}}]$$

and all random variables Z_n are $\mathbb{P}$-a.s. nonnegative, it follows that

$$\mathbb{E}_{\mathbb{P}}[Z_{\tau_{t,\varepsilon}}\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j > \varepsilon\}}] \le \sum_{m=1}^{t} \mathbb{E}_{\mathbb{P}}[Z_m\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j > \varepsilon,\ Z_m > \varepsilon\}}] = \sum_{m=1}^{t} \mathbb{E}_{\mathbb{P}}[Z_m\, \mathbf{1}_{\{Z_m > \varepsilon\}}].$$

Making use of the last inequality in (5), we deduce

$$\varepsilon - \varepsilon\mathbb{P}\Big(\max_{1\le j\le t} Z_j > \varepsilon\Big) \ge \mathbb{E}_{\mathbb{P}}[Z_1] - \sum_{m=1}^{t} \mathbb{E}_{\mathbb{P}}[Z_m\, \mathbf{1}_{\{Z_m > \varepsilon\}}];$$

thereof obtaining an alternative expression for (2).

Assume now, in addition, that there exists $\zeta > 0$ such that $Z_j \le \zeta$ for each $j \in \mathbb{N}$. Consider the random variable $\ddot{\tau}_{t,\varepsilon} \colon \Omega \longrightarrow \mathbb{R}$ defined as $\tau_{t,\varepsilon}$ in (3) but with $\ge$ and $<$ in the place of $>$ and $\le$, respectively.

Then following the same reasoning as in the proof of (2), we obtain

$$\mathbb{E}_{\mathbb{P}}[Z_1] - \mathbb{E}_{\mathbb{P}}[Z_t\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j < \varepsilon\}}] \le \mathbb{E}_{\mathbb{P}}[Z_{\ddot{\tau}_{t,\varepsilon}}\, \mathbf{1}_{\{\max_{1\le j\le t} Z_j \ge \varepsilon\}}] \le \zeta\mathbb{P}\Big(\max_{1\le j\le t} Z_j \ge \varepsilon\Big),$$

which completes the proof. □

Another alternative to Theorem 2.1 from [7] is next presented.

Proposition 2. *If $\{Z_j\}_{j\in\mathbb{N}}$ is an N-demisupermartingale under $\mathbb{P}$ such that $Z_j \geq 0$, $\mathbb{P}$-a.s. for each $j \in \mathbb{N}$, then for any fixed $t \in \mathbb{N}$ and for each $\varepsilon > 0$ the following inequality holds:*

$$\varepsilon\mathbb{P}\Big(\max_{1\leq j\leq t} Z_j \geq \varepsilon\Big) \leq \sum_{m=1}^{t} \mathbb{E}_{\mathbb{P}}[Z_m \mathbf{1}_{\{\max_{1\leq j\leq m} Z_j\geq\varepsilon\}}].$$

Proof. First fix on arbitrary $t \in \mathbb{N}$ and $\varepsilon > 0$. Define next the random variable $\widetilde{\tau}_{t,\varepsilon} : \Omega \longrightarrow \mathbb{R}$ by means of

$$\widetilde{\tau}_{t,\varepsilon}(\omega) := \widetilde{\tau}_{t,\varepsilon}(Z_1, \dots, Z_t)(\omega)$$
$$:= \begin{cases} \sup\{j \in \{1, \dots, t\}\colon Z_j(\omega) \geq \varepsilon\} & \text{if } \omega \in \bigcup_{j=1}^{t}\{Z_j \geq \varepsilon\}, \\ 1 & \text{if } \omega \in \bigcap_{j=1}^{t}\{Z_j < \varepsilon\}, \end{cases}$$

for each $\omega \in \Omega$. Clearly, $\widetilde{\tau}_{t,\varepsilon}$ is an $\mathcal{F}_t^{(Z)}$-measurable function and $R_{\widetilde{\tau}_{t,\varepsilon}} = \{1, \dots, t\}$.

Moreover, $\widetilde{\tau}_{t,\varepsilon}$ is a coordinatewise nondecreasing function of $Z_1, \dots, Z_t$. This can be verified by following similar arguments as those used for proving the corresponding assertion for $\tau_{t,\varepsilon}$ in the proof of Proposition 1. Consequently, $\mathbf{1}_{[0,t]}(\widetilde{\tau}_{t,\varepsilon})$ is a coordinatewise nonincreasing function of $Z_1, \dots, Z_t$.

The latter, together with our assumption that $\{Z_j\}_{j\in\mathbb{N}}$ is an N-demisupermartingale under $\mathbb{P}$, yields $\mathbb{E}_{\mathbb{P}}[Z_1] \geq \mathbb{E}_{\mathbb{P}}[Z_{\widetilde{\tau}_{t,\varepsilon}\wedge t}]$ (see [13] or better, see, e.g. [18, Theorem 3.1.7]), and so we obtain

$$\mathbb{E}_{\mathbb{P}}[Z_1] \geq \mathbb{E}_{\mathbb{P}}[Z_{\widetilde{\tau}_{t,\varepsilon}} \mathbf{1}_{\{\max_{1\leq j\leq t} Z_j\geq\varepsilon\}}] + \mathbb{E}_{\mathbb{P}}[Z_1 \mathbf{1}_{\{\max_{1\leq j\leq t} Z_j<\varepsilon\}}]$$

or, equivalently,

$$\mathbb{E}_{\mathbb{P}}[Z_1 \mathbf{1}_{\{\max_{1\leq j\leq t} Z_j\geq\varepsilon\}}] \geq \mathbb{E}_{\mathbb{P}}[Z_{\widetilde{\tau}_{t,\varepsilon}} \mathbf{1}_{\{\max_{1\leq j\leq t} Z_j\geq\varepsilon\}}].$$

But since all random variables Z_n are $\mathbb{P}$-a.s. nonnegative, it follows that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[Z_{\widetilde{\tau}_{t,\varepsilon}} \mathbf{1}_{\{\max_{1\leq j\leq t} Z_j\geq\varepsilon\}}] &\geq \mathbb{E}_{\mathbb{P}}[Z_t \mathbf{1}_{\{\max_{1\leq j\leq t} Z_j\geq\varepsilon\}\cap\{\max_{1\leq j\leq t-1} Z_j<\varepsilon,\, Z_t\geq\varepsilon\}}] \\ &\geq \varepsilon\mathbb{P}\Big(\max_{1\leq j\leq t-1} Z_j < \varepsilon, Z_t \geq \varepsilon\Big);\end{aligned}$$

hence, setting $\psi(t) := \psi(t; \varepsilon) := \mathbb{P}(\max_{1\leq j\leq t} Z_j \geq \varepsilon)$, we obtain

$$\varepsilon[\psi(t) - \psi(t-1)] \leq \mathbb{E}_{\mathbb{P}}[Z_1 \mathbf{1}_{\{\max_{1\leq j\leq t} Z_j\geq\varepsilon\}}] \quad \text{for each } t \in \mathbb{N},$$

since t was chosen arbitrarily. Then by induction our proposition follows. □

4. An application related to scan statistics

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

In what follows, we set $X_0 := 0$ and assume that every sum over an empty index set is equal to 0. For each $n \in \mathbb{N}$ and $k \in \mathbb{N}_0$, consider the random variable $Y_{n,k}$ on Ω defined by

$$Y_{n,k} := \sum_{j=\max\{n-k+1,1\}}^{n} X_j. \tag{6}$$

For any fixed $k \in \mathbb{N}$, the sequence $\{Y_{n,k}\}_{n\in\mathbb{N}}$ will be called a *scan enumerating process* of width k for the sequence of binary trials $\{X_n\}_{n\in\mathbb{N}}$. The random variable $T_r^{(k)}$, defined on Ω by means of

$$T_r^{(k)} := \min\{n \in \mathbb{N} \colon Y_{n,k} \geq r\},$$

is said to be *the waiting time for the first occurrence of a scan of type r/k.*

It is clear that for any fixed $k \in \mathbb{N}$ the sequence $\{Y_{n,k}\}_{n\in\mathbb{N}}$ is not $\mathbb{P}$-independent. On the other hand, it can be easily proven that for any fixed $k \in \mathbb{N}$ and for all $r, k, t \in \mathbb{N}$ with $r \leq k$, the equality $\{T_r^{(k)} \leq t\} = \{\max_{1\leq n\leq t} Y_{n,k} \geq r\}$ holds; therefore, it seems reasonable to wonder whether maximal inequalities can be exploited for obtaining some upper and lower bounds for the CDF $F_{r:k}(t;\, p) := \mathbb{P}(T_r^{(k)} \leq t)$; see also, e.g. [18, Chapters 2 and 3].

Motivated by the above question, the membership of $\{Y_{n,k}\}_{n\in\mathbb{N}}$, for $k \in \mathbb{N}$ fixed, in the classes of demi(sub)martingales and N-demi(super)martingales is first explored below and then a relevant N-demimartingale result is given.

In what follows, unless it is stated otherwise, we assume that $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathbb{P}$-independent and identically distributed ($\mathbb{P}$-i.i.d.) binary trials.

The next result is an immediate consequence of the definition of conditional expectation and the monotonicity of the functions used in the statement of the outcomes.

Lemma 1. *Let $k \in \mathbb{N}$ be arbitrary but fixed. For each $n \in \mathbb{N}$ with $n \geq k$, also let f be a coordinatewise nondecreasing real-valued function on $\mathbb{R}^n$ as well as $\{h_{i,k}\}_{i\in\{1,\dots,n\}}$ be a sequence of such functions on $\mathbb{R}^{i\wedge k}$. Then the following holds:*

$$\eta_{n,k,0}(h_{i,k};\, f) \leq \eta_{n,k,1}(h_{i,k};\, f),$$

where

$$\eta_{n,k,x}(h_{i,k};\, f) := \mathbb{E}_{\mathbb{P}}[f(h_1(X_1), \dots, h_n(X_{n-k+1}, \dots, X_n)) \mid \{X_{n-k+1} = x\}], \qquad x \in \{0, 1\}.$$

In particular,

$$\mathbb{E}_{\mathbb{P}}[f(Y_{1,k}, \dots, Y_{n,k}) \mid \{X_{n-k+1} = 0\}] \leq \mathbb{E}_{\mathbb{P}}[f(Y_{1,k}, \dots, Y_{n,k}) \mid \{X_{n-k+1} = 1\}].$$

It can be easily seen that the sequence $\{X_n\}_{n\in\mathbb{N}}$ is an N-demimartingale under $\mathbb{P}$ but it is not a $\mathbb{P}$-demisubmartingale. The latter can be also obtained as a special case of the next result.

Lemma 2. *Let $k \in \mathbb{N}$ be arbitrary but fixed. Then the sequence $\{Y_{n,k}\}_{n\in\mathbb{N}}$ is neither a $\mathbb{P}$-demisubmartingale nor an N-demisupermartingale under $\mathbb{P}$.*

Proof. First fix on an arbitrary $k \in \mathbb{N}$. Then note that, by (6), we have

$$Y_{n,k} - Y_{n-1,k} = X_n - X_{(n-k)^+} \quad \text{for each } n \in \mathbb{N}; \tag{7}$$

hence, for each $n \in \mathbb{N}$ and for every measurable function f on $\mathbb{R}^n$ such that each expectation

$$H_{n,k}(f) := H_{n,k}(Y_{1,k}, \dots, Y_{n,k};\, f) := \mathbb{E}_{\mathbb{P}}[(Y_{n+1,k} - Y_{n,k}) f(Y_{1,k}, \dots, Y_{n,k})]$$

exists, we obtain

$$H_{n,k}(f) = \mathbb{E}_{\mathbb{P}}[(p - X_{(n-k+1)^+}) f(Y_{1,k}, \dots, Y_{n,k})].$$

If in addition, f is coordinatewise nondecreasing, it follows by Lemma 1 that

$$H_{n,k}(f) = pq[\eta_{n,k,0}(h_{i,k}; f) - \eta_{n,k,1}(h_{i,k}; f)] \le 0 \quad \text{for each } n \in \mathbb{N} \text{ with } n \ge k, \tag{8}$$

where

$$h_{i,k}(x_{(i-k+1)\vee 1}, \dots, x_i) := \sum_{j=(i-k+1)\vee 1}^{i} x_j$$

for each $k \in \mathbb{N}$ and for each $(x_{(i-k+1)\vee 1}, \dots, x_i) \in \mathbb{R}^{i\wedge k}$, while it is immediate that

$$H_{n,k}(f) = p\mathbb{E}_{\mathbb{P}}[f(Y_{1,k}, \dots, Y_{n,k})] \ge 0 \quad \text{for each } n \in \mathbb{N} \text{ with } n < k. \tag{9}$$

Suppose now that $\{Y_{n,k}\}_{n\in\mathbb{N}}$ is a $\mathbb{P}$-demisubmartingale. It then follows that for each $n \in \mathbb{N}$ and for every nonnegative coordinatewise nondecreasing function f on $\mathbb{R}^n$ such that $H_{n,k}(f)$ exists, we have $H_{n,k}(f) \ge 0$. The latter, together with an application of condition (8) for $f = \pi_n$, yields $H_{n,k}(\pi_n) = 0$ for each $n \in \mathbb{N}$ with $n \ge k$. But since by assumption $\{X_n\}_{n\in\mathbb{N}}$ is $\mathbb{P}$-independent, it follows that

$$\mathbb{P}_{Y_{n,k}} = \operatorname{bin}(n \wedge k, p) \quad \text{for each } n \in \mathbb{N}, \tag{10}$$

implying, in conjunction with conditions (7) and (10), that

$$H_{n,k}(\pi_n) = \mathbb{E}_{\mathbb{P}}[(Y_{n+1,k} - Y_{n,k})Y_{n,k}] = \mathbb{E}_{\mathbb{P}}[X_{n+1}Y_{n,k}] - \mathbb{E}_{\mathbb{P}}[X_{n-k+1}Y_{n,k}] = -pq;$$

hence, $0 = H_{n,k}(\pi_n) = -pq$ for each $n \in \mathbb{N}$ with $n \ge k$, a contradiction. Thus, $\{Y_{n,k}\}_{n\in\mathbb{N}}$ cannot be a $\mathbb{P}$-demisubmartingale.

Moreover, suppose that $\{Y_{n,k}\}_{n\in\mathbb{N}}$ is an N-demisupermartingale under $\mathbb{P}$. Then applying similar arguments as above and considering $f = 1$ instead of $f = \pi_n$, we infer, by condition (9), that $0 = H_{n,k}(1) = p$ for each $n < k$, which is not valid; hence, $\{Y_{n,k}\}_{n\in\mathbb{N}}$ cannot be an N-demisupermartingale under $\mathbb{P}$ either. This completes the proof. □

As a result of Lemma 2, none of the maximal inequalities that are valid either for N-demi(super)martingales or for demi(sub)martingales can be exploited in the case of the enumerating process $\{Y_{n,k}\}_{n\in\mathbb{N}}$, where $k \in \mathbb{N}$ is arbitrary but fixed. To overcome this difficulty, consider, for any fixed $k \in \mathbb{N}$, the sequence $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}_0}$ of random variables on Ω defined by means of

$$\widetilde{Y}_{n,k} := \begin{cases} Y_{n,k} & \text{if } n \in \{k, k+1, \dots\}, \\ Y_{k,k} & \text{if } n \in \{1, \dots, k-1\}, \\ 0 & \text{if } n = 0, \end{cases} \tag{11}$$

and note that $R_{\widetilde{Y}_{n,k}} = \{0, \dots, k\}$ for each $n \in \mathbb{N}$.

Proposition 3. *For any fixed $k \in \mathbb{N}$, the sequence $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is an N-demimartingale under $\mathbb{P}$. Moreover, it is not a $\mathbb{P}$-demisubmartingale.*

Proof. First fix on an arbitrary $k \in \mathbb{N}$. Then note that the sequence $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is adapted to $\{\mathcal{F}_n^{(X)}\}_{n\in\mathbb{N}}$, since $\{Y_{n,k}\}_{n\in\mathbb{N}}$ is so by virtue of (6). Furthermore, for each $n \in \mathbb{N}$ and for every coordinatewise nondecreasing function f on $\mathbb{R}^n$ such that each expectation

$$\widetilde{H}_{n,k}(f) := \widetilde{H}_{n,k}(\widetilde{Y}_{1,k}, \ldots, \widetilde{Y}_{n,k}; f) := \mathbb{E}_{\mathbb{P}}[(\widetilde{Y}_{n+1,k} - \widetilde{Y}_{n,k}) f(\widetilde{Y}_{1,k}, \ldots, \widetilde{Y}_{n,k})]$$

exists, it follows, by (11), that

$$\widetilde{H}_{n,k}(f) = \begin{cases} H_{n,k}(f) & \text{if } n \in \{k, k+1, \ldots\}, \\ 0 & \text{if } n \in \{1, \ldots, k-1\}, \end{cases}$$

implying, together with condition (8), that $\widetilde{H}_{n,k}(f) \le 0$ for each $n \in \mathbb{N}$. The latter, in view of the fact that $\mathbb{E}_{\mathbb{P}}[\widetilde{Y}_{n,k}] = kp < \infty$ for each $n \in \mathbb{N}$, yields that the sequence $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is an N-demimartingale under $\mathbb{P}$.

Suppose now that $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is a $\mathbb{P}$-demisubmartingale as well. It then follows that $\widetilde{H}_{n,k}(f) = 0$ for each $n \in \mathbb{N}$ and for every nonnegative coordinatewise nondecreasing function f on $\mathbb{R}^n$ such that each expectation $\widetilde{H}_{n,k}(f)$ exists. But then, by the proof of Lemma 2, we infer, for each $n \in \mathbb{N}$ with $n \ge k$, that $0 = \widetilde{H}_{n,k}(\pi_n) = H_{n,k}(\pi_n) = -pq$, a contradiction. So, our statement that $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is not a $\mathbb{P}$-demisubmartingale follows. □

In the remainder of this section, we focus on illustrating how Propositions 1 and 2 can be exploited for obtaining bounds for the CDF of the waiting time $T_r^{(k)}$; see Corollaries 1 and 2, respectively. Note that bounds and approximations are widely used in the study of scans and runs as an extensive literature (including among others the fundamental contribution of Glaz and Naus [11], [4], and [5] as well as [1, Chapters 9 and 11], where an overview on this subject can be found) witnesses.

Corollary 1. *For any fixed $k \in \mathbb{N}$ and for all $r, t \in \mathbb{N}$ with $r \le k < t$, the following holds:*

$$\ell(r, k, p) \le F_{r:k}(t; p) \le u_1(r, k, t, p),$$

where

$$u_1(r, k, t, p) := 1 - \frac{kp}{r-1} + \frac{t}{r-1} M_b(r, k, p), \qquad \ell(r, k, p) := \frac{1}{k} M_b(r, k, p).$$

Proof. First fix on arbitrary $r, k, t \in \mathbb{N}$ with $r \le k < t$. We shall now proceed by carrying out the next three steps.

(i) We shall prove the inequality $F_{r\,:\,k}(t; p) \le u_1(r, k, t, p)$.

By virtue of Proposition 3 we may apply Proposition 1 for $\{Z_j\}_{j\in\mathbb{N}} = \{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ and $\varepsilon = r - 1$; hence, (2) becomes

$$(r-1)\mathbb{P}\Big(\max_{1\le n\le t} \widetilde{Y}_{n,k} \ge r\Big) \le (r-1) - \mathbb{E}_{\mathbb{P}}[\widetilde{Y}_{1,k}] + \sum_{n=1}^{k} \mathbb{E}_{\mathbb{P}}[\widetilde{Y}_{n,k}\, \mathbf{1}_{\{\widetilde{Y}_{n,k} \ge r\}}] + \sum_{n=k+1}^{t} \mathbb{E}_{\mathbb{P}}[\widetilde{Y}_{n,k}\, \mathbf{1}_{\{\widetilde{Y}_{n,k} \ge r\}}],$$

which together with conditions (10) and (11) as well as the fact that

$$\{T_r^{(k)} \le t\} = \Big\{\max_{k\le n\le t} Y_{n,k} \ge r\Big\} = \bigcup_{n=k}^{t} \{Y_{n,k} \ge r\}, \tag{12}$$

https://doi.org/10.1017/jpr.2017.5 Published online by Cambridge University Press

yields

$$F_{r\,:\,k}(t;\,p) \le 1 - \frac{kp}{r-1} + \frac{t}{r-1}\sum_{y=r}^{k} y b(y;\,k,\,p).$$

(ii) If we set $\varpi_{r,k,t}(y) := \mathbb{P}(Y_{t,k} = y, \max_{k\le n\le t} Y_{n,k} \le r-1)$ for $y \in \mathbb{N}_0$ then

$$\varpi_{r,k,t}(y) \le b(y;\,k,\,p) \quad \text{for each } y \in \{0,\dots,r-1\}.$$

For proving the above fact, note that since $Y_{n,k} = \sum_{j=1}^{n}(Y_j - Y_{j-1,k})$ for each $n \in \mathbb{N}$, we obtain, by (7),

$$Y_{m,k} - Y_{n,k} = \sum_{j=n+1}^{m}(X_j - X_{(j-k)^+}) = \sum_{j=n+1}^{m} X_j - \sum_{j=n-k+1}^{m-k} X_{j^+} \tag{13}$$

for each $n, m \in \mathbb{N}$ such that $m > n$. So, by the $\mathbb{P}$-independence of the binary trials $\{X_n\}_{n\in\mathbb{N}}$, it follows, for each $y \in \{0,\dots,r-1\}$, that

$$\begin{aligned}
\varpi_{r,k,t}(y) &= \mathbb{P}\biggl(Y_{t,k} = y,\ \max_{k\le n\le t}\biggl(\sum_{j=n-k+1}^{t-k} X_j - \sum_{j=n+1}^{t} X_j\biggr) \le r-1-y\biggr)\\
&\le \mathbb{P}\biggl(Y_{t,k} = y,\ \max_{k\le n\le t}\biggl(\sum_{j=n-k+1}^{t-k} X_j - t + k\biggr) \le r-1-y\biggr)\\
&\le \mathbb{P}\biggl(Y_{t,k} = y, \sum_{j=1}^{t-k} X_j \le t-k+r-1-y\biggr)\\
&= b(y;\,k,\,p)F_b(t-k+r-1-y;\,t-k,\,p)\\
&= b(y;\,k,\,p),
\end{aligned}$$

which completes (ii).

(iii) Finally, we shall prove that $F_{r:k}(t;\,p) \ge \ell(r,k,p)$.

Because of Proposition 3 and the fact that the process $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is upper bounded by k, we may apply the 'in addition' part of Proposition 1 for $\{Z_j\}_{j\in\mathbb{N}} = \{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ and $(\varepsilon,\zeta) = (r,k)$ to obtain

$$k\mathbb{P}\Bigl(\max_{1\le n\le t} \widetilde{Y}_{n,k} \ge r\Bigr) \ge \mathbb{E}_{\mathbb{P}}[\widetilde{Y}_{1,k}] - \mathbb{E}_{\mathbb{P}}[\widetilde{Y}_{t,k}\,\mathbf{1}_{\{\max_{1\le n\le t}\widetilde{Y}_{n,k}\le r-1\}}],$$

which together with conditions (10)–(12), yields

$$F_{r\,:\,k}(t;\,p) \ge p - \frac{1}{k}\mathbb{E}_{\mathbb{P}}[Y_{t,k}\,\mathbf{1}_{\{\max_{k\le n\le t} Y_{n,k}\le r-1\}}].$$

But since $\mathbb{E}_{\mathbb{P}}[Y_{t,k}\,\mathbf{1}_{\{\max_{k\le n\le t} Y_{n,k}\le r-1\}}] = \sum_{y=1}^{r-1} y\varpi_{r,k,t}(y)$, we obtain, by (ii),

$$\mathbb{E}_{\mathbb{P}}[Y_{t,k}\,\mathbf{1}_{\{\max_{k\le n\le t} Y_{n,k}\le r-1\}}] \le \sum_{y=1}^{r-1} y b(y;\,k,\,p) = kp - M_b(r,k,p),$$

and so (iii) follows.

Steps (i) and (iii) prove the corollary. □

We shall next derive three additional upper bounds for $F_{r;k}(t;\,p)$. In order to arrive at these we need to prove first the following lemma.

https://doi.org/10.1017/jpr.2017.5 Published online by Cambridge University Press

Lemma 3. *For any fixed $k \in \mathbb{N}$, for all $l, m \in \mathbb{N}$ such that $l \geq m \geq k$ and for each $B_m \in \mathfrak{B}$, the quantities $d_{l;m} := d_{l;m}(Y_{l,k}; Y_{m,k}, B_m) := \mathbb{E}_{\mathbb{P}}[Y_{l,k} \mathbf{1}_{Y_{m,k}^{-1}(B_m)}]$ satisfy the following equality:*

$$d_{l;m} = d_{m;m} + [(m-l) \wedge k] p \sum_{y \in B_m} [b(y; k, p) - b(y-1; k-1, p)].$$

Proof. First note that the validity of the above equality for $m = l$ is obvious. Fix now on arbitrary $k, l \in \mathbb{N}$ such that $k \leq l$.

Condition (13) yields, for each $m \in \mathbb{N}$ with $m > l$ and for each $B_m \in \mathfrak{B}$,

$$d_{l;m} = d_{m;m} + \sum_{j=l-k+1}^{m-k} \mathbb{E}_{\mathbb{P}}[X_j \mathbf{1}_{Y_{m,k}^{-1}(B_m)}] - \sum_{j=l+1}^{m} \mathbb{E}_{\mathbb{P}}[X_j \mathbf{1}_{Y_{m,k}^{-1}(B_m)}]. \tag{14}$$

By the $\mathbb{P}$-independence of $\{X_n\}_{n \in \mathbb{N}}$, we obtain, for each $m \in \mathbb{N}$ with $m > l$ and for each $B_m \in \mathfrak{B}$,

$$\sum_{j=l-k+1}^{m-k} \mathbb{E}_{\mathbb{P}}[X_j \mathbf{1}_{Y_{m,k}^{-1}(B_m)}] = (m-l) p \mathbb{P}_{Y_{m,k}}(B_m); \tag{15}$$

in addition, for each $j \in \{m-k+1, \dots, m\}$, we may write

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[X_j \mathbf{1}_{Y_{m,k}^{-1}(B_m)}] &= p \mathbb{P}\Big(\Big\{1 + \sum_{j \neq i = m-k+1}^{m} X_i \in B_m\Big\} \,\Big|\, \{X_j = 1\}\Big) \\ &= p \sum_{y \in B_m} b(y-1; k-1, p). \end{aligned} \tag{16}$$

Also note that, for each $m \in \mathbb{N}$ with $m > l$ and for each $B_m \in \mathfrak{B}$, the second sum of (14) can be rewritten as

$$\sum_{j=1+(m-k)\vee l}^{m} \mathbb{E}_{\mathbb{P}}[X_j \mathbf{1}_{Y_{m,k}^{-1}(B_m)}] + \sum_{j=l+1}^{(m-k)\vee l} \mathbb{E}_{\mathbb{P}}[X_j \mathbf{1}_{Y_{m,k}^{-1}(B_m)}];$$

hence, by virtue of (16) and the $\mathbb{P}$-independence of $\{X_n\}_{n \in \mathbb{N}}$, we infer that the sum

$$\sum_{j=l+1}^{m} \mathbb{E}_{\mathbb{P}}[X_j \mathbf{1}_{Y_{m,k}^{-1}(B_m)}]$$

reduces to $(m-l-k)^+ p \mathbb{P}_{Y_{m,k}}(B_m) + [(m-l) \vee k] p \sum_{y \in B_m} b(y-1; k-1, p)$.

The latter together with conditions (14) and (15) yields the requested equality. □

Corollary 2. *For any fixed $k \in \mathbb{N}$ and for all $r, t \in \mathbb{N}$ with $r \leq k < t$, the CDF $F_{r\,:\,k}(t; p)$ is upper bounded by each of the following three functions:*

$$\begin{aligned} u_2(r, k, t, p) &:= \frac{t}{r} M_b(r, k, p) \\ &\quad + \frac{1}{r} \sum_{m=k+1}^{t} \sum_{y=r+k-m}^{r-1} y b(y; k, p)[1 - F_b(r-1-y; m-k, p)], \end{aligned}$$

$$u_3(r, k, t, p) := \frac{k}{r} M_b(r, k, p) + \frac{(t-k)(t-k+3)k}{2r}[1 - F_b(r-1; k, p)],$$

and

$$u_4(r,k,t,p) := \Big[\frac{k}{r} + \frac{(t-k)(t-k+3)}{2r}\Big] M_b(r,k,p) + \frac{p}{r}[F_b(r-2;k-1,p) - F_b(r-1;k,p)] \sum_{m=k+1}^{t}\sum_{n=k}^{m}[(n-k)\wedge k].$$

Proof. First fix on arbitrary $r,k,t \in \mathbb{N}$ with $r \le k < t$.

Since by Proposition 3 the sequence $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is an N-demimartingale, it is also an N-demisupermartingale; hence, we may apply Proposition 2 to obtain the inequality

$$r\mathbb{P}\Big(\max_{1\le n\le t} \widetilde{Y}_{n,k} \ge r\Big) \le \sum_{m=1}^{t} \mathbb{E}_{\mathbb{P}}[\widetilde{Y}_{1,k}\,\mathbf{1}_{\{\max_{1\le n\le m}\widetilde{Y}_{n,k}\ge r\}}],$$

which taking into account conditions (10)–(12), entails that

$$rF_{r:k}(t;p) \le \sum_{m=1}^{k} \mathbb{E}_{\mathbb{P}}[Y_{k,k}\,\mathbf{1}_{\{Y_{k,k}\ge r\}}] + \sum_{m=k+1}^{t} \mathbb{E}_{\mathbb{P}}[Y_{k,k}\,\mathbf{1}_{\{\max_{k\le n\le m} Y_{n,k}\ge r\}}]$$

or, equivalently,

$$rF_{r:k}(t;p) \le k\sum_{y=r}^{k} yb(y;k,p) + \sum_{m=k+1}^{t} \mathbb{E}_{\mathbb{P}}[Y_{k,k}\,\mathbf{1}_{\{\max_{k\le n\le m} Y_{n,k}\ge r\}}]. \tag{17}$$

Next we bound the expectation

$$U(r,k,m,p) := \mathbb{E}_{\mathbb{P}}[Y_{k,k}\,\mathbf{1}_{\{\max_{k\le n\le m} Y_{n,k}\ge r\}}]$$

from above in three different ways.

(i) Set $\widetilde{\varpi}_{r,k,m}(y) := \mathbb{P}(Y_{k,k} = y, \max_{k\le n\le m} Y_{n,k} \ge r)$ for any fixed $m \in \mathbb{N}$ with $m > k$ and for each $y \in \mathbb{N}_0$. Then condition (13) together with (6) and the $\mathbb{P}$-independence of the binary trials $\{X_n\}_{n\in\mathbb{N}}$, yields, for each $y \in \{0,\dots,k\}$,

$$\widetilde{\varpi}_{r,k,m}(y) \le \mathbb{P}\Big(Y_{k,k} = y, \max_{k\le n\le m}\sum_{j=k+1}^{n} X_j \ge r-y\Big) = b(y;k,p)[1-F_b(r-1-y;m-k,p)].$$

Therefore,

$$U(r,k,m,p) \le \sum_{y=r+k-m}^{r-1} yb(y;k,p)[1-F_b(r-1-y;m-k,p)] + M_b(r,k,p),$$

which in turn yields

$$\sum_{m=k+1}^{t} U(r,k,m,p) \le \sum_{m=k+1}^{t}\sum_{y=r+k-m}^{r-1} yb(y;k,p)[1-F_b(r-1-y;m-k,p)] + (t-k)M_b(r,k,p). \tag{18}$$

(ii) Since $\{\widetilde{Y}_{n,k}\}_{n\in\mathbb{N}}$ is upper bounded by k, it follows, by (12), that

$$\begin{aligned}\sum_{m=k+1}^{t} U(r,k,m,p) &\le \sum_{m=k+1}^{t} k\mathbb{P}\bigg(\bigcup_{n=k}^{m}\{Y_{n,k}\ge r\}\bigg)\\ &\le \sum_{m=k+1}^{t}\sum_{n=k}^{m}\mathbb{P}(Y_{n,k}\ge r)\\ &\overset{(6)}{=} \sum_{m=k+1}^{t}(m-k+1)[1-F_b(r-1;k,p)]\\ &= [1-F_b(r-1;k,p)]\frac{(t-k)(t-k+3)}{2}. \end{aligned} \tag{19}$$

(iii) Setting now $d_{l;m}(r) := d_{l;m}(Y_{l,k}; Y_{m,k}, [r,\infty)) = \mathbb{E}_{\mathbb{P}}[Y_{l,k}\,\mathbf{1}_{\{Y_{m,k}\ge r\}}]$ for all positive integers l, m in $\{k, k+1, \dots\}$, we obtain

$$U(r,k,m,p) = \mathbb{E}_{\mathbb{P}}[Y_{k,k}\,\mathbf{1}_{\bigcup_{n=k}^{m}\{Y_{n,k}\ge r\}}] \le \sum_{n=k}^{m} d_{k;n}(r) \tag{20}$$

for all $m, n \in \mathbb{N}$ with $m \ge n \ge k$.

Applying Lemma 3 and taking into account the clear facts that $d_{n;n}(r) = M_b(r,k,p)$ for each $n\in\mathbb{N}$, and $\sum_{y=r}^{k} b(y-1;k-1,p) = 1-F_b(r-2;k-1,p)$, we obtain, for all $m,n\in\mathbb{N}$ with $m\ge n\ge k$,

$$d_{k;n}(r) = M_b(r,k,p) + p[F_b(r-2;k-1,p) - F_b(r-1;k,p)][(n-k)\wedge k],$$

implying

$$\sum_{n=k}^{m} d_{k;n}(r) = (m-k+1)M_b(r,k,p) + p[F_b(r-2;k-1,p) - F_b(r-1;k,p)][(n-k)\wedge k].$$

So, it follows, by (20), that the sum $\sum_{m=k+1}^{t} U(r,k,m,p)$ is upper bounded by

$$\frac{(t-k)(t-k+3)}{2}M_b(r,k,p) + p[F_b(r-2;k-1,p) - F_b(r-1;k,p)]\sum_{m=k+1}^{t}\sum_{n=k}^{m}[(n-k)\wedge k].$$

The latter, in conjunction with (17), yields $F_{r\,:\,k}(t;p) \le u_4(r,k,t,p)$. Moreover, condition (17) together with (18) and (19) implies that the CDF $F_{r\,:\,k}(t;p)$ is upper bounded by $u_2(r,k,t,p)$ and $u_3(r,k,t,p)$, respectively. This completes the proof. □

The following result is an immediate consequence of Corollaries 1 and 2.

Proposition 4. *For any fixed $k\in\mathbb{N}$ and for all $r,t\in\mathbb{N}$ such that $r\le k<t$, the following holds true:*

$$\ell(r,k,p) \le F_{r\,:\,k}(t;p) \le u_\star(r,k,t,p) := \min_{i\in\{1,2,3,4\}} u_i(r,k,t,p),$$

where the functions ℓ, u_1 and u_2, u_3, u_4 are as in Corollaries 1 and 2, respectively.

Remark 1. A first look at Proposition 4 suggests that the lower bound $\ell(r,k,p)$ seems to be of restricted practical value since, due to the fact that it is independent of t, its performance is not expected to be good for all t. Moreover, based on the following facts one might argue that $u_4(r, k, t, p)$ is expected to outperform the other three upper bounds.

(i) Condition $\lim_{p\to 0^+} F_{r:k}(t; p) \le \lim_{p\to 0^+} u_i(r, k, t, p) = 0$ holds for $i \in \{2, 4\}$ but not for $i \in \{1, 3\}$. In fact, $\lim_{p\to 0^+} u_3(r, k, t, p) = (t-k)(t-k+3)(k/2r)$, while $u_3(r, k, t, p)$ tends to 1 as p decreases to 0, fully opposing the intuitively natural fact that as the success probability p of each binary trial tends to 0, the probability $\mathbb{P}(T_r^{(k)} \le t)$ should do so as well. Also note that $\lim_{p\to 0^+} \ell(r, k, p) = 0$.

(ii) The upper bounds $u_1(r, k, t, p)$ and $u_3(r, k, t, p)$ may be used for all other values of p if they do not differ much from the other two bounds, since they are very easily attainable. On the other hand, the computation of $u_2(r, k, t, p)$ and $u_4(r, k, t, p)$ becomes quite cumbersome as t increases.

(iii) The upper bounds $u_2(r, k, t, p)$ and $u_4(r, k, t, p)$ seem to be more appropriate than the other two, not only because of (i) but also because they are more sophisticated as the proof of Proposition 4 reveals. Note that a key element in that proof is the technique applied for bounding from above the expectation $\mathbb{E}_{\mathbb{P}}[Y_{k,k}\,\mathbf{1}_{\{\max_{k\le n\le m} Y_{n,k}\ge r\}}]$. Clearly, the technique used for obtaining $u_3(r, k, t, p)$ is the simplest one (see (ii) of the above proof). On the contrary, $u_2(r, k, t, p)$ and especially $u_4(r, k, t, p)$ is extracted by less 'naive' approaches. More precisely, in the case of $u_2(r, k, t, p)$, i.e. in (i) of the same proof, we follow a reasoning similar to the one applied in Corollary 1, while for establishing $u_4(r, k, t, p)$ the implementation of Lemma 3 was essential.

(iv) The first three upper bounds possess an additional disadvantage: they are all nondecreasing functions of t, something that cannot be claimed for $u_4(r, k, t, p)$.

Concluding this section, we deem it necessary to stress that its main objective was to serve as an illustration for the applicability of the new maximal inequalities provided in Propositions 1 and 2, which are of independent interest. The bounds given in Proposition 4 are of some value only if a rough estimate of $F_{r:k}(t; p)$ is needed and the computation complexity of obtaining that is of major importance. Otherwise, one may resort to a bundle of more efficient bounds that are available in the literature (see, e.g. [11]).

In Tables 1–3, the computed values for the bounds of $F_{r:k}(t; p)$ provided in Proposition 4, together with the exact or simulated value of the CDF of $T_r^{(k)}$ (depending on the examined type of scan) are presented. Note that for *almost perfect runs*, i.e. for scans of type either $(k-1)/k$ or $(k-2)/k$, the exact CDF is computed by exploiting Proposition 1 of Bersimis *et al.* [2]; otherwise, the computation of the empirical CDF of $T_r^{(k)}$ is performed via simulation. More precisely, a sequence of i.i.d. binary trials with success probability p is generated each time and the number of trials needed until a scan of type r/k appears is recorded; the number of simulated sequences used for calculations was 1000.

Moreover, the values in **bold** denote the tightest upper bound (at each t), which is also recorded in the sixth column of each table. In the last column we have computed the *relative error*

$$\frac{u_\star(r, k, t, p) - \kappa(r, k, t, p)}{\kappa(r, k, t, p)}$$

for the upper bound (r.e.u. for short), where $\kappa(r, k, t, p)$ is the exact or simulated value of the CDF (sim. for short). Since the lower bound $\ell(r, k, p)$ does not depend on t, its value is given in the caption of each table.

https://doi.org/10.1017/jpr.2017.5 Published online by Cambridge University Press

TABLE 1: Computed bounds versus exact CDF: $(r, k, p) = (6, 7, 0.1)$; $\ell(6, 7, 0.1) = 5.5 \times 10^{-6}$.

t	Exact	u_1	u_2	u_3	u_4	$u_\star$	r.e.u.
8	0.000 011	0.860 062	0.000 066	0.000 583	**0.000 057**	0.000 057	4.057
13	0.000 038	0.860 100	0.001 240	0.007 303	**0.000 173**	0.000 173	3.977
18	0.000 058	0.860 139	0.009 118	0.020 743	**0.000 324**	0.000 324	4.561
23	0.000 082	0.860 177	0.034 884	0.040 903	**0.000 493**	0.000 493	5.039
28	0.000 105	0.860 216	0.092 585	0.067 783	**0.000 681**	0.000 681	5.481
33	0.000 129	0.860 254	0.195 339	0.101 383	**0.000 888**	0.000 888	5.907
38	0.000 152	0.860 293	0.352 587	0.141 703	**0.001 113**	0.001 113	6.324
43	0.000 175	0.860 331	0.568 994	0.188 743	**0.001 357**	0.001 357	6.736
48	0.000 199	0.860 370	0.844 688	0.242 503	**0.001 620**	0.001 620	7.145
53	0.000 222	0.860 408	1.176 250	0.302 983	**0.001 901**	0.001 901	7.551

TABLE 2: Computed bounds versus exact CDF: $(r, k, p) = (4, 6, 0.15)$; $\ell(4, 6, 0.15) = 0.004$.

t	Exact	u_1	u_2	u_3	u_4	$u_\star$	r.e.u.
7	0.008 994	0.755 885	**0.046 577**	0.3184	0.047 124	0.046 577	4.179
12	0.023 292	0.795 803	0.208 762	3.8495	**0.154 067**	0.154 067	5.615
17	0.037 355	0.835 721	0.604 470	1.9117	**0.301 887**	0.301 887	7.082
22	0.051 214	0.875 638	1.230 840	21.5050	**0.482 810**	0.482 810	8.427
27	0.064 874	0.915 556	2.038 740	35.6294	**0.696 837**	0.696 837	9.742
32	0.078 337	0.955 474	2.972 430	53.2848	**0.943 969**	0.943 969	11.050
37	0.091 606	**0.995 392**	3.986 510	74.4714	1.224 200	0.995 392	9.866
⋮							
42	0.104 684	1.035 310	5.049 160	99.1890	1.537 540	1.035 310	8.890
47	0.117 574	1.075 230	6.139 950	127.4380	1.883 990	1.075 230	8.145
52	0.130 278	1.115 150	7.246 560	159.2180	2.263 530	1.115 150	7.559

TABLE 3: Computed bounds versus simulated CDF: $(r, k, p) = (14, 28, \frac{1}{4})$; $\ell(14, 28, \frac{1}{4}) = 0.002$.

t	Sim.	u_1	u_2	u_3	u_4	$u_\star$	r.e.u.
30	0.046	0.587 144	**0.122 368**	7.520	0.128 012	0.122 368	1.660
32	0.047	0.595 518	**0.147 594**	2.861	0.161 861	0.147 594	2.140
33	0.049	0.599 705	**0.166 470**	29.755	0.184 118	0.166 470	2.397
37	0.052	0.616 453	**0.305 068**	8.153	0.307 031	0.305 068	4.867
39	0.053	0.624 826	0.426 376	114.246	**0.387 820**	0.387 820	6.317
41	0.054	0.633 200	0.593 177	154.268	**0.480 737**	0.480 737	7.903
45	0.055	**0.649 947**	1.091 080	252.100	0.700 671	0.649 947	10.817
47	0.056	**0.658 321**	1.432 420	309.910	0.826 547	0.658 321	10.756
50	0.057	**0.670 882**	2.067 290	407.741	1.033 640	0.670 882	10.767
52	0.061	**0.679 255**	2.572 470	48.374	1.183 060	0.679 255	10.135
54	0.062	**0.687 629**	3.141 010	558.936	1.340 900	0.687 629	10.091

Summarizing the conclusions of our numerical study, we may state the following.

- *For the case of almost perfect runs*, the r.e.u. usually increases as t does so and the other three parameters r, k, p are kept fixed. However, for some time intervals the r.e.u. may behave as follows: up to a trial, say t_1, increases with t, then it becomes a nonincreasing

https://doi.org/10.1017/jpr.2017.5 Published online by Cambridge University Press

function of t until another trial, say t_2, and so forth (see Tables 1 and 2); in other words, its graph versus time may resemble a multimodal function. Concerning now the way the r.e.u. evolves as p varies and r, k, t remain fixed, we mention that we obtain mixed messages for different values of t as far as scans of type $(k-2)/k$ are concerned. However, it should be mentioned that for extremely small values of p (i.e. for $p < 0.01$) the r.e.u. behaves as a nondecreasing function of p. The same applies for the other type of almost perfect runs and for small values of p ($p \leq 0.15$). Our numerical evidence (see Tables 1 and 2) indicates that again, for small values of p, the tightest upper bound is attained by $u_4(r, k, t, p)$, verifying in this way the main assertion of Remark 1. As far as the relative error for $\ell(r, k, p)$ is concerned, it becomes smaller as p increases, while it tends to 1 as t increases. The above facts are confirmed not only by the entries in Tables 1 to 2 but also by a similar numerical study conducted for the 3/5-, 4/5-, and 5/6-almost perfect runs as well as for $(r, k, p) = (6, 7, 0.15)$.

- *For other types of scan*, the comments made in the previous paragraph do not apply in general. In fact, in many cases, even the tightest of the upper bounds produces values greater than 1 (this may happen since all bounds of Proposition 4 do not correspond to probabilities). The scan of type 4/14 with $p = 0.1$ (as well as the 4/6-almost perfect run for $p = 0.15$) is such a case. However, exceptions as the one presented in Table 3 may occur. It is also worth noting that for scans other than almost perfect runs the simplest bounds to compute, i.e. $u_1(r, k, t, p)$ and $u_3(r, k, t, p)$, are usually proved to be more efficient than the other available upper bounds.

Acknowledgements

D. P. L. would like to dedicate this work in memory of his father Panagiotis (Takis) who recently passed away. His support during this research endeavor is one of the many moving memories which the co-author will always recall, full of love and gratitude.

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the operational program 'Education and Lifelong Learning' of the National Strategic Reference Framework research funding program: Aristeia II. Investing in knowledge society through the European Social Fund.

References

[1] Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications.* John Wiley, New York.

[2] Bersimis, S., Koutras, M. V. and Papadopoulos, G. K. (2014). Waiting time for an almost perfect run and applications in statistical process control. *Methodology Comput. Appl. Prob.* **16,** 207–222.

[3] Boutsikas, M. V. and Koutras, M. V. (2002). Modeling claim exceedances over thresholds. *Insurance Math. Econom.* **30,** 67–83.

[4] Boutsikas, M. V. and Koutras, M. V. (2006). On the asymptotic distribution of the discrete scan statistic. *J. Appl. Prob.* **43,** 1137–1154.

[5] Chen, J. and Glaz, J. (1999). Approximations for the distribution and the moments of discrete scan statistics. In *Scan Statistics and Applications*, eds. J. Glaz and N. Balakrishnan, Birkhäuser, Boston, MA, pp. 27–66.

[6] Chen, C. and Karlin, S. (2007). r-scans statistics of a Poisson process with events transformed by duplications, deletions, and displacements. *Anal. Appl. Prob.* **39,** 799–825.

[7] Christofides, T. C. (2003). Maximal inequalities for N-demimartingales. *Arch. Inequal. Appl.* **1,** 387–397.

[8] Dai, P., Shen, Y., Hu, S. and Yang, W. (2014). Some results for demimartingales and N-demimartingales. *J. Inequal. Appl.* **2014:489**, 12pp.

[9] Fu, J. C. (2001). Distribution of the scan statistic for a sequence of bistate trials. *J. Appl. Prob.* **38,** 908–916.

[10] Fu, J. C., Wu, T.-L. and Lou, W. Y. W. (2012). Continuous, discrete, and conditional scan statistics. *J. Appl. Prob.* **49,** 199–209.

[11] Glaz, J. and Naus, J. I. (1991). Tight bounds and approximations for scan statistic probabilities for discrete data. *Ann. Appl. Prob.* **1,** 306–318.
[12] Glaz, J., Pozdnyakov, V. and Wallenstein, S. (2009). *Scan Statistics: Methods and Applications.* Birkhäuser, Boston.
[13] Hu, S. H., Yang, W. Z., Wang, X. J. and Shen, Y. (2010). A note on the inequalities for N-demimartingales and demimartingales. *J. Systems Sci. Math. Sci.* **30,** 1052–1058 (in Chinese).
[14] Newman, C. M. and Wright, A. L. (1982). Associated random variables and martingale inequalities. *Z. Wahrscheinlichkeitsth.* **59,** 361–371.
[15] Papastavridis, S. G. and Koutras, M. V. (1993). Bounds for reliability of consecutive k-within-m-out-of-n:F systems. *IEEE Trans. Reliab.* **42,** 156–160.
[16] Pozdnyakov, V. and Steele, J. M. (2009). Martingale methods for patterns and scan statistics. In *Scan Statistics: Methods and Applications*, eds. J. Glaz *et al.* Birkhäuser, Boston, MA, pp. 289–317.
[17] Pozdnyakov, V., Glaz, J., Kulldorff, M. and Steele, J. M. (2005). A martingale approach to scan statistics. *Ann. Inst. Statist. Math.* **57,** 21–37.
[18] Prakasa Rao, B. L. S. (2012). *Associated Sequences, Demimartingales and Nonparametric Inference.* Birkhäuser/Springer, Basel.
[19] Wu, T.-L., Glaz, J. and Fu, J. C. (2013). Discrete, continuous and conditional multiple window scan statistics. *J. Appl. Prob.* **50,** 1089–1101.