On Waiting Time Distributions for the Occurrence of Patterns

Spiros D. Dafnis and Andreas N. Philippou *

Abstract—**In the present paper we study the waiting time distributions of patterns of two successes separated by at most or exactly $k-2$ failures ($k \geq 3$) in the case of first-order dependent trials. Employing both non-overlapping and overlapping counting schemes we obtain closed formulas for the probability generating functions and effective recursive schemes for the evaluation of the probability mass functions of the waiting time random variables. Finally, we give applications of our results to the moving window detection problem and a biomedical engineering one.**

Keywords: string, Markov chain embedding, probability mass function, moving window detection, biomedical engineering

1. Introduction

Recent research in applied probability has focused on the waiting time distributions for the r-th ($r \geq 1$) appearance of patterns. Such distributions are applicable to DNA matching (see [14]), meteorology and agriculture (see [5]), ecology, psychology, statistical quality control, start-up demonstration tests and the moving window detection problem (see [15] and [3]). The latter applications are related to the appearance of two successes separated by a string of failures.

For any given integer $k \geq 3$ we consider the patterns $\mathcal{E}_1$: two successes are separated by at most $k-2$ failures and $\mathcal{E}_2$: two successes are separated by exactly $k-2$ failures. Let $\{Z_t,\ t \geq 0\}$ be a time-homogeneous Markov chain with states labeled as 0 and 1, transition probabilities $p_{ij} = \Pr(Z_t = i \mid Z_{t-1} = j), \quad t \geq 1,\ 0 \leq i, j \leq 1$, and initial probabilities $p_j = P(Z_0 = j)$, $j = 0, 1$. Denote by $T^{(i)}_{r,k}$ and $W^{(i)}_{r,k}$ the waiting time for the $r-th$ occurrence of the pattern $\mathcal{E}_i$, $i = 1, 2$, in $Z_0, Z_1, Z_2, \ldots$ according to the non-overlapping and overlapping counting scheme, respectively.

In the case of i.i.d trials, $T^{(1)}_{r,k}$ was studied by Koutras [15] for $k \geq 2$. For $k = 2$ the distribution of $T^{(1)}_{r,k}$ is a special case of the negative binomial distribution of order k, studied earlier by Philippou et al. [19]. The pmf's of $T^{(2)}_{r,k}$ and $W^{(2)}_{r,k}$ were derived for $k \geq 3$ by Sen and Goyal [21] using combinatorial methods. In the case of first order dependent trials, the pgf of $W^{(1)}_{r,k}$ was derived by Antzoulakos [1]. In the case of higher order Markov chains, Sarkar et al. [20] studied $W^{(i)}_{r,k}$ ($i = 1, 2$) and derived a system of equations satisfied by its pgf.

Presently, we employ the Markov chain embedding technique to obtain closed formulas for the probability generating functions and recursive schemes for the probability mass functions of $T^{(i)}_{r,k}$ and $W^{(i)}_{r,k}$, $i = 1, 2$. This is done in Section 2. In Section 3 we apply our results on $T^{(1)}_{r,k}$ and $W^{(1)}_{r,k}$. We first give numerics on a variation of the moving window detection problem. Then, we present a possible application of $W^{(1)}_{r,k}$ in Biomedical Engineering. We suggest that the random variable $W^{(1)}_{r,k}$ could enrich an approach currently investigated by a group of Biomedical Engineers and work as a decision criterion for decreasing or increasing a patient's mechanical support. We review the research on this area and provide a few illustrative numerics.

2. Waiting time distributions

Let X_n denote the number of appearances of a pattern $\mathcal{E}$ in a sequence of n trials $Z_1, Z_2, ..., Z_n$ and Y_r denote the waiting time for the r-th occurrence of $\mathcal{E}$ in $Z_1, Z_2, ...$ Many formulas have been proved relating the distributions of X_n and Y_r (see, for example, [16]). Thus, studying the distribution of a waiting time variable through the distribution of the associated binomial type one and vice versa, is a procedure often followed in the literature.

In the present paper we shall adopt a similar approach. To study the waiting time random variables $T^{(i)}_{r,k}$ and $W^{(i)}_{r,k}$, $i = 1, 2$, we shall first define the associated binomial type ones. Let $\{Z_t,\ 0 \leq t \leq n\}$ be a time-homogeneous Markov chain with states labeled as 0 and 1, transition probabilities $p_{ij} = \Pr(Z_t = i \mid Z_{t-1} = j), \quad t \geq 1,\ 0 \leq i, j \leq 1$ and initial probabilities $p_j = P(Z_0 = j)$, $j = 0, 1$. Denote by $N^{(i)}_{n,k}$ the number of occurrences of the pattern $\mathcal{E}_i$ ($i = 1, 2$) when the patterns do not overlap, and by

*Date of the manuscript submission: 01/03/10. Corresponding author: Andreas N. Philippou. First author address: Department of Mathematics, University of Patras, 26500 Patras, Greece, Email:dafnisspyros@gmail.com. Second author address: Department of Mathematics, University of Patras, 26500 Patras, Greece, Email:anphilip@math.upatras.gr.

$M_{n,k}^{(i)}$ the number of occurrences of the pattern $\mathcal{E}_i$ when the patterns may overlap. The patterns $\mathcal{E}_i$ $(i = 1, 2)$ are defined as in the Introduction. For the derivation of the results we will make use of the Markov chain embedding technique introduced by Fu and Koutras [10] and subsequently enriched, among others, by Koutras and Alexandrou [17], Han and Aki [13], and Antzoulakos et al. [2] (see also [11]). We shall first show, that under the appropriate set-up, these random variables can be treated as Markov chain embeddable variables of binomial type. Before advancing to get our results, we deem necessary to recall the definition of the Markov chain embeddable variable of binomial type (MVB) from [17].

Let X_n (n a non-negative integer) be a non-negative finite integer-valued random variable and let $\ell_n = \sup\{x : \Pr(X_n = x) > 0\}$ its upper end point.

Definition 2.1. *The random variable X_n will be called Markov chain embeddable variable of binomial type if*

(a) *there exists a Markov chain $\{Y_t,\ t \geq 0\}$ defined on a discrete state space Ω which can be partitioned as*

$$\Omega = \bigcup_{x \geq 0} C_x, \quad C_x = \{c_{x0}, c_{x1}, ..., c_{x,s-1}\},$$

(b) $\Pr(Y_t \in C_y | Y_{t-1} \in C_x) = 0$, *for all $y \neq x,\ x+1$ and $t \geq 1$,*

(c) *the event $X_n = x$ is equivalent to $Y_n \in C_x$, i.e.*

$$\Pr(X_n = x) = \Pr(Y_n \in C_x), \quad n \geq 0,\ x \geq 0.$$

It follows from condition (b) of Definition 2.1 that for $\{Y_t, t \geq 0\}$ there are only transitions within the same substate set C_x and transitions from set C_x to set C_{x+1}. Those two types of transitions give birth to the next two $s \times s$ transition probability matrices

$$A_t(x) = (\Pr(Y_t = c_{xj} | Y_{t-1} = c_{xi})),$$
$$B_t(x) = (\Pr(Y_t = c_{x+1,j} | Y_{t-1} = c_{xi})).$$

In the present section we derive the probability generating function and establish recursive schemes for the evaluation of the probability mass function of $T_{r,k}^{(i)}$ and $W_{r,k}^{(i)}$, $i = 1, 2$.

2.1. Distribution of $T_{r,k}^{(1)}$

To study the distribution of the random variable $T_{r,k}^{(1)}$ we need to give the appropriate set-up under which the associated binomial type random variable $N_{n,k}^{(1)}$ can be treated as an MVB. We set $\ell_n = [n/2]$ and define $C_x = \{c_{x0}, c_{x1}, ..., c_{x,k}\}$, $x = 0, 1, \ldots, \ell_n$, $c_{xi} = (x, i)$, $0 \leq i \leq k$. We introduce a Markov chain $\{Y_t,\ t \geq 0\}$ on $\Omega = \bigcup_{x=0}^{\ell_n} C_x$ according to the following conditions:

(1) $Y_t = (0, 0)$ if $Z_1 = Z_2 = ... = Z_t = 0$;

(2) $Y_t = (x, 0)$, $x \geq 0$, if in the first t, t_1 and $t_1 - 1$ outcomes ($t_1 < t-k+2$) the pattern $\mathcal{E}_1$ has occurred x times, $Z_{t_1} = 1$ and $Z_{t_1+1} = Z_{t_1+2} = ... = Z_t = 0$;

(3) $Y_t = (x, 0)$, $x \geq 1$, if in the first t outcomes the pattern $\mathcal{E}_1$ has occurred x times, the $x - th$ occurrence of $\mathcal{E}_1$ occurred at the $t_1 - th$ trial $(2 \leq t_1 < t)$ and $Z_{t_1+1} = Z_{t_1+2} = ... = Z_t = 0$;

(4) $Y_t = (x, 1)$, $x \geq 0$, if in the first t and $t-1$ outcomes the pattern $\mathcal{E}_1$ has occurred x times and $Z_t = 1$;

(5) $Y_t = (x, i)$, $x \geq 0$ and $2 \leq i \leq k-1$, if in the first t, $t-i+1$ and $t-i$ outcomes $(t \geq i)$ the pattern $\mathcal{E}_1$ has occurred x times, $Z_{t-i+1} = 1$ and $Z_{t-i+2} = Z_{t-i+3} = \ldots = Z_t = 0$.

(6) $Y_t = (x, k)$, $x \geq 0$, if in the first t outcomes the pattern $\mathcal{E}_1$ has occurred x times, the $x - th$ occurrence of $\mathcal{E}_1$ occurred at the $t - th$ outcome (consequently $Z_t = 1$).

With this set-up, the random variable $N_{n,k}^{(1)}$ becomes an MVB with

$$\boldsymbol{\pi}_0 = (p_0,\ p_1,\ 0,\ ...,\ 0)_{1 \times (k+1)},$$

$$A = \begin{array}{c} \begin{matrix} (\cdot,0) & (\cdot,1) & (\cdot,2) & (\cdot,3) & \cdot & (\cdot,k-2) & (\cdot,k-1) & (\cdot,k) \end{matrix} \\ \begin{pmatrix} p_{00} & p_{01} & 0 & 0 & \cdot & 0 & 0 & 0 \\ 0 & 0 & p_{10} & 0 & \cdot & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{00} & \cdot & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & 0 & p_{00} & 0 \\ p_{00} & 0 & 0 & 0 & \cdot & 0 & 0 & 0 \\ p_{10} & p_{11} & 0 & 0 & \cdot & 0 & 0 & 0 \end{pmatrix} \end{array}$$

and the matrix B has all its entries 0 except for the entry $(2, k+1)$ which equals p_{11} and the entries $(i, k+1)$, $3 \leq i \leq k$, which are equal to p_{01}.

Since $N_{n,k}^{(1)}$ is an MVB, the double generating function of $T_{r,k}^{(1)}$, is given by the following relation (see [2]).

$$H(z, w) = 1 + wz\boldsymbol{\pi}_0[I - z(A + wB)^{-1}]B\mathbf{1}'. \qquad (2.1)$$

Set

$$D = p_{01}p_{10} - p_{00}p_{11},$$
$$L_1 = 1 - p_{00}z + p_1 p_{11} wz + wz^2(Dp_1 + p_0 p_{01} p_{11} - {p_{11}}^2)$$
$$+ wz^3(p_0 {p_{01}}^2 p_{10} - 2p_{01}p_{10}p_{11} + p_{00}{p_{11}}^2),$$
$$L_2 = {p_{00}}^{k-4} p_{01} p_{10} z^k(-{p_{00}}^2 + w(p_{01}(p_0 p_{00} - p_{10}) -$$
$$p_1 {p_{00}}^2 - Dp_{00}z)) \ \text{ and}$$
$$L_3 = 1 - p_{00}z - {p_{11}}^2 wz^2 + wz^3({p_{11}}^2 p_{00} - 2p_{01}p_{10}p_{11}) -$$
$${p_{00}}^{k-3} p_{01} p_{10} z^k (p_{00} - Dwz).$$

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Using (2.1), the set-up of the MVB $N^{(1)}_{n,k}$ and some algebra we get

$$\sum_{r=0}^{\infty}\sum_{x=0}^{\infty}\Pr(T^{(1)}_{r,k}=x)z^x w^r=\frac{P_1(z,w)}{P_2(z,w)}, \qquad (2.2)$$

where, for $k \geq 5$,

$$P_1(z,w)=L_1+L_2+p_{01}{}^2p_{10}wz^4(p_{10}-p_0p_{00})\times$$
$$(1+p_{00}z+\cdots+(p_{00}z)^{k-5}) \text{ and}$$
$$P_2(z,w)=L_3-p_{01}{}^2p_{10}{}^2wz^4(1+p_{00}z);$$

for $k=4$,

$$P_1(z,w)=L_1+L_2 \quad \text{and} \quad \mathrm{P_2(z,w)=L_3-p_{01}{}^2p_{10}{}^2wz^4}$$

and for $k=3$,

$$P_1(z,w)=L_1-p_{00}{}^{k-4}p_{01}p_{10}z^k-p_{01}{}^2p_{10}{}^2wz^k$$
$$+p_{00}p_{01}p_{10}p_{11}wz^{k+1} \quad \text{and} \quad \mathrm{P_2(z,w)=L_3}.$$

Equating the coefficients of w^r on both sides of (2.2) in all cases, we get

Lemma 2.1. *The probability generating function $H_r(z)$ of $T^{(1)}_{r,k}$ is given by*

$$H_r(z)=$$
$$\left(\frac{z^2(p_{11}{}^2(1-p_{00}z)+p_{01}{}^2p_{10}{}^2z^2C(z)-E(z))}{1-p_{00}z-p_{00}{}^{k-2}p_{01}p_{10}z^k}\right)^{r-1}\times$$
$$\frac{z(p_1+p_0p_{01}z-p_{00}p_1z)(p_{11}+p_{01}p_{10}zC(z))}{1-p_{00}z-p_{00}{}^{k-2}p_{01}p_{10}z^k},$$

where $C(z)=1+p_{00}z+\cdots+(p_{00}z)^{k-3}$ and $E(z)=p_{01}p_{10}p_{11}z((p_{00}z)^{k-2}-2)$.

We use Lemma 2.1 to derive the following recurrence.

Theorem 2.1. *The probability mass function $h_r(x)$ of $T^{(1)}_{r,k}$ satisfies, for $r \geq 2$ and $x \geq 2r-1$, the recursive scheme*

$$\begin{aligned}h_r(x)=&p_{00}h_r(x-1)+p_{11}(p_{11}(h_{r-1}(x-2)\\&-p_{00}h_{r-1}(x-3))+2p_{01}p_{10}h_{r-1}(x-3))\\&+p_{01}{}^2p_{10}{}^2(p_{00}{}^{k-3}h_{r-1}(x-k-1)\\&+p_{00}{}^{k-4}h_{r-1}(x-k)+\cdots+h_{r-1}(x-4))\\&+p_{00}{}^{k-2}p_{01}p_{10}(h_r(x-k)-p_{11}h_{r-1}(x-k-1)),\end{aligned}$$

with initial conditions

$$\begin{aligned}h_1(0)&=0,\\h_1(1)&=p_1p_{11},\\h_1(2)&=p_1p_{01}p_{10}+p_0p_{01}p_{11},\\h_1(x)&=p_{00}h_1(x-1)+p_0p_{00}{}^{x-3}p_{01}{}^2p_{10},\\&\quad 3\leq x\leq k-1,\\h_1(k)&=p_{00}h_1(k-1)+p_{00}{}^{k-3}p_{01}p_{10}(p_0p_{01}-p_1p_{00}),\\h_1(x)&=p_{00}h_1(x-1)+p_{00}{}^{k-2}p_{01}p_{10}h_1(x-k),\\&\quad x>k,\end{aligned}$$

$h_0(x)=\delta_{x,0},$ *and* $h_r(x)=0$ *for* $r\geq 1$ *and* $x<2r-1.$

Proof. It follows from Lemma 2.1 that, for $r \geq 2$,

$$\begin{aligned}&(1-p_{00}z-p_{00}{}^{k-2}p_{01}p_{10}z^k)H_r(z)=H_{r-1}(z)\times\\&z^2(p_{11}{}^2(1-p_{00}z)+p_{01}{}^2p_{10}{}^2z^2C(z)-E(z)).\end{aligned} \qquad (2.3)$$

Replacing $H_r(z)$ by the power series $H_r(z)=\sum_{x=0}^{\infty}h_r(x)z^x$ into (2.3) and equating coefficients of z^x on both sides of the resulting identity we get the recursive relation of Theorem 2.1.

We follow the same procedure for $r=1$.

2.2. Distribution of $T^{(2)}_{r,k}$

We set $\ell_n=[n/k]$ and define $C_x=\{c_{x0},c_{x1},...,c_{x,k}\}$, $x=0,1,\ldots,\ell_n$, where $c_{xi}=(x,i)$, $0\leq i\leq k$. We introduce a Markov chain $\{Y_t,\ t\geq 0\}$ on $\Omega=\bigcup_{x=0}^{\ell_n}C_x$ according to the conditions (1)-(6) of Section 2.1 ($\mathcal{E}_1$ is now replaced by $\mathcal{E}_2$ and the inequality $2\leq t_1<t$ of condition (3) becomes $k\leq t_1<t$).).

With this set-up, the random variable $N^{(2)}_{n,k}$ becomes an MVB with

$$\boldsymbol{\pi}_0=(p_0,\ p_1,\ 0,\ ...,\ 0)_{1\times(k+1)},$$

the matrix A is equal to

$$\left(\begin{array}{cccccccc}(\cdot,0)&(\cdot,1)&(\cdot,2)&(\cdot,3)&\cdot&(\cdot,k-2)&(\cdot,k-1)&(\cdot,k)\\p_{00}&p_{01}&0&0&\cdot&0&0&0\\0&p_{11}&p_{10}&0&\cdot&0&0&0\\0&p_{01}&0&p_{00}&\cdot&0&0&0\\0&p_{01}&0&0&\cdot&0&0&0\\\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\0&p_{01}&0&0&\cdot&0&p_{00}&0\\p_{00}&0&0&0&\cdot&0&0&0\\p_{10}&p_{11}&0&0&\cdot&0&0&0\end{array}\right)$$

and the matrix B has all its entries 0 except for the entry $(k,k+1)$ which equals p_{01}.

Using (2.1), the set-up of the MVB $N^{(2)}_{n,k}$ and some algebra we get

$$\sum_{r=0}^{\infty}\sum_{x=0}^{\infty}\Pr(T^{(2)}_{r,k}=x)z^x w^r=\frac{P_1(z,w)}{P_2(z,w)}, \qquad (2.4)$$

where

$$P_1(z,w)=1-z(p_{00}+p_{11})-Dz^2+p_{00}{}^{k-3}p_{01}p_{10}z^{k-1}\times$$
$$(1+p_1w-p_{00}z+wz(p_0p_{01}-p_1p_{00}-p_{11})-Dwz^2)$$

and

$$P_2(z,w)=1-z(p_{00}+p_{11})-Dz^2+p_{00}{}^{k-3}p_{01}p_{10}z^{k-1}\times$$
$$(1-p_{00}z-p_{11}wz-Dwz^2).$$

Using (2.4) and following the methodology of Section 2.1, we get the following results.

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Lemma 2.2. *The probability generating function $H_r(z)$ of $T_{r,k}^{(2)}$ is given by*

$$H_r(z) = \frac{[{p_{00}}^{k-3} p_{01} p_{10} z^{k-1}]^r [z(p_{11}(p_{00}z - 1) - p_{01}p_{10}z)]^{r-1} F(z)}{[-1 + z(p_{00} + p_{11}) + Dz^2 - {p_{00}}^{k-3} p_{01} p_{10} z^{k-1}(1 - p_{00}z)]^r},$$

where $F(z) = p_1(p_{00}z - 1) - p_0 p_{01} z$.

Theorem 2.2. *The probability mass function $h_r(x)$ of $T_{r,k}^{(2)}$ satisfies , for $r \geq 2$, the recursive scheme*

$$\begin{aligned} h_r(x) =& {p_{00}}^{k-3} p_{01} p_{10}((p_{00} h_r(x-k) - h_r(x-k+1) \\ &+ p_{01}p_{10} h_{r-1}(x-k-1)) + p_{11}(h_{r-1}(x-k) \\ &- p_{00} h_{r-1}(x-k-1))) + D h_r(x-2) \\ &+ (p_{00} + p_{11}) h_r(x-1),\ x \geq rk-1, \end{aligned}$$

with initial conditions

$$\begin{aligned} h_1(x) &= 0, \quad x < k-1, \\ h_1(k-1) &= p_1 p_{10} p_{00} p_{01}, \\ h_1(k) &= {p_{00}}^{k-3} p_{01} p_{10} (p_0 p_{01} - p_1 p_{00}) \\ &\quad + p_1 p_{10} p_{00} p_{01} (p_{00} + p_{11}), \\ h_1(x) &= p_{11} h_1(x-1) + p_{01} p_{10} (h_1(x-2) \\ &\quad - {p_{00}}^{k-3} h_1(x-k+1) + {p_{00}}^{k-2} h_1(x-k)) \\ &\quad + p_{00}(h_1(x-1) - p_{11} h_1(x-2)),\ x > k, \end{aligned}$$

$h_0(x) = \delta_{x,0}$, and $h_r(x) = 0$ for $r \geq 1$ and $x < rk-1$.

2.3. Distribution of $W_{r,k}^{(1)}$

We set $\ell_n = n - 1$ and define $C_x = \{c_{x0}, c_{x1}, ..., c_{x,k-1}\}$, $x = 0, 1, \ldots, \ell_n$, where $c_{xi} = (x, i)$, $0 \leq i \leq k-1$. We introduce a Markov chain $\{Y_t,\ t \geq 0\}$ on $\Omega = \bigcup_{x=0}^{\ell_n} C_x$ according to the following conditions:

(1) $Y_t = (0,0)$ if $Z_1 = Z_2 = ... = Z_t = 0$;

(2) $Y_t = (x, 0)$, $x \geq 0$, if in the first t outcomes the pattern $\mathcal{E}_1$ has occurred x times, $Z_{t_1} = 1$ and $Z_{t_1+1} = Z_{t_1+2} = ... = Z_t = 0$ $(t_1 < t - k + 2)$;

(3) $Y_t = (x, 1)$, $x \geq 0$, if in the first t outcomes the pattern $\mathcal{E}_1$ has occurred x times and $Z_t = 1$;

(4) $Y_t = (x, i)$, $x \geq 0$ and $2 \leq i \leq k-1$, if in the first t outcomes $(t \geq i)$ the pattern $\mathcal{E}_1$ has occurred x times, $Z_{t-i+1} = 1$ and $Z_{t-i+2} = Z_{t-i+3} = \ldots = Z_t = 0$.

With this set-up, the random variable $M_{n,k}^{(1)}$ becomes an MVB with

$$\boldsymbol{\pi}_0 = (p_0,\ p_1,\ 0,\ ...,\ 0)_{1 \times k},$$

$$A = \begin{array}{c} \begin{array}{ccccccc} (\cdot,0) & (\cdot,1) & (\cdot,2) & (\cdot,3) & \cdot & (\cdot,k-2) & (\cdot,k-1) \end{array} \\ \left(\begin{array}{ccccccc} p_{00} & p_{01} & 0 & 0 & \cdot & 0 & 0 \\ 0 & 0 & p_{10} & 0 & \cdot & 0 & 0 \\ 0 & 0 & 0 & p_{00} & \cdot & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & 0 & p_{00} \\ p_{00} & 0 & 0 & 0 & \cdot & 0 & 0 \end{array} \right) \end{array}$$

and the matrix B has all its entries 0 except for the entry $(2,2)$ which equals p_{11} and the entries $(i,2)$, $3 \leq i \leq k$, which are equal to p_{01}.

Using (2.1), the set-up of the MVB $M_{n,k}^{(1)}$ and some algebra we get that the double generating function of $W_{r,k}^{(1)}$ is equal to

$$H(z,w) = \frac{P_1(z,w)}{P_2(z,w)}, \tag{2.5}$$

where, for $k > 3$,

$$\begin{aligned} P_1(z,w) &= 1 - p_{00}z - p_0 p_{11} wz - p_0 p_{01} p_{10} wz^2 \\ &+ p_{11} wz^2 (p_{00} + p_0 p_{01} - p_1 p_{00}) - {p_{00}}^{k-3} p_{01} p_{10} z^k (p_{00}(1+w) \\ &+ p_0 p_{01} w - p_1 p_{00} w) + p_0 {p_{01}}^2 p_{10} wz^3 \times \\ &(1 + p_{00}z + \cdots + (p_{00}z)^{k-4}) \end{aligned}$$

and

$$P_2(z,w) = 1 - p_{00}z - p_{11}wz - Dwz^2 - {p_{00}}^{k-2} p_{01} p_{10} z^k (1-w);$$

for $k = 3$,

$$\begin{aligned} P_1(z,w) =& 1 - p_{00}z - p_0 p_{11} wz - p_0 p_{01} p_{10} wz^2 + \\ & p_{11} wz^2 (p_{00} + p_0 p_{01} - p_1 p_{00}) - {p_{00}}^{k-3} p_{01} p_{10} z^k \times \\ & (p_{00}(1+w) + p_0 p_{01} w - p_1 p_{00} w) \end{aligned}$$

and

$$P_2(z,w) = 1 - p_{00}z - p_{11}wz - Dwz^2 - {p_{00}}^{k-2} p_{01} p_{10} z^k (1-w).$$

Using (2.5) and following the methodology of Section 2.1, we get the following results.

Lemma 2.3. *The probability generating function $H_r(z)$ of $W_{r,k}^{(1)}$ is given by*

$$H_r(z) = \left(\frac{p_{11}z + Dz^2 - {p_{00}}^{k-2} p_{01} p_{10} z^k}{1 - p_{00}z - {p_{00}}^{k-2} p_{01} p_{10} z^k} \right)^{r-1} \times \frac{z(p_1 + p_0 p_{01} z - p_{00} p_1 z)(p_{11} + p_{01} p_{10} z C(z))}{1 - p_{00}z - {p_{00}}^{k-2} p_{01} p_{10} z^k}.$$

Theorem 2.3. *The probability mass function $h_r(x)$ of $W_{r,k}^{(1)}$ satisfies , for $r \geq 2$, the recursive scheme*

$$\begin{aligned} h_r(x) =& p_{11} h_{r-1}(x-1) + p_{01} p_{10} h_{r-1}(x-2) \\ &+ p_{00}(h_r(x-1) - p_{11} h_{r-1}(x-2)) \\ &+ {p_{00}}^{k-2} p_{01} p_{10} (h_r(x-k) - h_{r-1}(x-k)),\ x \geq r, \end{aligned}$$

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

with initial conditions

$$\begin{aligned}
h_1(0) &= 0,\\
h_1(1) &= p_1p_{11},\\
h_1(2) &= p_1p_{01}p_{10}+p_0p_{01}p_{11},\\
h_1(x) &= p_{00}h_1(x-1)+p_0{p_{00}}^{x-3}{p_{01}}^2p_{10},\\
&\quad 3\le x\le k-1,\\
h_1(k) &= p_{00}h_1(k-1)+{p_{00}}^{k-3}p_{01}p_{10}(p_0p_{01}-p_1p_{00}),\\
h_1(x) &= p_{00}h_1(x-1)+{p_{00}}^{k-2}p_{01}p_{10}h_1(x-k),\\
&\quad x>k,
\end{aligned}$$

$h_0(x)=\delta_{x,0}$, *and* $h_r(x)=0$ *for* $r\ge 1$ *and* $x<r$.

2.4. Distribution of $W^{(2)}_{r,k}$

We set $\ell_n = [(n-1)/(k-1)]$ and define $C_x = \{c_{x0}, c_{x1}, ..., c_{x,k-1}\}$, $x=0,1,\dots,\ell_n$, where $c_{xi}=(x,i)$, $0\le i\le k-1$. We introduce a Markov chain $\{Y_t,\ t\ge 0\}$ on $\Omega=\bigcup_{x=0}^{\ell_n} C_x$ according to the conditions (1)-(4) of Section 2.3 ($\mathcal{E}_1$ is now replaced by $\mathcal{E}_2$).

With this set-up, the random variable $M^{(2)}_{n,k}$ becomes an MVB with

$$\boldsymbol{\pi}_0=(p_0,\ p_1,\ 0,\ ...,\ 0)_{1\times k},$$

$$A=\begin{array}{c} \begin{array}{ccccccc} (\cdot,0) & (\cdot,1) & (\cdot,2) & (\cdot,3) & \cdot & (\cdot,k-2) & (\cdot,k-1) \end{array} \\ \left(\begin{array}{ccccccc}
p_{00} & p_{01} & 0 & 0 & \cdot & 0 & 0\\
0 & p_{11} & p_{10} & 0 & \cdot & 0 & 0\\
0 & p_{01} & 0 & p_{00} & \cdot & 0 & 0\\
0 & p_{01} & 0 & 0 & \cdot & 0 & 0\\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\
0 & p_{01} & 0 & 0 & \cdot & 0 & p_{00}\\
p_{00} & 0 & 0 & 0 & \cdot & 0 & 0
\end{array}\right)\end{array}$$

and the matrix B has all its entries 0 except for the entry $(k,2)$ which equals p_{01}.

Using (2.1), the set-up of the MVB $M^{(2)}_{n,k}$ and some algebra we get that the double generating function of $W^{(2)}_{r,k}$ is equal to

$$H(z,w)=\frac{P_1(z,w)}{P_2(z,w)}, \qquad (2.6)$$

where

$$\begin{aligned}P_1(z,w)=&1-z(p_{00}+p_{11})-Dz^2+{p_{00}}^{k-3}p_{01}p_{10}z^{k-1}\times\\ &(1-p_{00}(w+z)+p_0wz)\end{aligned}$$

and

$$\begin{aligned}P_2(z,w)=&1-z(p_{00}+p_{11})-Dz^2+{p_{00}}^{k-3}p_{01}p_{10}z^{k-1}\times\\ &(1-w-p_{00}z(1-w)).\end{aligned}$$

Lemma 2.4. *The probability generating function* $H_r(z)$ *of* $W^{(2)}_{r,k}$ *is given by*

$$H_r(z)=\frac{[p_{00}^{k-3}p_{01}p_{10}z^{k-1}]^r[p_{00}z-1]^{r-1}[p_1(p_{00}z-1)-p_0p_{01}z]}{[-1+z(p_{00}+p_{11})+Dz^2-p_{00}^{k-3}p_{01}p_{10}z^{k-1}(1-p_{00}z)]^r}.$$

Theorem 2.4. *The probability mass function* $h_r(x)$ *of* $W^{(2)}_{r,k}$ *satisfies the recursive scheme*

$$\begin{aligned}h_r(x)=&{p_{00}}^{k-3}p_{01}p_{10}(p_{00}(h_r(x-k)-h_{r-1}(x-k))\\
&+h_{r-1}(x-k+1)-h_r(x-k+1))+Dh_r(x-2)\\
&+(p_{00}+p_{11})h_r(x-1),\quad x\ge r(k-1),\end{aligned}$$

with initial conditions

$$\begin{aligned}
h_1(x) &= 0,\quad x<k-1,\\
h_1(k-1) &= p_1p_{10}p_{00}p_{01},\\
h_1(k) &= {p_{00}}^{k-3}p_{01}p_{10}(p_0p_{01}-p_1p_{00})\\
&\quad +p_1p_{10}p_{00}p_{01}(p_{00}+p_{11}),\\
h_1(x) &= p_{11}h_1(x-1)+p_{01}p_{10}(h_1(x-2)\\
&\quad -{p_{00}}^{k-3}h_1(x-k+1)+{p_{00}}^{k-2}h_1(x-k))\\
&\quad +p_{00}(h_1(x-1)-p_{11}h_1(x-2)),\ x>k,
\end{aligned}$$

$h_0(x)=\delta_{x,0}$, *and* $h_r(x)=0$ *for* $r\ge 1$ *and* $x<r(k-1)$.

In closing this section we mention that setting $p_0=1$, $p_1=0$, $p_{00}=p_{10}=q$ and $p_{01}=p_{11}=p$, our results capture the case of i.i.d trials. in particular, Lemmas 2.1-2.4 reduce to respective results of [6].

3. Applications

In this section, we apply our results regarding the random variables $T^{(1)}_{r,k}$ and $W^{(1)}_{r,k}$. In 3.1 we generalize a variation of the moving window detection problem. In 3.2 we propose an important application on the determination of the level of mechanical support provided to a patient in the intensive care unit. Our results can be combined with a current approach, still in research, and work as a decision criterion.

3.1. On the moving window detection problem

The moving window detection problem appears in [18] and [12]. We recall its variant proposed in [15]. Consider a radar sweep with a quantizer transmitting to the detector the digit 1 or 0 according to whether the signal-plus-noise waveform exceeds a predetermined threshold. The detectors' memory keeps track of the last k (at most) transmitted digits and generates a pulse when two $1's$ are observed. Should this happen, the contents of the detectors' memory are erased and the next transmitted digit is the first to be registered. The occurrence of the r-th pulse indicates an alarm. The transmissions are considered to be i.i.d.

We generalize the above variant considering transmissions that are first order dependent. Indeed, the occurrence of a value exceeding the threshold increases the probability of the occurrence of another one. In this case, $T^{(1)}_{r,k}$ denotes the waiting time for the r-th occurrence of a

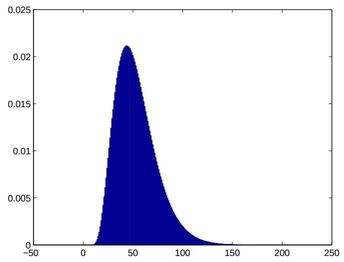

Figure 1: Distribution of $T^{(1)}_{5,6}$ for $p_0 = 0.75$, $p_{01} = 0.2$ and $p_{11} = 0.3$

pulse. For illustration purposes consider the case $r = 5$, $k = 6$, $p_0 = 0.75$, $p_{01} = 0.2$ and $p_{11} = 0.3$. Then, using Theorem 2.1, we calculate the distribution of $T^{(1)}_{5,6}$ (see Figure 1).

We observe, for example, that the occurrence of the 5-th pulse (alarm) will happen at the transmission of at most 80 digits with probability $P(T^{(1)}_{5,6} \leq 80) = 0.8768$.

3.2. On patients' mechanical support

Although the majority of patients receiving mechanical ventilation can be successfully disconnected after passing a trial of spontaneous breathing, approximately 20% of ventilated patients need a gradual reduction of mechanical support while they resume spontaneous breathing (see [9]).

This slow decrease in the amount of ventilator support with the patient gradually assuming a greater proportion of overall ventilation is called weaning from mechanical ventilation. The weaning is often taken to mean any method of discontinuing mechanical ventilation. In any case, a very significant fraction of a patient's time in the intensive care unit (ICU) is typically taken up with weaning.

For the majority of mechanically ventilated patients this process can be accomplished quickly and easily. There is, however, a significant percentage of patients in whom weaning fails. Part of the problem probably results from the fact that even excellent physicians often do not accurately judge when a patient is ready to wean but it is also true that the clinical approach to weaning, if poorly organized, adds additional time to the duration of mechanical ventilation.

Currently, weaning tends to be dictated by the experience and intuition of the attending physician who tries to maintain the patient in a state of 'comfort'. However, there is evidence that weaning may proceed more efficiently if directed according to some specified protocol.

Indeed, attempts have been made to formulate the weaning process as an algorithm, which could be automated on a computer (see, for example, [8]). In a new system, firstly proposed in [4], further developed in [7] and still in research, the researchers monitor, investigate, and take into account larger number of respiratory and cardio- circulatory parameters, namely: respiratory rate (RR), tidal volume (VT), the ratio RR/VT, pulse oxygen saturation (SpO2), end-tidal CO2 partial pressure (PETCO2), heart rate (HR), systolic arterial blood pressure (BP SAP), mean arterial blood pressure (BP MAP), and the end-expiratory pressure (PEEP). For each of these parameters a comfort zone (CZ) is defined by the physicians, specifying the range within the values of the monitored parameters should lie in order for the patient to be in a comfortable state. All the data are fed in a Fuzzy Logic Controller, which depending on the state of the patient decides if the patient's support level should be decreased or increased. This decision, within other parameters, is also taken based whether one or more of the monitored parameters are in or out of the CZ.

In the present paper we suggest that the random variable $W^{(1)}_{r,k}$ could be used as a decision criterion and enrich the system described above. We focus on one of the parameters. Suppose we get values of the parameter every 20 sec. Each value may be in the CZ with probability p_{00} or p_{10} if the previous value was in or out of the CZ, respectively. Each value may be out of the CZ with probability p_{01} or p_{11} if the previous value was in or out of the CZ, respectively. We denote by 0 and 1 the occurrence of a value in and out of the CZ, respectively. The occurrence of two consecutive 1's is a sign of a stabilized bad condition and could speak for the increase of mechanical support. This is the case even if the two consecutive 1's are separated by at most $k-2$ 0's (the number of values in the CZ is not enough to compensate for the others). What k should be depends on the clinical case. It is also clear that, for example, the occurrence of 111 implies a worse condition for the patient than the occurrence of 11. Thus, overlapping counting must be adopted.

To get some numerics, suppose that $k = 4$, $p_0 = 0.85$, $p_{01} = 0.1$ and $p_{11} = 0.2$. The probability p_{11} is greater than p_{01} since the occurrence of a value out of the CZ implies a less stable patients' condition. We are interested in the distributions of $W^{(1)}_{r,4}$, for various values of

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

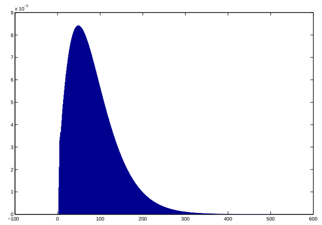

Figure 2: Distributions of $W_{3,4}^{(1)}$ and $W_{4,4}^{(1)}$ for $p_0 = 0.85$, $p_{01} = 0.1$ and $p_{11} = 0.2$

r. We get the above distributions using Theorem 2.3 (see Figure 2). The estimation of the parameters and the determination of the exact decision criteria require further combined research by Biomedical Engineers and Statisticians.

References

[1] Antzoulakos, D.L., "Waiting times for patterns in a sequence of multistate trials," *J Appl Probab*, V38, pp. 508-518, 1/01

[2] Antzoulakos, D.L., Bersimis, S., Koutras, M.V., "On the distribution of the total number of run lengths," *Ann Inst Stat Math*, V55, pp. 865-884, 4/03

[3] Balakrishnan, N., Koutras, M.V., *Runs and Scans with Applications*, Wiley, 2002.

[4] Buliev, I, Pallikarakis, N., Filos, K., "Computer-based system for automated weaning of conscious patients from assisted mechanical ventilation: preliminary tests". Annuals of Technical University of Varna, Bulgaria, 2006

[5] Dafnis, S.D., Antzoulakos D.L., Philippou, A.N., "Distributions related to (k_1, k_2) events," *J Statist Plann Inference*, V140, pp. 1691-1700, 7/10

[6] Dafnis, S.D., Philippou, A.N., Antzoulakos D.L., "Distributions of patterns of two successes separated by a string of $k-2$ failures". Submitted for publication

[7] Dermitzakis, A., Buliev, I., Filos, K., "Fuzzy logic controller for weaning of conscious patients from mechanical ventilation: a simulation study," 6th European Symposium on Biomedical Engineering, Chania, Greece, Code: s3.08, 2008

[8] Dojat, M., Harf, A., Touchard, D., Lemaire, F., Brochard L., "Clinical evaluation of a computer controlled pressure support mode," *American Journal of Critical Care in Medice, V161, pp. 1161-1166, 4/00*

[9] Esteban, A., Alia I., Ibanez, J., Benito, S., Tobin, M.J., "Modes of mechanical ventilation. A national survey of Spanish hospitals," Chest, V106, pp. 1188-1193, 4/94

[10] Fu, J.C., Koutras, M.V., "Distribution theory of runs: a Markov chain approach," J Am Stat Assoc V89, pp. 1050-1058, 427/94

[11] Fu, J.C., Lou, W.Y.W., Distribution theory of runs and patterns and its applications: a finite Markov imbedding approach, World Scientific Publishing, 2003.

[12] Glaz, J., "Moving window detection for discrete data," IEEE Trans Inform Theory, V29, pp. 457-462, 1983

[13] Han, Q., Aki, S., "Joint distributions of runs in a sequence of multi-state trials," Ann Inst Statist Math, V51, pp. 419-447, 3/99

[14] Huang, W., Tsai, C., "On a modified binomial distribution of order k, Statist Probab Lett, V11, pp. 125-131, 2/91

[15] Koutras M.V., "On a waiting time distribution in a sequence of Bernoulli trials," Ann Inst Statist Math, V48, pp. 789-806, 4/96

[16] Koutras, M.V., "Waiting times and number of appearances of events in a sequence of discrete random variables," Statistics for Industry and Technology, Advances in combinatorial methods and applications to probability and statistics, Boston, pp. 363-384, 1997

[17] Koutras, M.V., Alexandrou, V.A., "Runs, scans and urn model distributions: a unified Markov chain approach," Ann Inst Statist Math V47, pp. 743-766, 4/95

[18] Nelson, J.B., "Minimal-order models for false-alarm calculations on sliding windows," IEEE Trans Aerospace Electron Systems, V14, pp. 351-363, 2/78

[19] Philippou A.N., Georghiou C., Philippou G.N., "A generalized geometric distribution and some of its properties," Stat Probab Lett, V4, pp. 211-215, 1/83

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[20] *Sarkar, A., Sen, K., Anuradha, "Waiting time distributions of runs in higher order Markov chains," Ann Inst Statist Math, V56 , pp. 317-349, 2/04*

[21] *Sen, K., Goyal, B., "Distributions of patterns of two failures separated by success runs of length k, J Kor Stat Soc, V33, pp. 35-58, 1/04*

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)