

## Research Article

# On the Expected Number of Limited Length Binary Strings Derived by Certain Urn Models

Frossos S. Makri<sup>1</sup> and Zaharias M. Psillakis<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Patras, 26500 Patras, Greece

<sup>2</sup>Department of Physics, University of Patras, 26500 Patras, Greece

Correspondence should be addressed to Frossos S. Makri; makri@math.upatras.gr

Received 27 July 2014; Accepted 7 October 2014; Published 27 October 2014

Academic Editor: Tae-Sung Kim

Copyright © 2014 F. S. Makri and Z. M. Psillakis. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The expected number of 0-1 strings of a limited length is a potentially useful index of the behavior of stochastic processes describing the occurrence of critical events (e.g., records, extremes, and exceedances). Such model sequences might be derived by a Hoppe-Polya or a Polya-Eggenberger urn model interpreting the drawings of white balls as occurrences of critical events. Numerical results, concerning average numbers of constrained length interruptions of records as well as how on the average subsequent exceedances are separated, demonstrate further certain urn models.

## 1. Introduction and Preliminaries

Recently, some researches associated with the number of patterns which consist of runs of zeros (0s) between subsequent ones (1s) in sequences of binary random variables (RVs) have appeared in the literature. The 0-1 sequences  $\{X_n\}_{n \geq 1}$  may have several internal structures including among others sequences of independent but not necessarily identically distributed (INID) RVs, with  $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n p_i$ ,  $x_i \in \{0, 1\}$ , and  $p_i = P(X_i = 1) = 1 - P(X_i = 0) = 1 - q_i$ ,  $i = 1, 2, \dots, n$ , and sequences of exchangeable (EXCH) or symmetrically dependent RVs, the joint distribution of which is invariant under any permutation of its arguments; that is, for any  $n > 0$  and any vector  $(x_1, x_2, \dots, x_n)$ ,  $x_i \in \{0, 1\}$ , it holds that  $p_n(s) = P(X_{\pi_1} = x_1, X_{\pi_2} = x_2, \dots, X_{\pi_n} = x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$  for any permutation  $(\pi_1, \pi_2, \dots, \pi_n)$  of the set  $\{1, 2, \dots, n\}$  and  $s = \sum_{i=1}^n x_i$ . A common ground for both INID and EXCH sequences is sequences of independent and identically distributed (IID) RVs with probability of 1s  $p$ ,  $0 < p < 1$ , since an IID sequence is an INID sequence with  $p_i = p = 1 - q$  or an EXCH sequence with  $p_n(s) = p^s q^{n-s}$ .

In population genetics and evolution of species, urn models are frequently used as probabilistic models/devices

to explain/apply some theories. Among the plethora of such models (see, e.g., Johnson and Kotz [1], Blom et al. [2], and Mahmoud [3]) we consider in the sequel two of them: the Hoppe-Polya urn model (HPUM) and the Polya-Eggenberger urn model (PEUM). The first one, introduced by Holst [4, 5] as a generalization of the Hoppe urn model, is a device to produce certain INID binary sequences whereas the second one supplies a mechanism for producing particular EXCH binary sequences.

Special cases of PEUM are models of a (F/R – TM) fixed/random threshold (see, e.g., Eryilmaz and Yalcin [6], Makri and Psillakis [7], and Eryilmaz et al. [8]), whereas a special case of HPUM is the (RIM) record indicator model (see, e.g., Holst [5, 9, 10], Demir and Eryilmaz [11], and Makri and Psillakis [7]). F/R – TM and RIM find potential applications in the frequency analysis and risk managing of the occurrence of critical events (records, extremes, and exceedances) in several scientific disciplines like physical sciences (e.g., seismology, meteorology, and hydrology) and stochastic financial analysis (e.g., insurance and financial engineering). For details see the cited works.

In this paper we wish to use the expected number of an overlapping enumerative statistic (RV) as an index of the average occurrence of some special events over time or

space. The statistic counts 0-1 strings of a limited length on sequences derived by the urn models which are discussed in Section 2. In Section 3 we first define the statistic on any 0-1 sequence and then we provide its expected value for any INID and EXCH 0-1 sequence. Accordingly, the latter fact covers as particular cases the HPUM and PEUM, respectively. Finally, in Section 4 certain HPUM and PEUM associated with interpretations of special events, specifically the RIM and F/R-TM, are considered. Extensive numerics, concerning the average number of limited length interruptions of records and how on the average subsequent exceedances are separated, demonstrate further 0-1 sequences related to RIM and F/R-TM.

Throughout the paper  $I(A)$  is the indicator of an event  $A$ ; that is,  $I(A) = 1$  if the event  $A$  occurs, and  $I(A) = 0$  otherwise and  $\lfloor x \rfloor$  stands for the greatest integer less than or equal to a real number  $x$ . Also, we apply the conventions  $\sum_{i=a}^b = 0$ ,  $\prod_{i=a}^b = 1$ , for  $a > b$ ; that is, an empty sum (product) is to be interpreted as a zero (unity).

## 2. 0-1 Sequences Derived by Certain Urn Models

An urn that initially contains  $w$  white and  $b$  black balls,  $w > 0$ ,  $b \geq 0$ , is considered. A ball is drawn at random (all balls being equal likely) from the urn and its color is observed. We describe two drawing (or sampling) schemes that yield binary sequences different in structure. To this end we suppose the following.

- (i) Every time a white ball or a black ball is drawn from the urn, it is replaced together with a new ball of a color not already in the urn; otherwise the ball is replaced together with a copy of it. This drawing scheme is called Hoppe-Polya urn model and it is denoted as HPUM( $w, b$ ). If  $b = 0$ , that is, there is no black ball, the so-called Hoppe urn model HPUM( $w, 0$ ) is obtained.
- (ii) For  $w > 0$ ,  $b > 0$  we assume that all balls emanating from a draw of a white or a black ball are colored white or, respectively, black. Every time a ball is drawn from the urn, it is replaced together with  $c$  ( $c$ : an integer) balls of the same color of the ball drawn. This scheme is the well-known Polya-Eggenberger urn model denoted as PEUM( $w, b, c$ ). Certain values of the parameter  $c$  that define various sampling schemes of special interest are  $c = -1$  (sampling without replacement),  $c = 0$  (sampling with replacement),  $c = 1$  (Polya sampling), and  $c = w = b > 0$ .

We note in passing that in PEUM( $w, b, c$ ) there is a fixed set of admissible colors that appear in the urn, whereas in HPUM( $w, b$ ) an infinite set of admissible colors is considered. Interpreting colors as species, the Darwinian theory and the Mendelian theory of evolution of species may be captured, respectively (for details see Mahmoud [3]).

A  $\{0, 1\}$ -valued sequence of RVs  $\{X_n\}_{n \geq 1}$  where 1 denotes the drawing of a white ball and 0 anything else in a

HPUM( $w, b$ ) or a PEUM( $w, b, c$ ), is considered. Then  $\{X_n\}_{n \geq 1}$  constitutes an INID sequence with

$$p_n = P(X_n = 1) = \frac{w}{w + b + n - 1}, \quad n = 1, 2, \dots \quad (1)$$

for HPUM( $w, b$ ) and an EXCH sequence with

$$\begin{aligned} p_n(s) &= \sum_{j=0}^{n-s} (-1)^j \binom{n-s}{j} \lambda_{s+j}, \quad s = 0, 1, \dots, n, \quad n = 1, 2, \dots, \\ \lambda_i &= p_i(i) = \frac{\prod_{j=0}^{i-1} (w + jc)}{\prod_{j=0}^{i-1} (w + b + jc)}, \quad i = 0, 1, \dots, n \end{aligned} \quad (2)$$

for PEUM( $w, b, c$ ), respectively. For  $c = 0$ , it holds

$$\begin{aligned} \lambda_i &= \lambda_1^i, \quad i = 1, 2, \dots, n, \\ \lambda_1 &= p = \frac{w}{w+b}, \quad \lambda_0 = 1, \\ p_n(s) &= p^s(1-p)^{n-s}, \end{aligned} \quad (3)$$

and hence, an IID sequence with probability of 1s  $p$  is derived.

Next, we consider certain 0-1 stochastic processes  $\{X_i\}_{i \geq 1}$  describing the occurrences of critical events which are associated with particular HPUM and PEUM. The connection between the urn models and the model processes is achieved by interpreting the occurrence of a critical event ( $X_i = 1$ ) as a drawing of a white ball from the related urn model. More explicitly we have the following models.

- (A) Let  $\{Y_i\}_{i \geq 1}$  be a sequence of IID RVs with continuous cumulative distribution function (CDF). Then the (upper) records of the sequence are those elements with values strictly larger than all previous ones. We define the sequence  $\{X_i\}_{i \geq 1}$  of record indicators or record indicator model (RIM)  $X_1 = 1$  and  $X_i = I(Y_i > \max\{Y_1, Y_2, \dots, Y_{i-1}\})$ ,  $i = 2, 3, \dots$ . Then  $\{X_i\}_{i=1}^n$  constitutes an INID sequence derived by  $n$  repetitions of a HPUM( $1, 0$ ). By symmetry lower records of  $\{Y_i\}_{i \geq 1}$  are obtained by considering the (upper) records of the sequence  $\{-Y_i\}_{i \geq 1}$ . Namely, they are  $Y_i$ 's with  $X_i = 1$ .
- (B) We assume that being associated with a process  $\{Y_i\}_{i \geq 1}$  with continuous CDF, 1 occurs at a step  $i$  if a value of the process exceeds a threshold (i.e., a critical level)  $u$  at that step. In this way, a  $\{0, 1\}$ -valued sequence  $X_i = I(Y_i > u)$ ,  $i = 1, 2, \dots$ , is generated depending on the type of the threshold (fixed or random) used. Examples of model of exceedances over fixed or random threshold are among others as follows.  
 (a) If  $u$  is fixed and  $Y_i$ 's are IID RVs, then the so-called fixed threshold model (FTM) is derived.  
 (b) Suppose that two independent samples of size  $m$  ( $\geq 1$ ) and  $n$  ( $\geq 1$ ), each one of independent elements, of the process  $\{Y_i\}_{i \geq 1}$ , are considered, specifically the samples  $(Y_1, Y_2, \dots, Y_m)$  and  $(Y_{m+1}, Y_{m+2}, \dots, Y_{m+n})$ . The  $j$ th order statistic  $Y_{j:m}$  of the first sample is chosen

as a random threshold; that is,  $u = Y_{j:m}$ ,  $j = 1, 2, \dots, m$ . Then the model sequence which is generated by this random threshold model (RTM) is  $X_i = I(Y_{m+i} > Y_{j:m})$ ,  $i = 1, 2, \dots, n$ .

Both models, FTM and RTM, are derived by  $n$  repetitions of a PEUM( $w, b, c$ ) with initial composition  $w$  and  $b$  such that  $\lambda_1 = w/(w + b)$ . Specifically, FTM corresponds to  $c = 0$  (drawings with replacements so that the urn composition remains constant) and  $\lambda_1 = p$ , whereas RTM corresponds to  $c = 1$  (drawings such that the number of balls in the urn increases) and  $w = m+1-j$ ,  $b = j$ , so that  $\lambda_1 = 1-j/(m+1)$ . If  $m \rightarrow \infty$ ,  $j \rightarrow \infty$  so that  $j/(m+1) \rightarrow 1-p$ ,  $0 < p < 1$ , with  $n$  remaining fixed, then  $n$  repetitions of a RTM( $m, j$ ) give same results as  $n$  repetitions of a FTM( $p$ ). This is so because as  $w \rightarrow \infty$ ,  $b \rightarrow \infty$  so that  $w/(w+b) \rightarrow p$ , then for fixed  $n > 0$  the Polya-Eggenberger distribution has as a limit the Binomial distribution with parameters  $n$  and  $p$  (cf. Johnson and Kotz [1], page 371).

### 3. Counting 0-1 Strings of a Limited Length

In frequency analysis of 0-1 sequences  $\{X_n\}_{n \geq 1}$ , like the ones derived by the models (A) and (B) or more generally of that associated with HPUM and PEUM, may be useful to define a statistic (RV) denoting the distance (or the time) between two subsequent drawings of white balls. Measuring this distance as the number of consecutive 0s (nonwhite balls) between subsequent 1s (white balls) we equivalently face the problem of counting runs of 0s of a limited length between subsequent 1s. The number,  $M_{n;k,\ell}$ , of such 0-1 strings of limited (or constrained) length  $d+2$ ,  $0 \leq k \leq d \leq \ell \leq n-2$ , in the first  $n$  ( $\geq 2$ ) 0-1 trials (i.e., ball drawings) can be written as

$$\begin{aligned} M_{n;k,\ell} &= \sum_{i=k+2}^n I_i, \\ I_i &= \sum_{r=1}^{\beta_i-k} X_{i-k-r} \prod_{j=i-k-r+1}^{i-1} (1 - X_j) X_i, \end{aligned} \quad (4)$$

where

$$\beta_i = i-1 \quad \text{for } i = k+2, k+3, \dots, \ell+1; \quad (5)$$

$$\ell+1 \quad \text{for } i = \ell+2, \ell+3, \dots, n. \quad (6)$$

Readily, for  $n < k+2$ ,  $M_{n;k,\ell} = 0$ . The support (range set) of  $M_{n;k,\ell}$  is

$$\mathcal{R}(M_{n;k,\ell}) = \left\{ 0, 1, \dots, \left\lfloor \frac{n-1}{k+1} \right\rfloor \right\}, \quad 0 \leq k \leq \ell \leq n-2. \quad (7)$$

It should be mentioned that the counting of 0-1 strings via  $M_{n;k,\ell}$  is considered in the overlapping sense; that is, 1 which is not at either end of the sequence may contribute towards counting two possible strings, the one which ends with the occurrence of it and the next one which starts with it.

The study, via  $M_{n;k,\ell}$ , of constrained  $(k, \ell)$  strings of the form  $\underbrace{100 \dots 01}_{k \leq d \leq \ell}$  where 1 is followed by at least  $k$  and at most  $\ell$

0s before the next 1 in a 0-1 sequence covers as particular cases strings that have been considered recently by several authors. Specifically,  $M_{n;0,\ell}$ ,  $M_{n;k,k}$ , and  $M_{n;k,n-2}$  enumerate strings of two subsequent 1s separated by a run of 0s of length at most, exactly, and at least equal to a nonnegative integer number, that is, strings of the form  $\underbrace{100 \dots 01}_{d \leq \ell}$ ,  $\underbrace{100 \dots 01}_{d=k}$ , and  $\underbrace{100 \dots 01}_{d \geq k}$  respectively. Some relevant contributions on the subject are the works of Sarkar et al. [12], Sen and Goyal [13], Holst [4], Dafnis et al. [14], Huffer and Sethuraman [15], and Makri and Psillakis [16].

Applications of constrained  $(k, \ell)$  strings of the general  $(k \leq d \leq \ell)$  or the restricted forms ( $d \leq \ell$ ,  $d = k$ ,  $d \geq k$ ) were found in information theory and data compression (see Zehavi and Wolf [17], Jacquet and Szpankowski [18], and Stefanov and Szpankowski [19]) in urn models, record models, and random permutations (see Chern et al. [20], Joffe et al. [21], Chern and Hwang [22], Holst [5, 9, 10, 23], and Huffer et al. [24]) in system reliability (see Eryilmaz and Zuo [25], Eryilmaz and Yalcin [6], and Makri [26]) and in biomedical engineering (see Dafnis and Philippou [27]).

Recently, Makri and Psillakis [28] obtained the probability mass function  $f_{n;k,\ell}(x) = P(M_{n;k,\ell} = x)$ ,  $x \in \mathcal{R}(M_{n;k,\ell})$  as well as the expected value  $\mu_{n;k,\ell} = E(M_{n;k,\ell})$  of  $M_{n;k,\ell}$  defined on any 0-1 INID or EXCH sequence. Accordingly, with  $\beta_i$  given by (5) and (6), it holds that

$$\begin{aligned} E(M_{n;k,\ell}) &= \sum_{i=k+2}^n \sum_{r=1}^{\beta_i-k} c_{i,r}, \\ c_{i,r} &= p_{i-k-r} \left( \prod_{j=i-k-r+1}^{i-1} q_j \right) p_i \end{aligned} \quad (8)$$

for a HPUM( $w, b$ ) and

$$E(M_{n;k,\ell}) = \sum_{i=k+2}^n \sum_{j=k+2}^{\beta_i+1} p_j(2), \quad (9)$$

for a PEUM( $w, b, c$ ), respectively.

Since  $M_{n;k,\ell} = M_{n;0,\ell} - M_{n;0,k} + M_{n;k,k}$ , it holds that

$$\mu_{n;k,\ell} = \mu_{n;0,\ell} - \mu_{n;0,k} + \mu_{n;k,k}, \quad 0 \leq k \leq \ell \leq n-2. \quad (10)$$

Equation (10) offers an alternative way to compute  $\mu_{n;k,\ell}$  in cases for which we have available  $\mu_{n;0,\ell}$ ,  $0 \leq \ell \leq n-2$ , and  $\mu_{n;k,k}$ ,  $0 \leq k \leq n-2$ . In this direction the result of Holst [4] for a HPUM( $w, 0$ ) which is as follows:

$$\mu_{n;k,k} = \frac{w(n-1)^{k+1}}{(k+1)(w+n-1)^{k+1}}, \quad (11)$$

where  $x^m = \prod_{r=1}^m (x - r + 1)$  is helpful. Furthermore, he reestablished that  $M_{n;k,k}$  is a Poisson RV with  $\mu_{n;k,k} = w/(k+1)$  when  $n$  tends to infinity.

As it is observed from (8) and (9), both expressions of  $\mu_{n;k,\ell}$  are closed formulae proper for applications concerning large 0-1 sequences where  $\mu_{n;k,\ell}$  may be used to extract information for their stochastic behavior. For instance,  $\mu_{n;k,\ell}$  is a

TABLE 1: Expected values  $\mu_{n;k,\ell}$  of  $M_{n;k,\ell}$  defined on a RIM.

$n$	$k$	$\ell = 0$	$\ell = 1$	$\ell = 10$	$\ell = 98$	$\ell = 998$	$\ell = 9998$
$10^2$	0	0.990000	1.480000	2.909877	4.187378		
$10^3$		0.999000	1.498000	3.008877	5.078378	6.485471	
$10^4$		0.999900	1.499800	3.018777	5.167478	7.384571	8.787606
$10^5$		0.999990	1.499980	3.019767	5.176388	7.474481	9.687516
$10^2$	1		0.490000	1.919877	3.197378		
$10^3$			0.499000	2.009877	4.079378	5.486471	
$10^4$			0.499900	2.018877	4.167578	6.384671	7.787706
$10^5$			0.499990				
$10^2$	10			0.080909	1.358409		
$10^3$				0.089909	2.159409	3.566503	
$10^4$				0.090809	2.239509	4.456603	5.859638
$10^5$				0.090899			
$10^2$	98				0.000101		
$10^3$					0.009101	1.416194	
$10^4$					0.010001	2.227094	3.630130
$10^5$					0.010091		
$10^3$	998					0.000001	
$10^4$						0.000901	1.403936
$10^5$						0.000991	
$10^4$	9998						$0.1 \times 10^{-7}$
$10^5$							0.000090

potentially useful device to a trader (a human or a computer system) making transactions in a certain capital market (e.g., a stock market, a bond market, or an exchange market). It is assumed that the trader relies his/her/its buy-sell strategy on the expected value of  $M_{n;k,\ell}$ , whereas the interpretation of what  $M_{n;k,\ell}$  counts depends on the particular sequence on which  $M_{n;k,\ell}$  is defined.

#### 4. Numerical Results

In this section we consider two numerical examples that clarify further the theoretical results of Sections 2 and 3. In both examples 0-1 sequences of large or very large size are concerned. Moreover, the selected values of the urn parameters demonstrate the flexibility of the models as well as their applicability for a variety of situations.

*Example 1.* In a RIM sequence of length  $n$ ,  $n \gg 1$ ,  $M_{n;k,\ell}$  helps in understanding how the rather few records  $S_n = \sum_{i=1}^n X_i$ , with  $E(S_n) \approx \ln n + \gamma$  (see, e.g., Blom et al. [2], page 110) where  $\gamma$  ( $\approx 0.577216$ ) is the Euler's constant, are distributed among the many nonrecord elements of the sequence. Consequently,  $\mu_{n;k,\ell}$  indicates how many on the average strings of two subsequent records interrupted by at least  $k$  and at most  $\ell$  nonrecords occur in such a sequence.

In Table 1 we compute, via (8),  $\mu_{n;k,\ell}$  for various values of  $0 \leq k \leq \ell \leq n - 2$ . The RIM sequences used have lengths  $n = 10^2, 10^3, 10^4$ , and  $10^5$ . They approximately contain on the

average 5, 7, 10, and 12 records, respectively. The entries  $k = \ell$  of the table confirm that for  $n \gg 1$  it holds  $\mu_{n;k,k} \approx 1/(k+1)$ ,  $k \geq 0$ . Hence,  $\mu_{n;k,\ell} \approx \tilde{\mu}_{n;k,\ell} = \mu_{n;0,\ell} - \mu_{n;0,k} + (k+1)^{-1}$ ,  $n \gg 1$ , because of (10) and the Poisson approximation of (11). Consequently, a precompiled computation of  $\mu_{n;0,\ell}$ ,  $0 \leq \ell \leq n - 2$ , for large enough  $n$ , yields approximately  $\mu_{n;k,\ell}$  for  $0 < k < \ell \leq n - 2$ , too. For instance, let  $n = 10^5$ , let  $k = 10$ , and let  $\ell = 998$ . Then by the respective entries of the table we find  $\tilde{\mu}_{10^5;10,998} = 4.545623$ , whereas a direct computation, via (8) or the usage of the exact value  $\mu_{10^5;10,10} = 0.090899$  in (10), gives  $\mu_{10^5;10,998} = 4.545613$ . That is, the relative error between  $\mu_{10^5;10,998}$  and  $\tilde{\mu}_{10^5;10,998}$  is 0.0002%. The value  $\mu_{10^5;10,998} \approx 5$  means that in a 0-1 sequence of  $10^5$  elements containing on the average 12 records occur on the average 5 strings of 2 subsequent records separated by at least 10 and at most 998 nonrecord elements.

*Example 2.* We consider a F/R – TM sequence of length  $n$ ,  $n \gg 1$ , with an exceedance probability  $\lambda_1$ ,  $0 < \lambda_1 < 1$ . In a RTM there are infinitely many different selections of the parameters  $m$  and  $j$  which give the same value of  $\lambda_1$ . Let  $N$ ,  $\Lambda_N$  be positive integers such that  $1 - \lambda_1 = \Lambda_N/10^N$  and  $\Delta = \gcd(10^N, \Lambda_N)$  be their greatest common divisor. Then for every pair  $(m, j)$  such that  $m = (10^N/\Delta)(1-r) - 1$ ,  $j = (\Lambda_N/\Delta)(1-r)$ , for  $r = 0, -1, -2, \dots$ ,  $\lambda_1$  is constant. Therefore, for a given  $\lambda_1$  what values (small, moderate, or large) of the size  $m$  and the threshold  $j$  should be used depends on the demanded consideration (short, middle, or long) of the past

TABLE 2: Expected values  $\mu_{n;k,\ell}^{(\alpha)}$  of  $M_{n;k,\ell}$  defined on a FTM ( $\alpha = F$ ) and on a RTM ( $\alpha = R$ ).

$n$	$k$	$\ell$	$\mu_{n;k,\ell}^{(F)}$	$\mu_{n;k,\ell}^{(R)}$		
			$r$	0	-10	-50
			$m$	1	21	101
			$j$	1	11	51
						501
$\lambda_1 = 0.50$	$10^2$	0	24.750000	33.000000	25.826087	24.990291
		10	48.978760	48.512821	48.921634	48.968795
		98	49.000000	49.009901	49.000000	49.000000
	$10^3$	10	0.021729	0.103730	0.046343	0.027280
		98	0.042969	0.600810	0.124709	0.058486
	$10^4$	0	249.750000	333.000000	260.608696	252.174757
		10	498.759033	492.743590	498.101519	498.645178
		10	0.241455	1.152681	0.514980	0.303150
$\lambda_1 = 0.85$	$10^2$	0	2499.750000	3333.000000	2608.434783	2524.019417
		10	4996.571768	4935.051282	4989.900378	4995.409015
		10	2.438721	11.642191	5.201346	3.061847
	$10^3$	0	71.527500	72.128571	71.584615	71.539863
		10	84.000000	83.999932	84.000000	84.000000
		98	84.000000	80.000000	84.000000	84.000000
$\lambda_1 = 0.85$	$10^4$	0	721.777500	727.842857	722.353846	721.902253
		10	848.999999	848.999227	848.999998	848.999999
		10	0.000004	0.001072	0.000011	0.000005
	$10^4$	0	7224.277500	7284.985714	7230.046154	7225.526151
		10	8498.999993	8498.992180	8498.999975	8498.999990
		10	0.000042	0.010830	0.000109	0.000052

sequence's data. Readily, in a FTM sequence there is not such a need since  $p = \lambda_1$ . In both sequences (FTM or RTM) the expected value of  $M_{n;k,\ell}$  might be helpful in predicting how many on the average strings of two subsequent exceedances (e.g., of a stock prices) over a threshold (fixed or random) are separated by at least  $k$  and at most  $\ell$  nonthreshold exceedances.

In Table 2 we compute, via (9), the expected value  $\mu_{n;k,\ell}^{(\alpha)}$  of  $M_{n;k,\ell}$  defined on a FTM( $\alpha = F$ ) and on a RTM( $\alpha = R$ ) with the same exceedance probability  $\lambda_1 = 0.50, 0.85$  for various values of  $0 \leq k \leq \ell \leq n - 2$ . The F/R – TM sequences which have been used have lengths  $n = 10^2, 10^3$ , and  $10^4$  and sets of pairs  $(m, j)$  that give a selected value of  $\lambda_1$  are depicted for  $r = 0, -10, -50$ , and  $-500$ , as well. The entries of the table confirm that, for large  $m, j$  the dependence in structure of a RTM relaxes loosing its sensitivity with respect to the initial sample sequence's size and threshold. Eventually, it degenerates becoming a FTM and therefore  $E(M_{n;k,\ell}^{(R)})$  tends to  $E(M_{n;k,\ell}^{(F)})$ . This situation is more apparent if  $n$  is smaller than or equal to  $m$  and  $j$ . If  $n$  is larger than  $m$  and  $j$ , the expected value  $E(M_{n;k,\ell}^{(R)})$  departs, even a little, from that of  $E(M_{n;k,\ell}^{(F)})$ .

with the same  $\lambda_1$  showing that in this case the dependency of RTM on its  $m$  and  $j$  counts. For instance, this is the case for  $\lambda_1 = 0.50, n = 10^4$ , and  $(m, j) = (1, 1), (21, 11)$ , and  $(101, 51)$ .

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] N. L. Johnson and S. Kotz, *Urn Models and Their Applications*, John Wiley & Sons, New York, NY, USA, 1977.
- [2] G. Blom, L. Holst, and D. Sandell, *Problems and Snapshots from the World of Probability*, Springer, New York, NY, USA, 1994.
- [3] H. M. Mahmoud, *Polya Urn Models*, CRC Press, Boca Raton, Fla, USA, 2009.
- [4] L. Holst, "Counts of failure strings in certain Bernoulli sequences," *Journal of Applied Probability*, vol. 44, no. 3, pp. 824–830, 2007.
- [5] L. Holst, "The number of two consecutive successes in a Hoppe-Pólya urn," *Journal of Applied Probability*, vol. 45, no. 3, pp. 901–906, 2008.

- [6] S. Eryilmaz and F. Yalcin, "On the mean and extreme distances between failures in Markovian binary sequences," *Journal of Computational and Applied Mathematics*, vol. 236, no. 6, pp. 1502–1510, 2011.
- [7] F. S. Makri and Z. M. Psillakis, "On success runs of length exceeded a threshold," *Methodology and Computing in Applied Probability*, vol. 13, no. 2, pp. 269–305, 2011.
- [8] S. Eryilmaz, O. L. Gebizlioglu, and F. Tank, "Modeling of claim exceedances over random thresholds for related insurance portfolios," *Insurance: Mathematics & Economics*, vol. 49, no. 3, pp. 496–500, 2011.
- [9] L. Holst, "On consecutive records in certain Bernoulli sequences," *Journal of Applied Probability*, vol. 46, no. 4, pp. 1201–1208, 2009.
- [10] L. Holst, "A note on records in a random sequence," *Arkiv for Matematik*, vol. 49, no. 2, pp. 351–356, 2011.
- [11] S. Demir and S. Eryilmaz, "Run statistics in a sequence of arbitrarily dependent binary trials," *Statistical Papers*, vol. 51, no. 4, pp. 959–973, 2010.
- [12] A. Sarkar, K. Sen, and Anuradha, "Waiting time distributions of runs in higher order Markov chains," *Annals of the Institute of Statistical Mathematics*, vol. 56, pp. 317–349, 2004.
- [13] K. Sen and B. Goyal, "Distributions of patterns of two failures separated by success runs of length  $k$ ," *Journal of the Korean Statistical Society*, vol. 33, no. 1, pp. 35–58, 2004.
- [14] S. D. Dafnis, A. N. Philippou, and D. L. Antzoulakos, "Distributions of patterns of two successes separated by a string of  $k$ -2 failures," *Statistical Papers*, vol. 53, no. 2, pp. 323–344, 2012.
- [15] F. W. Huffer and J. Sethuraman, "Joint distributions of counts of strings in finite Bernoulli sequences," *Journal of Applied Probability*, vol. 49, no. 3, pp. 758–772, 2012.
- [16] F. S. Makri and Z. M. Psillakis, "Counting certain binary strings," *Journal of Statistical Planning and Inference*, vol. 142, no. 4, pp. 908–924, 2012.
- [17] E. Zehavi and J. K. Wolf, "On runlength codes," *IEEE Transactions on Information Theory*, vol. 34, no. 1, pp. 45–54, 1988.
- [18] P. Jacquet and W. Szpankowski, "On  $(d,k)$  sequences not containing a given word," in *Proceedings of the International Symposium on Information Theory*, pp. 1486–1489, 2006.
- [19] V. T. Stefanov and W. Szpankowski, "Waiting time distributions for pattern occurrence in a constrained sequence," *Discrete Mathematics & Theoretical Computer Science*, vol. 9, no. 1, pp. 305–320, 2007.
- [20] H. H. Chern, H. K. Hwang, and Y. N. Yeh, "Distribution of the number of consecutive records," *Random Structures and Algorithms*, vol. 17, pp. 169–196, 2000.
- [21] A. Joffe, E. Marchand, F. Perron, and P. Popadiuk, "On sums of products of Bernoulli variables and random permutations," *Journal of Theoretical Probability*, vol. 17, no. 1, pp. 285–292, 2004.
- [22] H.-H. Chern and H.-K. Hwang, "Limit distribution of the number of consecutive records," *Random Structures & Algorithms*, vol. 26, no. 4, pp. 404–417, 2005.
- [23] L. Holst, "A note on embedding certain Bernoulli sequences in marked Poisson processes," *Journal of Applied Probability*, vol. 45, no. 4, pp. 1181–1185, 2008.
- [24] F. W. Huffer, J. Sethuraman, and S. Sethuraman, "A study of counts of Bernoulli strings via conditional Poisson processes," *Proceedings of the American Mathematical Society*, vol. 137, no. 6, pp. 2125–2134, 2009.
- [25] S. Eryilmaz and M. Zuo, "Constrained  $(k, d)$ -out-of- $n$  systems," *International Journal of Systems Science*, vol. 41, no. 6, pp. 679–685, 2010.
- [26] F. S. Makri, "Minimum and maximum distances between failures in binary sequences," *Statistics & Probability Letters*, vol. 81, no. 3, pp. 402–410, 2011.
- [27] S. D. Dafnis and A. N. Philippou, "Distributions of patterns with applications in engineering," *IAENG International Journal of Applied Mathematics*, vol. 41, no. 1, pp. 68–75, 2011.
- [28] F. S. Makri and Z. M. Psillakis, "Exact distributions of constrained  $(k, \ell)$  strings of failures between subsequent successes," *Statistical Papers*, vol. 54, no. 3, pp. 783–806, 2013.

