

Counting Runs of Ones with Overlapping Parts in Binary Strings Ordered Linearly and Circularly

Frosso S. Makri¹, Zaharias M. Psillakis² & Anastasios N. Arapis¹

¹ Department of Mathematics, University of Patras, Patras, Greece

² Department of Physics, University of Patras, Patras, Greece

Correspondence: Frosso S. Makri, Department of Mathematics, University of Patras, Patras 26500, Greece. Tel: 30-2610-996-738. E-mail: makri@math.upatras.gr

Received: April 21, 2013 Accepted: May 20, 2013 Online Published: July 9, 2013

doi:10.5539/ijsp.v2n3p50

URL: <http://dx.doi.org/10.5539/ijsp.v2n3p50>

Abstract

On a binary (0 – 1) string of length n the ℓ -overlapping counting scheme of runs of 1 s of a fixed length k is considered. According to this scheme, a run of 1s of length k which is counted may have overlapping part of length at most ℓ , $0 \leq \ell < k \leq n$, with the previous run of 1 s of length k that has been enumerated. The numbers of all ℓ -overlapping runs of 1 s of length k in all 2^n binary strings (linearly or circularly ordered) of length n are examined, and simple and easy to compute closed explicit expressions are provided via the probability mass function and the expected value of properly defined random variables. The numbers of binary strings of length n , ordered on a line or on a circle, with a specific number of ℓ -overlapping runs of 1s of length k are also provided via closed expressions. The numbers which are studied, are potentially useful in several scientific areas like applied probability, engineering and bioinformatics. The study is illustrated by extensive numerical examples.

Keywords: binary strings, overlapping counting, run statistics

1. Introduction and Preliminaries

Nowadays, the increasing use of computer science in diverse applications including encoding, compression and transmission of digital information calls for understanding the distribution of runs of 1 s or 0 s. For instance, such knowledge helps in analyzing, and also in comparing, several techniques used in communication networks (wired or wireless). In such networks binary data, ranging from a few bytes (e.g. e-mails) to many gigabytes of greedy multimedia applications (e.g. video on demand), are highly processed. For details as well as for additional real-life applications see Sinha (2007), K. Sinha and B. P. Sinha (2009, 2012) and the references therein.

Another area where the study of distribution of runs of 1 s and 0 s has become increasingly useful is the field of bioinformatics or computational biology. For instance, in the context of hypothesis testing molecular biologists design similarity tests between two DNA (DeoxyriboNucleic Acid) sequences where a 1 is interpreted as a match of the sequences at a given position (Benson, 1999; Lou, 2003; Nuel et al., 2010; Makri & Psillakis, 2011a).

In such applications, as the indicative ones mentioned above, a key point is the understanding how 1 s and 0 s are distributed and combined as elements of a binary sequence (finite or infinite, memoryless or not) and eventually forming runs of 1 s and 0 s which are enumerated according to certain counting schemes. Each counting scheme defines how runs of same symbols (i.e. 1 s or 0 s) are formed and consequently counted. A scheme may depend on, among other considerations, whether overlapping counting is allowed or not as well as if the counting starts or not from scratch when a run of a certain size has been so far enumerated. For extensive reviews of the runs literature we refer to Balakrishnan and Koutras (2002), Fu and Lou (2003), and Koutras (2003). The topic is still active and also attractive, because of the wide range of its application in many areas of applied probability and engineering including hypothesis testing, quality control, system reliability, computer science and financial engineering. Some recent contributions on the subject, among others, are the works of Eryilmaz (2011), Makri and Psillakis (2012), Demir Atalay and Zeybek (2013), and Mytalas and Zazanis (2013).

Consider a sequence of n binary (zero-0, one-1) random variables (RVs) with values (i.e. outcomes of binary trials) ordered on a line or on a circle. In the circular case we assume that the first outcome is adjacent to (and follows)

the n -th outcome.

For integers $k, \ell, 0 \leq \ell < k \leq n$, a generalized way of counting runs of ones (1-runs) of length k , say ℓ -overlapping, was introduced by Aki and Hirano (2000) for the linear case (see also Han & Aki, 2000; Inoue & Aki, 2003; Makri & Philippou, 2005; Makri et al., 2007a) and by Makri and Philippou (2005) for the circular case (see also Makri et al., 2007a). It covers as particular cases the non-overlapping counting ($\ell = 0$; see Feller, 1968; Hirano, 1986; Philippou & Makri, 1986 for the linear case, and Charalambidis, 1994; Makri & Philippou, 1994; Koutras et al., 1995 for the circular case) and the overlapping counting ($\ell = k - 1$; see Ling, 1988; Hirano et al., 1991; Holst, 2009 for the linear case, and Inoue & Aki, 2010 for the circular case).

According to this counting scheme, the number $X_{n,k,\ell}^{(\alpha)}$ of ℓ -overlapping 1-runs of length k in n binary trials ordered on a line ($\alpha = L$) or on a circle ($\alpha = C$) is the number of 1-runs of length k ($k = 1, 2, \dots, n$) each of which may have overlapping (common) part of length at most ℓ ($\ell = 0, 1, \dots, k - 1$) with the previous 1-run of length k that has been enumerated. For trials arranged on a circle if there is (at least) one 0 in the sequence the counting of 1-runs starts from the first 1 following a 0 and if there is no 0s then it starts from any 1 in the sequence. For a given set of n, k, ℓ it holds $X_{n,k,\ell}^{(L)} \leq X_{n,k,\ell}^{(C)}$. This is so, because as the n trials are bent into a circle additional 1-runs of length k may be formed by combining 1s at the beginning and the end of the linear sequence. Readily, $X_{n,k,\ell_1}^{(\alpha)} \leq X_{n,k,\ell_2}^{(\alpha)}$, $0 \leq \ell_1 \leq \ell_2 < k \leq n$, $X_{n,k_1,\ell}^{(\alpha)} \geq X_{n,k_2,\ell}^{(\alpha)}$, $0 \leq \ell \leq k_1 - 1$, $1 \leq k_1 \leq k_2 \leq n$, and the events $\{X_{n,k,\ell}^{(\alpha)} = 0\}$, $0 < \ell < k \leq n$, are equivalent to the event $\{X_{n,k,0}^{(\alpha)} = 0\}$, $\alpha = L, C$.

Example 1. As an illustration, let us assume that $n = 15$ binary trials, numbered from 1 to 15, are performed and we get the following outcomes 11111011110111. Then, the ℓ -overlapping 1-runs of length 4 are: (a) 1, 2, 3, 4; 3, 4, 5, 6; 8, 9, 10, 11 for $\ell = 2$, and 1, 2, 3, 4; 2, 3, 4, 5; 3, 4, 5, 6; 8, 9, 10, 11 for $\ell = 3$, in the linear case and (b) 8, 9, 10, 11; 13, 14, 15, 1; 15, 1, 2, 3; 2, 3, 4, 5 for $\ell = 2$, and 8, 9, 10, 11; 13, 14, 15, 1; 14, 15, 1, 2; 15, 1, 2, 3; 1, 2, 3, 4; 2, 3, 4, 5; 3, 4, 5, 6 for $\ell = 3$, in the circular case. Hence, $X_{15,4,2}^{(L)} = 3$, $X_{15,4,3}^{(L)} = 4$ and $X_{15,4,2}^{(C)} = 4$, $X_{15,4,3}^{(C)} = 7$.

The distributions of $X_{n,k,\ell}^{(\alpha)}$ defined on a sequence of Bernoulli trials (that is, a sequence of independent and identically distributed RVs with a common probability of 1s p) ordered on a line or a circle have been called binomial and circular binomial distributions of order k in the ℓ -overlapping case with parameter vector (n, p) , respectively (see, Makri & Philippou, 2005). They have been denoted as $B_{k,\ell}^{(\alpha)}(n, p)$ (or $B_{k,\ell}(n, p)$ for $\alpha = L$ and $B_{k,\ell}^C(n, p)$ for $\alpha = C$) and cover as particular cases, the extensively studied in runs literature and been used in various applications of runs, Type I (circular Type I) and Type III (circular Type III) distributions of order k . The latter distributions are derived for $k > 1$ and for $\ell = 0, k - 1$. For $k = 1$ and $\ell = 0$ all these distributions reduce to the well known binomial distribution $B(n, p)$ with parameter vector (n, p) . For details see, e.g. Balakrishnan and Koutras (2002, pp. 140, 193).

According to Makri and Psillakis (2013), in order to study formally ℓ -overlapping counting in any sequence of 0-1 RVs $\{Z_i\}_{i=1}^n$, the run statistics $X_{n,k,\ell}^{(\alpha)}$, $0 \leq \ell < k \leq n$, $\alpha = L, C$, can be defined as follows

$$X_{n,k,\ell}^{(L)} = \sum_{j=k}^n I_j^{(L)}, \quad I_j^{(L)} = \sum_{s=1}^{\lfloor \frac{j-\ell}{k-\ell} \rfloor} \left(\prod_{i=j-\ell-s(k-\ell)+1}^j Z_i \right) (1 - Z_{j-\ell-s(k-\ell)}), \tag{1}$$

for $j = k, k + 1, \dots, n$, with the convention $Z_0 \equiv 0$ and

$$X_{n,k,\ell}^{(C)} = \sum_{j=1}^n I_j^{(C)}, \quad I_j^{(C)} = \sum_{s=1}^{\lfloor \frac{n-1-\ell}{k-\ell} \rfloor} \left(\prod_{i=j-\ell-s(k-\ell)+1}^j Z_i \right) (1 - Z_{j-\ell-s(k-\ell)}) + \sum_{r=1}^{\lfloor \frac{n}{k-\ell} \rfloor} \delta_{j,r(k-\ell)} \prod_{i=1}^n Z_i, \tag{2}$$

for $j = 1, 2, \dots, n$, with the conventions $Z_0 \equiv Z_n, Z_{-i} \equiv Z_{n-i}$, $1 \leq i \leq n - 1$. The supports (range sets) of $X_{n,k,\ell}^{(L)}$ and $X_{n,k,\ell}^{(C)}$ are

$$S_{n,k,\ell}^{(L)} = \{0, 1, \dots, \lfloor \frac{n-\ell}{k-\ell} \rfloor\} \quad \text{and} \quad S_{n,k,\ell}^{(C)} = \{0, 1, \dots, \lfloor \frac{n-\ell-1}{k-\ell} \rfloor, \lfloor \frac{n}{k-\ell} \rfloor\}, \tag{3}$$

respectively. In the above formulae, as well as in the whole article $\lfloor x \rfloor$ stands for the greatest integer less than or equal to a real number x and $\delta_{i,j}$ denotes the Kronecker delta function of the integer arguments i and j . Also, we apply the conventions $\sum_{i=r}^s a_i = 0, \prod_{i=r}^s a_i = 1$, for $r > s$.

The setup of indicators $I_j^{(\alpha)}$ in (1) and (2) holds true for any 0 – 1 sequence $\{Z_i\}_{i=1}^n$ and it is the main tool to derive the expected value, $E(X_{n,k,\ell}^{(\alpha)})$, of $X_{n,k,\ell}^{(\alpha)}$, $\alpha = L, C$ for such a sequence. Furthermore, it is useful to determine numeric

values of $X_{n,k,\ell}^{(\alpha)}$ and therefore empirical frequencies and moments in studies concerning applications like processing of computer files or DNA sequences.

In this point we should mention that the ℓ -overlapping counting scheme, as a natural generalization of both Feller's and Ling's counting, differs in concept from the other fundamental counting scheme in runs literature; specifically the Mood's (1940) counting scheme (see, also Koutras, 2003; Makri et al., 2007b; Eryilmaz, 2008; K. Sinha & B. P. Sinha, 2009; Makri & Psillakis, 2011a, 2011b). According to the latter scheme, a run of 1s (1-run) is defined to be a sequence of consecutive 1s preceded and succeeded by 0s or by nothing. The number of 1s in a 1-run is referred to as its length (or size). For integers n and k with $n \geq k \geq 1$, let $Y_{n,k}^{(\beta,\alpha)}$ denote the numbers of 1-runs of length [exactly equal to k ($\beta = E$), greater than or equal to k ($\beta = G$)] in n binary trials $\{Z_i\}_{i=1}^n$ ordered on a line ($\alpha = L$) or on a circle ($\alpha = C$). These run statistics can be formulated (see, also Makri et al., 2007b) by the indicators

$$I_j^{(E,\alpha)} = (1 - Z_{j-k})(1 - Z_{j+1}) \prod_{i=j-k+1}^j Z_i \quad (4)$$

for $1 \leq k \leq n$ if $\alpha = L$ and for $1 \leq k \leq n - 2$ if $\alpha = C$. For $k = n - 1$, $I_j^{(E,C)} = (1 - Z_{j-k}) \prod_{i=j-k+1}^j Z_i$. (Conventions: $Z_0 = Z_{n+1} \equiv 0$ for $\alpha = L$ and $Z_0 = Z_n, Z_{-i} = Z_{n-i}, Z_{n+1} = Z_1$ for $\alpha = C$). Then, the statistics $Y_{n,k}^{(E,\alpha)}$ can be expressed as

$$Y_{n,k}^{(E,L)} = \sum_{j=k}^n I_j^{(E,L)}, \quad 1 \leq k \leq n, \quad Y_{n,k}^{(E,C)} = \sum_{j=1}^n I_j^{(E,C)}, \quad 1 \leq k \leq n - 1, \quad (5)$$

and $Y_{n,n}^{(E,C)} = \prod_{i=1}^n Z_i$. Consequently, the statistics $Y_{n,k}^{(G,\alpha)}$ can be expressed as

$$Y_{n,k}^{(G,\alpha)} = \sum_{i=k}^n Y_{n,i}^{(E,\alpha)}. \quad (6)$$

Readily it holds

$$Y_{n,k}^{(E,\alpha)} \leq Y_{n,k}^{(G,\alpha)} \leq X_{n,k,0}^{(\alpha)} \leq X_{n,k,\ell}^{(\alpha)} \leq X_{n,k,k-1}^{(\alpha)}, \quad 0 \leq \ell < k \leq n, \quad \alpha = L, C. \quad (7)$$

Example 2. In order to make the distinction among the run statistics $X_{n,k,\ell}^{(\alpha)}$ and $Y_{n,k}^{(\beta,\alpha)}$ for $\beta = E, G, \alpha = L, C$ clear, let us consider the following 0 – 1 sequence of 21 outcomes, numbered from 1 to 21, 111011001111110101111. Then for $k = 3$ there is one 1-run of length exactly equal to 3, i.e. the run 1, 2, 3 in the linear case whereas there are no such 1-runs in the circular case. Consequently, $Y_{21,3}^{(E,L)} = 1$ and $Y_{21,3}^{(E,C)} = 0$. Furthermore, the 1-runs of length greater than or equal to 3 are the runs 1, 2, 3; 9, 10, 11, 12, 13, 14; 18, 19, 20, 21 in the linear case and the runs 9, 10, 11, 12, 13, 14; 18, 19, 20, 21, 1, 2, 3 in the circular case. Hence, $Y_{21,3}^{(G,L)} = 3$ and $Y_{21,3}^{(G,C)} = 2$. Finally, using ℓ -overlapping counting we find that $X_{21,3,0}^{(L)} = 4, X_{21,3,1}^{(L)} = 4, X_{21,3,2}^{(L)} = 7$ and $X_{21,3,0}^{(C)} = 4, X_{21,3,1}^{(C)} = 5, X_{21,3,2}^{(C)} = 9$.

In real-life applications like the interesting ones proposed by K. Sinha and B. P. Sinha (2009, 2012), which considered processing of computer files of various formats commonly encountered for computing and communication purposes, the replacement of $Y_{n,k}^{(E,L)}$ with $X_{n,k,\ell}^{(L)}$ might offer a useful alternative. The same rather holds for circularly connected networks.

The use of $X_{n,k,\ell}^{(\alpha)}$ instead of $Y_{n,k}^{(\beta,\alpha)}$ possibly offers a more informative testing, via (7), whereas the selection of the overlapping parameter ℓ provides, in turns, flexibility depending on the testing requirements of such files. Specifically, in stringent testing of binary sequences ℓ could be made as big as $k - 1$ and then almost all the binary digits participating in 1-runs of length k are tested more than once. Readily, ℓ could be made as low as 0, and then all the binary digits will be tested for 1-runs only once. The latter fact corresponds to less demanding testing of the binary sequences.

Associated with $Y_{n,k}^{(\beta,\alpha)}$, $\beta = E, G$ and $\alpha = L, C$ are the numbers $Q_{n,k}^{(\beta,\alpha)}$ of all 1-runs (in the Mood's sense) of length (equal, greater than or equal to) k in all 2^n possible binary strings (that is, symmetric Bernoulli sequences; $p = 1/2$) of length n ordered on a line or on a circle. Recently, Makri and Psillakis (2011b) and Makri et al. (2012) provided explicit expressions for the numbers $Q_{n,k}^{(\beta,L)}$, $\beta = E, G$. These works simplify, extend and unify too, the results of K. Sinha and B. P. Sinha (2009, 2012) referring to $Q_{n,k}^{(E,L)}$.

In the present paper we consider, in all 2^n binary strings of length n ordered on a line or on a circle, first the number $N_{x;n,k,\ell}^{(\alpha)}$ of all binary strings of length n which contain, $x, x \in S_{n,k,\ell}^{(\alpha)}$, ℓ -overlapping 1-runs of length k , and second the number $R_{n,k,\ell}^{(\alpha)}$ of all ℓ -overlapping 1-runs of length k .

Example 3. To clarify $S_{n,k,\ell}^{(\alpha)}$, $N_{x;n,k,\ell}^{(\alpha)}$ and $R_{n,k,\ell}^{(\alpha)}$ let us consider $\alpha = L$, $n = 4$, $k = 2$ and $\ell = 1$. Then all 2^4 binary strings ordered on a line and numbered from (0) to (15) are: (0) 0000, (1) 0001, (2) 0010, (3) 0011, (4) 0100, (5) 0101, (6) 0110, (7) 0111, (8) 1000, (9) 1001, (10) 1010, (11) 1011, (12) 1100, (13) 1101, (14) 1110, (15) 1111. Hence, $S_{4,2,1}^{(L)} = \{0, 1, 2, 3\}$ and the number of all 1-overlapping runs of length 2 is $R_{4,2,1}^{(L)} = 12$ which appear in strings (3), (6), (7), (11), (12), (13), (14) and (15). The strings (3), (6), (11), (12) and (13) contain one (i.e. $x = 1$) 1-overlapping run of length 2 and therefore $N_{1;4,2,1}^{(L)} = 5$.

Additional numerics illustrating the numbers $N_{x;n,k,\ell}^{(\alpha)}$ and $R_{n,k,\ell}^{(\alpha)}$ are computed and presented in the forthcoming Tables 1 to 4 as well as in the Examples 4 and 5.

The rest of the paper is organized as follows. In Section 2, we state the problem and we develop closed expressions for the numbers $N_{x;n,k,\ell}^{(\alpha)}$ and $R_{n,k,\ell}^{(\alpha)}$ for $\alpha = L, C$. The expressions so provided are relied on a new interpretation of recent results for the statistic $X_{n,k,\ell}^{(\alpha)}$ of Makri et al. (2007a) and Makri and Psillakis (2013). However, the introduction and study of the numbers $N_{x;n,k,\ell}^{(\alpha)}$ and $R_{n,k,\ell}^{(\alpha)}$, have not been addressed previously. In Section 3, we summarize our main results and we discuss the methodology used to derive them.

2. Mathematical Formulation

In this Section we present our main results. In Section 2.1 we consider Bernoulli sequences whereas in Section 2.2 we restrict our study to symmetric Bernoulli sequences.

2.1 Bernoulli Sequences

We consider a sequence $\{Z_i\}_{i=1}^n$ of length n of independent (i.e. derived by a memoryless source) and identically distributed 0 – 1 RVs with a common probability of 1 s p ; i.e. $p = P(Z_i = 1) = 1 - P(Z_i = 0) = 1 - q$, $i = 1, 2, \dots, n$. Such a sequence, called a finite Bernoulli sequence, is of particular importance in studies of applied probability because of its simplicity and its help in understanding the notion of randomness, and also since it may be considered as a special case of a sequence with dependent elements; e.g. a Markovian or an exchangeable one (see, e.g. Makri & Psillakis, 2012). The latter feature also serves as a valuable crosscheck of results referring to several random sources and obtained by various methods.

For a Bernoulli sequence of length n the expected value of $X_{n,k,\ell}^{(\alpha)}$, $E(X_{n,k,\ell}^{(\alpha)}; p)$, $n \geq k > \ell \geq 0$, $\alpha = L, C$, is determined, via the indicators in (1) and (2), as follows.

Let $A = \{k, \ell + 2(k - \ell), \dots, \ell + \lfloor \frac{n-\ell}{k-\ell} \rfloor (k - \ell)\}$ and $B = \{k - \ell, 2(k - \ell), \dots, \lfloor \frac{n}{k-\ell} \rfloor (k - \ell)\}$. Since the expected values, $E(I_j^{(\alpha)})$, of the indicators $I_j^{(\alpha)}$, $\alpha = L, C$, depend on j , the sets A and B help in expressing them. Specifically, by the independence of Z_i s we have

$$E(I_j^{(L)}) = p^j I_A(j) + qp^\ell \sum_{s=1}^{\lfloor \frac{j-\ell}{k-\ell} \rfloor - I_A(j)} p^{s(k-\ell)}, j = k, k + 1, \dots, n$$

and

$$E(I_j^{(C)}) = p^n I_B(j) + qp^\ell \sum_{s=1}^{\lfloor \frac{n-1-\ell}{k-\ell} \rfloor} p^{s(k-\ell)}, j = 1, 2, \dots, n$$

where $I_A(j) = 1$, if $j \in A$; 0, otherwise and $I_B(j)$ is defined in a similar manner. Substituting $E(I_j^{(\alpha)})$ in

$$E(X_{n,k,\ell}^{(\alpha)}; p) = \sum_{j \in J^{(\alpha)}} E(I_j^{(\alpha)})$$

where $J^{(L)} = \{k, k + 1, \dots, n\}$, $J^{(C)} = \{1, 2, \dots, n\}$, we conclude after some algebraic manipulations that

$$E(X_{n,k,\ell}^{(L)}; p) = \frac{p^k}{1 - p^{k-\ell}} \{q[(n - k + 1) - (k - \ell) \frac{p^{k-\ell} - p^{(k-\ell)r}}{1 - p^{k-\ell}} - (s + 1)p^{(k-\ell)r}] + p(1 - p^{(k-\ell)r})\}, \tag{8}$$

where $r = \lfloor (n - \ell) / (k - \ell) \rfloor$ and $s = n - \ell - r(k - \ell)$, and

$$E(X_{n,k,\ell}^{(C)}; p) = nqp^k \frac{1 - p^{\lfloor \frac{n-\ell}{k-\ell} \rfloor (k-\ell)}}{1 - p^{k-\ell}} + \lfloor \frac{n}{k-\ell} \rfloor p^n. \tag{9}$$

Equations (8)-(9) have been derived by Makri and Psillakis (2013, Corollaries 2.1 and 3.1) using a slightly different formulation. These authors also provide expressions for $E(X_{n,k,\ell}^{(\alpha)})$ for sequences of independent not identically distributed binary RVs ordered on a line or on a circle. For $0 \leq \ell \leq k-1$, Equation (8) improves the expression given by Proposition 2.1 of Makri and Philippou (2005) which provides $E(X_{n,k,\ell}^{(L)}; p)$ by means of a summation. For $\ell = 0$ it provides a formula that is simpler than that of Aki and Hirano (1988) and Antzoulakos and Chadjiconstantinidis (2001). For $0 \leq \ell \leq k-1$, Equation (9) recaptures Proposition 3.1 of Makri and Philippou (2005) which for $\ell = 0$ coincides with a result of Charalambides (1994) and Makri and Philippou (1994).

For the particular case $\ell = 0$, an even simpler new expression for $E(X_{n,k,\ell}^{(L)}; p)$ is obtained by solving the recurrence relation provided by Balakrishnan and Koutras (2002, p. 163); i.e.

$$E(X_{n,k,0}^{(L)}; p) = p^k [E(X_{n-k,k,0}^{(L)}; p) + 1 + (n-k)q], \quad n \geq k. \quad (10)$$

Specifically, using standard techniques (see, e.g. Rosen, 1995, pp. 318-325) we obtain that

$$E(X_{n,k,0}^{(L)}; p) = \frac{p^k}{1-p^k} [(1-p^n)(1 - \frac{kq}{1-p^k}) + nq], \quad n \geq k. \quad (11)$$

For $\ell = k-1$, Equations (8)-(9) imply that

$$E(X_{n,k,k-1}^{(L)}; p) = (n-k+1)p^k, \quad E(X_{n,k,k-1}^{(C)}; p) = np^k. \quad (12)$$

That is, by Equation (12) the formula for $E(X_{n,k,k-1}^{(L)}; p)$ provided by Ling (1988) and the expression for $E(X_{n,k,k-1}^{(C)}; p)$ given in Proposition 2.2 of Makri and Philippou (1994), are recaptured.

For $k=1, \ell=0, X_{n,k,\ell}^{(\alpha)}, \alpha=L, C$ is a $B(n, p)$ with $P(X_{n,1,0}^{(\alpha)} = x) = \binom{n}{x} p^x (1-p)^{n-x}, x \in S_{n,1,0}^{(\alpha)} = \{0, 1, \dots, n\}$, hence, $E(X_{n,1,0}^{(\alpha)}; p) = np$. The latter one is confirmed as a special case of Equations (8)-(9), (11)-(12).

2.2 Symmetric Bernoulli Sequences

For a Bernoulli sequence of length n with probability of 1 s p , let $f_{n,k,\ell}^{(\alpha)}(x; p)$ denote the probability mass function (PMF) of the RV $X_{n,k,\ell}^{(\alpha)}$; i.e.

$$f_{n,k,\ell}^{(\alpha)}(x; p) = P(X_{n,k,\ell}^{(\alpha)} = x), \quad x \in S_{n,k,\ell}^{(\alpha)}, \quad \alpha = L, C. \quad (13)$$

Then, the expected value of $X_{n,k,\ell}^{(\alpha)}$ is

$$E(X_{n,k,\ell}^{(\alpha)}; p) = \sum_{x \in S_{n,k,\ell}^{(\alpha)}} x f_{n,k,\ell}^{(\alpha)}(x; p), \quad \alpha = L, C. \quad (14)$$

In the sequel, we consider a symmetric ($p = 1/2$) finite Bernoulli sequence (i.e. a finite binary string) of length n for which we obtain our main results. Since the cardinality of a proper sample space is 2^n (i.e. there are 2^n binary strings that are equally likely to occur) the classical definition of probability implies that

$$f_{n,k,\ell}^{(\alpha)}(x; 1/2) = N_{x;n,k,\ell}^{(\alpha)} / 2^n, \quad x \in S_{n,k,\ell}^{(\alpha)}, \quad \alpha = L, C. \quad (15)$$

In the above formula, $N_{x;n,k,\ell}^{(\alpha)}$ is the number of all binary strings of length n with exactly x , $x \in S_{n,k,\ell}^{(\alpha)}$, ℓ -overlapping 1-runs of length k , among all 2^n (possible) binary strings of length n , $n \geq k > \ell \geq 0$, ordered linearly ($\alpha = L$) or circularly ($\alpha = C$). In other words, $N_{x;n,k,\ell}^{(\alpha)}$ is the number of 0-1 strings of length n so that $X_{n,k,\ell}^{(\alpha)} = x$. That is,

$$N_{x;n,k,\ell}^{(\alpha)} = \sum_{\mathbf{z}=(z_1, z_2, \dots, z_n), z_i \in \{0,1\}} I(\mathbf{z}), \quad x \in S_{n,k,\ell}^{(\alpha)}, \quad (16)$$

where

$$I(\mathbf{z}) = \begin{cases} 1, & \text{if } X_{n,k,\ell}^{(\alpha)} = x \text{ on } \mathbf{z} \\ 0, & \text{otherwise,} \end{cases}$$

and the values of $X_{n,k,\ell}^{(\alpha)}$ are determined via (1)-(2).

Equation (16) implies that if we wish to empirically determine $N_{x;n,k,\ell}^{(\alpha)}$ we have to generate (by a computer) all 2^n 0 – 1 strings of length n and then count among all 2^n strings those for which it holds $X_{n,k,\ell}^{(\alpha)} = x$. This approach, i.e. first listing and then counting, although possible and useful too in several applications like the ones mentioned in the Introduction, does not determine theoretically the numbers $N_{x;n,k,\ell}^{(\alpha)}$. It is the combinatorial analysis that gives answers in such problems since it counts arrangements of things without listing them.

Next closed expressions for $N_{x;n,k,\ell}^{(\alpha)}$, $x \in S_{n,k,\ell}^{(\alpha)}$, $\alpha = L, C$, in terms of sums of binomial coefficients, are provided. They are directly obtained by Equation (15) and by Theorems 2.1 and 4.1 of Makri et al. (2007a). The latter Theorems provide the PMF of $X_{n,k,\ell}^{(\alpha)}$ defined on a binary sequence derived by a Polya-Eggenberger urn model of which a Bernoulli sequence and in turns a binary string are particular cases. Their method was based on the solution of a combinatorial problem; specifically, the allocation of indistinguishable balls into distinguishable cells under certain restrictions about the capacity of the cells. Accordingly, we have:

$$N_{0;n,k,\ell}^{(\alpha)} = \begin{cases} \sum_{y=\lfloor n/k \rfloor}^n C_R(n-y, y+1, k-1), & \alpha = L \\ \sum_{y=\lfloor n/k \rfloor}^n \binom{n}{y} C_R(n-y, y, k-1), & \alpha = C \end{cases}$$

and for $x \in S_{n,k,\ell}^{(\alpha)} - \{0\}$

$$N_{x;n,k,\ell}^{(\alpha)} = \begin{cases} \sum_{y=y_1^L}^{y_2^L} \sum_{i=1}^{\lfloor \frac{n-y}{k} \rfloor} \binom{y+1}{i} \binom{x-1}{i-1} C_M(\gamma_i, i, y+1-i, k-\ell-1, k-1), & \alpha = L \\ \sum_{y=y_1^C}^{y_2^C} \binom{n}{y} \sum_{i=1}^y \binom{y}{i} \binom{x-1}{i-1} C_M(\gamma_i, i, y-i, k-\ell-1, k-1) + \delta_{x, \lfloor \frac{n}{k} \rfloor}, & \alpha = C \end{cases} \tag{17}$$

with $y_1^L = \lfloor \frac{n+x\ell}{k} \rfloor - x$, $y_1^C = 1$, $y_2 = n - k - (x - 1)(k - \ell)$, $\gamma_i = n - y - ik - (x - i)(k - \ell)$. It is noticed that the RHS of (17) for $x = 0$ is the same for every ℓ , $0 \leq \ell \leq k - 1$, since $P(X_{n,k,\ell}^{(\alpha)} = 0) = P(X_{n,k,0}^{(\alpha)} = 0)$, hence $N_{0;n,k,\ell}^{(\alpha)} = N_{0;n,k,0}^{(\alpha)}$, $0 < \ell < k$, $\alpha = L, C$, by (15).

The coefficients C_M and C_R are

$$C_M(\gamma, i, r-i, m_1-1, m_2-1) = \begin{cases} 0, & \gamma < 0 \\ 1, & \gamma = 0 \\ \sum_{j_1=0}^{\lfloor \frac{\gamma}{m_1} \rfloor} \binom{i}{j_1} \sum_{j_2=0}^{\lfloor \frac{\gamma-m_1 j_1}{m_2} \rfloor} (-1)^{j_1+j_2} \binom{r-i}{j_2} \binom{\gamma-m_1 j_1-m_2 j_2+r-1}{\gamma-m_1 j_1-m_2 j_2}, & \gamma > 0 \end{cases} \tag{18}$$

and

$$C_R(\gamma, r, m_1-1) \equiv C_M(\gamma, i, r-i, m_1-1, m_1-1) = \begin{cases} 0, & \gamma < 0 \\ 1, & \gamma = 0 \\ \sum_{j=0}^{\lfloor \frac{\gamma}{m_1} \rfloor} (-1)^j \binom{r}{j} \binom{\gamma-m_1 j+r-1}{\gamma-m_1 j}, & \gamma > 0 \end{cases}$$

by Lemma 2.1 and Corollary 2.1 of Makri et al. (2007a), respectively.

The number $C_M(\gamma, i, r-i, m_1-1, m_2-1)$ is the number of integer solutions of the equation $x_1 + x_2 + \dots + x_r = \gamma$, s.t. $0 \leq x_j < m_1$, $1 \leq j \leq i$; $0 \leq x_{i+j} < m_2$, $1 \leq j \leq r-i$. Equivalently, it gives the number of allocations of γ indistinguishable balls into r distinguishable cells, of which i specified cells have a capacity of $m_1 - 1$, and each of the remaining $r - i$ cells have a capacity of $m_2 - 1$. Consequently, $C_R(\gamma, r, m_1 - 1)$ represents the number of allocations of γ indistinguishable balls into r distinguishable cells, each cell with capacity $m_1 - 1$. This is so because, in the case of C_R , there are not specified cells and all $i + (r - i) = r$ cells have the same capacity $m_1 - 1$.

In Tables 1 and 2 we compute, via (17)-(18), and present the numbers $N_{x;n,k,\ell}^{(\alpha)}$, $x \in S_{n,k,\ell}^{(\alpha)}$, for $n = 4$, $1 \leq k \leq n$, $0 \leq \ell \leq k - 1$ and $\alpha = L, C$. The value $n = 4$ was chosen small so that the computation can also be done by hand like the calculation of $N_{1;4,2,1}^{(L)} = 5$ in the indicative Example 3. Tables 1 and 2 also provide the sets $S_{n,k,\ell}^{(\alpha)}$, determined by (3), by just looking at the non empty entries of them and then notifying the respective x s. For instance, $S_{4,2,1}^{(L)} = \{0, 1, 2, 3\}$ (Example 3) and $S_{4,2,1}^{(C)} = \{0, 1, 2, 4\}$. Furthermore, the entries of the columns of the Tables labeled $x = 0$ confirm that $N_{0;4,k,\ell}^{(\alpha)} = N_{0;4,k,0}^{(\alpha)}$, for $1 \leq k \leq 4$, $\ell = 1, 2, \dots, k - 1$, $\alpha = L, C$ as it is expected.

Table 1. Distribution of $N_{x;4,k,\ell}^{(L)}$, $x \in S_{4,k,\ell}^{(L)}$ for run length size k , $1 \leq k \leq 4$ and overlapping (common) size ℓ , $0 \leq \ell \leq k - 1$

k	ℓ	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
1	0	1	4	6	4	1
2	0	8	7	1		
	1	8	5	2	1	
3	0	13	3			
	1	13	3			
	2	13	2	1		
4	0	15	1			
	1	15	1			
	2	15	1			
	3	15	1			

Table 2. Distribution of $N_{x;n,k,\ell}^{(C)}$, $x \in S_{4,k,\ell}^{(C)}$ for run length size k , $1 \leq k \leq 4$ and overlapping (common) size ℓ , $0 \leq \ell \leq k - 1$

k	ℓ	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
1	0	1	4	6	4	1
2	0	7	8	1		
	1	7	4	4		1
3	0	11	5			
	1	11	4	1		
	2	11	4			1
4	0	15	1			
	1	15	1			
	2	15		1		
	3	15				1

The calculation of the entries of Tables 1 and 2, showing the whole class of $N_{x;n,k,\ell}^{(\alpha)}$ for $n = 4$, has been done by a computer. The same approach can easily be followed for larger ns . But, because of space limitations it is not possible to depict all these numbers in the paper for large ns . Instead of it, and simultaneously to give a sense of the values of $N_{x;n,k,\ell}^{(\alpha)}$ even for moderate ns , let us consider Example 4.

Example 4. Let $k = 3$, $\ell = 0, 1, 2$, $x = 0, 1$, $\alpha = L, C$ and $n = 16, 32$. Then, by Equations (17)-(18), we easily compute that: $N_{0;16,3,\ell}^{(L)} = 19513$ for $\ell = 0, 1, 2$, $N_{1;16,3,0}^{(L)} = 27820$, $N_{1;16,3,1}^{(L)} = 23780$, $N_{1;16,3,2}^{(L)} = 15776$; $N_{0;16,3,\ell}^{(C)} = 17155$ for $\ell = 0, 1, 2$, $N_{1;16,3,0}^{(C)} = 27280$, $N_{1;16,3,1}^{(C)} = 22896$, $N_{1;16,3,2}^{(C)} = 14832$ and $N_{0;32,3,\ell}^{(L)} = 334745777$ for $\ell = 0, 1, 2$, $N_{1;32,3,0}^{(L)} = 1009569304$, $N_{1;32,3,1}^{(L)} = 854712776$, $N_{1;32,3,2}^{(L)} = 560052736$; $N_{0;32,3,\ell}^{(C)} = 294294531$ for $\ell = 0, 1, 2$, $N_{1;32,3,0}^{(C)} = 935981600$, $N_{1;32,3,1}^{(C)} = 785556960$, $N_{1;32,3,2}^{(C)} = 508882912$.

The (total) number, $R_{n,k,\ell}^{(\alpha)}$, of occurrences of all ℓ -overlapping 1-runs of length k , in all possible 2^n binary strings of length n , $n \geq k > \ell \geq 0$, ordered linearly ($\alpha = L$) or circularly ($\alpha = C$) can be defined via $N_{x;n,k,\ell}^{(\alpha)}$ as

$$R_{n,k,\ell}^{(\alpha)} = \sum_{x \in S_{n,k,\ell}^{(\alpha)}} x N_{x;n,k,\ell}^{(\alpha)}, \quad n \geq k > \ell \geq 0, \quad \alpha = L, C. \tag{19}$$

Alternatively, $R_{n,k,\ell}^{(\alpha)}$ can be obtained without the usage of $N_{x;n,k,\ell}^{(\alpha)}$. This is so, because

$$E(X_{n,k,\ell}^{(\alpha)}; 1/2) = \sum_{x \in S_{n,k,\ell}^{(\alpha)}} x f_{n,k,\ell}^{(\alpha)}(x; 1/2),$$

hence by (15) and (19),

$$R_{n,k,\ell}^{(\alpha)} = 2^n E(X_{n,k,\ell}^{(\alpha)}; 1/2). \tag{20}$$

Therefore, for $n \geq k > \ell \geq 0$, Equations (8)-(9), (11)-(12) and (20) imply that

$$R_{n,k,\ell}^{(L)} = \begin{cases} n2^{n-1}, & k = 1, \ell = 0 \\ \frac{2^{n-k}}{1-2^{-k}} \left\{ (1-2^{-n}) \left[1 - \frac{k}{2(1-2^{-k})} \right] + \frac{n}{2} \right\}, & 2 \leq k \leq n, \ell = 0 \\ (n-k+1)2^{n-k}, & 2 \leq k \leq n, \ell = k-1 \\ \frac{2^{n-k}}{1-2^{-(k-\ell)}} \left\{ 2^{-1} \left[(n-k+1) - (k-\ell) \frac{2^{-(k-\ell)} - 2^{-(k-\ell)r}}{1-2^{-(k-\ell)}} \right] - (s+1)2^{-(k-\ell)r} \right\} + 2^{-1} (1-2^{-(k-\ell)r}), & \text{otherwise} \end{cases} \quad (21)$$

with $r = \lfloor (n-\ell)/(k-\ell) \rfloor$, $s = n-\ell-r(k-\ell)$, and

$$R_{n,k,\ell}^{(C)} = \begin{cases} n2^{n-1}, & k = 1, \ell = 0 \\ n2^{n-k}, & 2 \leq k \leq n, \ell = k-1 \\ n2^{n-k-1} \frac{1-2^{-\lfloor \frac{n-\ell-1}{k-\ell} \rfloor (k-\ell)}}{1-2^{-(k-\ell)}} + \lfloor \frac{n}{k-\ell} \rfloor, & \text{otherwise.} \end{cases} \quad (22)$$

In brief, the numbers $N_{x;n,k,\ell}^{(\alpha)}$ and $R_{n,k,\ell}^{(\alpha)}$ are associated with the PMF and the mean of $X_{n,k,\ell}^{(\alpha)}$. Accordingly, they are interrelated and $R_{n,k,\ell}^{(\alpha)}$ can be computed by (19), which uses (17)-(18), or directly and efficiently too (even with the use of a normal pocket calculator) by the explicit expressions (21)-(22).

By symmetry, $R_{n,k,\ell}^{(\alpha)}$ provide the respective numbers associated with ℓ -overlapping 0-runs of length k , in all 2^n binary strings of length n , $n \geq k > \ell \geq 0$, $\alpha = L, C$.

It is noticed that $R_{n,k,\ell}^{(\alpha)}$, $\alpha = L, C$, is decreasing in k and increasing in n and ℓ . That is, $R_{n,k,\ell_1}^{(\alpha)} \leq R_{n,k,\ell_2}^{(\alpha)}$, $0 \leq \ell_1 \leq \ell_2 < k \leq n$, $R_{n,k_1,\ell}^{(\alpha)} \geq R_{n,k_2,\ell}^{(\alpha)}$, $0 \leq \ell \leq k_1 - 1$, $1 \leq k_1 \leq k_2 \leq n$ and $R_{n_1,k,\ell}^{(\alpha)} \leq R_{n_2,k,\ell}^{(\alpha)}$, $0 \leq \ell < k \leq n_1 \leq n_2$. This is so, because the numbers $R_{n,k,\ell}^{(\alpha)}$ follow in general the behavior of statistics $X_{n,k,\ell}^{(\alpha)}$, $\alpha = L, C$. The latter feature immediately implies that $R_{n,k,\ell}^{(L)} \leq R_{n,k,\ell}^{(C)}$, $0 \leq \ell < k \leq n$.

Tables 3 and 4 present the numbers $R_{n,k,\ell}^{(L)}$ and $R_{n,k,\ell}^{(C)}$ for binary strings of length $n = 2^t$, $t = 1, 2, 3$ and for $1 \leq k \leq n$, $0 \leq \ell \leq k - 1$. As an illustration notice that (see also Example 3) $R_{4,2,1}^{(L)} = 12$ which is directly computed by the third of (21). The entries of Tables 3 and 4, efficiently computed by (21)-(22), also confirm the previously noticed ordering among the depicted numbers. Furthermore, they suggest that the numbers $R_{n,k,\ell}^{(L)}$ and $R_{n,k,\ell}^{(C)}$ decrease exponentially as k increases tending to n . They eventually tend to $R_{n,n,\ell}^{(L)} = 1$, $\ell = 0, 1, \dots, n - 1$ and $R_{n,n,\ell}^{(C)} = \lfloor \frac{n}{n-\ell} \rfloor$, $\ell = 0, 1, \dots, n - 1$, respectively.

Finally, in order to get a sense of the values of the numbers $R_{n,k,\ell}^{(\alpha)}$ even for moderate ns , we compute and present some such numbers in Example 5 which continues Example 4.

Example 5. Let $k = 3$, $\ell = 0, 1, 2$, $\alpha = L, C$ and $n = 16, 32$. Then, Equations (21)-(22) directly provide that: $R_{16,3,0}^{(L)} = 68211$, $R_{16,3,1}^{(L)} = 78279$, $R_{16,3,2}^{(L)} = 114688$; $R_{16,3,0}^{(C)} = 74901$, $R_{16,3,1}^{(C)} = 87384$, $R_{16,3,2}^{(C)} = 131072$ and $R_{32,3,0}^{(L)} = 9378806136$, $R_{32,3,1}^{(L)} = 10856722887$, $R_{32,3,2}^{(L)} = 16106127360$; $R_{32,3,0}^{(C)} = 9817068106$, $R_{32,3,1}^{(C)} = 11453246128$, $R_{32,3,2}^{(C)} = 17179869184$.

Table 3. Number of occurrences of ℓ -overlapping 1-runs of length k , $R_{n,k,\ell}^{(L)}$ in binary strings of length n ordered on a line

n	k	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$
2	1	4							
	2	1	1						
4	1	32							
	2	9	12						
	3	3	3	4					
	4	1	1	1	1				
8	1	1024							
	2	313	448						
	3	120	135	192					
	4	49	51	57	80				
	5	20	20	21	23	32			
	6	8	8	8	8	9	12		
	7	3	3	3	3	3	3	4	
	8	1	1	1	1	1	1	1	1

Table 4. Number of occurrences of ℓ -overlapping 1-runs of length k , $R_{n,k,\ell}^{(C)}$ in binary strings of length n ordered on a circle

n	k	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$
2	1	4							
	2	1	2						
4	1	32							
	2	10	16						
	3	5	6	8					
	4	1	1	2	4				
8	1	1024							
	2	340	512						
	3	146	172	256					
	4	66	74	84	128				
	5	33	34	34	44	64			
	6	17	17	18	18	20	32		
	7	9	9	9	10	10	12	16	
	8	1	1	1	1	2	2	4	8

3. Concluding Review

In this paper we concerned 0 – 1 sequences of finite length n , ordered on a line or on a circle. For such sequences we considered run statistics, $X_{n,k,\ell}^{(\alpha)}$, $\alpha = L(\text{line}), C(\text{circle})$, associated with counting of 1-runs of a fixed length k , assuming that these runs may have overlapping parts of a given length ℓ , $0 \leq \ell < k \leq n$. Other schemes, along with the associated statistics $Y_{n,k}^{(\beta,\alpha)}$, counting runs of 1s of length exactly equal to k ($\beta = E$) and greater than or equal to k ($\beta = G$) are also mentioned. Relationships among the statistics $X_{n,k,\ell}^{(\alpha)}$ and $Y_{n,k}^{(\beta,\alpha)}$ are discussed. $X_{n,k,\ell}^{(\alpha)}$ and $Y_{n,k}^{(\beta,\alpha)}$ are defined on any binary sequence ordered on a line or on a circle via appropriate indicators. Using the latter ones, in Section 2.1 the expected values, $E(X_{n,k,\ell}^{(\alpha)}; p)$, of $X_{n,k,\ell}^{(\alpha)}$ for Bernoulli sequences with probability of 1s p were expressed. The formulae (8)-(9) for $E(X_{n,k,\ell}^{(\alpha)}; p)$, $\alpha = L, C$ are simple, computationally efficient and also general. According to the latter feature some known results for $E(X_{n,k,\ell}^{(\alpha)}; p)$, $\alpha = L, C$ and $\ell = 0, k - 1$ are recaptured as special cases of the unified approach. An alternative expression of $E(X_{n,k,0}^{(L)}; p)$, given by (11), which is not implied for $\ell = 0$ by the general formula (8) for $E(X_{n,k,\ell}^{(L)}; p)$ is obtained.

After the review in Section 2.1 on some results concerning $E(X_{n,k,\ell}^{(\alpha)}; p)$, the numbers $N_{x;n,k,\ell}^{(\alpha)}$ and $R_{n,k,\ell}^{(\alpha)}$, $\alpha = L, C$ are introduced and studied on binary strings (symmetric Bernoulli sequences) in Section 2.2. $R_{n,k,\ell}^{(\alpha)}$ is the number of the ℓ -overlapping runs of 1s of length k in all 2^n binary strings of length n and is associated with $E(X_{n,k,\ell}^{(\alpha)}; 1/2)$. $N_{x;n,k,\ell}^{(\alpha)}$ is the number of those strings, among the 2^n strings, which contain x ℓ -overlapping runs of 1s of length k and is related with the PMF of $X_{n,k,\ell}^{(\alpha)}$. Accordingly, for $N_{x;n,k,\ell}^{(\alpha)}$ we give, by Equations (17)-(18), closed expressions in terms of sums of binomial coefficients whereas for $R_{n,k,\ell}^{(\alpha)}$ we provide, by Equations (21)-(22), explicit closed expressions in addition to those, i.e. Equations (19), that can be derived using $N_{x;n,k,\ell}^{(\alpha)}$. The expressions so obtained clarify further the interdependencies among the numbers $N_{x;n,k,\ell}^{(\alpha)}$ and $R_{n,k,\ell}^{(\alpha)}$ which might be potentially useful in applications like those mentioned in the Introduction.

Acknowledgements

The authors would like to thank the Editor and the two anonymous reviewers for the thorough reading, useful comments and suggestions which helped to improve the paper.

References

- Aki, S., & Hirano, K. (2000). Numbers of success-runs of specified length until certain stopping time rules and generalized binomial distributions of order k . *Annals of the Institute of Statistical Mathematics*, 52, 767-777. <http://dx.doi.org/10.1023/A:1017585512412>
- Antzoulakos, D. L., & Chadjiconstantinidis, S. (2001). Distribution of numbers of success runs of fixed length in Markov dependent trials. *Annals of the Institute of Statistical Mathematics*, 53, 599-619.

<http://dx.doi.org/10.1023/A:1014689617413>

- Balakrishnan, N., & Koutras, M. V. (2002). *Runs and Scans with Applications*. New York: Wiley.
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, 27, 573-580. <http://dx.doi.org/10.1093/nar/27.2.573>
- Charalambides, C. A. (1994). Success runs in a circular sequence of independent Bernoulli trials. In A. P. Godbole & S. G. Papastavridis (Eds.), *Runs and Patterns in Probability* (pp. 15-30). Amsterdam: Kluwer. http://dx.doi.org/10.1007/978-1-4613-3635-8_2
- Demir Atalay, S., & Zeybek, M. (2013). Circular success and failure runs in a sequence of exchangeable binary trials. *Journal of Statistical Planning and Inference*, 143, 621-629. <http://dx.doi.org/10.1016/j.jspi.2012.07.016>
- Eryilmaz, S. (2008). Run statistics defined on a multicolor urn model. *Journal of Applied Probability*, 45, 1007-1023. <http://dx.doi.org/10.1239/jap/1231340230>
- Eryilmaz, S. (2011). Joint distribution of run statistics in partially exchangeable processes. *Statistics and Probability Letters*, 81, 163-168. <http://dx.doi.org/10.1016/j.spl.2010.10.014>
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (Vol. I, 3rd ed.). New York: John Wiley.
- Fu, J. C., & Lou, W. Y. W. (2003). *Distribution Theory of Runs and Patterns and its Applications: A Finite Markov Chain Imbedding Approach*. New Jersey: World Scientific.
- Han, S., & Aki, S. (2000). A unified approach to binomial-type distributions of order k . *Communications in Statistics-Theory and Methods*, 29, 1929-1943. <http://dx.doi.org/10.1080/03610920008832586>
- Hirano, K. (1986). Some properties of the distributions of order k . In A. N. Philippou, A. F. Horadam, & G. E. Bergum (Eds.), *Fibonacci Numbers and Their Applications* (pp. 43-53). Dordrecht: Reidel. http://dx.doi.org/10.1007/978-94-009-4311-7_4
- Hirano, K., Aki, S., Kashiwagi, N., & Kuboki, H. (1991). On Ling's binomial and negative binomial distributions of order k . *Statistics and Probability Letters*, 11, 503-509. [http://dx.doi.org/10.1016/0167-7152\(91\)90115-8](http://dx.doi.org/10.1016/0167-7152(91)90115-8)
- Holst, L. (2009). On consecutive records in certain Bernoulli sequences. *Journal of Applied Probability*, 46, 1201-1208. <http://dx.doi.org/10.1239/jap/1261670698>
- Inoue, K., & Aki, S. (2003). Generalized binomial and negative binomial distributions of order k by the ℓ -overlapping enumeration scheme. *Annals of the Institute of Statistical Mathematics*, 55, 153-167. <http://dx.doi.org/10.1007/BF02530491>
- Inoue, K., & Aki, S. (2010). On the conditional and unconditional distributions of the number of success runs on a circle with applications. *Statistics and Probability Letters*, 80, 874-885. <http://dx.doi.org/10.1016/j.spl.2010.01.022>
- Koutras, M. V. (2003). Applications of Markov chains to the distribution theory of runs and patterns. In D. N. Shanbhag & C. R. Rao (Eds.), *Stochastic Processes: Modelling and Simulation*, Handbook Statist., 21, 431-472. Amsterdam: North-Holland.
- Koutras, M. V., Papadopoulos, G. K., & Papastavridis, S. G. (1995). Runs on a circle. *Journal of Applied Probability*, 32, 396-404. <http://dx.doi.org/10.2307/3215295>
- Ling, K. D. (1988). On binomial distributions of order k . *Statistics and Probability Letters*, 6, 247-250. [http://dx.doi.org/10.1016/0167-7152\(88\)90069-7](http://dx.doi.org/10.1016/0167-7152(88)90069-7)
- Lou, W. Y. W. (2003). The exact distribution of the k -tuple statistic for sequence homology. *Statistics and Probability Letters*, 61, 51-59. [http://dx.doi.org/10.1016/S0167-7152\(02\)00337-1](http://dx.doi.org/10.1016/S0167-7152(02)00337-1)
- Makri, F. S., & Philippou, A. N. (1994). Binomial distributions of order k on the circle. In A. P. Godbole & S. G. Papastavridis (Eds.), *Runs and Patterns in Probability* (pp. 65-81). Amsterdam: Kluwer. http://dx.doi.org/10.1007/978-1-4613-3635-8_5
- Makri, F. S., & Philippou, A. N. (2005). On binomial and circular binomial distributions of order k for ℓ -overlapping success runs of length k . *Statistical Papers*, 46, 411-432. <http://dx.doi.org/10.1007/BF02762842>

- Makri, F. S., Philippou, A. N., & Psillakis, Z. M. (2007a). Polya, inverse Polya, and circular Polya distributions of order k for ℓ -overlapping success runs. *Communications in Statistics-Theory and Methods*, *36*, 657-668. <http://dx.doi.org/10.1080/03610920601033942>
- Makri, F. S., Philippou, A. N., & Psillakis, Z. M. (2007b). Success run statistics defined on an urn model. *Advances in Applied Probability*, *39*, 991-1019. <http://dx.doi.org/10.1239/aap/1198177236>
- Makri, F. S., & Psillakis, Z. M. (2011a). On runs of length exceeding a threshold: normal approximation. *Statistical Papers*, *52*, 531-551. <http://dx.doi.org/10.1007/s00362-009-0268-y>
- Makri, F. S., & Psillakis, Z. M. (2011b). On success runs of a fixed length in Bernoulli sequences: exact and asymptotic results. *Computers and Mathematics with Applications*, *61*, 761-772. <http://dx.doi.org/10.1016/j.camwa.2010.12.023>
- Makri, F. S., & Psillakis, Z. M. (2012). Exact distributions of constrained (k, ℓ) strings of failures between subsequent successes. *Statistical Papers*. <http://dx.doi.org/10.1007/s00362-012-0462-1>
- Makri, F. S., & Psillakis, Z. M. (2013). On ℓ -overlapping runs of ones of length k in sequences of independent binary random variables. *Communications in Statistics-Theory and Methods*. Accepted for publication.
- Makri, F. S., Psillakis, Z. M., & Kollas, N. (2012). Counting runs of ones and ones in runs of ones in binary strings. *Open Journal of Applied Sciences*. Supplement: 2012 World Congress in Engineering and Technology, 44-47.
- Mood, A. M. (1940). The distribution theory of runs. *Annals of Mathematical Statistics*, *11*, 367-392. <http://dx.doi.org/10.1214/aoms/1177731825>
- Mytalas, G. C., & Zazanis, M. A. (2013). Central limit theorem approximations for the number of runs in Markov-dependent binary sequences. *Journal of Statistical Planning and Inference*, *143*, 321-333. <http://dx.doi.org/10.1016/j.jspi.2012.07.014>
- Nuel, G., Regad, L., Martin, J., & Camproux, A-C. (2010). Exact distribution of a pattern in a set of random sequences generated by a Markov source: applications to biological data. *Algorithms for Molecular Biology*, *5*, 1-18. <http://dx.doi.org/10.1186/1748-7188-5-15>
- Philippou, A. N., & Makri, F. S. (1986). Successes, runs and longest runs. *Statistics and Probability Letters*, *4*, 211-215. [http://dx.doi.org/10.1016/0167-7152\(86\)90069-6](http://dx.doi.org/10.1016/0167-7152(86)90069-6)
- Rosen, K. H. (1995). *Discrete Mathematics and Its Applications*. New York: McGraw-Hill, Inc.
- Sinha, K. (2007). Location and communication issues in mobile networks. (Ph.D. dissertation, Department of Computer Science and Engineering, Jadavpur University, Calcutta, India).
- Sinha, K., & Sinha, B. P. (2009). On the distribution of runs of ones in binary strings. *Computers and Mathematics with Applications*, *58*, 1816-1829. <http://dx.doi.org/10.1016/j.camwa.2009.07.057>
- Sinha, K., & Sinha, B. P. (2012). Energy efficient communication: understanding the distribution of runs in binary strings. *1st International Conference on Recent Advances in Information Technology (RAIT-2012)*, 177-181.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).