

## ON OCCURRENCES OF $F$ - $S$ STRINGS IN LINEARLY AND CIRCULARLY ORDERED BINARY SEQUENCES

FROSSO S. MAKRI,\* *University of Patras*

### Abstract

Consider a sequence of exchangeable or independent binary trials ordered on a line or on a circle. The statistics denoting the number of times an  $F$ - $S$  string of length (at least)  $k_1 + k_2$ , that is, (at least)  $k_1$  failures followed by (at least)  $k_2$  successes in  $n$  such trials, are studied. The associated waiting time for the  $r$ th occurrence of an  $F$ - $S$  string of length (at least)  $k_1 + k_2$  in linearly ordered trials is also examined. Exact formulae, lower/upper bounds and approximations are derived for their distributions. Mean values and variances of the number of occurrences of  $F$ - $S$  strings are given in exact formulae too. Particular exchangeable and independent sequences of binary random variables, used in applied research, combined with numerical examples clarify further the theoretical results.

*Keywords:* Distribution of order  $(k_1, k_2)$ ; linear and circular binary sequences; exchangeable trial; Poisson trial; urn model; records; runs

2000 Mathematics Subject Classification: Primary 60E05; 62E15

Secondary 60C05; 60G09; 60F05

### 1. Introduction

The study of runs and run-related statistics has attracted much attention in the literature because of their wide range of applications in many areas, including statistical hypothesis testing, reliability theory and quality control, molecular biology, and computer science. In particular, special attention has been devoted to the number of runs, under certain enumerative schemes, defined on binary sequences of several internal structures as well as to the variety of techniques that have been used to derive their probability distributions. Since Feller (1968), Philippou and Muwaffi (1982), and Philippou *et al.* (1983) shed light on the number of runs of fixed length  $k$  and the associated waiting times defined on binary sequences, that is, the discrete distributions of order  $k$ , a series of articles has been published on the subject. Past and current developments in the area along with the relevant nomenclature are well documented in Balakrishnan and Koutras (2002) as well as in Fu and Lou (2003). Recent studies on the topic are included among others in the works of Antzoulakos *et al.* (2003), Koutras (2003), Makri and Philippou (2005), Eryilmaz and Demir (2007), Makri *et al.* (2007a), (2007b), Demir and Eryilmaz (2008), and Eryilmaz (2008), (2009).

Let  $X_1, X_2, \dots$ , be a sequence of binary trials resulting in either a success (denoted by  $S$  or 1) or a failure (denoted by  $F$  or 0). According to Feller's nonoverlapping enumeration scheme, once  $k$  ( $k > 0$ ) consecutive  $S$ s show up a success run of length  $k$  is counted and the enumeration procedure starts anew (from scratch). The number of nonoverlapping success runs of length  $k$  in  $n$  ( $n > 0$ ) trials is denoted by  $N_{n,k}$ , whereas the waiting time for the  $r$ th ( $r \geq 1$ ) appearance

Received 17 November 2009; revision received 25 January 2010.

\* Postal address: Department of Mathematics, University of Patras, 26500 Patras, Greece.

Email address: makri@math.upatras.gr

of a success run of length  $k$  is denoted by  $T_{r,k}$ . A generalization of the pre-mentioned scheme was introduced in Huang and Tsai (1991), who also addressed two possible applications: the security installation of a safe box and the process of random mating. Let  $k_1$  and  $k_2$  be two nonnegative integers with  $(k_1, k_2) \neq (0, 0)$ . We say that an  $F$ - $S$  string of length (at least)  $k_1 + k_2$ , a  $(k_1, k_2)$ -event in their terminology, has occurred if (at least)  $k_1$  consecutive  $F$ s are followed by (at least)  $k_2$  consecutive  $S$ s; that is, an

$$\underbrace{FF \dots F}_{\geq k_1} \underbrace{SS \dots S}_{\geq k_2}$$

pattern appears. Let  $N_{n;k_1,k_2}$  denote the number of occurrences of such strings in  $n$  binary trials. Then, the support of  $N_{n;k_1,k_2}$  is the set  $S(n, k_1, k_2) = \{0, 1, \dots, \lfloor n/(k_1 + k_2) \rfloor\}$ , where by  $\lfloor x \rfloor$  we denote the greatest integer less than or equal to  $x$ . The latter authors also suggested looking at the waiting time, say  $T_{r;k_1,k_2}$ , of the  $r$ th ( $r \geq 1$ ) occurrence of an  $F$ - $S$  string of length (at least)  $k_1 + k_2$ . Readily,  $N_{n;0,k} = N_{n,k}$  and  $T_{r;0,k} = T_{r,k}$ . Furthermore, for  $k_1 = 1$  and  $k_2 = k$ , we obtain the random variable (RV) that counts the number of occurrences of the pattern  $FSS \dots S$  (at least  $k$   $S$ s) in  $n$  trials. It has been studied in Godbole and Schaffner (1993) and plays, in addition to its independent merit, an important role in obtaining Poisson approximations of other run-related statistics (see, e.g. Balakrishnan and Koutras (2002, pp. 170–179)).

The random variables  $T_{r;k_1,k_2}$  and  $N_{n;k_1,k_2}$  are associated and they are related via the dual relationship

$$T_{r;k_1,k_2} > n \quad \text{if and only if} \quad N_{n;k_1,k_2} < r, \quad r \in S(n, k_1, k_2) - \{0\}. \quad (1.1)$$

Hence, (1.1) offers an alternative way of obtaining results for  $T_{r;k_1,k_2}$  through formulae established for  $N_{n;k_1,k_2}$ , and vice versa.

For Bernoulli trials (independent and identically distributed binary trials), Huang and Tsai (1991) called the distribution of  $N_{n;k_1,k_2}$  a binomial distribution of order  $(k_1, k_2)$  and denoted it by  $B_{k_1,k_2}(n, p)$ . Apparently,  $B_{0,k}(n, p)$  is identical to  $B_k(n, p)$ , that is, the binomial distribution of order  $k$  (see, e.g. Hirano (1986) and Philippou and Makri (1986)), which generalized the ordinary binomial distribution  $B(n, p)$  or  $B_1(n, p)$ . For Bernoulli trials, they recursively provided the probability mass function (PMF) as well as a Poisson limit theorem for  $N_{n;k_1,k_2}$ . They also derived the probability generating function of  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$ . Vellaisamy (2004) employed the Stein–Chen method to obtain a total variation upper bound for the rate of convergence of  $N_{n;k_1,k_2}$ , defined on Bernoulli trials, to a suitable Poisson RV. As a special case of his approach, the limit theorem of Huang and Tsai (1991) is re-established. He also presented Poisson approximation results for the occurrences of  $F$ - $S$  strings of length at least  $k_1 + k_2$  under stationary Markov-dependent binary trials. Sen *et al.* (2006) considered binary trials derived according to the Pólya–Eggenberger urn model (sampling scheme) and established the PMFs of  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$  on such a model. The corresponding distributions of  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$  were called Pólya and inverse Pólya distributions of order  $(k_1, k_2)$ , respectively. Depending on the specific urn sampling procedure, several discrete distributions of order  $(k_1, k_2)$  are obtained, along with the binomial distribution of order  $(k_1, k_2)$ , as special cases.

If we assume that the outcomes of the binary trials  $X_1, X_2, \dots, X_n$  are ordered on a circle, in such a way that the first outcome is adjacent to (and follows) the  $n$ th outcome, then we denote by  $N_{n;k_1,k_2}^c$  the number of occurrences of  $F$ - $S$  strings of length (at least)  $k_1 + k_2$ , i.e. (at least)  $k_2$  successes are preceded by (at least)  $k_1$  failures, on a circle. Its support is again the set  $S(n, k_1, k_2)$ . Readily,  $N_{n;0,k}^c = N_{n,k}^c$ , denoting the number of nonoverlapping success runs of length  $k$  on a circle. The PMF of this RV has been studied formally in Charalambides (1994),

Makri and Philippou (1994), and Koutras *et al.* (1995). Specific probabilities of  $N_{n,k}^c$  have appeared previously in the works of Philippou and Makri (1990) and Alevizos *et al.* (1993) in connection with the longest success run on a circle and the reliability of cyclic  $m$ -consecutive- $k$ -out-of- $n$ :  $F$  systems.

The forgoing definitions are illustrated using the following example. Let the outcomes of the first 15 binary trials be  $SSFFFSFSFFSSSF$ . Then,  $N_{15;1,1} = 3$ ,  $N_{15;1,2} = 2$ ,  $N_{15;2,2} = 2$ , and  $N_{15;3,2} = 2$ ;  $N_{15;1,1}^c = 4$ ,  $N_{15;1,2}^c = 3$ ,  $N_{15;2,2}^c = 2$ , and  $N_{15;3,2}^c = 2$ ;  $T_{2;1,1} = 9$ ,  $T_{1;2,2} = 7$ ,  $T_{2;2,2} = 14$ , and  $T_{3;2,2} > 15$ . We note that, since  $N_{15;2,2} = 2 < 3$ ,  $T_{3;2,2} > 15$ , and vice versa.

In this paper we study the RVs  $N_{n;k_1,k_2}$ ,  $N_{n;k_1,k_2}^c$ , and  $T_{r;k_1,k_2}$ , with  $\min(k_1, k_2) \geq 1$ , defined on binary sequences of exchangeable or independent RVs. Specifically, our work is organized as follows. In Section 2 we derive the exact PMFs of  $N_{n;k_1,k_2}$ ,  $N_{n;k_1,k_2}^c$ , and  $T_{r;k_1,k_2}$  for exchangeable sequences of binary RVs and the exact PMF of  $N_{n;k_1,k_2}$  for Poisson (independent but not necessarily identically distributed) sequences of binary RVs. The PMFs are given via sums of binomial coefficients and recursively for exchangeable and Poisson sequences, respectively. The PMF of  $T_{r;k_1,k_2}$  for Poisson sequences is derived via its relation with the PMF of  $N_{n;k_1,k_2}$ . In Section 3, exact formulae for the means and variances of  $N_{n;k_1,k_2}$  and  $N_{n;k_1,k_2}^c$  are given for both exchangeable and Poisson sequences. Their expressions are obtained via representations of these RVs as sums of indicator RVs holding for any binary sequence. The same setup of the indicator RVs is also used to establish the asymptotic normality of  $N_{n;k_1,k_2}$  and, consequently, of  $T_{r;k_1,k_2}$  defined on Bernoulli sequences (Poisson sequences with a common success probability). The means and variances, in addition to their independent merit, are also used to derive lower/upper bounds and approximations for the probability distributions of  $N_{n;k_1,k_2}$ ,  $N_{n;k_1,k_2}^c$ , and  $T_{r;k_1,k_2}$ , which hold for both types of sequences. Finally, in Section 4 our results are illustrated for various, widely used in applied probability, exchangeable and Poisson sequences of RVs, e.g. the Pólya–Eggenberger urn model, the record threshold model, and the record indicator model.

We end this section by noting that the vast majority of the presented results are new. In the paper we introduce the RV  $N_{n;k_1,k_2}^c$  and generalize, unify, and/or provide alternative formulae for particular results on the RVs  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$  due to Huang and Tsai (1991), Vellaisamy (2004), and Sen *et al.* (2006).

## 2. Exact distributions

In this section we provide the exact PMFs of the RVs  $N_{n;k_1,k_2}$ ,  $T_{r;k_1,k_2}$ , and  $N_{n;k_1,k_2}^c$  defined on sequences of exchangeable (Theorems 2.1, 2.2, and 2.3) or independent (Theorem 2.4 and Remark 2.4) binary trials. First we consider a sequence,  $X_1, X_2, \dots$ , of exchangeable RVs.

Let the outcomes of  $n$  exchangeable binary trials  $X_1, X_2, \dots, X_n$ ,  $n > 0$ , be arranged on a line or on a circle, and let  $Y_n$  denote the number of  $F$ s (0s) in the  $n$  trials. The elements  $\omega$  of the appropriate sample space  $\Omega$  are linear or circular permutations  $(i_1, \dots, i_n)$  with  $i_j \in \{S, F\}$ ,  $j = 1, \dots, n$ . For convenience, we use  $X_{n;k_1,k_2}$  to represent either the RV  $N_{n;k_1,k_2}$  or the RV  $N_{n;k_1,k_2}^c$ . An element of the event

$$\Gamma_{X_{n;k_1,k_2}}(x, y) = \{\omega \in \Omega: X_{n;k_1,k_2}(\omega) = x, Y_n(\omega) = y\}$$

is a sequence of  $y$   $F$ s and  $n - y$   $S$ s. Exchangeability implies that all finite sequences with the same length and the same number of failures, and, hence, the same number of successes, are

equally likely. By Theorem 2.1 of George and Bowman (1995) we obtain

$$\begin{aligned}
 p_n(y) &= P(X_1 = X_2 = \dots = X_{n-y} = 1, X_{n-y+1} = X_{n-y+2} = \dots = X_n = 0) \\
 &= \sum_{i=0}^y (-1)^i \binom{y}{i} \lambda_{n-y+i}, \quad y = 0, 1, \dots, n,
 \end{aligned}
 \tag{2.1}$$

where

$$\lambda_i = P(X_1 = X_2 = \dots = X_i = 1), \quad i = 1, 2, \dots, n \text{ and } \lambda_0 = 1.$$

Then

$$P(X_{n;k_1,k_2} = x) = \sum_y |\Gamma_{X_{n;k_1,k_2}}(x, y)| p_n(y)
 \tag{2.2}$$

and

$$P(X_{n;k_1,k_2} = x \mid Y_n = y) = \binom{n}{y}^{-1} |\Gamma_{X_{n;k_1,k_2}}(x, y)|,$$

where  $|\Gamma_{X_{n;k_1,k_2}}|$  denotes the number of binary sequences in  $\Gamma_{X_{n;k_1,k_2}}(x, y)$ , i.e. its cardinality. Now, it is evident that the problem of establishing the exact PMF of  $X_{n;k_1,k_2}$  and the conditional PMF of  $X_{n;k_1,k_2}$ , given the number of failures, is a combinatorial one, namely the computation of the number  $|\Gamma_{X_{n;k_1,k_2}}|$ . The same argument holds for the PMF of  $T_{r;k_1,k_2}$  by a slight modification on the definition of  $\Gamma$ . We will give a preliminary result.

**Lemma 2.1.** (Makri *et al.* (2007b).) *The number of allocations of  $\alpha$  indistinguishable balls into  $r$  distinguishable urns, where each of the  $m$ ,  $0 \leq m \leq r$ , specified urns is occupied by at most  $k$  balls, is given by*

$$H_m(\alpha, r, k) = \sum_{j=0}^{\lfloor \alpha/(k+1) \rfloor} (-1)^j \binom{m}{j} \binom{\alpha - (k+1)j + r - 1}{\alpha - (k+1)j}.$$

Setting  $k = 0$ ,  $H_m(\alpha, r, 0)$  represents the number of allocations of  $\alpha$  indistinguishable balls into  $r - m$  distinguishable urns, and is given by

$$H_m(\alpha, r, 0) = \binom{\alpha + r - m - 1}{\alpha}.$$

**Theorem 2.1.** *The PMF of  $N_{n;k_1,k_2}$  defined on a sequence of exchangeable binary RVs  $X_1, X_2, \dots, X_n$  ordered on a line, for  $n \geq k_1 + k_2$ , is given by*

(a) *for  $k_1 > 1$  and  $k_2 > 1$ ,*

$$\begin{aligned}
 P(N_{n;k_1,k_2} = x) &= \sum_{y=k_1x}^{n-k_2x} p_n(y) \sum_{m=x}^{m_1} \sum_{\ell=x}^m \binom{m}{\ell} \binom{\ell}{x} H_{m-\ell}(y - \ell(k_1 - 1) - m, m + 1, k_1 - 2) \\
 &\quad \times H_{\ell-x}(n - y - x(k_2 - 1) - m, m + 1, k_2 - 2);
 \end{aligned}$$

(b) *for  $k_1 = 1$  and  $k_2 > 1$ ,*

$$\begin{aligned}
 P(N_{n;k_1,k_2} = x) &= \sum_{y=x}^{n-k_2x} p_n(y) \sum_{m=x}^{m_1} \binom{m}{x} \binom{y}{m} H_{m-x}(n - y - x(k_2 - 1) - m, m + 1, k_2 - 2);
 \end{aligned}$$

(c) for  $k_1 > 1$  and  $k_2 = 1$ ,

$$P(N_{n;k_1,k_2} = x) = \sum_{y=k_1x}^{n-x} p_n(y) \sum_{m=x}^{m_1} \binom{m}{x} \binom{n-y}{m} H_{m-x}(y-x(k_1-1)-m, m+1, k_1-2);$$

(d) for  $k_1 = k_2 = 1$ ,

$$P(N_{n;k_1,k_2} = x) = \sum_{y=x}^{n-x} p_n(y) \binom{y}{x} \binom{n-y}{x};$$

where  $m_1 = \min\{y-x(k_1-1), n-y-x(k_2-1)\}$ .

*Proof.* We first note that a pattern

$$\underbrace{FF \dots F}_{\geq 1} \underbrace{SS \dots S}_{\geq 1}$$

appears in a sequence of  $\Gamma_{N_{n;k_1,k_2}}$  in one of

$$\underbrace{FF \dots F}_{\geq k_1} \underbrace{SS \dots S}_{\geq k_2}, \quad \underbrace{FF \dots F}_{\geq k_1} \underbrace{SS \dots S}_{\leq k_2-1}, \quad \text{or} \quad \underbrace{FF \dots F}_{\leq k_1-1} \underbrace{SS \dots S}_{\geq 1},$$

which we refer to as type  $A$ ,  $B$ , or  $C$ , respectively. The pattern  $A$ ,  $B$ , or  $C$  may be considered as an urn with two distinguishable cells. Then, a sequence in  $\Gamma_{N_{n;k_1,k_2}}(x, y)$  can be visualized as an arrangement of  $m+2$  distinguishable urns, say  $U_1, U_2, \dots, U_{m+2}$ , with each of  $U_1$  and  $U_{m+2}$  having one cell receiving  $S$ s and  $F$ s only, respectively, and each of  $U_2, \dots, U_{m+1}$  being of type  $A$ ,  $B$ , or  $C$ , having two cells, the first of which receives only  $F$ s and the second only  $S$ s. For example, for  $k_1 = k_2 = 2$ , the sequence  $SSFFFSSFSFFFSSFFSF$  may be visualized as  $SSACABF$ . The number of permutations of  $x$  urns of type  $A$ ,  $\ell-x$  urns of type  $B$ , and  $m-\ell$  urns of type  $C$ ,  $m = x, x+1, \dots, m_1$  and  $\ell = x, x+1, \dots, m$ , equals  $\binom{m}{\ell} \binom{\ell}{x}$ . For each specified permutation of  $A$ s,  $B$ s, and  $C$ s,  $k_1 F$ s are placed in the first cell of every urn of types  $A$  and  $B$ , and one  $F$  is placed in the first cell of every urn of type  $C$ . The number of distributions of the remaining  $y-\ell k_1-(m-\ell) F$ s into  $U_{m+2}$  and the first cells of the  $m$  urns  $U_2, \dots, U_{m+1}$  so that each urn of type  $C$  receives no more than  $k_1-2 F$ s equals  $H_{m-\ell}(y-\ell k_1-(m-\ell), m+1, k_1-2)$ , by Lemma 2.1. For the same permutation of  $A$ s,  $B$ s, and  $C$ s,  $k_2 S$ s are placed in the second cell of every urn of type  $A$  and one  $S$  in the second cell of every urn of types  $B$  and  $C$ . The number of distributions of the remaining  $n-y-xk_2-(m-x) S$ s into  $U_1$  and the second cells of the  $m$  urns  $U_2, U_3, \dots, U_{m+1}$  so that each urn of type  $B$  receives no more than  $k_2-2 S$ s equals  $H_{\ell-x}(n-y-xk_2-m+x, m+1, k_2-2)$ . Thus, according to the multiplicative principle and summing with respect to  $m$  and  $\ell$  we conclude that the cardinality of  $\Gamma_{N_{n;k_1,k_2}}(x, y)$  equals

$$\sum_{m=x}^{m_1} \sum_{\ell=x}^m \binom{m}{\ell} \binom{\ell}{x} H_{m-\ell}(y-\ell k_1-m+\ell, m+1, k_1-2) H_{\ell-x}(n-y-xk_2-m+x, m+1, k_2-2).$$

Part (a) of the theorem is then implied by (2.2). Parts (b), (c), and (d) follow by employing an analogous method of distributing indistinguishable balls into distinguishable urns.

**Theorem 2.2.** *The PMF of  $T_{r;k_1,k_2}$  defined on a sequence of exchangeable binary RVs for  $n = r(k_1 + k_2), r(k_1 + k_2) + 1, \dots$  is given by*

(a) *for  $k_1 > 1$  and  $k_2 > 1$ ,*

$$\begin{aligned}
 &P(T_{r;k_1,k_2} = n) \\
 &= \sum_{y=rk_1}^{n-rk_2} p_n(y) \sum_{m=r}^{m_1} \sum_{\ell=r}^m \binom{m-1}{r-1} \binom{m-r}{\ell-r} H_{m-\ell}(y - \ell(k_1 - 1) - m, m, k_1 - 2) \\
 &\quad \times H_{\ell-r}(n - y - r(k_2 - 1) - m, m, k_2 - 2);
 \end{aligned}$$

(b) *for  $k_1 = 1$  and  $k_2 > 1$ ,*

$$\begin{aligned}
 P(T_{r;k_1,k_2} = n) &= \sum_{y=r}^{n-rk_2} p_n(y) \sum_{m=r}^{m_1} \binom{m-1}{r-1} \binom{y-1}{m-1} \\
 &\quad \times H_{m-r}(n - y - r(k_2 - 1) - m, m, k_2 - 2);
 \end{aligned}$$

(c) *for  $k_1 > 1$  and  $k_2 = 1$ ,*

$$\begin{aligned}
 P(T_{r;k_1,k_2} = n) &= \sum_{y=k_1r}^{n-r} p_n(y) \sum_{m=r}^{m_1} \binom{m-1}{r-1} \binom{n-y-1}{m-1} \\
 &\quad \times H_{m-r}(y - r(k_1 - 1) - m, m, k_1 - 2);
 \end{aligned}$$

(d) *for  $k_1 = k_2 = 1$ ,*

$$P(T_{r;k_1,k_2} = n) = \sum_{y=r}^{n-r} p_n(y) \binom{y-1}{r-1} \binom{n-y-1}{r-1},$$

where  $m_1 = \min\{y - r(k_1 - 1), n - y - r(k_2 - 1)\}$ .

*Proof.* A typical element of the event  $(T_{r;k_1,k_2} = n, Y_n = y)$  is (as in the proof of Theorem 2.1) an arrangement of  $y$  failures and  $n - y$  successes, which we consider form  $m + 1$  distinguishable urns numbered from 1 to  $m + 1$ . The first urn has one cell receiving only  $S$ s, the  $(m + 1)$ th urn is of type  $A$  and the 2nd, 3rd,  $\dots$ ,  $m$ th urns are of type  $A, B$ , or  $C$ . We recall that each urn of type  $A, B$ , or  $C$  has two distinguishable cells. Then, the number of permutations of the first  $r - 1$  urns of type  $A, \ell - r$  urns of type  $B$ , and  $m - \ell$  urns of type  $C, m = r, r + 1, \dots, m_1$  and  $\ell = r, r + 1, \dots, m$ , equals  $\binom{m-1}{r-1} \binom{m-r}{m-\ell}$ . For each such permutation,  $k_1$   $F$ s are placed in the first cell of each urn of types  $A$  and  $B$ , and one  $F$  is placed in the first cell of each urn of type  $C$ . The number of distributions of the remaining  $y - k_1\ell - (m - \ell)$   $F$ s in the first cells of the  $m$  urns of types  $A, B$ , and  $C$  so that each cell of an urn of type  $C$  receives no more than  $k_1 - 2$   $F$ s equals  $H_{m-\ell}(y - \ell k_1 - m + \ell, m, k_1 - 2)$ . Next, we place  $k_2$   $S$ s in the second cell of each urn of type  $A$  and one  $S$  in the second cell of each urn of types  $B$  and  $C$ . The number of distributions of the  $n - y - k_2r - (m - r)$  remaining  $S$ s into the first urn and the second cells of the  $m - 1$  urns of types  $A, B$ , and  $C$  (except the  $(m + 1)$ th one) so that each cell of an urn of type  $B$  receives no more than  $k_2 - 2$   $S$ s equals  $H_{\ell-r}(n - y - rk_2 - m + r, m, k_2 - 2)$ . Applying the multiplicative principle and summing with respect to  $m$  and  $\ell$  we get the cardinality of  $(T_{r;k_1,k_2} = n, Y_n = y), y = rk_1, \dots, n - rk_2$ . The results follow by proceeding as in Theorem 2.1.

**Remark 2.1.** Sen *et al.* (2006, Theorems 2.1, 3.1) obtained the PMFs of  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$  defined on a particular (in our approach) exchangeable sequence derived by the Pólya–Eggenberger urn model (see Section 4.1). Although the method we have used for the proof of Theorems 2.1 and 2.2 is based on a similar combinatorial approach to theirs, their formulae are more complicated than ours since the PMFs of  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$  are expressed as a sum of two sums each with eight summations involving binomial coefficients, instead of a sum of five summations in our formulae. The latter improvement is due to the unified consideration of  $F$ - $S$  strings as urns of types  $A$ ,  $B$ , and  $C$  each with two cells, as well as, to the employment of the coefficient  $H_m(\alpha, r, k)$ .

Next we proceed to derive the PMF of the RV  $N_{n;k_1,k_2}^c$ . We will first give a preliminary result.

Let  $B_{N_{n;k_1,k_2}}(x, y)$ ,  $y \geq 1$ , be a subset of  $\Gamma_{N_{n;k_1,k_2}}(x, y)$  defined by

$$B_{N_{n;k_1,k_2}}(x, y) = \{\omega \in \Omega : N_{n;k_1,k_2}(\omega) = x, Y_n(\omega) = y, X_1 = 0, X_n = 1\}.$$

Then the number of its elements is given by the following lemma.

**Lemma 2.2.** *The cardinality  $|B_{N_{n;k_1,k_2}}(x, y)|$  of the set  $B_{N_{n;k_1,k_2}}(x, y)$  is given by*

(a) for  $n < k_1 + k_2$ ,

$$|B_{N_{n;k_1,k_2}}(0, y)| = \binom{n-2}{y-1} \quad \text{and} \quad |B_{N_{n;k_1,k_2}}(x, y)| = 0, \quad x \neq 0;$$

(b) for  $n \geq k_1 + k_2$  and  $x = 0, 1, \dots, \lfloor n/(k_1 + k_2) \rfloor$ ,

$$|B_{N_{n;k_1,k_2}}(x, y)| = \begin{cases} \sum_{m=x}^{m_1} \sum_{\ell=x}^m \binom{m}{\ell} \binom{\ell}{x} H_{m-\ell}(y - \ell(k_1 - 1) - m, m, k_1 - 2) \\ \quad \times H_{\ell-x}(n - y - x(k_2 - 1) - m, m, k_2 - 2) & \text{for } k_1 > 1, k_2 > 1, \\ \sum_{m=x}^{m_1} \binom{y-1}{m-1} \binom{m}{x} \\ \quad \times H_{m-x}(n - y - x(k_2 - 1) - m, m, k_2 - 2) & \text{for } k_1 = 1, k_2 > 1, \\ \sum_{m=x}^{m_1} \binom{n-y-1}{m-1} \binom{m}{x} \\ \quad \times H_{m-x}(y - x(k_1 - 1) - m, m, k_1 - 2) & \text{for } k_1 > 1, k_2 = 1, \\ \binom{y-1}{x-1} \binom{n-y-1}{x-1} & \text{for } k_1 = k_2 = 1, \end{cases}$$

where  $m_1 = \min\{y - x(k_1 - 1), n - y - x(k_2 - 1)\}$ .

*Proof.* Part (a) of the lemma is apparent. To prove part (b), we note that a typical element of  $B_{N_{n;k_1,k_2}}(x, y)$  is a sequence of  $n$  binary trials that starts with an  $F$  and ends with an  $S$ . Such a sequence is considered as an arrangement of patterns of type  $A$ ,  $B$ , or  $C$  (we consider only urns with two cells), as defined in the proof of Theorem 2.1. The proof is completed along the lines of the proof of Theorem 2.1.

**Theorem 2.3.** *The PMF of  $N_{n;k_1,k_2}^c$  defined on a sequence of exchangeable binary RVs  $X_1, X_2, \dots, X_n$  ordered on a circle is given by*

$$\begin{aligned}
 P(N_{n;k_1,k_2}^c = x) &= \sum_{y=y_1}^{y_2} p_n(y) \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} (j_1 + j_2) |B_{N_{n-j_1-j_2;k_1,k_2}}(x - I_A((j_1, j_2)), y - j_1)| \\
 &\quad + n \sum_{y=1}^{n-1} p_n(y) [I_A((y, n - y))\delta_{x,1} + I_{A^c}((y, n - y))\delta_{x,0}] \\
 &\quad + [p_n(0) + p_n(n)]\delta_{x,0},
 \end{aligned}$$

where  $y_1 = \max\{k_1x, 1\}$ ,  $y_2 = \min\{n - k_2x, n - 2\}$ ,  $r_1 = \min\{n - 3, y - 1\}$ ,  $r_2 = \min\{n - 2 - j_1, n - y - 1\}$ , and  $A = \{(j_1, j_2) : (j_1, j_2) \geq (k_1, k_2)\}$  with  $I_A(x) = 1$  if  $x \in A$  and 0 otherwise.

*Proof.* Let  $x = 0, 1, \dots, \lfloor n/(k_1 + k_2) \rfloor$ . First we note that every element of the appropriate sample space is a circular arrangement of  $n$  Ss and Fs. The sample space is considered as the union of (a) the set of sequences including at least two patterns of type

$$\underbrace{FF \dots F}_{\geq 1} \underbrace{SS \dots S}_{\geq 1};$$

(b) the set of sequences including only one such pattern; and (c) the set of sequences with no such patterns. Then, we define the following events. For  $j_1 = 1, \dots, r_1$ ,  $j_2 = 1, \dots, r_2$ , and  $i = n - j_1 - j_2 + 1, \dots, n$ ,

$$A_{j_1, j_2, i} = \left\{ X_i \prod_{j=i+1}^{i+j_1} (1 - X_j) \prod_{j=i+j_1+1}^{i+j_1+j_2} X_j (1 - X_{i+j_1+j_2+1}) = 1 \right\};$$

for  $i = 1, \dots, n$ ,

$$C_{y,i} = \left\{ \prod_{j=i}^{i+y-1} (1 - X_j) \prod_{j=i+y}^{i+n-1} X_j = 1 \right\},$$

and  $D = \{\prod_{j=1}^n X_j = 1\}$  and  $E = \{\prod_{j=1}^n (1 - X_j) = 1\}$ . (Here we have used the convention that  $X_{n+i} \equiv X_i$ .)

Readily,

$$\begin{aligned}
 (N_{n;k_1,k_2}^c = x, Y_n = y) &= \bigcup_{j_1=1}^{r_1} \bigcup_{j_2=1}^{r_2} \bigcup_{i=n-j_1-j_2+1}^n [(N_{n;k_1,k_2}^c = x, Y_n = y) \cap A_{j_1, j_2, i}] \\
 &\quad \cup \left[ (N_{n;k_1,k_2}^c = x, Y_n = y) \cap \left( \bigcup_{i=1}^n C_{y,i} \right) \right] \\
 &\quad \cup [(N_{n;k_1,k_2}^c = x, Y_n = y) \cap D] \\
 &\quad \cup [(N_{n;k_1,k_2}^c = x, Y_n = y) \cap E],
 \end{aligned}$$

from which we obtain

$$\begin{aligned} &P(N_{n;k_1,k_2}^c = x, Y_n = y) \\ &= \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{i=n-j_1-j_2+1}^n p_n(y) |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap A_{j_1,j_2,i}| \\ &\quad + \sum_{i=1}^n p_n(y) |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap C_{y,i}| + p_n(0) |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap D| \\ &\quad + p_n(n) |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap E|. \end{aligned}$$

Fix  $j_1, j_2$ , and  $i$ . Readily,

$$\begin{aligned} |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap A_{j_1,j_2,i}| &= |B_{N_{n-j_1-j_2;k_1,k_2}}(x - I_A((j_1, j_2)), y - j_1)|, \\ |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap C_{y,i}| &= \delta_{x,1} I_A((y, n - y)) + \delta_{x,0} I_{A^c}((y, n - y)), \\ |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap D| &= \delta_{x,0} \delta_{y,0}, \quad \text{and} \quad |(N_{n;k_1,k_2}^c = x, Y_n = y) \cap E| = \delta_{x,0} \delta_{y,n}. \end{aligned}$$

Summing with respect to  $y$ , the result follows directly.

**Remark 2.2.** For Bernoulli trials with a common success probability  $p$  ( $0 < p = 1 - q < 1$ ),  $p_n(y)$  reduces to

$$p_n(y) = p^{n-y} q^y.$$

Therefore, its usage in Theorems 2.1, 2.2, and 2.3 provides the PMFs of the corresponding RVs for such trials.

Next we replace the assumption of exchangeability with that of independence. Let  $N_{n;k_1,k_2}$  be an RV denoting the number of  $F$ - $S$  strings of length at least  $k_1 + k_2$  in a sequence of  $n$  Poisson trials  $X_i, i = 1, 2, \dots, n$ , with  $P(X_i = 1) = p_i = 1 - q_i = 1 - P(X_i = 0)$ , ordered on a line. For such sequences we establish the following theorem.

**Theorem 2.4.** *The PMF of the RV  $N_{n;k_1,k_2}$  satisfies the recursive scheme  $P(N_{n;k_1,k_2} = 0) = 1$  and  $P(N_{n;k_1,k_2} = x) = 0, x \neq 0$ , for  $n < k_1 + k_2$ ;  $P(N_{n;k_1,k_2} = x) = 0$  for  $x < 0$  or  $x > \lfloor n/(k_1 + k_2) \rfloor$ ;*

$$\begin{aligned} P(N_{n+1;k_1,k_2} = x) &= P(N_{n;k_1,k_2} = x) \\ &\quad + \beta [P(N_{n+1-k_1-k_2;k_1,k_2} = x - 1) - P(N_{n+1-k_1-k_2;k_1,k_2} = x)], \end{aligned} \tag{2.3}$$

for  $x = 0, 1, \dots, \lfloor n/(k_1 + k_2) \rfloor, n \geq k_1 + k_2$ , where  $\beta = (\prod_{i=n-k_1-k_2+2}^{n-k_2+1} q_i) (\prod_{i=n-k_2+2}^{n+1} p_i)$ .

*Proof.* Obviously, for  $x < 0$  or  $x > \lfloor n/(k_1 + k_2) \rfloor$  and  $n < k_1 + k_2$ , the theorem holds. For  $n \geq k_1 + k_2$ , following Huang and Tsai (1991), we first observe that, for  $r \geq 1$ ,

$$\begin{aligned} &P(N_{n+1;k_1,k_2} = r, N_{n;k_1,k_2} = r - 1) \\ &= P(N_{n+1-k_1-k_2;k_1,k_2} = r - 1) \prod_{i=n-k_1-k_2+2}^{n-k_2+1} P(X_i = 0) \prod_{i=n-k_2+2}^{n+1} P(X_i = 1), \end{aligned} \tag{2.4}$$

by the independence of the  $X_i$ s. Then, for  $x \geq 1$ , we have

$$\begin{aligned} P(N_{n+1;k_1,k_2} = x) &= P(N_{n+1;k_1,k_2} = x, N_{n;k_1,k_2} \neq x) + P(N_{n+1;k_1,k_2} = x, N_{n;k_1,k_2} = x) \\ &= P(N_{n+1;k_1,k_2} = x, N_{n;k_1,k_2} \neq x) + P(N_{n;k_1,k_2} = x) \\ &\quad - P(N_{n+1;k_1,k_2} \neq x, N_{n;k_1,k_2} = x) \\ &= P(N_{n+1;k_1,k_2} = x, N_{n;k_1,k_2} = x - 1) + P(N_{n;k_1,k_2} = x) \\ &\quad - P(N_{n+1;k_1,k_2} = x + 1, N_{n;k_1,k_2} = x), \\ P(N_{n+1;k_1,k_2} = 0) &= P(N_{n+1;k_1,k_2} = 0, N_{n;k_1,k_2} = 0) \\ &= P(N_{n;k_1,k_2} = 0) - P(N_{n+1;k_1,k_2} = 1, N_{n;k_1,k_2} = 0). \end{aligned}$$

Using (2.4), we obtain the result of the theorem.

**Remark 2.3.** For  $p_i = p = 1 - q$ , that is, for Bernoulli trials, (2.3) reduces to a recursive scheme of Huang and Tsai (1991).

**Remark 2.4.** Using the dual relationship (1.1) and Theorem 2.4, we can determine the PMF of the waiting time  $T_{r;k_1,k_2}$  defined on Poisson trials  $X_1, X_2, \dots$  through the relation

$$P(T_{r;k_1,k_2} = n) = \sum_{x=0}^{r-1} [P(N_{n-1;k_1,k_2} = x) - P(N_{n;k_1,k_2} = x)], \quad r \in S(n, k_1, k_2) - \{0\},$$

for  $n = r(k_1 + k_2), r(k_1 + k_2) + 1, \dots$

### 3. Means, variances, bounds, and approximations

In this section we first give a formal definition of the RVs  $N_{n;k_1,k_2}$  and  $N_{n;k_1,k_2}^c$  via suitable sets of indicator RVs. This setup, holding for any binary sequence, is the main tool used to derive exact means and variances of these RVs defined on exchangeable (Propositions 3.1 and 3.2) and Poisson (Propositions 3.3 and 3.4) sequences, as well as to establish the asymptotic normality of  $N_{n;k_1,k_2}$  defined on Bernoulli sequences (Theorem 3.1). The expressions of the means and variances are also used to obtain lower/upper bounds and approximations for the cumulative distribution functions of  $N_{n;k_1,k_2}, T_{r;k_1,k_2}$ , and  $N_{n;k_1,k_2}^c$ , which hold for both types of sequences.

Let  $I_j, j \in J$ , be indicator RVs defined on a binary sequence  $\{X_i\}_{i=1}^n$  with  $J = \{k_1 + k_2, k_1 + k_2 + 1, \dots, n\}$  or  $J = \{1, 2, \dots, n\}$  if the sequence is linearly or circularly ordered, respectively. Again, let  $X_{n;k_1,k_2}$  stand for either  $N_{n;k_1,k_2}$  or  $N_{n;k_1,k_2}^c$ . Then, for  $\min(k_1, k_2) \geq 1$ ,

$$X_{n;k_1,k_2} = \sum_{j \in J} I_j \quad \text{with} \quad I_j = \left( \prod_{i=1}^{k_2} X_{j-i+1} \right) \left( \prod_{i=1}^{k_1} (1 - X_{j-k_2+1-i}) \right) \quad (3.1)$$

(using the conventions that  $X_0 \equiv X_n$  and  $X_{-i} \equiv X_{n-i}$ ). Accordingly, we have

$$\begin{aligned} E(X_{n;k_1,k_2}) &= \sum_{j \in J} P(I_j = 1), \\ V(X_{n;k_1,k_2}) &= \sum_{j \in J} \{P(I_j = 1) - \{P(I_j = 1)\}^2\} \\ &\quad + 2 \sum_{\substack{j_1 < j_2 \\ j_1, j_2 \in J}} \{P(I_{j_1} = 1, I_{j_2} = 1) - P(I_{j_1} = 1)P(I_{j_2} = 1)\}. \end{aligned}$$

First we consider exchangeable trials. Readily, it holds that  $P(I_j = 1) = p_{k_1+k_2}(k_1)$  and  $P(I_{j_1} = 1, I_{j_2} = 1) = 0$  for any combination  $\{j_1, j_2\}$  possessing the property that there are fewer than  $k_1 + k_2 - 1$  points between  $j_1$  and  $j_2$ , and  $P(I_{j_1} = 1, I_{j_2} = 1) = p_{2(k_1+k_2)}(2k_1)$  for any combination  $\{j_1, j_2\}$  with the property that there are at least  $k_1 + k_2 - 1$  points between  $j_1$  and  $j_2$  ( $j, j_1, j_2 \in J$ ).

**Proposition 3.1.** For  $n \geq k_1 + k_2$ , the mean value  $E(N_{n;k_1,k_2})$  and the variance  $V(N_{n;k_1,k_2})$  of  $N_{n;k_1,k_2}$  are given by

$$E(N_{n;k_1,k_2}) = (n - k_1 - k_2 + 1)p_{k_1+k_2}(k_1),$$

$$V(N_{n;k_1,k_2}) = \begin{cases} (n - k_1 - k_2 + 1)p_{k_1+k_2}(k_1)[1 - (n - k_1 - k_2 + 1)p_{k_1+k_2}(k_1)] & \text{for } n < 2(k_1 + k_2), \\ (n - k_1 - k_2 + 1)p_{k_1+k_2}(k_1)[1 - (n - k_1 - k_2 + 1)p_{k_1+k_2}(k_1)] \\ + (n - 2k_1 - 2k_2 + 2)(n - 2k_1 - 2k_2 + 1)p_{2(k_1+k_2)}(2k_1) & \text{for } n \geq 2(k_1 + k_2). \end{cases}$$

*Proof.* For  $n < 2(k_1 + k_2)$ , it holds that, for every 2-combination  $\{j_1, j_2\}$  of the  $n - k_1 - k_2 + 1$  numbers of the index set  $J = \{k_1 + k_2, \dots, n\}$ ,  $j_2 - j_1 < k_1 + k_2$ . For  $n \geq 2(k_1 + k_2)$ , the number of 2-combinations  $\{j_1, j_2\}$  of the  $n - k_1 - k_2 + 1$  numbers  $\{k_1 + k_2, \dots, n\}$  for which it holds that  $j_2 - j_1 \geq k_1 + k_2$  equals

$$\binom{n - k_1 - k_2 + 1 - (2 - 1)(k_1 + k_2 - 1)}{2} = \binom{n - 2(k_1 + k_2) + 2}{2}$$

(see Charalambides (2002, p. 99)), and the number for which it holds that  $j_2 - j_1 < k_1 + k_2$  equals

$$\binom{n - k_1 - k_2 + 1}{2} - \binom{n - 2(k_1 + k_2) + 2}{2}.$$

Hence, on using the above expressions the results follow after some algebraic manipulations.

**Proposition 3.2.** For  $n \geq k_1 + k_2$ , the mean value  $E(N_{n;k_1,k_2}^c)$  and the variance  $V(N_{n;k_1,k_2}^c)$  of  $N_{n;k_1,k_2}^c$  are given by

$$E(N_{n;k_1,k_2}^c) = np_{k_1+k_2}(k_1),$$

$$V(N_{n;k_1,k_2}^c) = \begin{cases} np_{k_1+k_2}(k_1)(1 - np_{k_1+k_2}(k_1)) & \text{for } n < 2(k_1 + k_2), \\ np_{k_1+k_2}(k_1)(1 - np_{k_1+k_2}(k_1)) \\ + n(n - 2(k_1 + k_2) + 1)p_{2(k_1+k_2)}(2k_1) & \text{for } n \geq 2(k_1 + k_2). \end{cases}$$

*Proof.* First, we observe that, for  $n < 2(k_1 + k_2)$ , there are no 2-combinations  $\{j_1, j_2\}$  of the  $n$  numbers of the index set  $\{1, 2, \dots, n\}$  possessing the property that between them there are at least  $k_1 + k_2 - 1$  numbers. For  $n \geq 2(k_1 + k_2)$ , there are

$$\frac{n}{n - 2(k_1 + k_2 - 1)} \binom{n - 2(k_1 + k_2 - 1)}{2}$$

2-combinations  $\{j_1, j_2\}$  of the  $n$  numbers  $\{1, 2, \dots, n\}$ , displayed on a circle, with at least  $k_1 + k_2 - 1$  points between them and

$$\binom{n}{2} - \frac{n}{n - 2(k_1 + k_2 - 1)} \binom{n - 2(k_1 + k_2 - 1)}{2}$$

with fewer than  $k_1 + k_2 - 1$  points between them. The results follow after some algebraic manipulations.

Next we obtain exact formulae for the mean values and variances of the RVs  $N_{n;k_1,k_2}$  and  $N_{n;k_1,k_2}^c$  defined on a sequence of independent binary trials (i.e. Poisson trials).

**Proposition 3.3.** For  $n \geq k_1 + k_2$ , the mean value  $E(N_{n;k_1,k_2})$  and the variance  $V(N_{n;k_1,k_2})$  of  $N_{n;k_1,k_2}$  are given by

$$E(N_{n;k_1,k_2}) = \sum_{j=k_1+k_2}^n \mu_j,$$

$$V(N_{n;k_1,k_2}) = \begin{cases} \sum_{j=k_1+k_2}^n \mu_j(1 - \mu_j) - 2 \sum_{j=k_1+k_2}^{n-1} \mu_j \sum_{i=1}^{n-j} \mu_{j+i} & \text{for } n < 2(k_1 + k_2), \\ \sum_{j=k_1+k_2}^n \mu_j(1 - \mu_j) - 2 \sum_{j=k_1+k_2}^{n-k_1-k_2+1} \mu_j \sum_{i=1}^{k_1+k_2-1} \mu_{j+i} \\ - 2 \sum_{j=n-k_1-k_2+2}^{n-1} \mu_j \sum_{i=1}^{n-j} \mu_{j+i} & \text{for } n \geq 2(k_1 + k_2), \end{cases}$$

where  $\mu_j = \prod_{i=j-k_2+1}^j p_i \prod_{i=j-k_1-k_2+1}^{j-k_2} q_i$ ,  $j = k_1 + k_2, k_1 + k_2 + 1, \dots, n$ .

*Proof.* We first note that, for  $j = k_1 + k_2, \dots, n$ , the RVs  $I_j$  are not (in general) identically distributed and it holds that

$$P(I_j = 1) = P(X_{j-k_1-k_2+1} = X_{j-k_1-k_2+2} = \dots = X_{j-k_2} = 0, \\ X_{j-k_2+1} = X_{j-k_2+2} = \dots = X_j = 1) = \mu_j,$$

by the independence of the binary sequence. Next, because of the internal structure of the RVs  $I_{j_1}$  and  $I_{j_2}$ , we observe that  $P(I_{j_1} = 1, I_{j_2} = 1) = 0$  if  $0 < j_2 - j_1 < k_1 + k_2$ , so that  $\text{cov}(I_{j_1}, I_{j_2}) = -\mu_{j_1}\mu_{j_2}$ , whereas  $I_{j_1}$  and  $I_{j_2}$  are independent RVs for  $j_2 - j_1 \geq k_1 + k_2$ . The proposition follows.

**Proposition 3.4.** For  $n \geq k_1 + k_2$ , the mean value  $E(N_{n;k_1,k_2}^c)$  and the variance  $V(N_{n;k_1,k_2}^c)$  of  $N_{n;k_1,k_2}^c$  are given by

$$E(N_{n;k_1,k_2}^c) = \sum_{j=1}^n \mu_j,$$

$$V(N_{n;k_1,k_2}^c) = \begin{cases} \sum_{j=1}^n \mu_j(1 - \mu_j) - 2 \sum_{j=1}^{n-1} \mu_j \sum_{i=1}^{n-j} \mu_{j+i} & \text{for } n < 2(k_1 + k_2), \\ \sum_{j=1}^n \mu_j(1 - \mu_j) - 2 \sum_{j=1}^n \mu_j \sum_{i=1}^{k_1+k_2-1} \mu_{j+i} & \text{for } n \geq 2(k_1 + k_2), \end{cases}$$

with  $\mu_j = \prod_{i=j-k_2+1}^j p_i \prod_{i=j-k_1-k_2+1}^{j-k_2} q_i$ ,  $j = 1, 2, \dots, n$  (using the conventions that  $p_0 \equiv p_n$ ,  $p_{-i} \equiv p_{n-i}$  and  $\mu_{n+i} = \mu_i$ ).

*Proof.* For  $n < 2(k_1 + k_2)$ ,  $j = 1, 2, \dots, n - 1$  and  $i = 1, 2, \dots, n - j$ , we have  $P(I_j = 1, I_{j+i} = 1) = 0$ , so that  $\text{cov}(I_j, I_{j+i}) = -\mu_j\mu_{j+i}$ . Noting that, for  $n \geq 2(k_1 + k_2)$ ,

$\text{cov}(I_{j_1}, I_{j_2}) = 0$  for every 2-combination  $\{j_1, j_2\}$  of the  $n$  numbers of the set  $\{1, 2, \dots, n\}$  displayed on a circle possessing the property that between them there are at least  $k_1 + k_2 - 1$  numbers, we obtain  $\sum_{1 \leq j_1 < j_2 \leq n} \text{cov}(I_{j_1}, I_{j_2}) = -\sum_{i=1}^n \sum_{j=1}^{k_1+k_2-1} \mu_j \mu_{j+i}$ . The proposition follows.

For large  $n$ , calculating the probability

$$F(x) = P(X_{n;k_1,k_2} < x), \quad x \in S(n, k_1, k_2) - \{0\},$$

is often a hard task, because of the computational effort needed to calculate the required recursions or the sums of the binomial coefficients involved. Therefore, the need for easily computed bounds and approximations is apparent.

Let  $m = E(X_{n;k_1,k_2})$  and  $v^2 = V(X_{n;k_1,k_2})$ . Employing Markov's inequality and the one-sided Chebychev inequality, as well as Equations (34)–(36) and (38) of Makri and Psillakis (2009a), we obtain the bounds

$$F(x) \geq L_{MC}(x) \quad \text{for } x \geq m, \quad F(x) \leq U_C(x) \quad \text{for } x < m + 1, \quad (3.2)$$

where

$$L_{MC}(x) = \begin{cases} 0 & \text{if } x = m, \\ 1 - \frac{m}{x} & \text{if } m < x \leq m + v^2/m, \\ 1 - \frac{v^2}{v^2 + (x - m)^2} & \text{if } x > m + v^2/m, \end{cases}$$

and  $U_C(x) = v^2/[v^2 + (1 + m - x)^2]$ . For an  $x$  such that both bounds  $L_{MC}$  and  $U_C$  can be used, we obtain the approximation

$$\hat{F}(x) = \frac{1}{2}(L_{MC}(x) + U_C(x)) \quad (3.3)$$

of  $F(x)$ , with an upper bound

$$\hat{B}(x) = \frac{U_C(x) - L_{MC}(x)}{2L_{MC}(x)}, \quad L_{MC}(x) > 0, \quad (3.4)$$

of the relative error between  $F(x)$  and  $\hat{F}(x)$ ,  $B(x) = |F(x) - \hat{F}(x)|/F(x)$  for  $F(x) > 0$ . We note that the estimate  $\hat{B}(x)$  of the relative error  $B(x)$  does not assume the knowledge of the exact value of  $F(x)$ . It depends only on the lower and upper bounds. Therefore, it gives an advantage in cases for which the exact value is difficult to compute. Furthermore, (3.2)–(3.4) provide lower/upper bounds and approximations of the tail probability of the waiting time  $T_{r;k_1,k_2}$  too, since  $P(T_{r;k_1,k_2} > n) = P(N_{n;k_1,k_2} < r) = F(r)$ .

We recall that  $B_{k_1,k_2}(n, p)$  is a generalization of  $B_k(n, p)$ , which is well approximated by a Poisson distribution, as well as by a normal distribution (see, e.g. Balakrishnan and Koutras (2002, pp. 174–176, 178–180)). It is expected that analogous approximations also hold for  $B_{k_1,k_2}(n, p)$ . This turns out to be true. As in the case of the ordinary Bernoulli distribution  $B_1(n, p)$ , in these approaches two different setups might be used. In the first setup (Poisson limit law (PLL)) the lengths  $k_1$  and  $k_2$  are assumed fixed, whereas the success probability  $p$  depends on  $n$  and tends to 0 as  $n$  tends to  $\infty$ . In the second setup (normal limit law or central limit theorem (CLT)) we suppose that not only  $k_1$  and  $k_2$  are fixed but also that  $p$  is fixed while  $n$  (the number of trials) tends to  $\infty$ .

Following the first setup, Huang and Tsai (1991), using a result of Kendall (1967), and Vellaisamy (2004), using the Stein–Chen method (see, e.g. Arratia *et al.* (1989)), established a PLL accompanied by an upper bound on the rate of convergence of this approximation. But, to the author’s knowledge, there are no results on the asymptotic normality of  $N_{n;k_1,k_2}$  with  $\min(k_1, k_2) \geq 1$ . In the sequel, using the second setup, we establish a CLT for  $N_{n;k_1,k_2}$  based on a theorem of Hoeffding and Robbins (1948) for dependent stationary sequences of RVs.

**Theorem 3.1.** *Let  $N_{n;k_1,k_2}$  have a  $B_{k_1,k_2}(n, p)$  distribution with  $\min(k_1, k_2) \geq 1$  and  $0 < p = 1 - q < 1$ . Then, for fixed  $k_1, k_2$  ( $k_1 + k_2 \leq n$ ),*

$$\frac{N_{n;k_1,k_2} - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z \sim N(0, 1) \quad \text{as } n \rightarrow \infty, \tag{3.5}$$

where  $\mu = q^{k_1} p^{k_2}$ ,  $\sigma^2 = \mu\{1 - (2(k_1 + k_2) - 1)\mu\}$ , and ‘ $\xrightarrow{D}$ ’ denotes convergence in distribution.

*Proof.* In the proof all limits are taken as  $n \rightarrow \infty$ . For  $j \in \{1, 2, \dots, n - k_1 - k_2 + 1\}$ , let  $Z_j = I_{(k_1+k_2-1)+j}$ , with the  $I$ s defined as in (3.1). Since the RVs  $X_i$  are independent and identically distributed (i.i.d.) so that the RVs  $Z_j$  are identically distributed, it follows that the sequence of RVs  $Z_1, Z_2, \dots$  is stationary and also, by their definition,  $k_1 + k_2 - 1$  dependent. Then, noting that  $E(Z_1^3) < \infty$ , the CLT of Hoeffding and Robbins (1948, Theorem 2) holds for the RV  $N_{n;k_1,k_2} = \sum_{j=1}^{n-k_1-k_2+1} Z_j$ , with  $\sigma^2 = V(Z_1) + 2 \sum_{j=2}^{k_1+k_2} \text{cov}(Z_1, Z_j) = q^{k_1} p^{k_2} - (q^{k_1} p^{k_2})^2 - 2(k_1 + k_2 - 1)(q^{k_1} p^{k_2})^2$ . That is,

$$Y_{n;k_1,k_2} = \frac{N_{n;k_1,k_2} - E(N_{n;k_1,k_2})}{\sigma\sqrt{n - k_1 - k_2 + 1}} = \frac{N_{n;k_1,k_2} - (n - k_1 - k_2 + 1)q^{k_1} p^{k_2}}{\sigma\sqrt{n - k_1 - k_2 + 1}} \xrightarrow{D} Z \sim N(0, 1),$$

or, equivalently,

$$Y_{n;k_1,k_2} = \beta_n \left[ \frac{N_{n;k_1,k_2} - nq^{k_1} p^{k_2}}{\sigma\sqrt{n}} \right] + \gamma_n \xrightarrow{D} Z \sim N(0, 1),$$

where

$$\beta_n = \sqrt{\frac{n}{n - k_1 - k_2 + 1}} \quad \text{and} \quad \gamma_n = \frac{(k_1 + k_2 - 1)q^{k_1} p^{k_2}}{\sigma\sqrt{n - k_1 - k_2 + 1}}.$$

Then, since  $\beta_n \rightarrow 1$  and  $\gamma_n/\beta_n \rightarrow 0$ , we obtain

$$\frac{N_{n;k_1,k_2} - n\mu}{\sigma\sqrt{n}} = \frac{1}{\beta_n} Y_{n;k_1,k_2} - \frac{\gamma_n}{\beta_n} \xrightarrow{D} Z \sim N(0, 1).$$

**Remark 3.1.** A practical interpretation of (3.5) is that, for sufficiently large  $n$ , i.e.  $n \gg 1$ , and  $z \in \mathbb{R}$ , we have the approximation

$$P(N_{n;k_1,k_2} \leq z) \simeq \Phi\left(\frac{z - n\mu}{\sigma\sqrt{n}}\right),$$

where  $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x e^{-t^2/2} dt$ ,  $x \in \mathbb{R}$ . Clearly, the cumulative distribution function of  $N_{n;k_1,k_2}$  defined on Bernoulli trials depends on the parameter vector  $(n, p, k_1, k_2)$  with  $1 \leq k_1 \leq n$ ,  $1 \leq k_2 \leq n$ ,  $k_1 + k_2 \leq n$ , and  $0 < p < 1$ . Following Makri and Psillakis (2009b), we suggest that, in order for the distribution of  $N_{n;k_1,k_2}$  to be well approximated by that of a

normal RV with mean  $n\mu$  and variance  $n\sigma^2$ , it has to be checked whether, for the parameter vector  $(n, p, k_1, k_2)$ , the conditions

$$\mu - \frac{3\sigma}{\sqrt{n}} > 0 \quad \text{and} \quad \mu + \frac{3\sigma}{\sqrt{n}} < \frac{\lfloor n/(k_1 + k_2) \rfloor}{n} \tag{3.6}$$

are satisfied.

#### 4. Applications and numerics

The knowledge of probabilities  $\lambda_i$  and  $p_i$  is sufficient for someone to investigate the probabilistic behavior of the RVs  $N_{n;k_1,k_2}$ ,  $T_{r;k_1,k_2}$ , and  $N_{n;k_1,k_2}^c$  on actual exchangeable and Poisson sequences used in applied research. Next, some examples of such binary sequences are considered. They are indicative of real situations and have recently appeared in numerous fields of applications of run-related statistics. See, e.g. Demir and Eryilmaz (2008) and Makri and Psillakis (2009a). For these sequences, we first present a summary of the involved concepts to describe their internal structure. After that, we select some parametric configurations which we use in numerical examples. The latter providing a sense of the involved numerics and an insight into the formulae presented in Sections 2 and 3.

##### 4.1. Pólya–Eggenberger urn model

An exchangeable sequence of special interest in applied probability is a sequence derived according to a Pólya–Eggenberger sampling scheme (cf. Johnson and Kotz (1977, pp. 176–178)). In this scheme a ball is drawn at random from an urn initially containing  $w$  white balls and  $b$  black balls, its color is observed, and it is then returned to the urn along with  $s$  additional balls of the same color as the ball drawn. Drawing a white ball is considered a success and drawing a black ball is considered a failure. We denote this scheme as  $PE(w, b, s)$ . It is clear that  $n$  repetitions of the scheme derive an exchangeable binary sequence with

$$p_n(y) = \frac{\prod_{j=0}^{n-y-1} (w + js) \prod_{j=0}^{y-1} (b + js)}{\prod_{j=0}^{n-1} (w + b + js)}, \quad 0 \leq y \leq n,$$

or, alternatively, with  $p_n(y)$  given by (2.1) and

$$\lambda_i = \prod_{j=0}^{i-1} \frac{w + js}{w + b + js}, \quad i = 0, 1, \dots, n. \tag{4.1}$$

The corresponding distributions of  $N_{n;k_1,k_2}$  ( $N_{n;k_1,k_2}^c$ ) and  $T_{r;k_1,k_2}$  are called Pólya (circular Pólya) and inverse Pólya distributions of order  $(k_1, k_2)$ , respectively, and they can be studied using Theorems 2.1 to 2.3. For this particular exchangeable sequence, i.e. the  $PE(w, b, s)$  scheme, alternative formulae for the PMFs of  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$  were given in Sen *et al.* (2006); see Remark 2.1.

For Bernoulli trials which correspond to a Pólya–Eggenberger sampling scheme with replacements, i.e.  $s = 0$ , (4.1) gives  $\lambda_i = \lambda_1^i$ ,  $i = 1, 2, \dots, n$ , and  $\lambda_0 = 1$  with  $\lambda_1 = w/(w + b) = p$ , where  $p, 0 < p < 1$ , is the common success probability of the trials. For such trials,  $p_n(y)$  reduces to  $p^{n-y}q^y$ ,  $p + q = 1$  (see also Remark 2.2), and the distributions of  $N_{n;k_1,k_2}$  and  $T_{r;k_1,k_2}$  are called binomial and negative binomial of order  $(k_1, k_2)$ , respectively. Other particular values of  $s$  of special interest are  $s = -1$ ,  $s = w = b$ , and  $s = 1$ . The

respective distributions of  $N_{n;k_1,k_2}$  are called hypergeometric, uniform, and beta binomial of order  $(k_1, k_2)$ , whereas those of  $T_{r;k_1,k_2}$  are called negative hypergeometric, factorial, and beta Pascal of order  $(k_1, k_2)$ , respectively. However, the case in which  $s = -1$  for the RV  $T_{r;k_1,k_2}$  requires special attention because the sampling procedure might terminate without observing  $r$   $F$ - $S$  strings of length  $k_1 + k_2$  when the sampling is done without replacement. Its exact distribution can be derived using similar arguments as in Makri *et al.* (2007a, p. 663), (2007b, Proposition 4.5).

**4.2. Record models**

In this subsection we assume that  $\{Y_i\}_{i \geq 1}$  is a sequence of i.i.d. RVs with continuous distribution function  $F$ . For such sequences, we define record times  $K_j$  and record values  $V_j$  as follows (cf. Nevzorov (2001, pp. 56–57)):

$$K_1 = 1, \quad K_j = \min\{i > K_{j-1} : Y_i > \max(Y_1, Y_2, \dots, Y_{i-1})\}, \quad j = 2, 3, \dots, \\ \text{and } V_j = Y_{K_j}, \quad j = 1, 2, \dots,$$

that is,  $K_j$  is the index (or the position) of the  $j$ th record, the value of which is  $V_j$ . By convention,  $Y_1$  is a record (since  $K_1 = 1$ ).

Let  $Y'_1, Y'_2, \dots, Y'_n$  be i.i.d. RVs with continuous distribution function  $G$  and independent of  $\{Y_i\}_{i \geq 1}$ . If the  $j$ th record value  $V_j$  is chosen as a random threshold then the sequence associated with this record threshold model (RTM) defined by

$$X_i = \begin{cases} 1 & \text{if } Y'_i > V_j, \\ 0 & \text{otherwise,} \end{cases} \quad i = 1, 2, \dots, n, \quad j > 1, \tag{4.2}$$

is exchangeable and, under the hypothesis  $H_0: F = G$ , it holds (see Demir and Eryilmaz (2008)) that

$$p_n(y) = \sum_{i=0}^y (-1)^i \binom{y}{i} \frac{1}{(n - y + i + 1)^j}, \quad j > 1, \quad 0 \leq y \leq n. \tag{4.3}$$

Next, again considering the sequence  $\{Y_i\}_{i \geq 1}$ , we define (cf. Nevzorov (2001, pp. 57–58)) the record indicator model (RIM) sequence by

$$X_1 = 1 \quad \text{and} \quad X_i = \begin{cases} 1 & \text{if } M_i > M_{i-1}, \\ 0 & \text{otherwise,} \end{cases} \quad i = 2, 3, \dots, \tag{4.4}$$

where

$$M_i = \max(Y_1, Y_2, \dots, Y_i), \quad i = 1, 2, \dots$$

That is,  $X_i = 1$  if  $Y_i$  is an upper record (or simply a record) and  $X_i = 0$  otherwise. 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  $X_i = 1$ .

For sequences like  $\{Y_i\}_{i \geq 1}$ , the record indicators  $X_i$  have two important properties. They are independent RVs and

$$p_i = P(X_i = 1) = 1 - P(X_i = 0) = 1/i, \quad i = 1, 2, \dots \tag{4.5}$$

The sequences defined by (4.2) and (4.4) have been used as model underlined sequences of run-related statistics, referring to several enumerative schemes, in the works of Eryilmaz and

Tutuncu (2002), Chern and Hwang (2005), Demir and Eryilmaz (2008), and Makri and Psillakis (2009a). Occurrences of  $F$ - $S$  strings of length (at least)  $k_1 + k_2$  defined on the sequences (4.2) and (4.4) can be studied using Theorems 2.1, 2.2, and 2.4 with  $p_n(y)$  and the  $p_i$ s given by (4.3) and (4.5), respectively.

**4.3. Numerical examples**

In this section we clarify the theoretical results of Sections 2 and 3 by means of numerical examples. In Example 4.1 we present some comparative numerics referring to probabilities, means, and variances of the RVs  $N_{n;k_1,k_2}$  and  $N_{n;k_1,k_2}^c$  defined on several sequences. In Examples 4.2 and 4.3 we study the RVs  $N_{n;k_1,k_2}$ ,  $N_{n;k_1,k_2}^c$ , and  $T_{r;k_1,k_2}$  defined on the particular exchangeable and Poisson sequences discussed in Sections 4.1 and 4.2. Finally, in Example 4.4 we consider some asymptotic results for  $N_{n;k_1,k_2}$  defined on Bernoulli sequences.

**Example 4.1.** (*Numerical comparisons.*) In this example we give in Table 1 numerics concerning three model sequences of length  $n = 5$  which belong to the various kinds of binary sequences considered in the paper. They show a variety of possible configurations, by selecting several values of  $k_1$  and  $k_2$ , and shed some light on the similarities/discrepancies among the corresponding probabilities, means, and variances of the RVs  $N_{5;k_1,k_2}$  and  $N_{5;k_1,k_2}^c$ . The value  $n = 5$  was chosen to be small so that the required computations can also be carried out by hand. The sequences used in the table are as follows.

*Case I.* A Bernoulli sequence with a common success probability  $p = \frac{1}{2}$ .

*Case II.* An exchangeable sequence with  $\lambda_i = 1/(i + 1)$ ,  $i = 1, 2, \dots, 5$ .

*Case III.* A sequence of Poisson trials with  $p_i = 1/(i + 1)$ ,  $i = 1, 2, \dots, 5$ .

All the entries of Table 1 are exact values of probabilities, means, and variances, except the probabilities for the circular subcase of case III, which are approximating values computed via (3.2)–(3.4).

**Example 4.2.** (*Pólya–Eggenberger urn model.*) Let  $N_{n;k_1,k_2}$ ,  $N_{n;k_1,k_2}^c$ , and  $T_{r;k_1,k_2}$  be defined on a binary sequence derived by a Pólya–Eggenberger sampling scheme. In Table 2 we present the exact PMFs, means, and variances of  $N_{n;k_1,k_2}$  and  $N_{n;k_1,k_2}^c$  for  $n = 10$ ,  $k_1 = 2$ ,  $k_2 = 3$ ,  $w = b = 10$ , and some values of  $s$  of particular importance in the urn scheme  $PE(10, 10, s)$ ;

TABLE 1: Probabilities, means, and variances of  $N_{5;k_1,k_2}$  and  $N_{5;k_1,k_2}^c$ .

Case	$k_1$	$k_2$	$E(N_{5;k_1,k_2})$	$V(N_{5;k_1,k_2})$	$P(N_{5;k_1,k_2} < 1)$	$E(N_{5;k_1,k_2}^c)$	$V(N_{5;k_1,k_2}^c)$	$P(N_{5;k_1,k_2}^c < 1)$
I	1	1	1.0000	0.3750	0.1875	1.2500	0.3125	0.0625
	1	2	0.3750	0.2344	0.6250	0.6250	0.2344	0.3750
	2	2	0.1250	0.1094	0.8750	0.3125	0.2148	0.6875
II	2	3	0.0313	0.0303	0.9688	0.1563	0.1318	0.8438
	1	2	0.2500	0.1875	0.7500	0.4167	0.2431	0.5833
III	2	2	0.0667	0.0622	0.9333	0.1667	0.1389	0.8333
	2	3	0.0167	0.0164	0.9833	0.0833	0.0764	0.9167
	1	2	0.1000	0.0900	0.9000	0.3056	0.2122	0.6944
III	2	2	0.0333	0.0322	0.9667	0.2292	0.1767	0.7708
	2	3	0.0028	0.0028	0.9972	0.0625	0.0586	0.9375

TABLE 2: PMFs, means, and variances of  $N_{10;2,3}$  and  $N_{10;2,3}^c$  for a PE(10, 10,  $s$ ) model.

$s$	$x$	$P(N_{10;2,3} = x)$	$E(N_{10;2,3})$	$V(N_{10;2,3})$	$P(N_{10;2,3}^c = x)$	$E(N_{10;2,3}^c)$	$V(N_{10;2,3}^c)$
-1	0	0.792 158	0.208 978	0.167 580	0.657 386	0.348 297	0.238 353
	1	0.206 705			0.336 931		
	2	0.001 137			0.005 683		
0	0	0.813 477	0.187 500	0.154 297	0.692 383	0.312 500	0.224 609
	1	0.185 547			0.302 734		
	2	0.000 977			0.004 883		
1	0	0.830 043	0.170 807	0.143 333	0.719 575	0.284 679	0.212 145
	1	0.169 106			0.276 171		
	2	0.000 851			0.004 554		
10	0	0.900 433	0.100 000	0.090 866	0.835 498	0.166 667	0.143 218
	1	0.099 134			0.162 338		
	2	0.000 433			0.002 165		

TABLE 3: Stopping times, cumulative sums, means, and variances of  $T_{r;k_1,k_2}$  for a PE( $w, b, s$ ) model.

$s$	$w$	$b$	$k_1$	$k_2$	$r$	$t$	$\sum_{x=r(k_1+k_2)}^t P(T_{r;k_1,k_2} = x)$	$E(T_{r;k_1,k_2})$	$V(T_{r;k_1,k_2})$		
0	5	5	1	1	1	11	0.9941	3.92	3.49		
						17	0.9936	7.88	7.15		
						22	0.9916	11.79	10.62		
			2	2	1	25	0.9905	7.71	19.07		
						38	0.9904	15.58	40.25		
						59	0.9903	15.31	112.62		
	9	1	2	3	1	125	0.9883	30.22	555.38		
						47	0.9909	11.83	71.59		
						15	0.9915	4.33	5.71		
			1	5	5	1	1	24	0.9912	8.72	12.70
								32	0.9903	13.10	20.28
								59	0.9902	9.46	60.20
9	1	2			2	1	103	0.9901	19.33	254.96	
							125	0.9328	32.32	685.55	
							125	0.9323	17.38	484.83	
		1			2	1	1	125	0.8677	26.56	633.89
								125	0.9318	18.53	474.60
								125	0.9840	8.80	177.98
5	5	5	1	1	125	0.9677	15.54	284.83			
					125	0.9086	14.93	382.07			
					125	0.8218	23.34	576.70			

specifically, for  $s = -1, 0, 1, 10$ . In Table 3 we provide approximate mean values and variances of  $T_{r;k_1,k_2}$  for several PE( $w, b, s$ ) schemes with  $s \geq 0$  and for various values of  $k_1, k_2$ , and  $r$ . We also include the used (stopping) time  $t$  and  $\sum_{x=r(k_1+k_2)}^t P(T_{r;k_1,k_2} = x)$ , where  $t = \min\{125, \min\{n : \sum_{x=r(k_1+k_2)}^n P(T_{r;k_1,k_2} = x) \geq 0.99\}\}$ .

**Example 4.3.** (*Record model.*) For an illustration of the exact PMFs, means, and variances of  $N_{n;k_1,k_2}$  defined on an exchangeable sequence derived according to an RTM, we provide Table 4. In this table we selected the indicative values  $n = 5, 10$ ;  $k_1 = k_2 = 1, k_1 = 2, k_2 = 3$ ; and  $j = 2, 3, \text{ and } 5$ . We observe that an increase in  $j$  (the order of the record used as a threshold) leads to an increase in the probability  $P(N_{n;k_1,k_2} = 0)$ , as well as a decrease in the mean value and the variance of  $N_{n;k_1,k_2}$ . Table 5 provides exact waiting time probabilities for the  $r$ th occurrence of an  $F$ - $S$  string for several values of  $r, k_1, \text{ and } k_2$ . The RV  $T_{r;k_1,k_2}$  is defined on an RIM sequence. The data of the table admit the following interpretation. Let, for instance,  $r = 1$  and  $k_1 = k_2 = 2$ . Then the fourth column of the table gives the probabilities for the waiting time until two successive records following (at least) two successive values which are not records appear for the first time. Furthermore, we note that the first entries of every column of the table, i.e.  $P(T_{r;k_1,k_2} = r(k_1 + k_2))$ , equal 0. This is because, since  $X_1 = 1$  (or  $p_1 = 1$ ), it is impossible to have a  $(k_1, k_2)$  event in the remaining  $r(k_1 + k_2) - 1$  positions (after the first one) of the used RIM sequence  $\{X_i\}_{i \geq 1}$ .

**Example 4.4.** (*Limiting distributions.*) In order to illustrate the implementation of a normal approximation of  $N_{n;k_1,k_2}$  discussed in Remark 3.1, we present in Table 6 some indicative numerics. The Poisson approximation of  $N_{n;k_1,k_2}$  (see Huang and Tsai (1991)) as well as an upper bound,  $d_{TV}$ , of the total variation distance (see Vellaisamy (2004)) have also been used. The exact distributions may be computed using Theorem 2.1 or Theorem 2.4. The entries of Table 6 present some cases for which the root mean square error (RMSE) of the normal approximation of  $N_{n;k_1,k_2}$  is smaller than that of the corresponding Poisson approximation for the same values of  $p, n, k_1, \text{ and } k_2$ . Condition (3.6) is depicted as well. The numbers in parentheses are the alternative values of  $p, k_1, \text{ and } k_2$  for which the same distribution of  $N_{n;k_1,k_2}$  is derived. The RMSE between the exact  $f(x) = P(N_{n;k_1,k_2} = x)$  and an approximate PMF,

TABLE 4: PMFs, means, and variances of  $N_{n;1,1}$  and  $N_{n;2,3}$  for an RTM.

$x$	$n = 5$		$n = 10$		
	$j = 2$	$j = 3$	$j = 2$	$j = 3$	$j = 5$
	$P(N_{n;1,1} = x)$				
0	0.522 778	0.689 509	0.308 718	0.511 312	0.814 137
1	0.398 889	0.269 130	0.287 736	0.260 997	0.137 822
2	0.078 333	0.041 361	0.265 330	0.160 061	0.038 562
3			0.121 525	0.060 088	0.008 660
4			0.016 425	0.007 427	0.000 808
5			0.000 266	0.000 115	0.000 011
$E(N_{n;1,1})$	0.555 556	0.351 852	1.250 000	0.791 667	0.244 213
$V(N_{n;1,1})$	0.403 580	0.310 774	1.149 722	0.937 005	0.323 571
	$P(N_{n;2,3} = x)$				
0	0.989 722	0.995 745	0.938 580	0.974 557	0.997 214
1	0.010 278	0.004 255	0.061 173	0.025 388	0.002 781
2			0.000 247	0.000 085	0.000 005
$E(N_{n;2,3})$	0.010 278	0.004 255	0.061 667	0.025 528	0.002 791
$V(N_{n;2,3})$	0.010 172	0.004 237	0.058 357	0.025 045	0.002 794

TABLE 5: Waiting time probabilities of the  $r$ th occurrence of a  $(k_1, k_2)$  event for an RIM.

$x$	$P(T_{1;1,1} = x)$	$P(T_{1;1,2} = x)$	$P(T_{1;2,2} = x)$	$P(T_{1;2,3} = x)$	$P(T_{2;1,1} = x)$
2	0.000 000				
3	0.166 667	0.000 000			
4	0.166 667	0.041 667	0.000 000		0.000 000
5	0.125 000	0.033 333	0.016 667	0.000 000	0.025 000
6	0.088 889	0.025 000	0.016 667	0.002 778	0.044 444
7	0.064 484	0.018 254	0.014 286	0.002 381	0.051 587
8	0.048 512	0.013 765	0.011 905	0.001 786	0.051 190
9	0.037 751	0.010 714	0.009 755	0.001 323	0.047 704
10	0.030 203	0.008 573	0.008 056	0.000 992	0.043 377
11	0.024 712	0.007 014	0.006 734	0.000 755	0.039 113
12	0.020 593	0.005 845	0.005 700	0.000 586	0.035 217
13	0.017 425	0.004 946	0.004 881	0.000 463	0.031 764
14	0.014 936	0.004 239	0.004 225	0.000 372	0.028 737
15	0.012 944	0.003 674	0.003 691	0.000 302	0.026 093
16	0.011 326	0.003 215	0.003 251	0.000 249	0.023 783
17	0.009 994	0.002 837	0.002 885	0.000 208	0.021 758
18	0.008 883	0.002 521	0.002 577	0.000 175	0.019 979
19	0.007 948	0.002 256	0.002 315	0.000 149	0.018 408
20	0.007 153	0.002 030	0.002 092	0.000 127	0.017 016

TABLE 6: RMSE between exact and approximate distributions of  $N_{n;k_1,k_2}$  for indicative values of  $n, p, k_1$ , and  $k_2$ .

$p$	$k_1$	$k_2$	$n$	(3.6)	RMSE		$d_{TV}$		
					Normal	Poisson			
0.8(0.2)	1(3)	3(1)	24	False	0.0539	0.0915	0.6829		
			25	True	0.0526	0.0907	0.6905		
	1(2)	2(1)	25	False	0.0284	0.0628	0.6477		
			26	True	0.0276	0.0623	0.6514		
			1(1)	1(1)	29	False	0.0100	0.0281	0.5007
					30	True	0.0094	0.0270	0.5013
			1(2)	2(1)	100	True	0.0021	0.0113	0.4968
					100	True	0.0052	0.0235	0.6681
	1(3)	3(1)	100	True	0.0101	0.0353	0.7525		
			100	True	0.0101	0.0353	0.7525		
0.5(0.5)	1(1)	1(1)	9	False	0.0569	0.1166	0.6802		
			10	True	0.0492	0.1060	0.7024		
			100	True	0.0033	0.0226	0.7700		
	1(2)	2(1)	100	True	0.0051	0.0226	0.6529		
			100	True	0.0070	0.0174	0.4666		
	1(2,3)	3(2,1)	100	True	0.0070	0.0174	0.4666		

$f^*(x)$ , of  $N_{n;k_1,k_2}$  was computed using

$$RMSE(n, p, k_1, k_2) = \left\{ \frac{1}{1 + \lfloor n/(k_1 + k_2) \rfloor} \sum_{x=0}^{\lfloor n/(k_1+k_2) \rfloor} [f(x) - f^*(x)]^2 \right\}^{1/2} .$$

## Acknowledgement

The author would like to thank the anonymous referee for a thorough reading of the paper.

## References

- ALEVIZOS, P. D., PAPASTAVRIDIS, S. G. AND SYPSAS, P. (1993). Reliability of cyclic  $m$ -consecutive- $k$ -out-of- $n$ : $F$  systems. In *Proc. 2nd IASTED Int. Conf. Reliability, Quality Control, and Risk Assessment*, IASTED-ACTA Press, Anaheim, pp. 140–143.
- ANTZOULAKOS, D. L., BERSIMIS, S. AND KOUTRAS, M. V. (2003). On the distribution of the total number of run lengths. *Ann. Inst. Statist. Math.* **55**, 865–884.
- ARRATIA, R., GOLDSTEIN, L. AND GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17**, 9–25.
- BALAKRISHNAN, N. AND KOUTRAS, M. V. (2002). *Runs and Scans with Applications*. John Wiley, New York.
- CHARALAMBIDES, C. A. (1994). Success runs in a circular sequence of independent Bernoulli trials. In *Runs and Patterns in Probability*, eds A. P. Godbole and S. G. Papastavridis, Kluwer, Dordrecht, pp. 15–30.
- CHARALAMBIDES, C. A. (2002). *Enumerative Combinatorics*. Chapman and Hall/CRC, Boca Raton, FL.
- CHERN, H.-H. AND HWANG, H.-K. (2005). Limit distribution of the number of consecutive records. *Random Structures Algorithms* **26**, 404–417.
- DEMIR, S. AND ERYILMAZ, S. (2008). Run statistics in a sequence of arbitrarily dependent binary trials. To appear in *Statist. Papers*.
- ERYILMAZ, S. (2008). Run statistics defined on the multicolor urn model. *J. Appl. Prob.* **45**, 1007–1023.
- ERYILMAZ, S. (2009). Mean success run length. *J. Korean Statist. Soc.* **38**, 65–71.
- ERYILMAZ, S. AND DEMIR, S. (2007). Success runs in a sequence of exchangeable binary trials. *J. Statist. Planning Inference* **137**, 2954–2963.
- ERYILMAZ, S. AND TUTUNCU, G. Y. (2002). Success run model based on records. *J. Statist. Theory Appl.* **1**, 75–81.
- FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edn. John Wiley, New York.
- FU, J. C. AND LOU, W. Y. W. (2003). *Distribution Theory of Runs and Patterns and Its Applications*. World Scientific, River Edge, NJ.
- GEORGE, E. O. AND BOWMAN, D. (1995). A full likelihood procedure for analysing exchangeable binary data. *Biometrics* **51**, 512–523.
- GODBOLE, A. P. AND SCHAFFNER, A. A. (1993). Improved Poisson approximations for word patterns. *Adv. Appl. Prob.* **25**, 334–347.
- HIRANO, K. (1986). Some properties of the distributions of order  $k$ . In *Fibonacci Numbers and Their Applications*, eds A. N. Philippou, A. F. Horadam and G. E. Bergum, Reidel, Dordrecht, pp. 43–53.
- HOEFFDING, W. AND ROBBINS, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15**, 773–780.
- HUANG, W. T. AND TSAI, C. S. (1991). On a modified binomial distribution of order  $k$ . *Statist. Prob. Lett.* **11**, 125–131.
- JOHNSON, N. AND KOTZ, S. (1977). *Urn Models and Their Applications*. John Wiley, New York.
- KENDALL, D. G. (1967). On finite and infinite sequences of exchangeable events. *Studia Sci. Math. Hung.* **2**, 319–327.
- KOUTRAS, M. V. (2003). Applications of Markov chains to the distribution theory of runs and patterns. In *Stochastic Processes: Modelling and Simulation* (Handbook Statist. **21**), eds D. N. Shanbhag and C. R. Rao, North-Holland, Amsterdam, pp. 431–472.
- KOUTRAS, M. V., PAPADOPOULOS, G. K. AND PAPASTAVRIDIS, S. G. (1995). Runs on a circle. *J. Appl. Prob.* **32**, 396–404.
- MAKRI, F. S. AND PHILIPPOU, A. N. (1994). Binomial distributions of order  $k$  on the circle. In *Runs and Patterns in Probability*, eds A. P. Godbole and S. G. Papastavridis, Kluwer, Dordrecht, pp. 65–81.
- MAKRI, F. S. AND PHILIPPOU, A. N. (2005). On binomial and circular binomial distributions of order  $k$  for  $\ell$ -overlapping success runs of length  $k$ . *Statist. Papers* **46**, 411–432.
- MAKRI, F. S., PHILIPPOU, A. N. AND PSILLAKIS, Z. M. (2007a). Pólya, inverse Pólya, and circular Pólya distributions of order  $k$  for  $\ell$ -overlapping success runs. *Commun. Statist. Theory Meth.* **36**, 657–668.
- MAKRI, F. S., PHILIPPOU, A. N. AND PSILLAKIS, Z. M. (2007b). Success run statistics defined on an urn model. *Adv. Appl. Prob.* **39**, 991–1019.
- MAKRI, F. S. AND PSILLAKIS, Z. M. (2009a). On success runs of length exceeded a threshold. To appear in *Methodology Comput. Appl. Prob.*
- MAKRI, F. S. AND PSILLAKIS, Z. M. (2009b). On runs of length exceeding a threshold: normal approximation. To appear in *Statist. Papers*.
- NEVZOROV, V. B. (2001). *Records: Mathematical Theory*. American Mathematical Society, Providence, RI.
- PHILIPPOU, A. N. AND MAKRI, F. S. (1986). Successes, runs and longest runs. *Statist. Prob. Lett.* **4**, 211–215.

- PHILIPPOU, A. N. AND MAKRI, F. S. (1990). Longest circular runs with an application to reliability via Fibonacci-type polynomials of order  $k$ . In *Applications of Fibonacci Numbers*, eds E. G. Bergum *et al.*, Kluwer, Dordrecht, pp. 281–286.
- PHILIPPOU, A. N. AND MUWAFI, A. A. (1982). Waiting for the  $k$ th consecutive success and the Fibonacci sequence of order  $k$ . *Fibonacci Quart.* **20**, 28–32.
- PHILIPPOU, A. N., GEORGIU, C. AND PHILIPPOU, G. N. (1983). A generalized geometric distribution and some of its properties. *Statist. Prob. Lett.* **1**, 171–175.
- SEN, K., AGARWAL, M. AND BHATTACHARYA, S. (2006). Pólya–Eggenberger F-S models of order  $(k_1, k_2)$ . *Studia Sci. Math. Hung.* **43**, 1–31.
- VELLAISAMY, P. (2004). Poisson approximation for  $(k_1, k_2)$ -events via the Stein–Chen method. *J. Appl. Prob.* **41**, 1081–1092.