

RUN STATISTICS DEFINED ON THE MULTICOLOR URN MODEL

SERKAN ERYILMAZ,* *Izmir University of Economics*

Abstract

Recently, Makri, Philippou and Psillakis (2007b) studied the exact distribution of success run statistics defined on an urn model. They derived the exact distributions of various success run statistics for a sequence of binary trials generated by the Pólya–Eggenberger sampling scheme. In our study we derive the joint distributions of run statistics defined on the multicolor urn model using a simple unified combinatorial approach and extend some of the results of Makri, Philippou and Psillakis (2007b). As a consequence of our results, we obtain the joint distributions of success and failure runs defined on the two-color urn model. The results enable us to compute the characteristics of particular consecutive-type systems and start-up demonstration tests.

Keywords: Consecutive- k , r -out-of- n : DFM system; Dirichlet measure; multicolor Pólya sampling scheme; multistate exchangeable trial; runs; start-up demonstration test; waiting time

2000 Mathematics Subject Classification: Primary 60C05

Secondary 62E15

1. Introduction

Runs and run-related statistics are useful tools for analyzing sequential data appearing in many fields including statistical process control, reliability, and molecular biology. The distribution theory of runs has been well developed using various techniques such as combinatorial, probability generating function, and Markov chain approaches. Combinatorial methods have been successfully used to derive the distribution of runs in a sequence of independent, identically distributed (i.i.d.) binary trials (Philippou and Makri (1985), (1986), Makri and Philippou (2005)). For an extensive review of the topic, we refer the reader to Balakrishnan and Koutras (2002), as well as Fu and Lou (2003). The topic is still very attractive because of its wide range of applications. Some recent contributions to the topic include Inoue and Aki (2005), Makri *et al.* (2007a), Eryilmaz and Demir (2007), and Antzoulakos and Boutsikas (2007).

Early discussions on runs are based on a sequence of binary trials (Fu and Lou (2007), Wald and Wolfowitz (1940), Mosteller (1941), Wolfowitz (1943)). Runs of sequences consisting of multistate trials have also been studied in the literature using various techniques (Fu (1996), Han and Aki (1999), Vaggelatos (2003), Eryilmaz (2008b)). There are various definitions of runs for both binary and multistate trials. Various types of run statistics are used depending on the type of problem. For example, the numbers of runs of length exactly equal to and greater than or equal to a specified value, and the longest run are useful in statistical hypothesis testing (Lou (1996), Koutras and Alexandrou (1997)). The longest run random variable is also an important tool in reliability analysis (Philippou (1988), Chao *et al.* (1995), Eryilmaz (2008a))

Received 4 April 2008; revision received 3 September 2008.

* Postal address: Izmir University of Economics, Department of Mathematics, 35330 Izmir, Turkey.

Email address: serkan.eryilmaz@ieu.edu.tr

and molecular biology (Goldstein (1990), Lou (2003), Chang *et al.* (2003), Wang and Cheung (2004)). Waiting time random variables associated with runs are indicative in statistical process control applications (Fu *et al.* (2003), Chakraborti and Eryilmaz (2007)).

The problem of finding the distribution of runs in a sequence that consists of multistate trials might be tedious. However, this problem should be carefully studied because the outcomes of most of the random processes appear as a sequence of trials which can be classified into more than two groups. For example, deoxyribonucleic acid (DNA) sequences consist of four different nucleotides (with the bases A, T, C, and G).

In this paper we study the distribution of run statistics defined on the multicolor urn model. This model generates a sequence that consists of multistate trials whose outcomes are denoted by the type of ball selected from the urn. In this multicolor urn scheme, a ball is drawn from the urn initially containing m_j balls of color j , $j = 1, 2, \dots, t$, and its color is noted. If a ball of color j is drawn at a stage, s balls of color j , $j = 1, 2, \dots, t$, are added to the urn. Drawing a ball of color j is considered as a trial of type j , $j = 1, 2, \dots, t$. This scheme is repeated n times and a sequence consisting of trials, namely $\{1, 2, \dots, t\}$, is derived. For a review of the theory and applications of urn models, the reader is referred to Johnson and Kotz (1977).

In a recent paper, Makri *et al.* (2007b) studied the distribution of run statistics defined on a two-color ($t = 2$) version of the abovementioned urn model. In the present paper we extend some of the results of Makri *et al.* (2007b) to the multicolor urn model using a simple unified combinatorial approach. Our paper is organized as follows. In Section 2 we provide the definitions, notation, and some preliminary results that will be used throughout the paper. In Section 3 we present general expressions for the joint distributions of run statistics defined on the multicolor urn model. In Section 4 we provide an alternative definition for run statistics, which is helpful in obtaining bounds. Section 5 includes some applications of the material presented in this paper which can be helpful in the fields other than the urn model.

2. Definitions, notation, and preliminary results

Consider a generalized Pólya urn which initially contains m_j balls of color j , $j = 1, 2, \dots, t$, $m_1 + \dots + m_t = m$. At each stage a ball is drawn from the urn and its color is noted. If a ball of color j is drawn at a stage, s balls of color j , $j = 1, 2, \dots, t$, are added to the urn. This multicolor Pólya sampling scheme is repeated and a sequence $\{Z_i\}_{i \geq 1}$, whose elements represent the type of ball selected from the urn, is obtained. That is, $Z_i = j$ if the i th ball selected is of type j , $i = 1, 2, \dots$ and $j = 1, 2, \dots, t$.

Let Z_1, Z_2, \dots, Z_n denote the first n outcomes associated with this sampling scheme. It is obvious that the elements of Z_1, Z_2, \dots, Z_n are dependent. For $i = 1, 2, \dots$ and $j = 1, 2, \dots, t$, define the following random variables:

- $R_n^{(j)}$: the total number of runs of type j ,
- $\theta_i^{(j)}$: the length of the i th run of type j ,
- $E_{n,k_j}^{(j)}$: the total number of runs of type j with length exactly equal to k_j ,
- $G_{n,k_j}^{(j)}$: the total number of runs of type j with length at least k_j ,
- $L_n^{(j)}$: the length of the longest run of type j ,
- L_n : the length of the longest run of any type in Z_1, Z_2, \dots, Z_n ,

- $W_{k_j}^{(j)}$: the waiting time for the first ‘j’ run of length k_j ,
- W_k : the waiting time for the first run of any type with length k in Z_1, Z_2, \dots

These definitions can be illustrated with the following example. Consider an urn which contains three different colors of balls ($t = 3$). Let the first 15 trials be

$$221333112111231. \tag{1}$$

Then, $R_n^{(1)} = 4, R_n^{(2)} = 3, R_n^{(3)} = 2, \theta_1^{(1)} = 1, \theta_2^{(1)} = 2, \theta_3^{(1)} = 3, \theta_4^{(1)} = 1, \theta_1^{(2)} = 2, \theta_2^{(2)} = 1, \theta_3^{(2)} = 1, \theta_1^{(3)} = 3, \theta_2^{(3)} = 1, E_{15,1}^{(1)} = 2, E_{15,2}^{(1)} = 1, E_{15,3}^{(1)} = 1, E_{15,1}^{(2)} = 2, E_{15,2}^{(2)} = 1, E_{15,3}^{(2)} = 0, G_{15,2}^{(1)} = 2, G_{15,3}^{(2)} = 0, G_{15,2}^{(3)} = 1, L_{15}^{(1)} = 3, L_{15}^{(2)} = 2, L_{15}^{(3)} = 3, L_{15} = 3, W_3^{(1)} = 12, \text{ and } W_3 = 6.$

We readily have the following expressions:

$$E_{n,k_j}^{(j)} = \sum_{i=1}^{R_n^{(j)}} \mathbf{1}\{\theta_i^{(j)} = k_j\}, \quad G_{n,k_j}^{(j)} = \sum_{i=1}^{R_n^{(j)}} \mathbf{1}\{\theta_i^{(j)} \geq k_j\},$$

$$L_n^{(j)} = \max_{1 \leq i \leq R_n^{(j)}} \theta_i^{(j)}, \quad L_n = \max_{1 \leq j \leq t} L_n^{(j)},$$

$$W_k = \min(W_k^{(1)}, W_k^{(2)}, \dots, W_k^{(t)}),$$

where $\mathbf{1}\{A\} = 1$ if A occurs and $\mathbf{1}\{A\} = 0$ otherwise.

As is seen, most of the run statistics can be expressed as a compound random variable, which is a function of the run lengths and the total number of runs. That is, if $X_{n,k_j}^{(j)}$ represents any of the random variables $E_{n,k_j}^{(j)}, G_{n,k_j}^{(j)}$, or $L_n^{(j)}$, then it can be viewed mathematically as

$$X_{n,k_j}^{(j)} = \phi_{k_j}^{(j)}(\theta_1^{(j)}, \dots, \theta_{R_n^{(j)}}^{(j)}), \quad j = 1, 2, \dots, t. \tag{2}$$

Thus, the distribution of any run statistic based on a sequence Z_1, Z_2, \dots, Z_n can be evaluated using the distribution of the vector

$$(\theta_1^{(1)}, \dots, \theta_{R_n^{(1)}}^{(1)}, \theta_1^{(2)}, \dots, \theta_{R_n^{(2)}}^{(2)}, \dots, \theta_1^{(t)}, \dots, \theta_{R_n^{(t)}}^{(t)}).$$

It is known that the repetition of the multicolor Pólya sampling scheme defined above derives a sequence of exchangeable multistate trials (see, e.g. Hill *et al.* (1987)). If Z_1, Z_2, \dots is a sequence of multistate exchangeable trials with $P\{Z_i = j\} = p_j, j = 1, 2, \dots, t, i \geq 1, \sum_{j=1}^t p_j = 1$, then as a generalization of de Finetti’s representation from an infinite sequence of exchangeable binary variables to an infinite sequence of exchangeable multistate variables, we have

$$P\{Z_1 = 1, \dots, Z_{n_1} = 1, Z_{n_1+1} = 2, \dots, Z_{n_1+n_2} = 2, \dots, Z_{n_1+n_2+\dots} = t, \dots, Z_n = t\}$$

$$= \int_{\sum p_j=1} \prod_{j=1}^t p_j^{n_j} dG(p_1, \dots, p_{t-1}) \tag{3}$$

(see, e.g. Zabell (1982)), where $dG(p_1, \dots, p_{t-1})$ is the de Finetti measure that exists on the simplex.

The sequence (Z_1, Z_2, \dots) derived in the aforementioned multicolor Pólya urn scheme is exchangeable with the de Finetti measure, which is Dirichlet and given by

$$dG(p_1, \dots, p_{t-1}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \dots + \alpha_t)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\dots\Gamma(\alpha_t)} p_1^{\alpha_1-1} p_2^{\alpha_2-1} \dots p_{t-1}^{\alpha_{t-1}-1} \times (1 - p_1 - \dots - p_{t-1})^{\alpha_t-1} dp_1 \dots dp_{t-1}, \tag{4}$$

where $\sum_{j=1}^{t-1} p_j \leq 1$ and $\alpha_j = m_j/s, j = 1, 2, \dots, t, s > 0$. The Dirichlet measure given by (4) has beta marginals, i.e.

$$dG(p_j) = \frac{1}{B(\alpha_j, \sum_{i \neq j} \alpha_i)} p_j^{\alpha_j-1} (1 - p_j)^{(\sum_{i \neq j} \alpha_i)-1} dp_j \text{ for } j = 1, 2, \dots, t, \tag{5}$$

where $B(a, b)$ is the beta function defined by

$$B(a, b) = \int_0^1 x^{a-1} (1 - x)^{b-1} dx \text{ for } a > 0 \text{ and } b > 0.$$

For $s > 0$, using (4) in (3), the probability of obtaining n_1 balls of color 1, n_2 balls of color 2, etc. in n drawings is obtained as

$$p_n(n_1, \dots, n_t) = \frac{\Gamma(\sum_{i=1}^t \alpha_i)}{\Gamma(\sum_{i=1}^t \alpha_i + n)} \prod_{j=1}^t \frac{\Gamma(\alpha_j + n_j)}{\Gamma(\alpha_j)},$$

where $\sum_{j=1}^t n_j = n$ and $\Gamma(a)$ is the gamma function defined by

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \text{ for } a > 0.$$

The cases in which $s = 0$ (sampling with replacement) and $s = -1$ (sampling without replacement) are of special interest in urn models. For $s = 0, p_n(n_1, \dots, n_t)$ reduces to $p_1^{n_1} p_2^{n_2} \dots p_t^{n_t}$ with $p_j = m_j/m, j = 1, 2, \dots, t$. That is, this case corresponds to i.i.d. multistate trials with respective probabilities p_1, p_2, \dots, p_t . For $s = -1$, we have

$$p_n(n_1, \dots, n_t) = \frac{m_1^{(n_1)} m_2^{(n_2)} \dots m_t^{(n_t)}}{m^{(n_1)} (m - n_1)^{(n_2)} \dots (m_t - \sum_{i=1}^{t-1} n_i)^{(n_t)},}$$

where $x^{(a)} = x(x - 1) \dots (x - a + 1)$, denoting the a th falling factorial of x with $x^{(0)} = 1$.

Throughout the paper, for integers a and $b, [a]$ denotes the integer part of a and $\binom{a}{b}$ denotes the extended binomial coefficient (see Feller (1968, pp. 50, 63)). For convenience, we also apply $\sum_{i=a}^b = 0$ for $a > b$.

The following lemma will be helpful in our developments.

Lemma 1. (Eryilmaz (2008b).) *Let Z_1, Z_2, \dots be an infinite sequence of exchangeable multistate trials. Suppose that each trial has t ($t \geq 2$) possible outcomes labeled $\{1, 2, \dots, t\}$ with $P\{Z_i = j\} = p_j, j = 1, 2, \dots, t, i \geq 1$. Then*

$$P\{\theta_1^{(1)} = i_1^{(1)}, \dots, \theta_{r_1}^{(1)} = i_{r_1}^{(1)}, \dots, \theta_1^{(t)} = i_1^{(t)}, \dots, \theta_{r_t}^{(t)} = i_{r_t}^{(t)}, R_n^{(1)} = r_1, \dots, R_n^{(t)} = r_t\} = F_t(r_1, \dots, r_t) \int_{\sum p_j=1} \prod_{j=1}^t p_j^{n_j} dG(p_1, \dots, p_{t-1}),$$

where $n_j = \sum_{m=1}^{r_j} i_m^{(j)}$,

$$F_t(r_1, \dots, r_t) = (-1)^r \sum_{m_1=1}^{r_1} \dots \sum_{m_t=1}^{r_t} (-1)^m \binom{r_1-1}{m_1-1} \dots \binom{r_t-1}{m_t-1} \binom{m}{m_1, \dots, m_t},$$

$r = \sum_{j=1}^t r_j \leq n$, and $m = \sum_{j=1}^t m_j$.

Note that the number $F_t(r_1, \dots, r_t)$ denotes the total number of ways for getting r_1 runs of type 1, r_2 runs of type 2, and so on, without two adjacent runs being of the same type, and it has been derived by Kong (2006). When $t = 2$, $F_2(r_1, r_2)$ can be simplified as $\binom{2}{r_1-r_2+1}$.

In particular, for the multicolor Pólya sampling scheme, we have

$$\begin{aligned} P\{\theta_1^{(1)} = i_1^{(1)}, \dots, \theta_{r_1}^{(1)} = i_{r_1}^{(1)}, \dots, \theta_1^{(t)} = i_1^{(t)}, \dots, \theta_{r_t}^{(t)} = i_{r_t}^{(t)}, R_n^{(1)} = r_1, \dots, R_n^{(t)} = r_t\} \\ = F_t(r_1, \dots, r_t) p_n(n_1, \dots, n_t). \end{aligned} \tag{6}$$

3. Joint distribution of run statistics

In this section we provide the joint distribution of run statistics defined on the multicolor urn model described above. Let $X_{n,k_j}^{(j)}$ denote any run statistic which can be represented as a function of the run lengths and the total number of runs of the type with color j , $j = 1, 2, \dots, t$. The following theorem provides the joint distribution of the random variables $X_{n,k_1}^{(1)}, X_{n,k_2}^{(2)}, \dots, X_{n,k_t}^{(t)}$.

Theorem 1. *The joint distribution of $X_{n,k_1}^{(1)}, X_{n,k_2}^{(2)}, \dots, X_{n,k_t}^{(t)}$ is given by*

$$\begin{aligned} P\{X_{n,k_1}^{(1)} \in B_1, X_{n,k_2}^{(2)} \in B_2, \dots, X_{n,k_t}^{(t)} \in B_t\} \\ = \sum_{\mathbf{r}} \sum_{\mathbf{n}} \prod_{m=1}^t |I_{\phi_{k_m}^{(m)}}^{(m)}(B_m)| F_t(r_1, \dots, r_t) p_n(n_1, \dots, n_t), \end{aligned} \tag{7}$$

where $\mathbf{r} = (r_1, \dots, r_t)$, $\mathbf{n} = (n_1, \dots, n_t)$, and, for $m = 1, 2, \dots, t$, the B_m s are Borel sets,

$$I_{\phi_{k_m}^{(m)}}^{(m)}(B_m) = \{i_1^{(m)}, \dots, i_{r_m}^{(m)} : i_1^{(m)} + \dots + i_{r_m}^{(m)} = n_m; \phi_{k_m}^{(m)}(i_1^{(m)}, \dots, i_{r_m}^{(m)}) \in B_m\},$$

and $|A|$ shows the cardinality of the set A . The first and second sums are taken respectively over $r_1 + \dots + r_t \leq n$ and $n_1 + \dots + n_t = n$.

Proof. In view of (2) and conditioning on $(R_n^{(1)}, \dots, R_n^{(t)})$ we have

$$\begin{aligned} P\{X_{n,k_1}^{(1)} \in B_1, \dots, X_{n,k_t}^{(t)} \in B_t\} \\ = \sum_{r_1} \dots \sum_{r_t} P\{\phi_{k_1}^{(1)}(\theta_1^{(1)}, \dots, \theta_{r_1}^{(1)}) \in B_1, \dots, \phi_{k_t}^{(t)}(\theta_1^{(t)}, \dots, \theta_{r_t}^{(t)}) \in B_t, \\ R_n^{(1)} = r_1, \dots, R_n^{(t)} = r_t\}. \end{aligned}$$

Now, conditioning on the frequencies of each type of object, we obtain

$$\begin{aligned}
 & P\{X_{n,k_1}^{(1)} \in B_1, \dots, X_{n,k_t}^{(t)} \in B_t\} \\
 &= \sum_{r_1} \dots \sum_{r_t} \sum_{n_1} \dots \sum_{n_t} P\{\phi_{k_1}^{(1)}(\theta_1^{(1)}, \dots, \theta_{r_1}^{(1)}) \in B_1, \dots, \phi_{k_t}^{(t)}(\theta_1^{(t)}, \dots, \theta_{r_t}^{(t)}) \in B_t, \\
 &\quad R_n^{(1)} = r_1, \dots, R_n^{(t)} = r_t, S_n^{(1)} = n_1, \dots, S_n^{(t)} = n_t\} \\
 &= \sum_r \sum_n \sum_{i_1^{(1)} + \dots + i_{r_1}^{(1)} = n_1} \dots \sum_{\phi_{k_1}^{(1)}(i_1^{(1)}, \dots, i_{r_1}^{(1)}) \in B_1} \dots \\
 &\quad \times \sum_{i_1^{(t)} + \dots + i_{r_t}^{(t)} = n_t} \dots \sum_{\phi_{k_t}^{(t)}(i_1^{(t)}, \dots, i_{r_t}^{(t)}) \in B_t} P\{\theta_1^{(1)} = i_1^{(1)}, \dots, \theta_{r_1}^{(1)} = i_{r_1}^{(1)}, \dots, \theta_1^{(t)} = i_1^{(t)}, \dots, \theta_{r_t}^{(t)} = i_{r_t}^{(t)}, \\
 &\quad R_n^{(1)} = r_1, \dots, R_n^{(t)} = r_t\}, \tag{8}
 \end{aligned}$$

where $S_n^{(j)}$ denotes the total number of objects of type j . The contribution of the m th multiple sum ($m = 1, \dots, t$) to the expression inside the sums is the total number of integer solutions to the system

$$i_1^{(m)} + \dots + i_{r_m}^{(m)} = n_m \quad \text{such that} \quad \phi_{k_m}^{(m)}(i_1^{(m)}, \dots, i_{r_m}^{(m)}) \in B_m, i_1^{(m)}, \dots, i_{r_m}^{(m)} > 0. \tag{9}$$

This number coincides with the cardinality of the set $I_{\phi_m}^{(m)}(B_m)$. The proof is now completed using (6) in (8).

Joint distributions of various run statistics can be obtained using Theorem 1 with the appropriate selections of B_m s and the functions $\phi_{k_m}^{(m)}$, $m = 1, 2, \dots, t$. The cardinalities $|I_{\phi_m}^{(m)}(B_m)|$, $m = 1, 2, \dots, t$, completely depend on the selection of $\phi_{k_j}^{(j)}$ (and, hence, $X_{n,k_j}^{(j)}$), and the computation of $|I_{\phi_m}^{(m)}(B_m)|$ is enough to derive the corresponding distributions. Thus, the problem of finding the joint distributions of $X_{n,k_1}^{(1)}, X_{n,k_2}^{(2)}, \dots, X_{n,k_t}^{(t)}$ is a combinatorial one, specifically the determination of the total number of integer solutions to system (9).

Theorem 2. *The joint probability mass function of $E_{n,k_1}^{(1)}, E_{n,k_2}^{(2)}, \dots, E_{n,k_t}^{(t)}$ is given by*

$$\begin{aligned}
 & P\{E_{n,k_1}^{(1)} = x_1, E_{n,k_2}^{(2)} = x_2, \dots, E_{n,k_t}^{(t)} = x_t\} \\
 &= \sum_r \sum_n \prod_{m=1}^t \binom{r_m}{x_m} A(n_m - r_m - x_m(k_m - 1), r_m - x_m, k_m - 1) \\
 &\quad \times F_t(r_1, \dots, r_t) p_n(n_1, \dots, n_t),
 \end{aligned}$$

where

$$A(\alpha, r, k) = \sum_{j=0}^{[\alpha/k]} (-1)^j \binom{r}{j} \binom{\alpha - (k+1)j + r - 1}{\alpha - jk}. \tag{10}$$

Proof. For $m = 1, 2, \dots, t$, choosing $B_m = \{x_m\}$ and

$$\phi_{k_m}^{(m)}(i_1^{(m)}, \dots, i_{r_m}^{(m)}) = \sum_{j=1}^{r_m} \mathbf{1}\{i_j^{(m)} = k_m\},$$

we have

$$I_{\phi_{k_m}^{(m)}}^{(m)}(\{x_m\}) = \{(i_1^{(m)}, \dots, i_{r_m}^{(m)}) : i_1^{(m)} + \dots + i_{r_m}^{(m)} = n_m; x_m \text{ of } i_1^{(m)}, \dots, i_{r_m}^{(m)} \text{ equals } k_m \\ \text{and } r_m - x_m \text{ of } i_1^{(m)}, \dots, i_{r_m}^{(m)} \text{ differs from } k_m\}.$$

We readily have

$$|I_{\phi_{k_m}^{(m)}}^{(m)}(\{x_m\})| = \binom{r_m}{x_m} A(n_m - r_m - x_m(k_m - 1), r_m - x_m, k_m - 1),$$

where $A(n_m - r_m - x_m(k_m - 1), r_m - x_m, k_m - 1)$ is the total number of integer solutions to the system $i_1^{(m)} + \dots + i_{r_m - x_m}^{(m)} = n_m - k_m x_m$ such that $i_1^{(m)} \neq k_m, \dots, i_{r_m - x_m}^{(m)} \neq k_m; i_1^{(m)} > 0, \dots, i_{r_m - x_m}^{(m)} > 0$, or, equivalently, the number of allocations of $n_m - r_m - x_m(k_m - 1)$ indistinguishable balls into $r_m - x_m$ distinguishable cells, where no cell has exactly $k_m - 1$ balls. Equation (10) gives $A(n_m - r_m - x_m(k_m - 1), r_m - x_m, k_m - 1)$ with $\alpha = n_m - r_m - x_m(k_m - 1), r = r_m - x_m$, and $k = k_m - 1$ (see, e.g. Sen *et al.* (2003)). Thus, the proof is completed.

In the following example we compute the probability $P\{E_{5,2}^{(1)} = 1, E_{5,1}^{(2)} = 2, E_{5,1}^{(3)} = 1\}$. We choose $n = 5$ so that the computations can also be done by hand, by writing out all possible sequences from five trivariate trials.

Example 1. Let us consider an urn with initial composition $m_1 = 2, m_2 = 2, m_3 = 1$, and $s = 1$. Then

$$P\{E_{5,2}^{(1)} = 1, E_{5,1}^{(2)} = 2, E_{5,1}^{(3)} = 1\} \\ = \sum_{r_1 \geq 1} \sum_{r_2 \geq 2} \sum_{r_3 \geq 1} \sum_{n_1 \geq r_1} \sum_{n_2 \geq r_2} \binom{r_1}{1} \binom{r_2}{2} \binom{r_3}{1} A(n_1 - r_1 - 1, r_1 - 1, 1) \\ \times A(n_2 - r_2, r_2 - 2, 0) A(n - n_1 - n_2 - r_3, r_3 - 1, 0) \\ \times F_3(r_1, r_2, r_3) p_5(n_1, n_2, n - n_1 - n_2) \\ = F_3(1, 2, 1) p_5(2, 2, 1) \\ = 6p_5(2, 2, 1),$$

where $p_5(2, 2, 1) = \Gamma(5)\Gamma(4)\Gamma(4) / \Gamma(10)$.

Corollary 1. For the two-color ($t = 2$) urn model, if getting a ball of color 1 or 2 represents success or, respectively, failure, then the joint distribution of the total number of success runs of length exactly k_1 and the number of failure runs of length exactly k_2 is given by

$$P\{E_{n,k_1}^{(1)} = x_1, E_{n,k_2}^{(2)} = x_2\} \\ = \sum_{r_1} \sum_{r_2} \sum_{n_1} \binom{r_1}{x_1} \binom{r_2}{x_2} A(n_1 - r_1 - x_1(k_1 - 1), r_1 - x_1, k_1 - 1) \\ \times A(n - n_1 - r_2 - x_2(k_2 - 1), r_2 - x_2, k_2 - 1) \binom{2}{r_1 - r_2 + 1} p_n(n_1, n - n_1).$$

Theorem 3. *The joint probability mass function of $G_{n,k_1}^{(1)}, G_{n,k_2}^{(2)}, \dots, G_{n,k_t}^{(t)}$ is given by*

$$\begin{aligned} &P\{G_{n,k_1}^{(1)} = x_1, G_{n,k_2}^{(2)} = x_2, \dots, G_{n,k_t}^{(t)} = x_t\} \\ &= \sum_r \sum_n \prod_{m=1}^t \binom{r_m}{x_m} C(r_m, r_m - x_m, k_m - 1, n_m) \\ &\quad \times F_t(r_1, \dots, r_t) p_n(n_1, \dots, n_t), \end{aligned}$$

where

$$C(r, j, k, l) = \sum_{z=0}^j (-1)^z \binom{j}{z} \binom{r+l-j-(k+1)(r-j)-zk-1}{r-1}.$$

Proof. For $m = 1, 2, \dots, t$, choosing $B_m = \{x_m\}$ and

$$\phi_{k_m}^{(m)}(i_1^{(m)}, \dots, i_{r_m}^{(m)}) = \sum_{j=1}^{r_m} \mathbf{1}\{i_j^{(m)} \geq k_m\},$$

we have

$$\begin{aligned} I_{\phi_{k_m}^{(m)}}(\{x_m\}) &= \{(i_1^{(m)}, \dots, i_{r_m}^{(m)}) : i_1^{(m)} + \dots + i_{r_m}^{(m)} = n_m; x_m \text{ of } i_1^{(m)}, \dots, i_{r_m}^{(m)} \geq k_m \\ &\quad \text{and } r_m - x_m \text{ of } i_1^{(m)}, \dots, i_{r_m}^{(m)} < k_m\}. \end{aligned}$$

It is clear that

$$|I_{\phi_{k_m}^{(m)}}(\{x_m\})| = \binom{r_m}{x_m} C(r_m, r_m - x_m, k_m - 1, n_m),$$

where $C(r, j, k, l)$ is the total number of integer solutions to the system $i_1 + \dots + i_r = l$ such that $i_1 < k+1, \dots, i_j < k+1, i_{j+1} \geq k+1, \dots, i_r \geq k+1; i_1 > 0, \dots, i_r > 0$, or, equivalently, the total number of integer solutions to the system $y_1 + \dots + y_r = l - j - (k+1)(r-j)$ such that $0 \leq y_1 < k, \dots, 0 \leq y_j < k, y_{j+1} \geq 0, \dots, y_r \geq 0$. Let A_i be the event of $y_i \geq k, i = 1, \dots, j$. Then

$$C(r, j, k, l) = \left| \bigcap_{t=1}^j A_t^c \right| = s(r, j, k, l) - \left| \bigcup_{t=1}^j A_t \right|,$$

where $s(r, j, k, l)$ denotes the total number of integer solutions to $y_1 + \dots + y_r = l - j - (k+1)(r-j)$ such that $y_i \geq 0, i = 1, \dots, r$. The number $C(r, j, k, l)$ can now be easily obtained by using the inclusion–exclusion principle along with Theorem 2.12 of Charalambides (2002, p. 69). Thus, the proof is completed.

Corollary 2. *Choosing $\{x_m\} = \{0\}, m = 1, 2, \dots, t$, in Theorem 3 we obtain*

$$\begin{aligned} &P\{L_n^{(1)} < k_1, L_n^{(2)} < k_2, \dots, L_n^{(t)} < k_t\} \\ &= \sum_r \sum_n \prod_{m=1}^t C(r_m, r_m, k_m - 1, n_m) F_t(r_1, \dots, r_t) p_n(n_1, \dots, n_t), \end{aligned}$$

which coincides with Corollary 4 of Eryilmaz (2008b).

Remark 1. Sen *et al.* (2002, Theorem 6.2) obtained the probability mass function of the longest run of any type denoted by $M_n = \max(L_n^{(1)}, L_n^{(2)})$ for the two-color urn model. Their result is based on similar combinatorial formulae that we have used and can be readily obtained from Corollary 2, since $P\{M_n = x\} = P\{L_n^{(1)} < x + 1, L_n^{(2)} < x + 1\} - P\{L_n^{(1)} < x, L_n^{(2)} < x\}$.

The distribution of the waiting time random variable W_k can be easily obtained using the dual relationship

$$\begin{aligned} P\{W_k > n\} &= P\{W_k^{(1)} > n, W_k^{(2)} > n, \dots, W_k^{(t)} > n\} \\ &= P\{L_n^{(1)} < k, L_n^{(2)} < k, \dots, L_n^{(t)} < k\} \\ &= P\{L_n < k\}. \end{aligned}$$

Thus, we obtain the following corollary.

Corollary 3. *Choosing $k_1 = k_2 = \dots = k_t = k$ in Corollary 2 we have*

$$P\{W_k > n\} = \sum_r \sum_n \prod_{m=1}^t C(r_m, r_m, k - 1, n_m) F_t(r_1, \dots, r_t) p_n(n_1, \dots, n_t).$$

Corollary 4. *Choosing $k_1 = k_2 = \dots = k_t = 1$ in Theorem 3 we have*

$$\begin{aligned} P\{R_n^{(1)} = r_1, R_n^{(2)} = r_2, \dots, R_n^{(t)} = r_t\} \\ = \sum_n \prod_{m=1}^t \binom{n_m - 1}{r_m - 1} F_t(r_1, \dots, r_t) p_n(n_1, \dots, n_t), \end{aligned}$$

which coincides with Corollary 2 of Eryilmaz (2008b).

4. Alternative definition of run statistics

In this section we provide an alternative definition for run statistics which might be useful for various purposes. For a sequence Z_1, Z_2, \dots, Z_n that consists of multistate trials with possible outcomes $\{1, 2, \dots, t\}$, let $\eta_j^{(i)}$ denote the length of the i th run at the j th stage, $i = 1, 2, \dots, t$ and $j = 1, 2, \dots, n$. That is, we define a vector

$$\eta_j^\top = (\eta_j^{(1)}, \eta_j^{(2)}, \dots, \eta_j^{(t)})$$

of t components associated with the j th trial in Z_1, Z_2, \dots, Z_n such that $\{\eta_j^{(i)} = m\}$ if and only if $\{Z_j = i, Z_{j-1} = i, \dots, Z_{j-m+1} = i, Z_{j-m} \neq i\}$. For the sequence given by (1), we have $\eta_1^\top = (0, 1, 0)$, $\eta_2^\top = (0, 2, 0)$, $\eta_3^\top = (1, 0, 0)$, $\eta_4^\top = (0, 0, 1)$, $\eta_5^\top = (0, 0, 2)$, $\eta_6^\top = (0, 0, 3)$, etc. It is clear that exactly one of the components of η_j^\top is nonzero, i.e. $P\{\eta_j^\top = \mathbf{0}\} = 0$, where $\mathbf{0}$ is a $1 \times t$ vector whose entries are all 0. If Z_1, Z_2, \dots, Z_n is a sequence of i.i.d. multistate trials with $P\{Z_j = i\} = p_i, i = 1, 2, \dots, t$, then we have

$$\begin{aligned} P\{\eta_j^\top = \mathbf{I}_{i,m}\} &= P\{\eta_j^{(1)} = 0, \dots, \eta_j^{(i-1)} = 0, \eta_j^{(i)} = m, \eta_j^{(i+1)} = 0, \dots, \eta_j^{(t)} = 0\} \\ &= P\{Z_j = i, Z_{j-1} = i, \dots, Z_{j-m+1} = i, Z_{j-m} \neq i\} \\ &= \begin{cases} p_i^m (1 - p_i) & \text{if } m < j \leq n, \\ p_i^m & \text{if } m = j, \end{cases} \end{aligned}$$

where $I_{i,m} = (0, \dots, m, \dots, 0)$ denotes the vector whose i th entry is m (greater than 1) and the others are 0.

Run statistics can also be defined in terms of the vectors $\eta_j^\top, j \geq 1$. For example, the longest run of type i can be defined as $L_n^{(i)} = \max_{1 \leq j \leq n} \eta_j^{(i)}$. Similarly, the longest run of any type in Z_1, Z_2, \dots, Z_n can be represented as

$$L_n = \max_{1 \leq i \leq t} \max_{1 \leq j \leq n} \eta_j^{(i)}. \tag{11}$$

The total number of runs of type i with length exactly equal to k_i and the total number of runs of type i with length at least k_i can be respectively represented as

$$E_{n,k_i}^{(i)} = \sum_{j=k_i}^n \mathbf{1}\{\eta_j^{(i)} = k_i, \eta_{j+1}^{(i)} = 0\} \quad \text{and} \quad G_{n,k_i}^{(i)} = \sum_{j=k_i}^n \mathbf{1}\{\eta_j^{(i)} \geq k_i, \eta_{j+1}^{(i)} = 0\}. \tag{12}$$

Thus, the statistics $E_{n,k_i}^{(i)}$ and $G_{n,k_i}^{(i)}$ are represented as the sum of dependent Bernoulli variables with success probabilities $P\{\eta_j^{(i)} = k_i, \eta_{j+1}^{(i)} = 0\}$ and $P\{\eta_j^{(i)} \geq k_i, \eta_{j+1}^{(i)} = 0\}$, respectively, and we have

$$P\{\eta_j^{(i)} = k_i, \eta_{j+1}^{(i)} = 0\} = \begin{cases} p_i^{k_i} (1 - p_i)^2 & \text{if } k_i < j < n, \\ p_i^{k_i} (1 - p_i) & \text{if } j = k_i \text{ or } j = n. \end{cases}$$

In the following we compute the expectations of $E_{n,k_i}^{(i)}$ and $G_{n,k_i}^{(i)}$ using the representations given by (12). Propositions 1 and 2, below, hold for $s > 0$.

Proposition 1. *Let $\mu_{E_{n,k_i}^{(i)}}, i = 1, 2, \dots, t$, denote the mean of the total number of runs of type i with length exactly equal to k_i . Then*

$$\mu_{E_{n,k_i}^{(i)}} = \begin{cases} \frac{2B(\alpha_i + k_i, (\sum_{j \neq i} \alpha_j) + 1) + (n - k_i - 1)B(\alpha_i + k_i, (\sum_{j \neq i} \alpha_j) + 2)}{B(\alpha_i, \sum_{j \neq i} \alpha_j)}, & k_i = 1, \dots, n - 1, \\ \frac{B(\alpha_i + n, \sum_{j \neq i} \alpha_j)}{B(\alpha_i, \sum_{j \neq i} \alpha_j)}, & k_i = n, \end{cases}$$

where $\alpha_i = m_i/s$.

Proof. We can write

$$\mu_{E_{n,k_i}^{(i)}} = \int_0^1 \mu_{E_{n,k_i}^{(i)} | p_i} dG(p_i), \tag{13}$$

where $\mu_{E_{n,k_i}^{(i)} | p_i}$ denotes the mean of $E_{n,k_i}^{(i)}$ in a sequence of n i.i.d. multistate trials with $P\{Z_i = j\} = p_j$. Since (p_1, p_2, \dots, p_t) has a Dirichlet density for the considered multicolor urn model, p_i has a beta distribution with parameters α_i and $\sum_{j \neq i} \alpha_j$. Thus, using

$$\begin{aligned} \mu_{E_{n,k_i}^{(i)} | p_i} &= P\{\eta_{k_i}^{(i)} = k_i, \eta_{k_i+1}^{(i)} = 0\} + \sum_{j=k_i+1}^{n-1} P\{\eta_j^{(i)} = k_i, \eta_{j+1}^{(i)} = 0\} + P\{\eta_n^{(i)} = k_i\} \\ &= \begin{cases} p_i^{k_i} (1 - p_i)[2 + (n - k_i - 1)(1 - p_i)], & k_i = 1, \dots, n - 1, \\ p_i^n, & k_i = n, \end{cases} \end{aligned}$$

together with (5) in (13), the proof is completed.

Proposition 2. Let $\mu_{G_{n,k_i}^{(i)}}$, $i = 1, 2, \dots, t$, denote the mean of the random variable $G_{n,k_i}^{(i)}$. Then

$$\mu_{G_{n,k_i}^{(i)}} = \frac{B(\alpha_i + k_i, \sum_{j \neq i} \alpha_j) + (n - k_i)B(\alpha_i + k_i, \sum_{j \neq i} \alpha_j + 1)}{B(\alpha_i, \sum_{j \neq i} \alpha_j)}, \quad k_i \leq n.$$

Proof. The proof immediately follows using similar arguments as in Proposition 1 and

$$\mu_{G_{n,k_i}^{(i)} | p_i} = p_i^{k_i} (1 + (n - k_i)(1 - p_i))$$

(see, e.g. Balakrishnan and Koutras (2002, p. 170)).

As is seen from Propositions 1 and 2, we first find the mean of the corresponding run statistic for i.i.d. multistate trials and then, conditioning it on p_j , we obtain the means for the multistate (exchangeable) trials defined on the urn model. Below we obtain some bounds for the distribution of the longest run random variable using the same idea. Using (11), we can write

$$P\{L_n < k\} = P\{\eta_k < \mathbf{k}, \eta_{k+1} < \mathbf{k}, \dots, \eta_n < \mathbf{k}\}, \tag{14}$$

where $\mathbf{k} = (k, \dots, k)$ and, for the vectors $\mathbf{x}^\top = (x_1, \dots, x_t)$ and $\mathbf{y}^\top = (y_1, \dots, y_t)$, we write $\mathbf{x} < \mathbf{y}$ if $x_i < y_i$ for $i = 1, \dots, t$. The representation given by (14) enables us to obtain bounds for the distribution of L_n using Bonferroni- and product-type inequalities. Using the first-order Bonferroni-type inequality, we have

$$P\{L_n < k\} = P\left\{\bigcap_{j=k}^n E_j\right\} \geq 1 - \sum_{j=k}^n P\{E_j^c\}, \tag{15}$$

where $E_j \equiv \bigcap_{i=1}^t A_j^{k,i}$ and $A_j^{k,i} \equiv \{\eta_j^{(i)} < k\}$.

Proposition 3. Let Z_1, Z_2, \dots, Z_n be a sequence of i.i.d. multistate trials with $P\{Z_i = j\} = p_j$, $j = 1, 2, \dots, t$. Then

$$P\{L_n < k\} \geq 1 - (n - k + 1) \left(1 - \sum_{i=1}^t (p_i - p_i^k)\right).$$

Proof. It is obvious that

$$\begin{aligned} P\{E_j^c\} &= 1 - P\left\{\bigcap_{i=1}^t A_j^{k,i}\right\} \\ &= 1 - P\{\eta_j^{(1)} < k, \eta_j^{(2)} < k, \dots, \eta_j^{(t)} < k\} \\ &= 1 - \sum_{i=1}^t \sum_{m=1}^{k-1} P\{\eta_j^{(i)} = m, \eta_j^{(s)} = 0; s \neq i\} \\ &= 1 - \sum_{i=1}^t (p_i - p_i^k). \end{aligned}$$

The result now follows using $P\{E_j^c\}$ in (15).

Proposition 4. *Let Z_1, Z_2, \dots, Z_n be a sequence of i.i.d. multistate trials with $P\{Z_i = j\} = p_j, j = 1, 2, \dots, t$. Then*

$$P\{L_n < k\} \leq \left(\sum_{i=1}^t (p_i - p_i^k) \right)^{\lceil n/k \rceil}.$$

Proof. We have

$$P\{L_n < k\} = P\left\{ \bigcap_{j=k}^n E_j \right\} \leq P\{E_k \cap E_{2k} \cap \dots \cap E_{(r+1)k}\}, \tag{16}$$

where $r = \lceil (n - k)/k \rceil$. Since the events on the right-hand side of (16) are independent, we have

$$P\{L_n < k\} \leq (P\{E_k\})^{\lceil n/k \rceil}.$$

Thus, the proof is completed.

We obtain the following corollary as a consequence of Propositions 3 and 4.

Corollary 5. *Let Z_1, Z_2, \dots, Z_n be a sequence of i.i.d. binary trials with*

$$P\{Z_i = 1\} = p = 1 - P\{Z_i = 2\}.$$

Then

$$1 - (n - k + 1)(p^k + q^k) \leq P\{L_n < k\} \leq (1 - p^k - q^k)^{\lceil n/k \rceil},$$

where $q = 1 - p$.

Using the foregoing results, we can also obtain similar bounds for exchangeable multistate trials and, hence, for the longest run arising in the multicolor urn model that we have discussed. In this case we have

$$P\{E_j^c\} = 1 - \sum_{i=1}^t \int_0^1 (p_i - p_i^k) dG(p_i). \tag{17}$$

For the multicolor urn model, using (5) in (17), we obtain

$$P\{E_j^c\} = 1 - \sum_{i=1}^t \frac{B(\alpha_i + 1, \sum_{j \neq i} \alpha_j) - B(\alpha_i + k, \sum_{j \neq i} \alpha_j)}{B(\alpha_i, \sum_{j \neq i} \alpha_j)}.$$

Thus, we obtain

$$P\{L_n < k\} \geq 1 - (n - k + 1) \left(1 - \sum_{i=1}^t \frac{B(\alpha_i + 1, \sum_{j \neq i} \alpha_j) - B(\alpha_i + k, \sum_{j \neq i} \alpha_j)}{B(\alpha_i, \sum_{j \neq i} \alpha_j)} \right).$$

An upper bound can also be obtained similarly.

For an illustration, we provide some numerical results for the longest run distribution defined on the two-color urn model. Table 1 contains exact values, lower bounds (LBs), and upper bounds (UBs) for $P\{L_n < k\}$ defined on the urn model with various initial compositions m_1, m_2 , and $s = 1$. Since the probabilities in Table 1 depend on s, m_1 , and m_2 through the values of $\alpha_1 = m_1/s$ and $\alpha_2 = m_2/s$, we obtain the same numerical results for different selections of

TABLE 1: Exact values and bounds for the longest run distribution.

m_1	m_2	n	k	Exact	LB	UB
1	1	50	10	0.4708	–	0.6028
1	1	50	20	0.7889	–	0.8583
1	1	50	30	0.8952	–	0.9355
1	1	50	40	0.9396	0.4634	0.9512
1	1	100	30	0.8284	–	0.8828
1	1	100	40	0.8889	–	0.9271
1	1	100	50	0.9231	–	0.9414
3	1	50	10	0.3921	–	0.5203
3	1	50	20	0.7229	–	0.8073
3	3	50	20	0.9625	0.7304	0.9840
4	4	50	20	0.9794	0.8764	0.9924
4	4	100	20	0.9577	0.6769	0.9832
4	4	100	30	0.9916	0.9247	0.9971

$s, m_1,$ and m_2 when $s = m_1 = m_2$. Although the performance of the upper bound is generally good, the first-order Bonferroni-type inequality provides poor results, especially when $n - k$ is large. Undoubtedly, consideration of higher-order Bonferroni-type inequalities provides better results. These kinds of bound which have simple forms are preferred for approximating the reliability of consecutive-type systems.

5. Applications

In a Bayesian perspective, the probabilities (parameters) associated with the events are assumed to be random variables. These kinds of consideration are widely available in reliability studies (see, e.g. Singpurwalla (2006)). If, for example, p denotes the functioning probability of a component then it can be modeled as a random variable for the following reasons. Lau (1992) noted that ‘the randomness of p can arise from many different situations. For example, a complex system is operated in a random environment and the magnitude of p changes as the environment changes.’ The randomness of p can also arise from the production line of the components. Thus, it will be convenient to consider p as a random variable if one or more of these situations occur.

Replacing $p_n(n_1, \dots, n_t)$ by the probability given in (3), we obtain the distribution of runs whenever the set of probabilities $\mathbf{p} = (p_1, p_2, \dots, p_t)$ is a random vector having a measure $dG(\mathbf{p})$. This consideration might be useful for some reliability problems associated with the distribution of runs. In the following we provide two examples.

5.1. Consecutive- k, r -out-of- n : DFM systems

A consecutive- k, r -out-of- n : DFM system introduced by Koutras (1997) is a system which has two different kinds of failure and fails if and only if at least k consecutive components are failed-open or at least r consecutive components are failed-short. Let Z_i denote the state of the i th component, $i = 1, 2, \dots, n$. Suppose that states 1, 2, and 3 represent failed-open, failed-short, and working states, respectively, and that $P\{Z_i = j\} = p_j, j = 1, 2, 3$. If $\mathbf{p} = (p_1, p_2, p_3)$ is considered to be a random variable then the reliability of a consecutive- $k,$

TABLE 2: Reliability of a consecutive- k , r -out-of- n : DFM system.

k	r	a	b	c	n	R_n
2	2	0.5	0.5	3	2	0.9250
					5	0.8038
					10	0.6782
		0.5	0.5	4	2	0.9500
					5	0.8591
					10	0.7548
		1.0	0.5	4	2	0.9231
					5	0.7901
					10	0.6463
2	3	0.5	0.5	3	2	0.9625
					5	0.8695
					10	0.7709
		0.5	0.5	4	2	0.9750
					5	0.9099
					10	0.8337
		1.0	0.5	4	2	0.9441
					5	0.8335
					10	0.7106

r -out-of- n : DFM system can be calculated using

$$\begin{aligned}
 R_n &= P\{L_n^{(1)} < k, L_n^{(2)} < r\} \\
 &= P\{L_n^{(1)} < k, L_n^{(2)} < r, L_n^{(3)} < n + 1\} \\
 &= \sum_r \sum_n C(r_1, r_1, k - 1, n_1)C(r_2, r_2, r - 1, n_2)C(r_3, r_3, n, n_3) \\
 &\quad \times F_3(r_1, r_2, r_3) \int \int_{p_1+p_2 \leq 1} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n-n_1-n_2} dG(p_1, p_2),
 \end{aligned}$$

where first and second sums are taken respectively over $r_1 + r_2 + r_3 \leq n$ and $n_1 + n_2 + n_3 = n$.

In Table 2 we present some numerical values for R_n , assuming that

$$dG(p_1, p_2) = \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b)\Gamma(c)} p_1^{a-1} p_2^{b-1} (1 - p_1 - p_2)^{c-1} dp_1 dp_2$$

for $a, b, c > 0$ and $p_1 + p_2 < 1$.

5.2. Start-up demonstration tests

A start-up demonstration test is typically used in practice to elicit information about the reliability of equipment. Items such as gas lawn mowers, water pumps, car batteries, and outboard motors may be placed on this type of test which basically consists of attempting to start (start-up) the unit several times and observing the outcomes, a success meaning the unit turned on and a failure meaning that it did not. A decision regarding the reliability is made based on these outcomes and certain statistical rules. According to the consecutive successes and consecutive failures (CSCF) rule, a unit is accepted if k consecutive successful start-ups occur before d consecutive failures. See, e.g. Smith and Griffith (2008) for CSCF and other

TABLE 3: Probability of acceptance of a particular equipment.

k	d	a	b	$P_{k,d}^u$
5	3	1	1.00	0.3890
			0.75	0.4875
			0.50	0.6149
			0.25	0.7811
			0.10	0.9050
5	2	1	1.00	0.3002
			0.75	0.3969
			0.50	0.5318
			0.25	0.7230
			0.10	0.8764

types of start-up demonstration test. Let Z_1, Z_2, \dots denote the outcomes of successive start-up tests applied to a device, where Z_i is a binary random variable taking values 1 (successful start-up) and 2 (failure). Let p denote the probability of the event that $Z_i = 1$, i.e. the i th start-up test is a successful start-up. Characteristics of a CSCF start-up demonstration test can be computed using the formulae presented in the previous section. The test length $W_{k,d}$, which is the total number of attempts (or start-ups) until termination of the experiment, is defined by $W_{k,d} = \min(W_k^{(1)}, W_d^{(2)})$. Using Corollary 2, we have

$$\begin{aligned}
 &P\{W_{k,d} > n\} \\
 &= \sum_{r_1=1}^{[(n+1)/2]} \sum_{r_2=1}^{n-r_1} \sum_{n_1=r_1}^{n-r_2} C(r_1, r_1, k-1, n_1) C(r_2, r_2, d-1, n-n_1) \\
 &\quad \times \binom{2}{r_1-r_2+1} \int_0^1 p^{n_1} (1-p)^{n-n_1} dG(p).
 \end{aligned}$$

One of the most important quantities in start-up demonstration tests is the probability of accepting the equipment, which is defined as $P_{k,d} = P\{W_k^{(1)} < W_d^{(2)}\}$ for a CSCF test. In Table 3 we evaluate $P_{k,d}$, assuming beta distribution for p , i.e.

$$dG(p) = \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} dp, \quad 0 < p < 1, a, b > 0.$$

The beta distribution has been found to be useful because its domain is from 0 to 1, and since we can obtain distributions of very different shapes by the appropriate choices of parameters a and b . An upper bound for the probability $P_{k,d}$ can be obtained using Corollary 2 since

$$\begin{aligned}
 P_{k,d} &= P\{W_k^{(1)} < W_d^{(2)}\} \\
 &= \sum_{n \geq k} P\{W_k^{(1)} < W_d^{(2)}, W_k^{(1)} = n\} \\
 &\leq \sum_{n \geq k} (P\{L_n^{(1)} < k+1, L_n^{(2)} < d\} - P\{L_n^{(1)} < k, L_n^{(2)} < d\}) \\
 &= P_{k,d}^u.
 \end{aligned}$$

Acknowledgement

The author would like to express gratitude to the anonymous referee for a thorough review and valuable comments that led to some improvements in this paper.

References

- ANTZOULAKOS, D. L. AND BOUTSIKAS, M. V. (2007). A direct method to obtain the joint distribution of successes, failures and patterns in enumeration problems. *Statist. Prob. Lett.* **77**, 32–39.
- BALAKRISHNAN, N. AND KOUTRAS, M. V. (2002). *Runs and Scans with Applications*. John Wiley, New York.
- CHAKRABORTI, S. AND ERYILMAZ, S. (2007). A nonparametric Shewhart-type signed-rank control chart based on runs. *Commun. Statist. Simul. Comput.* **36**, 335–356.
- CHARALAMBIDES, C. A. (2002). *Enumerative Combinatorics*. Chapman and Hall/CRC, Boca Raton, FL.
- CHANG, C. J., FANN, C. S. J., CHOU, W. C. AND LIAN, I. B. (2003). On the tail probability of the longest well-matching run. *Statist. Prob. Lett.* **63**, 267–274.
- CHAO, M. T., FU, J. C. AND KOUTRAS, M. V. (1995). Survey of reliability studies of consecutive k -out-of- n : F and related systems. *IEEE Trans. Reliab.* **44**, 120–127.
- ERYILMAZ, S. (2008a). Consecutive k -out-of- n : G system in stress-strength setup. *Commun. Statist. Simul. Comput.* **37**, 579–589.
- ERYILMAZ, S. (2008b). Distribution of runs in a sequence of exchangeable multistate trials. *Statist. Prob. Lett.* **78**, 1505–1513.
- ERYILMAZ, S. AND DEMIR, S. (2007). Success runs in a sequence of exchangeable binary trials. *J. Statist. Planning Infer.* **137**, 2954–2963.
- FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edn. John Wiley, New York.
- FU, J. C. (1996). Distribution theory of runs and patterns associated with a sequence of multistate trials. *Statistica Sinica* **6**, 957–974.
- FU, J. C. AND LOU, W. Y. W. (2003). *Distribution Theory of Runs and Patterns and Its Applications*. World Scientific, River Edge, NJ.
- FU, J. C. AND LOU, W. Y. W. (2007). On the normal approximation for the distribution of the number of simple or compound patterns in a random sequence of multistate trials. *Methodology Comput. Appl. Prob.* **9**, 195–205.
- FU, J. C., SHMUELI, G. AND CHANG, Y. M. (2003). A unified Markov chain approach for computing the run length distribution in control charts with simple or compound rules. *Statist. Prob. Lett.* **65**, 457–466.
- GOLDSTEIN, L. (1990). Poisson approximations and DNA sequence matching. *Commun. Statist. Theory Meth.* **19**, 4167–4179.
- HAN, Q. AND AKI, S. (1999). Joint distributions of runs in a sequence of multistate trials. *Ann. Inst. Statist. Math.* **51**, 419–447.
- HILL, B. M., LANE, D. AND SUDDERTH, W. (1987). Exchangeable urn processes. *Ann. Prob.* **15**, 1586–1592.
- INOUE, K. AND AKI, S. (2005). Joint distributions of numbers of success runs of specified lengths in linear and circular sequences. *Ann. Inst. Statist. Math.* **57**, 353–368.
- JOHNSON, N. AND KOTZ, S. (1977). *Urn Models and Their Application*. John Wiley, New York.
- KOUTRAS, M. V. (1997). Consecutive- k , r -out-of- n : DFM systems. *Microelectron. Reliab.* **37**, 597–603.
- KOUTRAS, M. V. AND ALEXANDROU, V. A. (1997). Non-parametric tests based on success runs of fixed length. *Statist. Prob. Lett.* **32**, 393–404.
- KONG, Y. (2006). Distribution of runs and longest runs: a new generating function approach. *J. Amer. Statist. Assoc.* **101**, 1253–1263.
- LAU, T. S. (1992). The reliability of exchangeable binary systems. *Statist. Prob. Lett.* **13**, 153–158.
- LOU, W. Y. W. (1996). On runs and longest run tests: a method of finite Markov chain imbedding. *J. Amer. Statist. Assoc.* **91**, 1595–1601.
- LOU, W. Y. W. (2003). The exact distribution of the k -tuple statistic for sequence homology. *Statist. Prob. Lett.* **61**, 51–59.
- MAKRI, F. S. AND PHILIPPOU, A. N. (2005). On binomial and circular binomial distributions of order k for l -overlapping success runs of length k . *Statist. Papers* **46**, 411–432.
- MAKRI, F. S., PHILIPPOU, A. N. AND PSILLAKIS, Z. M. (2007a). Shortest and longest length of success runs in binary sequences. *J. Statist. Planning Infer.* **137**, 2226–2239.
- 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.
- MOSTELLER, F. (1941). Note on an application of runs to quality control charts. *Ann. Math. Statist.* **12**, 228–232.
- PHILIPPOU, A. N. (1988). Recursive theorems for success runs and reliability of consecutive- k -out-of- n : F systems. In *Applications of Fibonacci Numbers*, Kluwer, Dordrecht, pp. 149–161.

- PHILIPPOU, A. N. AND MAKRI, F. S. (1985). Longest success runs and Fibonacci-type polynomials. *Fibonacci Quart.* **23**, 338–346.
- PHILIPPOU, A. N. AND MAKRI, F. S. (1986). Successes, runs and longest runs. *Statist. Prob. Lett.* **4**, 101–105.
- SINGPURWALLA, N. D. (2006). *Reliability and Risk*. John Wiley, Chichester.
- SEN, K., AGARWAL, M. L. AND CHAKRABORTY, S. (2002). Lengths of runs and waiting time distributions by using Pólya–Eggenberger sampling scheme. *Studia Sci. Math. Hung.* **39**, 309–332.
- SEN, K., AGARWAL, M. AND BHATTACHARYA, S. (2003). On circular distributions of order k based on Pólya–Eggenberger sampling scheme. *J. Math. Sci.* **2**, 34–54.
- SMITH, M. L. AND GRIFFITH, S. W. (2008). The analysis and comparison of start-up demonstration tests. *Europ. J. Operat. Res.* **186**, 1029–1045.
- VAGGELATOU, E. (2003). On the length of the longest run in a multi-state Markov chain. *Statist. Prob. Lett.* **62**, 211–221.
- WALD, A. AND WOLFOWITZ, J. (1940). On a test whether two samples are from the same population. *Ann. Math. Statist.* **11**, 147–162.
- WANG, L. AND CHEUNG, K. (2004). Use of run statistics for pattern recognition in genomic DNA sequences. *J. Comput. Biol.* **11**, 107–124.
- WOLFOWITZ, J. (1943). On the theory of runs with some applications to quality control. *Ann. Math. Statist.* **14**, 280–288.
- ZABELL, S. L. (1982). Johnson’s “sufficientness” postulate. *Ann. Statist.* **10**, 1091–1099.