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), Vaggelatou (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))
Waiting time random variables associated with runs are indicative in statistical process control applications (Fu et al. (2003), Chakraborti and Eryılmaz (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 $j$, $j = 1, 2, \ldots, t$, and its color is noted. If a ball of color $j$ is drawn at a stage, $s$ balls of color $j$, $j = 1, 2, \ldots, t$, are added to the urn. Drawing a ball of color $j$ is considered as a trial of type $j$, $j = 1, 2, \ldots, t$. This scheme is repeated $n$ times and a sequence consisting of trials, namely $\{1, 2, \ldots, t\}$, 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 $j$, $j = 1, 2, \ldots, t$, $m_1 + \cdots + m_t = m$. 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 $j$, $j = 1, 2, \ldots, t$, 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 $j$, $i = 1, 2, \ldots$ and $j = 1, 2, \ldots, t$.

Let $Z_1, Z_2, \ldots, Z_n$ denote the first $n$ outcomes associated with this sampling scheme. It is obvious that the elements of $Z_1, Z_2, \ldots, Z_n$ are dependent. For $i = 1, 2, \ldots$ and $j = 1, 2, \ldots, t$, 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 $j$,
- $E^{(j)}_{n,k_j}$: the total number of runs of type $j$ with length exactly equal to $k_j$,
- $G^{(j)}_{n,k_j}$: the total number of runs of type $j$ with length at least $k_j$,
- $L^{(j)}_n$: the length of the longest run of type $j$,
- $L_n$: the length of the longest run of any type in $Z_1, Z_2, \ldots, Z_n$. 


Thus, the distribution of any run statistic based on a sequence of exchangeable multistate trials (see, e.g. Hill et al. (1987)). If \( Z_1, Z_2, \ldots \) is a sequence of multistate exchangeable trials with \( P[Z_j = j] = p_j, j = 1, 2, \ldots, t, i \geq 1, \sum_{j=1}^{t} p_j = 1 \), 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

\[
P[Z_1 = 1, \ldots, Z_{n_1} = 1, Z_{n_1+1} = 2, \ldots, Z_{n_1+n_2} = 2, \ldots, Z_{n-n_t+1} = t, \ldots, Z_n = t]
= \int \sum_{j=1}^{t} p_j^{n_j} dG(p_1, \ldots, p_{t-1})
\]

(see, e.g. Zabell (1982)), where \( dG(p_1, \ldots, p_{t-1}) \) is the de Finetti measure that exists on the simplex.
The sequence \((Z_1, Z_2, \ldots)\) derived in the aforementioned multicolor Pólya urn scheme is exchangeable with the de Finetti measure, which is Dirichlet and given by

\[
dG(p_1, \ldots, p_{t-1}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_t)}{\Gamma(\alpha_1)\Gamma(\alpha_2) \cdots \Gamma(\alpha_t)} p_1^{\alpha_1-1} p_2^{\alpha_2-1} \cdots p_{t-1}^{\alpha_{t-1}-1} \times (1 - p_1 - \cdots - p_{t-1})^{\alpha_t-1} \, dp_1 \cdots dp_{t-1},
\]

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

\[
dG(p_j) = \frac{1}{B(\alpha_j, \sum_{i \neq j} \alpha_i)} p_j^{\alpha_j-1} (1 - p_j)^{\sum_{i \neq j} \alpha_i-1} \, dp_j \quad \text{for } j = 1, 2, \ldots, t.
\]

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

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

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

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

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

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

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

\[
p_n(n_1, \ldots, n_t) = \frac{m_1^{(n_1)} m_2^{(n_2)} \cdots m_t^{(n_t)}}{m^{(\sum_{i=1}^t n_i)} (m - n_1)^{\alpha_2} \cdots (m - \sum_{i=1}^{t-1} n_i)^{\alpha_t}},
\]

where \(x^{(a)} = x(x-1) \cdots (x-a+1)\), 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.** (Eryılmaz (2008b).) Let \(Z_1, Z_2, \ldots\) be an infinite sequence of exchangeable multistate trials. Suppose that each trial has \(t\) \((t \geq 2)\) possible outcomes labeled \(1, 2, \ldots, t\) with \(P[Z_i = j] = p_j, \quad j = 1, 2, \ldots, t, \quad i \geq 1\). Then

\[
P[\theta_1^{(1)} = 1_1, \ldots, \theta_{r_1}^{(1)} = 1_{r_1}, \ldots, \theta_1^{(t)} = 1_1, \ldots, \theta_{r_t}^{(t)} = 1_{r_t}, \quad R_1^{(1)} = r_1, \ldots, \quad R_t^{(t)} = r_t] \equiv F_t(r_1, \ldots, r_t) \int \sum_{p_j=1}^t p_j^{n_j} \, dG(p_1, \ldots, p_{t-1}),
\]
Run statistics defined on the multicolor urn model

where \( n_j = \sum_{m=1}^{r_j} i_{m}^{(j)} \).

\[
F_t(r_1, \ldots, r_t) = (-1)^r \sum_{m_1=1}^{r_1} \cdots \sum_{m_t=1}^{r_t} (-1)^m \binom{r_1-1}{m_1-1} \cdots \binom{r_t-1}{m_t-1} \binom{m}{m_1, \ldots, m_t}.
\]

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

Note that the number \( F_t(r_1, \ldots, r_t) \) 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 \( \binom{2r_1 - r_2 + 1}{r_1} \).

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

\[
P_{\{\theta_1^{(1)} = i_1^{(1)}, \ldots, \theta_r^{(1)} = i_r^{(1)}, \ldots, \theta_1^{(t)} = i_1^{(t)}, \ldots, \theta_r^{(t)} = i_r^{(t)}, R_n^{(1)} = r_1, \ldots, R_n^{(t)} = r_t\}} = F_t(r_1, \ldots, r_t) p_n(n_1, \ldots, n_t).
\]

### 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 \( j \), \( j = 1, 2, \ldots, t \). The following theorem provides the joint distribution of the random variables \( X_{n,k_1}^{(1)}, X_{n,k_2}^{(2)}, \ldots, X_{n,k_t}^{(t)} \).

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

\[
P_{\{X_{n,k_1}^{(1)} \in B_1, X_{n,k_2}^{(2)} \in B_2, \ldots, X_{n,k_t}^{(t)} \in B_t\}} = \sum_{r} \sum_{n} \prod_{m=1}^{t} |I_{\psi_m}(B_m)| F_t(r_1, \ldots, r_t) p_n(n_1, \ldots, n_t),
\]

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

\[
I_{\psi_m}(B_m) = \{(i_1^{(m)}, \ldots, i_{r_m}^{(m)}): i_1^{(m)} + \cdots + i_{r_m}^{(m)} = n_m; \psi_m(i_1^{(m)}, \ldots, i_{r_m}^{(m)}) \in B_m\},
\]

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

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

\[
P_{\{X_{n,k_1}^{(1)} \in B_1, \ldots, X_{n,k_t}^{(t)} \in B_t\}} = \sum_{r_1} \cdots \sum_{r_t} P_{\{\theta_1^{(1)} = i_1^{(1)}, \ldots, \theta_{r_1}^{(1)} \in B_1, \ldots, \theta_1^{(t)} = i_1^{(t)}, \ldots, \theta_{r_t}^{(t)} \in B_t, R_n^{(1)} = r_1, \ldots, R_n^{(t)} = r_t\}}.
\]
Now, conditioning on the frequencies of each type of object, we obtain

\[
P\{X_{n,k}^{(1)} \in B_1, \ldots, X_{n,k}^{(t)} \in B_t\} = \sum_{r_1} \cdots \sum_{r_t} \sum_{n_1} \cdots \sum_{n_t} P\{\phi_{x_{1}^{(m)}}^{(1)}(\theta_{1}^{(1)}), \ldots, \phi_{x_{r_t}^{(m)}}^{(t)}(\theta_{r_t}^{(t)}) \in B_1, \ldots, \theta_{r_t}^{(t)} \} = \sum_{r} \sum_{n} \cdots \sum_{n_t}
\]

\[
R_n^{(1)} = r_1, \ldots, R_n^{(t)} = r_t, S_n^{(1)} = n_1, \ldots, S_n^{(t)} = n_t,
\]

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

\[
i_1^{(m)} + \cdots + i_{r_t}^{(m)} = n_m \text{ such that } \phi_{m}^{(m)}(i_1^{(m)}, \ldots, i_{r_t}^{(m)}) \in B_m, i_1^{(m)}, \ldots, i_{r_t}^{(m)} > 0.
\]

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 \(\phi_{m}^{(m)}\), \(m = 1, 2, \ldots, t\). The cardinalities \(|I_{\phi_{m}^{(m)}}(B_m)|\), \(m = 1, 2, \ldots, t\), completely depend on the selection of \(\phi_{m}^{(m)}\) (and, hence, \(X_{n,k}^{(m)}\)), 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 \(X_{n,k_1}^{(1)}, X_{n,k_2}^{(2)}, \ldots, X_{n,k}^{(t)}\) 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 \(E_{n,k}^{(1)} = E_{n,k}^{(2)} = \ldots = E_{n,k}^{(t)}\) is given by

\[
P\{E_{n,k}^{(1)} = x_1, E_{n,k}^{(2)} = x_2, \ldots, E_{n,k}^{(t)} = x_t\}
\]

\[
= \sum_{r} \sum_{n} \cdots \sum_{n_t} \prod_{m=1}^{t} \binom{r_m}{x_m} A(n_m - r_m - x_m(k_m - 1), r_m - x_m, k_m - 1)
\]

\[
\times F_1(r_1, \ldots, r_t) p_n(n_1, \ldots, n_t),
\]

where

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

(10)
Proof. For \( m = 1, 2, \ldots, t \), choosing \( B_m = \{s_m\} \) and

\[
\phi_{\text{i}(m)}^{(m)} (i_1^{(m)}, \ldots, i_{r(m)}^{(m)}) = \sum_{j=1}^{r_m} 1^{(m)} \equiv k_m \},
\]

we have

\[
I_{\phi_{\text{i}(m)}}(\{s_m\}) = \{ (i_1^{(m)}, \ldots, i_{r(m)}^{(m)}): i_1^{(m)} + \cdots + i_{r(m)}^{(m)} = n_m; x_m \text{ of } i_1^{(m)}, \ldots, i_{r(m)}^{(m)} \text{ equals } k_m
\]

and \( r_m = x_m \) of \( i_1^{(m)}, \ldots, i_{r(m)}^{(m)} \) differs from \( k_m \} \).

We readily have

\[
\left| I_{\phi_{\text{i}(m)}}^{(m)}(\{s_m\}) \right| = \left( \begin{array}{c} r_m \\ x_m \end{array} \right) A(n_m - r_m - x_m(k_m - 1), r_m - x_m, k_m - 1),
\]

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 \( i_1^{(m)} + \cdots + i_{r(m)}^{(m)} = n_m - k_m x_m \) such that \( i_1^{(m)} \neq k_m, \ldots, i_{r(m) - x_m} \neq k_m, i_{r(m)} > 0, \ldots, i_{r(m) - x_m} > 0 \), 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 \( A(n_m - r_m - x_1(k_m - 1), r_m - x_1, k_m - 1) \) with \( \alpha = n_m - r_m - x_1(k_m - 1) \), \( r = r_m - x_1, \) 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 \( P(E_{x, 2}^{(1)} = 1, E_{x, 1}^{(2)} = 2, E_{x, 1}^{(3)} = 1) \). 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

\[
P(E_{x, 2}^{(1)} = 1, E_{x, 1}^{(2)} = 2, E_{x, 1}^{(3)} = 1) = \sum \sum \sum \sum \left( \begin{array}{c} r_1 \\ 1 \end{array} \right) \left( \begin{array}{c} r_2 \\ 2 \end{array} \right) \left( \begin{array}{c} r_3 \\ 1 \end{array} \right) A(n_1 - r_1 - 1, r_1 - 1, 1)
\]

\[
\times A(n_2 - r_2, r_2 - 2, 0) A(n - n_1 - n_2 - r_3, r_3 - 1, 0)
\]

\[
\times F_3(r_1, r_2, r_3, p_5(n_1, n_2, n - n_1 - n_2)
\]

\[
= F_3(1, 2, 1) p_5(2, 2, 1)
\]

\[
= 6 p_5(2, 2, 1),
\]

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

\[
P(E_{n, k_1}^{(1)} = x_1, E_{n, k_2}^{(2)} = x_2)
\]

\[
= \sum \sum \sum \left( \begin{array}{c} r_1 \\ n_1 \end{array} \right) \left( \begin{array}{c} r_2 \\ x_2 \end{array} \right) A(n_1 - r_1 - x_1(k_1 - 1), r_1 - x_1, k_1 - 1)
\]

\[
\times A(n - n_1 - r_2 - x_2(k_2 - 1), r_2 - x_2, k_2 - 1)
\]

\[
\times \left( \begin{array}{c} 2 \\ r_1 \end{array} + 1 \right) p_5(n_1, n - n_1).}
Theorem 3. The joint probability mass function of \( G_{n,k_1}^{(1)}, G_{n,k_2}^{(2)}, \ldots, G_{n,k_t}^{(t)} \) is given by
\[
P\{G_{n,k_1}^{(1)} = x_1, G_{n,k_2}^{(2)} = x_2, \ldots, G_{n,k_t}^{(t)} = x_t \} = \sum \sum \prod_{m=1}^{t} \binom{r_m}{x_m} C(r_m, r_m - x_m, k_m - 1, n_m) \times F_t(r_1, \ldots, r_t) p_n(n_1, \ldots, n_t),
\]
where
\[
C(r, j, k, l) = \sum_{z=0}^{j} (-1)^z \binom{j}{z} \left( r + l - j - (k + 1)(r - j) - zk - 1 \right).
\]

Proof. For \( m = 1, 2, \ldots, t \), choosing \( B_m = \{x_m\} \) and
\[
\phi_{x_m}^{(m)}(i_1^{(m)}, \ldots, i_r^{(m)}) = \sum_{j=1}^{r_m} 1[i_j^{(m)} \geq k_m],
\]
we have
\[
I_{\phi_{x_m}}^{(m)}(\{x_m\}) = \{i_1^{(m)}, \ldots, i_r^{(m)}: \sum_{i=1}^{r} i_i^{(m)} = n_m; x_m of i_1^{(m)}, \ldots, i_r^{(m)} \geq k_m
\]
and \( r_m - x_m of i_1^{(m)}, \ldots, i_r^{(m)} < k_m \} \].

It is clear that
\[
|I_{\phi_{x_m}}^{(m)}(\{x_m\})| = \binom{r_m}{x_m} C(r_m, r_m - x_m, k_m - 1, n_m),
\]
where \( C(r, j, k, l) \) is the total number of integer solutions to the system \( i_1 + \cdots + i_r = l \) such that \( i_1 < k+1, \ldots, i_j < k+1, i_j+1 \geq k+1, \ldots, i_r \geq k+1; i_1 > 0, \ldots, i_r > 0 \), or, equivalently, the total number of integer solutions to the system \( y_1 + \cdots + y_r = l - j - (k + 1)(r - j) \) such that \( 0 \leq y_1 < k, \ldots, 0 \leq y_j < k, y_{j+1} \geq 0, \ldots, y_r \geq 0 \). Let \( A_i \) be the event of \( y_i \geq k, i = 1, \ldots, j \). Then
\[
C(r, j, k, l) = \bigcap_{i=1}^{j} A_i^c = s(r, j, k, l) - \bigcup_{i=1}^{j} A_i,
\]
where \( s(r, j, k, l) \) denotes the total number of integer solutions to \( y_1 + \cdots + y_r = l - j - (k + 1)(r - j) \) such that \( y_i \geq 0, i = 1, \ldots, r \). 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 \( \{x_m\} = \{0\}, m = 1, 2, \ldots, t \), in Theorem 3 we obtain
\[
P\{L_n^{(1)} < k_1, L_n^{(2)} < k_2, \ldots, L_n^{(t)} < k_t \} = \sum \sum \prod_{m=1}^{t} C(r_m, r_m - x_m, k_m - 1, n_m) F_t(r_1, \ldots, r_t) p_n(n_1, \ldots, n_t),
\]
which coincides with Corollary 4 of Eryilmaz (2008b).
Thus, we obtain the following corollary.

**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

$$P[W_k > n] = P[W_k^{(1)} > n, W_k^{(2)} > n, \ldots, W_k^{(r)} > n] = P[L_n^{(1)} < k, L_n^{(2)} < k, \ldots, L_n^{(r)} < k] = P[L_n < k].$$

Thus, we obtain the following corollary.

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

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

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

$$P[R_n^{(1)} = r_1, R_n^{(2)} = r_2, \ldots, R_n^{(t)} = r_t] = \sum_n \prod_{m=1}^{t} \binom{n_m - 1}{r_m - 1} F_i(r_1, \ldots, r_t) p_n(n_1, \ldots, n_t),$$

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 $Z_1, Z_2, \ldots, Z_n$ that consists of multistate trials with possible outcomes $\{1, 2, \ldots, t\}$, let $\eta_j^{(i)}$ denote the length of the $i$th run at the $j$th stage, $i = 1, 2, \ldots, t$ and $j = 1, 2, \ldots, n$. That is, we define a vector

$$\eta_j^T = (\eta_j^{(1)}, \eta_j^{(2)}, \ldots, \eta_j^{(t)})$$

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

$$P[\eta_j^T = I_{i,m}] = P[\eta_j^{(1)} = 0, \ldots, \eta_j^{(i-1)} = 0, \eta_j^{(i)} = m, \eta_j^{(i+1)} = 0, \ldots, \eta_j^{(t)} = 0] = P[Z_j = i, Z_{j-1} = i, \ldots, Z_{j-m+1} = i, Z_{j-m} \neq i] = \begin{cases} p_i^m(1 - p_i) & \text{if } m < j \leq n, \\ p_i^m & \text{if } m = j. \end{cases}$$
Thus, the statistics $E(i)$, $i = 1, 2, \ldots, t$, denote 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 $\eta(j) T$, $j \geq 1$. For example, the longest run of type $i$ can be defined as $L_n(i) = \max_{1 \leq j \leq n} \eta(j)$. Similarly, the longest run of any type in $Z_1, Z_2, \ldots, Z_n$ can be represented as

$$L_n = \max_{1 \leq j \leq n} \eta(i) .$$

(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

$$E(i, k_i) = \sum_{j=k_i}^{n} I[\eta(j) = k_i, \eta(j+1) = 0] \quad \text{and} \quad G(i, k_i) = \sum_{j=k_i}^{n} I[\eta(j) \geq k_i, \eta(j+1) = 0].$$

(12)

Thus, the statistics $E(i, k_i)$ and $G(i, k_i)$ are represented as the sum of dependent Bernoulli variables with success probabilities $P[\eta(j) = k_i, \eta(j+1) = 0]$ and $P[\eta(j) \geq k_i, \eta(j+1) = 0]$, respectively, and we have

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

In the following we compute the expectations of $E(n, k_i)$ and $G(n, k_i)$ using the representations given by (12). Propositions 1 and 2, below, hold for $s > 0$.

**Proposition 1.** Let $\mu(n, k_i)$, $i = 1, 2, \ldots, t$, denote the mean of the total number of runs of type $i$ with length exactly equal to $k_i$. Then

$$\mu(n, k_i) = \begin{cases} 2B(\alpha_i + k_i, (\sum_{j \neq i} \alpha_j) + 1) + (n - k_i - 1)B(\alpha_i + k_i, (\sum_{j \neq i} \alpha_j) + 2), & k_i = 1, \ldots, n - 1, \\ B(\alpha_i + n, (\sum_{j \neq i} \alpha_j) + 1), & k_i = n, \end{cases}$$

where $\alpha_i = m_i/s$.

**Proof.** We can write

$$\mu(n, k_i) = \int_{0}^{1} \mu(n, k_i) | p_i \text{ d}G(p_i),$$

(13)

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

$$\mu(n, k_i) | p_i = P[\eta(j) = k_i, \eta(j+1) = 0] + \sum_{j=k_i+1}^{n} P[\eta(j) = k_i, \eta(j+1) = 0] + P[\eta(j) = k_i]$$

$$= \begin{cases} p(k_i) (1 - p_i)[2 + (n - k_i - 1)(1 - p_i)], & k_i = 1, \ldots, n - 1, \\ p(k_i), & k_i = n, \end{cases}$$

together with (5) in (13), the proof is completed.
Proposition 2. Let $\mu_{G_{n,k_i}}^{(i)}$, $i = 1, 2, \ldots, t$, 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

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

(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

$$P\{L_n < k\} = P\{\eta_1 < k, \eta_{k+1} < k, \ldots, \eta_n < k\},$$

where $k = (k, \ldots, k)$ and, for the vectors $x^T = (x_1, \ldots, x_i)$ and $y^T = (y_1, \ldots, y_r)$, we write $x < y$ if $x_i < y_i$ for $i = 1, \ldots, t$. 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

$$P\{L_n < k\} \geq 1 - \sum_{j=k}^{n} P\{E_j^c\},$$

where $E_j = \bigcap_{i=1}^{t} A_{j}^{k,i}$ and $A_{j}^{k,i} = \{\eta_i^{(i)} < k\}$.

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

$$P\{E_j^c\} = 1 - P\left\{\bigcap_{i=1}^{t} A_{j}^{k,i}\right\}$$

$$= 1 - P\{\eta_1^{(1)} < k, \eta_2^{(2)} < k, \ldots, \eta_t^{(t)} < k\}$$

$$= 1 - \sum_{i=1}^{t} \sum_{m=1}^{k-1} P\{\eta_i^{(i)} = m, \eta_j^{(s)} = 0; s \neq i\}$$

$$= 1 - \sum_{i=1}^{t} (p_i - p_i^k).$$

The result now follows using $P\{E_j^c\}$ in (15).
Proposition 4. Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of i.i.d. multistate trials with $P(Z_i = j) = p_j, j = 1, 2, \ldots, t$. Then

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

Proof. We have

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

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

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

Thus, the proof is completed.

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

Corollary 5. Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of i.i.d. binary trials with $P(Z_i = 1) = p = 1 - P(Z_i = 2)$. Then

$$1 - (n - k + 1)(p^k + q^k) \leq P[L_n < k] \leq (1 - p^k - q^k)^{\lfloor n/k \rfloor},$$

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

$$P(E_j^c) = 1 - \sum_{i=1}^{t} \int_{0}^{1} (p_i - p_i^k) dG(p_i).$$

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
s, m₁, and m₂ when s = m₁ = m₂. 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 \( p_n(n_1, \ldots, n_t) \) by the probability given in (3), we obtain the distribution of runs whenever the set of probabilities \( p = (p_1, p_2, \ldots, p_t) \) is a random vector having a measure \( dG(p) \). 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, \( i = 1, 2, \ldots, n \). 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 \( p = (p_1, p_2, p_3) \) is considered to be a random variable then the reliability of a consecutive-k,

---

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

<table>
<thead>
<tr>
<th>m₁</th>
<th>m₂</th>
<th>n</th>
<th>k</th>
<th>Exact</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>50</td>
<td>10</td>
<td>0.4708</td>
<td>–</td>
<td>0.6028</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>50</td>
<td>20</td>
<td>0.7889</td>
<td>–</td>
<td>0.8583</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>50</td>
<td>30</td>
<td>0.8952</td>
<td>–</td>
<td>0.9355</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>50</td>
<td>40</td>
<td>0.9396</td>
<td>0.4634</td>
<td>0.9512</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>100</td>
<td>30</td>
<td>0.8284</td>
<td>–</td>
<td>0.8828</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>100</td>
<td>40</td>
<td>0.8889</td>
<td>–</td>
<td>0.9271</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>100</td>
<td>50</td>
<td>0.9231</td>
<td>–</td>
<td>0.9414</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>50</td>
<td>10</td>
<td>0.3921</td>
<td>–</td>
<td>0.5203</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>50</td>
<td>20</td>
<td>0.7229</td>
<td>–</td>
<td>0.8073</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>50</td>
<td>20</td>
<td>0.9625</td>
<td>0.7304</td>
<td>0.9840</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>50</td>
<td>20</td>
<td>0.9794</td>
<td>0.8764</td>
<td>0.9924</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>100</td>
<td>20</td>
<td>0.9577</td>
<td>0.6769</td>
<td>0.9832</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>100</td>
<td>30</td>
<td>0.9916</td>
<td>0.9247</td>
<td>0.9971</td>
</tr>
</tbody>
</table>
Table 2: Reliability of a consecutive-\(k, r\)-out-of-\(n\): DFM system.

<table>
<thead>
<tr>
<th>(k)</th>
<th>(r)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(n)</th>
<th>(R_n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>2</td>
<td>0.9250</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.8038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.6782</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>4</td>
<td>2</td>
<td>0.9500</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.8591</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.7548</td>
</tr>
<tr>
<td>1.0</td>
<td>0.5</td>
<td>4</td>
<td>2</td>
<td>0.9231</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.7901</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.6463</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>2</td>
<td>0.9625</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.8695</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.7709</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>4</td>
<td>2</td>
<td>0.9750</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.9099</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.8337</td>
</tr>
<tr>
<td>1.0</td>
<td>0.5</td>
<td>4</td>
<td>2</td>
<td>0.9441</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.8335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.7106</td>
</tr>
</tbody>
</table>

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

\[
R_n = P\{L_1^{(1)} < k, L_2^{(2)} < r\} = P\{L_1^{(1)} < k, L_2^{(2)} < r, L_3^{(3)} < n + 1\}
\]

\[
= \sum_{r_1} \sum_{r_2} C(r_1, r_1, k-1, n_1) C(r_2, r_2, r-1, n_2) C(r_3, r_3, n, n_3)
\times F_3(r_1, r_2, r_3) \int_{p_1+p_2 \leq 1} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n-n_1-n_2} dG(p_1, p_2),
\]

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 \(p_1 + p_2 < 1\).

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

\[
dG(p_1, p_2) = \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b)\Gamma(c)} p_1^{a-1} p_2^{b-1} (1 - p_1 - p_2)^{c-1} d p_1 d p_2
\]

for \(a, b, c > 0\) and \(p_1 + p_2 < 1\).
types of start-up demonstration test. Let \( Z_1, Z_2, \ldots \) 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

\[
W_{k,d} = \min(W(1)_k, W(2)_d).
\]

Using Corollary 2, we have

\[
P\{W_{k,d} > n\} = \left(\frac{n+1}{2}\right)^{n-r_1} C(r_1, r_1, k-1, n_1)C(r_2, r_2, d-1, n-n_1)
\]

\[
\times \left(\frac{2}{r_1 - r_2 + 1}\right) \int_0^1 p^{n_1}(1-p)^{n-n_1} dG(p).
\]

One of the most important quantities in start-up demonstration tests is the probability of accepting the equipment, which is defined as \( P_{k,d} = P\{W(1)_k < W(2)_d\} \) 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

\[
P_{k,d} = P\{W(1)_k < W(2)_d\}
\]

\[
= \sum_{n \geq k} P\{W(1)_k < W(2)_d \mid W(1)_k = n\}
\]

\[
\leq \sum_{n \geq k} \left( P\{L(1)_n < k, L(2)_n < d\} - P\{L(1)_n < k, L(2)_n < d\} \right)
\]

\[
= P_{k,d}.
\]
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


