Adv. Appl. Prob. **39,** 991–1019 (2007)
Printed in Northern Ireland
© *Applied Probability Trust* 2007

SUCCESS RUN STATISTICS DEFINED ON AN URN MODEL

FROSSO S. MAKRI,[*] [**]
ANDREAS N. PHILIPPOU [*] [***] AND
ZAHARIAS M. PSILLAKIS,[****] *University of Patras*

Abstract

Statistics denoting the numbers of success runs of length exactly equal and at least equal to a fixed length, as well as the sum of the lengths of success runs of length greater than or equal to a specific length, are considered. They are defined on both linearly and circularly ordered binary sequences, derived according to the Pólya–Eggenberger urn model. A waiting time associated with the sum of lengths statistic in linear sequences is also examined. Exact marginal and joint probability distribution functions are obtained in terms of binomial coefficients by a simple unified combinatorial approach. Mean values are also derived in closed form. Computationally tractable formulae for conditional distributions, given the number of successes in the sequence, useful in nonparametric tests of randomness, are provided. The distribution of the length of the longest success run and the reliability of certain consecutive systems are deduced using specific probabilities of the studied statistics. Numerical examples are given to illustrate the theoretical results.

Keywords: Success run; waiting time; urn model; Pólya–Eggenberger sampling scheme; linear and circular binary sequences; Bernoulli trial; nonparametric test of randomness; reliability of consecutive systems

2000 Mathematics Subject Classification: Primary 60C05
Secondary 62E15

1. Introduction

Runs are important in applied probability and statistical inference. As mentioned by Móri (1991), problems connected with runs and waiting times are popular in applied probability as they can be formulated without difficult notions or involved technical terms. Their solutions, however, are far from trivial, but they help us to understand the nature of randomness. Runs on a line are used in many areas such as meteorology, hypothesis testing, quality control, DNA sequences, psychology, radar astronomy, and system reliability (Cochran (1938), Mosteller (1941), Wolfowitz (1943), Schwager (1983), and Philippou (1986)). Runs on a circle are studied in relevant problems arising from oriented circles, circular arrays, distributions of balls in a ring, statistical run tests, and reliability theory (Barton and David (1958), Makri and Philippou (1994), and Koutras *et al.* (1995)). For a review of the theory and applications of runs we refer the reader to Balakrishnan and Koutras (2002), who provided excellent information on past and current developments in the area.

Received 29 November 2006; revision received 24 September 2007.

[*] Postal address: Department of Mathematics, University of Patras, 26500 Patras, Greece.
[**] Email address: makri@math.upatras.gr
[***] Email address: anphilip@math.upatras.gr
[****] Postal address: Department of Physics, University of Patras, 26500 Patras, Greece.
Email address: psillaki@physics.upatras.gr

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Consider a sequence of n two-state (success–failure) trials, arranged on a line or on a circle. A success run is defined as a sequence of consecutive successes (S) preceded and succeeded by failures (F) or by nothing. The number of successes in a (success) run is referred to as its length. The study of the number of success runs, the number of runs of a specified length using several counting schemes, the waiting time for the occurrence of a prespecified number of runs, and the shortest and the longest success run length have attracted the interest of many authors. Of statistical importance, among these studies, are the ones which consider the numbers of success runs of length exactly equal to and greater than or equal to a threshold length; see Mood (1940), Fu and Koutras (1994), Muselli (1996), Sen *et al.* (2002), (2003), and Eryilmaz and Demir (2007). Recently Fu *et al.* (2002), Antzoulakos *et al.* (2003a), and Lou (2003) studied a statistic denoting the sum of the lengths of the success runs (i.e. the total number of successes in all the success runs) of length greater than or equal to a prespecified length, and the waiting time for the first time that the abovementioned statistic equals or exceeds a predetermined level (Antzoulakos *et al.* (2003b)). These authors studied sequences of trials ordered on a line by exploiting a Markov chain embedding technique; see Fu and Koutras (1994), and Fu and Lou (2003).

In this article we adopt a simple unified combinatorial approach, through distributions of balls into cells (see, for instance, Riordan (1964) and Charalambides (2002)), to investigate the abovementioned statistics and also the bivariate random variable defined by the sum of lengths of all success runs of length at least equal to a threshold length paired with the number of success runs of length at least equal to the same threshold. The statistics are defined for binary sequences ordered both on a line or on a circle. The sequences considered are the outcomes of drawing balls from an urn model with stochastic replacements according to the Pólya–Eggenberger sampling scheme, which can be used as a probabilistic model in applied probability and statistics; see Johnson and Kotz (1977). 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 (S), and drawing a black ball is considered a failure (F). This sampling scheme is repeated n times and a binary sequence is derived, which is reduced to a Bernoulli sequence (independent and identically distributed binary trials) for $s = 0$. The sequences can be arranged on a line or on a circle. Owing to the generality of the approach, new simple exact formulae are established and known ones are rediscovered for special values of the parameter s.

Our article is organized as follows. In Section 2 we state the definitions of the studied statistics, referring to several types of nonoverlapping enumerating schemes, and we give a brief outline of a general framework for evaluating their probability distribution functions (PDFs) and means. In Section 3 we examine the numbers of success runs of length exactly equal to and greater than or equal to a threshold length. In Section 4 we consider the sum of the lengths of the success runs of length greater than or equal to a threshold length and the bivariate statistic defined by it paired with the number of success runs of length at least equal to the same threshold. We also examine the associated waiting time until the sum of lengths statistic equals or exceeds a predetermined value for the first time. Specifically, the exact PDF of all the abovementioned statistics are derived, via combinatorial analysis, as well as the mean values of the first three of them in closed form, for both linear and circular sequences. An efficient recursive scheme is also given for the PDF of the sum of lengths statistic defined on a linear sequence, and connecting relationships of its PDF for linearly and circularly ordered Bernoulli sequences are provided. New computationally tractable formulae for the conditional

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

distributions, given the number of successes in the sequence, are derived for both the number of success runs of length at least equal to a threshold length and the sum of lengths of the success runs of length at least equal to the same threshold. The latter distributions defined on linearly and circularly ordered sequences may be used in certain nonparametric tests of randomness. See Koutras and Alexandrou (1997), and Antzoulakos *et al.* (2003a) for a recent use of the linear distributions. Furthermore, in Section 4 we give the distributions of the length of the longest success run and of the waiting time until a specified number of consecutive successes is observed for the first time. They are deduced as a by-product of the study of the statistics examined in Sections 3 and 4, of which specific probabilities also give the reliability of a linear or circular consecutive-k-out-of-n:F system. Finally, in Section 5 numerical examples are given to illustrate our theoretical results.

We end this section by noting that the present paper generalizes, unifies, and/or provides alternative formulae for the results of Mood (1940), Philippou *et al.* (1983), Panaretos and Xekalaki (1986), Ling (1988), Godbole (1990), Goldstein (1990), Hirano and Aki (1993), Koutras and Alexandrou (1997), Tripsiannis and Philippou (1997), Sen *et al.* (2002), (2003), Antzoulakos *et al.* (2003a), (2003b), and Makri *et al.* (2007a), (2007b).

2. Notation, definitions, and general results

In Section 2.1 we give the basic definitions and the required notation that will be used throughout this article. In Section 2.2 we present a general framework for evaluating the probability distribution function and the mean values of the statistics (random variables (RVs)) that appear in Section 2.1. The details are provided in Sections 3 and 4.

2.1. Notation and definitions

The Pólya–Eggenberger sampling scheme, $\mathrm{PE}(w, b, s)$, is repeated n times and a binary sequence is derived. The sequence can be ordered on a line or on a circle. In the circular case we assume that the first outcome is adjacent to (and follows) the nth outcome. Hence, depending on the ordering of the outcomes, two kinds of sequences are defined. A linear sequence is defined if the outcomes are ordered on a line, and a circular sequence is defined if the outcomes are ordered circularly. In addition, for any linear or circular sequence the parameter s defines various (discrete) sampling schemes. Of special interest are the following values of s: $s = -1$ (sampling without replacement), $s = 0$ (sampling with replacement), $s = 1$, and $s = w = b > 0$.

Given the sampling scheme $\mathrm{PE}(w, b, s)$, the length of the binary sequence n, $n > 0$, the success run threshold length k, $0 < k \leq n$, and a predetermined value r, $r \geq k > 0$, we define the following variables.

(a) Let $E_{n,k}$ denote the number of success runs with length exactly equal to k in the sequence.

(b) Let $G_{n,k}$ denote the number of success runs of length at least k, i.e. $G_{n,k} = \sum_{i=k}^{n} E_{n,i}$.

(c) Let $S_{n,k}$ denote the sum of the lengths of the success runs (i.e. the total number of successes in all the success runs) of length greater than or equal to k, i.e. $S_{n,k} = \sum_{i=k}^{n} i E_{n,i}$.

In a study of a binary sequence it is natural for someone to be interested in $S_{n,k}$ and simultaneously in $G_{n,k}$. This is because together these two numbers provide a more refined view of the internal clustering structure of the sequence, compared with the information derived by each one alone. For instance, a large value of $S_{n,k}$ paired with a small value of $G_{n,k}$ indicates the existence of a large success cluster and therefore

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

a trend, whereas the same large value of $S_{n,k}$ paired with a large value of $G_{n,k}$ indicates a 'uniform' distribution of success runs of small sizes in the sequence. Therefore, the usefulness of the following bivariate RV is apparent.

(d) Let $\boldsymbol{M}_{n,k} = (S_{n,k}, G_{n,k})$.

When the trials are arranged on a circle we denote by $E^{\mathrm{c}}_{n,k}$, $G^{\mathrm{c}}_{n,k}$, $S^{\mathrm{c}}_{n,k}$, and $\boldsymbol{M}^{\mathrm{c}}_{n,k} = (S^{\mathrm{c}}_{n,k}, G^{\mathrm{c}}_{n,k})$ the corresponding RVs.

(e) Let $T_{r,k}$ denote the waiting time until the sum of the lengths of all success runs of length greater than or equal to k equals or exceeds the value r for the first time, i.e. $T_{r,k} = \min\{n \geq r \colon S_{n,k} \geq r\}$.

In addition, let L_n and L^{c}_n denote the lengths of the longest success runs in n trials ordered linearly and circularly, respectively, and W_k denote the number of draws until a sequence of k consecutive successes is observed for the first time.

It is clear that, given a sampling scheme $\mathrm{PE}(w, b, s)$ for any n and k, the relationships

$$\begin{gathered}
E_{n,k} \leq G_{n,k} \leq S_{n,k}, \qquad E_{n,k} = G_{n,k} - G_{n,k+1}; \\
E^{\mathrm{c}}_{n,k} \leq G^{\mathrm{c}}_{n,k} \leq S^{\mathrm{c}}_{n,k}, \qquad E^{\mathrm{c}}_{n,k} = G^{\mathrm{c}}_{n,k} - G^{\mathrm{c}}_{n,k+1}; \\
W_k > n \text{ if and only if } L_n < k \text{ if and only if } G_{n,k} = S_{n,k} = 0; \\
L^{\mathrm{c}}_n < k \text{ if and only if } G^{\mathrm{c}}_{n,k} = S^{\mathrm{c}}_{n,k} = 0; \\
W_k = T_{k,k}; \text{ and} \\
L_n = S_{n,k} = k \text{ if } W_k = n,
\end{gathered}$$

always hold.

The RVs $G_{n,1}$ and $G^{\mathrm{c}}_{n,1}$ denote the numbers of success runs in the sequence, and $S_{n,1}$ and $S^{\mathrm{c}}_{n,1}$, where $S_{n,1} \equiv S^{\mathrm{c}}_{n,1}$, denote the numbers of successes in n trials ordered linearly and circularly, respectively. Furthermore, for $s = 0$, $S_{n,1}$ is a binomial $\mathrm{B}(n, w/(w+b))$ RV, and $T_{r,1}$ is a negative binomial $\mathrm{NB}(r, w/(w+b))$ RV.

The foregoing definitions are illustrated using the following example. Let the first 10 binary trials be SSSFSFSSFS. Then, $E_{10,1} = 2$, $E^{\mathrm{c}}_{10,1} = 1$, $E_{10,2} = 1$, $E^{\mathrm{c}}_{10,2} = 1$, $E_{10,3} = 1$, $E^{\mathrm{c}}_{10,3} = 0$, $E_{10,4} = 0$, $E^{\mathrm{c}}_{10,4} = 1$, $E_{10,i} = E^{\mathrm{c}}_{10,i} = 0$, $i = 5, 6, \dots, 10$; $G_{10,2} = 2$, $G^{\mathrm{c}}_{10,2} = 2$; $S_{10,2} = 5$, $S^{\mathrm{c}}_{10,2} = 6$; $L_{10} = 3$, $L^{\mathrm{c}}_{10} = 4$; and $T_{2,2} = W_2 = L_2 = S_{2,2} = 2$, $T_{3,2} = 3$, $T_{4,2} = T_{5,2} = 8$, $T_{6,2} > 10$.

Throughout the article, for integers n, m, $\binom{n}{m}$ denotes the extended binomial coefficient; see Feller (1968, pp. 50, 63). Furthermore, in order to avoid repetitions we note here that (unless otherwise stated) $\delta_{i,j}$ denotes the Kronecker delta function of the integer arguments i and j;

$$\mathrm{B}(\alpha, \beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1}\,\mathrm{d}p \quad \text{for } \alpha > 0 \text{ and } \beta > 0;$$

$[x]$ denotes the greatest integer less than or equal to x; and, for $m = 0, 1, \dots,$

$$x^{(m)} = x(x-1)\cdots(x-m+1) \quad \text{with } x^{(0)} = 1,$$

denoting the mth falling factorial of x. Also, we apply the conventions $\sum_{i=a}^{b} = 0$ and $\prod_{i=a}^{b} = 1$ for $a > b$.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

2.2. General results

In this section we provide a brief outline of a simple unified combinatorial approach for the study of the above-defined statistics. The approach is based on the technique of allocating indistinguishable balls into distinguishable cells.

Let the sequence of outcomes in n repetitions of the PE(w, b, s) sampling scheme be arranged on a line or on a circle. The elements ω of the appropriate sample space Ω are linear or circular permutations $(i_1, \dots, i_n)$ with $i_j \in \{\mathrm{S}, \mathrm{F}\}$, $j = 1, \dots, n$. For convenience we use the RV $X_{n,k}$ to represent any of the following RVs: $E_{n,k}$, $E^{\mathrm{c}}_{n,k}$, $G_{n,k}$, $G^{\mathrm{c}}_{n,k}$, $\boldsymbol{M}_{n,k}$, $\boldsymbol{M}^{\mathrm{c}}_{n,k}$, $S_{n,k}$, or $S^{\mathrm{c}}_{n,k}$. Let Y_n denote the number of Fs in the sequence. An element of the event

$$\Gamma_{X_{n,k}}(x, y) = \{\omega \in \Omega \colon X_{n,k}(\omega) = x, Y_n(\omega) = y\}$$

is a sequence of n drawings with y Fs and $n - y$ Ss that has probability $p_n(y)$ given by

$$p_n(y) = \frac{\prod_{j=0}^{n-y-1}(w + js)\prod_{j=0}^{y-1}(b + js)}{\prod_{j=0}^{n-1}(w + b + js)}, \qquad 0 \le y \le n. \tag{2.1}$$

For $s = 0$, $p_n(y)$ reduces to $p^{n-y}q^y$ with $p = w/(w + b)$ and $q = 1 - p = b/(w + b)$, and it denotes the probability of the above sequence when the balls are drawn with replacement. This case corresponds to Bernoulli trials with constant success probability p.

Noting that

$$\mathrm{P}(Y_n = y) = \binom{n}{y} p_n(y), \qquad y = 0, 1, \dots, n,$$

and that

$$\mathrm{P}(X_{n,k} = x,\ Y_n = y) = \mathrm{P}(\Gamma_{X_{n,k}}(x, y)) = N_{X_{n,k}}(x, y)p_n(y),$$

where $N_{X_{n,k}}(x, y)$ denotes the number of permutations contained in $\Gamma_{X_{n,k}}(x, y)$, we obtain the following result.

Proposition 2.1. *Let $X_{n,k}$ and Y_n be as defined above. Then,*

$$\mathrm{P}(X_{n,k} = x) = \sum_{y} N_{X_{n,k}}(x, y)p_n(y) \tag{2.2}$$

and

$$\mathrm{P}(X_{n,k} = x \mid Y_n = y) = \binom{n}{y}^{-1} N_{X_{n,k}}(x, y). \tag{2.3}$$

Remark 2.1. According to Proposition 2.1, the problem of establishing the PDF of $X_{n,k}$ and the conditional PDF of $X_{n,k}$, given Y_n, is a combinatorial one; specifically, the computation of the number $N_{X_{n,k}}(x, y)$. For this task, we note that the y Fs form $y + 1$ $(U_j,\ j = 1, \dots, y + 1)$ or y $(U^{\mathrm{c}}_j,\ j = 1, \dots, y)$ cells if the outcomes are ordered linearly or circularly, respectively. Here U_1 is the cell formed before the first F, U_{y+1} is the cell formed after the last F, and U_j, $j = 2, \dots, y$, is the cell formed between the $(j - 1)$th and the jth Fs. In a circular sequence, labelling an F as the first one, U^{c}_j, $j = 1, \dots, y - 1$, is the cell formed between the jth and the $(j + 1)$th Fs, and U^{c}_y is the cell formed between the yth and the first Fs. So, the problem of establishing $N_{X_{n,k}}(x, y)$ is to enumerate the different allocations of $n - y$ Ss in the formed cells so that $\Gamma_{X_{n,k}}(x, y)$ occurs, depending on the internal structure of the RV $X_{n,k}$.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Remark 2.2. The repetition of the Pólya–Eggenberger sampling derives an exchangeable binary sequence. Proposition 2.1 expresses, in a formal way, an idea that is often used in the statistical analysis of exchangeable binary sequences; see, for instance, Schuster (1991, Lemma 2.1) and Eryilmaz and Demir (2007, Lemma 2.2). The exchangeability implies that all sequences with the same number of failures are equally likely. This elementary property establishes the truth of (2.3) for any exchangeable binary sequence. Furthermore, (2.2) is also valid for exchangeable binary sequences, provided that the probability $p_n(y)$ is properly given; see George and Bowman (1995, Theorem 2.1) or Eryilmaz and Demir (2007, Equation (2.2)).

Next we consider the mean value of the RV $X_{n,k}$, $s \ge 0$, where $X_{n,k}$ stands for any of the following RVs: $E_{n,k}$, $E^{\mathrm{c}}_{n,k}$, $G_{n,k}$, $G^{\mathrm{c}}_{n,k}$, $S_{n,k}$, or $S^{\mathrm{c}}_{n,k}$. Proposition 2.2, below, gives the mean value $\mathrm{E}(X_{n,k})$ for $s > 0$, provided that $\mathrm{E}(X_{n,k})$ is known for $s = 0$.

Let X and P be two RVs such that P is distributed as $\mathrm{Beta}(\alpha, \beta)$, $\alpha > 0$ and $\beta > 0$, and that the conditional PDF of X, given that $P = p$, is

$$f_{X \mid P}(x \mid p) = \sum_{y} N_{X_{n,k}}(x, y) p^{n-y}(1-p)^{y}.$$

Then,

$$f_X(x) = \mathrm{P}(X = x) = \sum_{y} N_{X_{n,k}}(x, y) \frac{\mathrm{B}(\alpha + n - y, \beta + y)}{\mathrm{B}(\alpha, \beta)}.$$

For $\alpha = w/s$ and $\beta = b/s$, $s > 0$, we have $\mathrm{B}(\alpha + n - y, \beta + y)/\mathrm{B}(\alpha, \beta) = p_n(y)$, so that $f_X(x)$ is the PDF of $X_{n,k}$ and $\mathrm{E}(X_{n,k}) = \mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X \mid P))$. Clearly, for $s = 0$, $\mathrm{E}(X_{n,k}) = \mathrm{E}(X \mid P = p)$. Hence, we have the following result.

Proposition 2.2. *If, for $s = 0$,*

$$\mathrm{E}(X_{n,k}) = \sum_{m,r} \lambda_{m,r} p^{m}(1-p)^{r},$$

with integers m, r and $\lambda_{m,r} \in \mathbb{R}$ then, for $s > 0$,

$$\mathrm{E}(X_{n,k}) = \frac{1}{\mathrm{B}(\alpha, \beta)} \sum_{m,r} \lambda_{m,r} \mathrm{B}(\alpha + m, \beta + r),$$

where $\alpha = w/s$, $\beta = b/s$, $\alpha + m > 0$, and $\beta + r > 0$.

Before we proceed further we give three preliminary lemmas.

Lemma 2.1. (Sen *et al.* (2003).) *The number of allocations of α indistinguishable balls into r distinguishable cells, where no cell has exactly k balls, is given by*

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

Lemma 2.2. *The number of allocations of α indistinguishable balls into r distinguishable cells, where each of the m, $0 \le m \le r$, specified cells is occupied by at most k balls, is given by*

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

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Proof. It is clear that $H_m(\alpha, r, k)$ is equal to the t^α coefficient of the generating function $g(t) = (1 - t^{k+1})^m(1 - t)^{-r}$, from which the result follows.

We note that $H_m(\alpha, r, 0)$ denotes the number of allocations of α indistinguishable balls into $r - m$ distinguishable cells. Therefore, setting $k = 0$ in Lemma 2.2 and employing the identity

$$\sum_{k=0}^{n}(-1)^{n-k}\binom{r}{n-k}\binom{s+k}{k} = \binom{s-r+n}{n}, \qquad s \geq r$$

(see Charalambides (2002, p. 128)), we obtain, as a corollary, the following well-known result.

Corollary 2.1. *Let $H_m(\alpha, r, k)$ be as given in Lemma 2.2. Then,*

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

Also, $H_r(\alpha, r, k)$ denotes the number of allocations of α indistinguishable balls into r distinguishable cells, where each cell is occupied by at most k balls. Therefore, replacing m by r in Lemma 2.2 we obtain, as a corollary, the following well-known result; see Riordan (1964, p. 104).

Corollary 2.2. *Let $H_m(\alpha, r, k)$ be as given in Lemma 2.2. Then,*

$$H_r(\alpha, r, k) \equiv C(\alpha, r, k) = \sum_{j=0}^{[a/(k+1)]}(-1)^j\binom{r}{j}\binom{a - (k+1)j + r - 1}{\alpha - (k+1)j}.$$

Lemma 2.3. *Let $J(\alpha, r, k)$ denote the number of allocations of α indistinguishable balls into r distinguishable cells, where each cell is occupied by at least k balls. Then,*

$$J(\alpha, r, k) = \binom{\alpha - (k-1)r - 1}{r - 1}.$$

3. Statistics referring to the number of success runs

In this section we deal with the numbers of success runs of length exactly equal to and greater than or equal to a threshold length k. Our study is carried out for both linearly and circularly ordered trials derived by n repetitions of a $\mathrm{PE}(w, b, s)$ sampling scheme. Specifically, in Section 3.1 we give the PDF and the mean values of $E_{n,k}$ and $E^{\mathrm{c}}_{n,k}$, whereas in Section 3.2 we give the PDF , the conditional PDF , given $S_{n,1}$, and the mean values of $G_{n,k}$ and $G^{\mathrm{c}}_{n,k}$. All the results are obtained by means of the method presented in Propositions 2.1 and 2.2.

3.1. The number of success runs of length exactly equal to k

For trials arranged on a line, Theorem 3.1 gives the PDF of the RV $E_{n,k}$ and Proposition 3.1 provides its mean value for $s \geq 0$. For trials arranged on a circle, Theorem 3.2 gives the PDF of the RV $E^{\mathrm{c}}_{n,k}$ and Proposition 3.2 provides its mean value for $s \geq 0$. First we consider trials arranged on a line.

Theorem 3.1. *The PDF of $E_{n,k}$ is given by*

$$\mathrm{P}(E_{n,k} = x) = \sum_{y=0}^{n-kx} p_n(y)\binom{y+1}{x}A(n - y - kx, y + 1 - x, k), \qquad x = 0, 1, \ldots, \left[\frac{n+1}{k+1}\right]. \tag{3.1}$$

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Proof. Note that x cells can be selected from the $y+1$ distinguishable cells in $\binom{y+1}{x}$ ways, $x = 0, 1, \dots, [(n+1)/(k+1)]$. Next, one success run of length k is placed in each selected cell in only one way. Furthermore, for each specified selection of x cells $\{U_{j_1}, \dots, U_{j_x}\}$ out of the $y+1$ cells $\{U_1, \dots, U_{y+1}\}$, the number of ways the remaining $n-y-kx$ Ss can be placed in the remaining $y+1-x$ cells (excluding the x specified cells) with no cell receiving exactly k Ss equals $A(n-y-kx, y+1-x, k)$ by Lemma 2.1. Thus, according to the multiplicative principle, the total number of allocations of $n-y$ Ss in the $y+1$ cells yielding x runs of Ss of length exactly equal to k is given by

$$N_{E_{n,k}}(x, y) = \binom{y+1}{x} A(n-y-kx, y+1-x, k).$$

The result then follows from Proposition 2.1.

A possible alternative formula, in terms of binomial and multinomial coefficients, for the PDF of $E_{n,k}$ may be obtained using Theorem 3.1 of Sen *et al.* (2002) and following the approach used in Corollary 2 of the same article. However, by this approach the PDF of $E_{n,k}$ should contain six consecutive summations involving binomial and multinomial coefficients instead of the two summations involving binomial coefficients given in (3.1).

Proposition 3.1. *Let $\mu_{E_{n,k}}$ denote the mean of the RV $E_{n,k}$. Then,*

(a) *for $s = 0$,*

$$\mu_{E_{n,k}} = \begin{cases} qp^k(2 + (n-k-1)q), & k = 1, \dots, n-1, \\ p^n, & k = n; \end{cases} \tag{3.2}$$

(b) *for $s > 0$, $\alpha = w/s$, and $\beta = b/s$,*

$$\mu_{E_{n,k}} = \begin{cases} \dfrac{2\mathrm{B}(\alpha+k, \beta+1) + (n-k-1)\mathrm{B}(\alpha+k, \beta+2)}{\mathrm{B}(\alpha, \beta)}, & k = 1, \dots, n-1, \\ \dfrac{\mathrm{B}(\alpha+n, \beta)}{\mathrm{B}(\alpha, \beta)}, & k = n. \end{cases}$$

Proof. (a) Obviously, $\mu_{E_{n,n}} = p^n$. Let Z_i, $1 \le i \le n$, be independent RVs with PDF

$$\mathrm{P}(Z_i = x) = p^x(1-p)^{1-x}, \qquad x = 0, 1,\ 0 < p < 1,\ 1 \le i \le n.$$

We define a binary RV U_j, $k \le j \le n$, as follows:

$$U_j = \begin{cases} 1 & \text{if } \prod_{i=j-k+1}^{j} Z_i = 1,\ Z_{j-k} = Z_{j+1} = 0, \\ 0 & \text{otherwise} \end{cases}$$

(convention: $Z_0 = Z_{n+1} = 0$). As $E_{n,k} = \sum_{j=k}^{n} U_j$, we have

$$\mu_{E_{n,k}} = \sum_{j=k}^{n} \mathrm{E}(U_j) = 2qp^k + (n-k-1)q^2p^k, \qquad k = 1, \dots, n-1.$$

(b) It follows from part (a) and Proposition 2.2.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

For an alternative derivation of (3.2), see Mood (1940). Next we consider trials arranged on a circle.

Theorem 3.2. *The PDF of $E^{\mathrm{c}}_{n,k}$ is given by*

(a) *for $n = k$, $\mathrm{P}(E^{\mathrm{c}}_{n,k} = 0) = 1 - p_n(0)$ and $\mathrm{P}(E^{\mathrm{c}}_{n,k} = 1) = p_n(0)$;*

(b) *for $n \geq k + 1$ and $x = 0, 1, \dots, [n/(k + 1)]$,*

$$\mathrm{P}(E^{\mathrm{c}}_{n,k} = x) = \sum_{y=1}^{n-kx} p_n(y)\frac{n}{y}\binom{y}{x}A(n - y - kx, y - x, k) + p_n(0)\delta_{x,0}. \tag{3.3}$$

Proof. The proof of part (a) is apparent. To prove part (b) first note that x cells can be chosen from the y distinguishable cells, $U^{\mathrm{c}}_1, \dots, U^{\mathrm{c}}_y$, in $\binom{y}{x}$ ways, $x = 0, 1, \dots, [n/(k + 1)]$, $y \geq 1$. For each specified selection of x cells, one success run of length k is placed in each selected cell, while the remaining $n - y - kx$ Ss can be placed in the remaining $y - x$ cells with no cell receiving exactly k Ss in $A(n - y - kx, y - x, k)$ ways. Therefore, the total number of allocations of $n - y$ Ss in the y cells yielding x runs of Ss of length exactly equal to k is given by $\binom{y}{x}A(n - y - kx, y - x, k)$. Furthermore, each of these arrangements gives rise to n arrangements of the $n - y$ Ss and y Fs by rotation. But the set of $n\binom{y}{x}A(n - y - kx, y - x, k)$ arrangements is partitioned into sets of y like arrangements. So, the total number of circular arrangements with $n - y$ Ss and y Fs yielding x runs of Ss of length exactly equal to k is given by

$$N_{E^{\mathrm{c}}_{n,k}}(x, y) = \frac{n}{y}\binom{y}{x}A(n - y - kx, y - x, k).$$

Furthermore, for $x = 0, 1, \dots, [n/(k + 1)]$, $N_{E^{\mathrm{c}}_{n,k}}(x, 0) = \delta_{x,0}$. The result then follows from Proposition 2.1.

For an alternative formula for the PDF of $E^{\mathrm{c}}_{n,k}$, see Sen *et al.* (2003, Theorem 3). Their formula is more complicated than ours as it contains four consecutive summations of binomial coefficients instead of the two summations given in (3.3).

Proposition 3.2. *Let $\mu_{E^{\mathrm{c}}_{n,k}}$ denote the mean of $E^{\mathrm{c}}_{n,k}$. Then,*

(a) *for $s = 0$,*

$$\mu_{E^{\mathrm{c}}_{n,k}} = \begin{cases} nq^2p^k, & k = 1, \dots, n - 2, \\ nqp^{n-1}, & k = n - 1, \\ p^n, & k = n; \end{cases} \tag{3.4}$$

(b) *for $s > 0$, $\alpha = w/s$, and $\beta = b/s$,*

$$\mu_{E^{\mathrm{c}}_{n,k}} = \begin{cases} \dfrac{n\mathrm{B}(\alpha + k, \beta + 2)}{\mathrm{B}(\alpha, \beta)}, & k = 1, \dots, n - 2, \\[2ex] \dfrac{n\mathrm{B}(\alpha + n - 1, \beta + 1)}{\mathrm{B}(\alpha, \beta)}, & k = n - 1, \\[2ex] \dfrac{\mathrm{B}(\alpha + n, \beta)}{\mathrm{B}(\alpha, \beta)}, & k = n. \end{cases}$$

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Proof. (a) Obviously, $\mu_{E^{\mathrm{c}}_{n,n}} = p^n$ and $\mu_{E^{\mathrm{c}}_{n,n-1}} = nqp^{n-1}$. Let Z_i, $1 \le i \le n$, be as defined in the proof of Proposition 3.1. For $1 \le k \le n-2$, we define a binary RV U_j, $1 \le j \le n$, as follows. For $j = 1, \dots, k$,

$$U_j = \begin{cases} 1 & \text{if } \prod_{i=1}^{j} Z_i \prod_{i=n-k+j+1}^{n} Z_i = 1,\ Z_{j+1} = Z_{n-k+j} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

For $j = k+1, \dots, n$,

$$U_j = \begin{cases} 1 & \text{if } \prod_{i=j-k+1}^{j} Z_i = 1,\ Z_{j+1} = Z_{j-k} = 0, \\ 0 & \text{otherwise} \end{cases}$$

($Z_{n+1} \equiv Z_1$). Then, $E^{\mathrm{c}}_{n,k} = \sum_{j=1}^{n} U_j$, so that

$$\mu_{E^{\mathrm{c}}_{n,k}} = \sum_{j=1}^{n} \mathrm{E}(U_j) = nq^2p^k, \qquad k = 1, \dots, n-2.$$

(b) It follows directly from part (a) and Proposition 2.2.

3.2. The number of success runs of length at least *k*

For trials arranged on a line, Theorem 3.3 gives the PDF of the RV $G_{n,k}$. For trials arranged on a circle, Theorem 3.4 gives the PDF of the RV $G^{\mathrm{c}}_{n,k}$. In Corollaries 3.1 and 3.2 we obtain the conditional PDFs of $G_{n,k}$ and $G^{\mathrm{c}}_{n,k}$, respectively, given the number of successes $S_{n,1}$. Finally, Propositions 3.3 and 3.4 provide the means of $G_{n,k}$ and $G^{\mathrm{c}}_{n,k}$ for $s \ge 0$.

Theorem 3.3. *The PDF of* $G_{n,k}$ *is given by*

$$\mathrm{P}(G_{n,k} = x) = \sum_{y=0}^{n-kx} p_n(y) \binom{y+1}{x} H_{y+1-x}(n-y-kx, y+1, k-1), \qquad x = 0, 1, \dots, \left[\frac{n+1}{k+1}\right]. \tag{3.5}$$

Proof. First note that the number of ways that x cells, in each one of which a success run of length k is placed, can be chosen from the $y+1$ cells equals $\binom{y+1}{x}$, $x = 0, 1, \dots, [(n+1)/(k+1)]$. Furthermore, for each specified selection of x cells, the number of ways the remaining $n-y-kx$ Ss can be placed in the $y+1$ cells with each one of the remaining $y+1-x$ cells receiving no more than $k-1$ Ss equals $H_{y+1-x}(n-y-kx, y+1, k-1)$ by Lemma 2.2. Therefore, the total number of allocations of $n-y$ Ss in the $y+1$ cells yielding x runs of Ss of length greater than or equal to k is given by

$$N_{G_{n,k}}(x, y) = \binom{y+1}{x} H_{y+1-x}(n-y-kx, y+1, k-1).$$

The result then follows from Proposition 2.1.

A possible alternative for the PDF of $G_{n,k}$ may be obtained using the same guidelines discussed after the proof of Theorem 3.1. This, however, should contain six consecutive summations involving binomial and multinomial coefficients instead of the two summations involving binomial coefficients given in (3.5). For $s = 0$, a single summation formula for $\mathrm{P}(G_{n,k} = x)$ has been given by Muselli (1996). Unfortunately, the use of this expression does not allow us to generalize to the case in which $s \neq 0$.

Setting $k = 1$ in (3.5) we derive, by means of Corollary 2.1, the PDF of $G_{n,1}$, the number of success runs in a sequence of n draws according to the $\mathrm{PE}(w, b, s)$ sampling scheme ordered linearly:

$$\mathrm{P}(G_{n,1} = x) = \sum_{y=0}^{n-x} p_n(y) \binom{y+1}{x} \binom{n-y-1}{n-y-x}, \qquad x = 0, 1, \dots, \left[\frac{n+1}{2}\right].$$

Corollary 3.1. *The conditional PDF* $\mathrm{P}(G_{n,k} = x \mid S_{n,1} = n - y)$ *for* $0 \le y \le n$ *and* $x = 0, 1, \dots, [(n-y)/k]$ *is given by*

$$\begin{aligned}\mathrm{P}(G_{n,k} &= x \mid S_{n,1} = n - y) \\ &= \binom{n}{y}^{-1} \binom{y+1}{x} \sum_{i=0}^{[(n-y-kx)/k]} (-1)^i \binom{y+1-x}{i} \binom{n-k(x+i)}{n-k(x+i)-y}. \end{aligned} \qquad (3.6)$$

Proof. It follows directly from Proposition 2.1.

Corollary 3.1 immediately yields the following result, by means of Corollary 2.1.

The conditional PDF $\mathrm{P}(G_{n,1} = x \mid S_{n,1} = n - y)$ for $0 \le y \le n$ and $0 \le x \le n - y$ is given by

$$\mathrm{P}(G_{n,1} = x \mid S_{n,1} = n - y) = \binom{n}{y}^{-1} \binom{y+1}{x} \binom{n-y-1}{n-y-x}. \qquad (3.7)$$

Equation (3.6) has been derived by Koutras and Alexandrou (1997) using generating functions, whereas Mood (1940) (see also Gibbons and Chakraborti (2003, p. 79)) has derived (3.7) using a different method.

Using the representation $G_{n,k} = \sum_{i=k}^{n} E_{n,i}$, we have a straightforward derivation of its mean.

Proposition 3.3. *Let* $\mu_{G_{n,k}}$ *denote the mean of the RV* $G_{n,k}$. *Then,*

(a) *for* $s = 0$,

$$\mu_{G_{n,k}} = p^k(1 + (n-k)q), \qquad n \ge k; \qquad (3.8)$$

(b) *for* $s > 0$, $\alpha = w/s$, *and* $\beta = b/s$,

$$\mu_{G_{n,k}} = \frac{\mathrm{B}(\alpha + k, \beta) + (n-k)\mathrm{B}(\alpha + k, \beta + 1)}{\mathrm{B}(\alpha, \beta)}, \qquad n \ge k.$$

Proof. (a) We have

$$\mu_{G_{n,k}} = \sum_{i=k}^{n} \mu_{E_{n,i}} = p^n + (2q + (n-1)q^2) \sum_{i=k}^{n-1} p^i - q^2 \sum_{i=k}^{n-1} i p^i = p^k(1 + (n-k)q).$$

(b) It follows from part (a) and Proposition 2.2.

For alternative derivations of (3.8), see Mood (1940), Goldstein (1990), and Hirano and Aki (1993). Next we consider trials arranged on a circle.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Theorem 3.4. *The PDF of* $G^{\mathrm{c}}_{n,k}$ *is given as follows:*

(a) *for* $n = k$, $\mathrm{P}(G^{\mathrm{c}}_{n,k} = 0) = 1 - p_n(0)$ *and* $\mathrm{P}(G^{\mathrm{c}}_{n,k} = 1) = p_n(0)$*;*

(b) *for* $n \geq k + 1$ *and* $x = 0, 1, \dots, [n/(k+1)]$,

$$\mathrm{P}(G^{\mathrm{c}}_{n,k} = x) = \sum_{y=1}^{n-kx} p_n(y) \frac{n}{y} \binom{y}{x} H_{y-x}(n - y - kx, y, k - 1) + p_n(0)\delta_{x,1}. \qquad (3.9)$$

Proof. The proof of part (a) is apparent. To prove part (b) first note that the number of ways that x cells, within each one of which a success run of length k is placed, can be chosen from the y cells equals $\binom{y}{x}$, $x = 0, 1, \dots, [n/(k+1)]$, $y \geq 1$. Furthermore, the number of ways the remaining $n - y - kx$ Ss can be placed in the y cells with each of the remaining $y - x$ cells receiving no more than $k - 1$ Ss is equal to $H_{y-x}(n - y - kx, y, k - 1)$. Therefore, the total number of allocations of $n - y$ Ss in the y cells yielding x runs of Ss of length at least equal to k is given by $\binom{y}{x} H_{y-x}(n - y - kx, y, k - 1)$. Continuing along the lines of the proof of Theorem 3.2 we find that, for $x = 0, 1, \dots, [n/(k+1)]$ and $y \geq 1$, the total number of circular arrangements with $n - y$ Ss and y Fs yielding x runs of Ss of length at least equal to k is given by

$$N_{G^{\mathrm{c}}_{n,k}}(x, y) = \frac{n}{y} \binom{y}{x} H_{y-x}(n - y - kx, y, k - 1).$$

Furthermore, for $x = 0, 1, \dots, [n/(k+1)]$, $N_{G^{\mathrm{c}}_{n,k}}(x, 0) = \delta_{x,1}$. The result then follows from Proposition 2.1.

Theorem 4 of Sen *et al.* (2003) gives an alternate formula for the PDF of $G^{\mathrm{c}}_{n,k}$. However, it contains five consecutive summations of binomial coefficients instead of the two summations given in (3.9).

Setting $k = 1$ in Theorem 3.4 we obtain, by means of Corollary 2.1, the PDF of $G^{\mathrm{c}}_{n,1}$, the number of success runs in n draws according to a PE(w, b, s) sampling scheme ordered circularly:

$$\mathrm{P}(G^{\mathrm{c}}_{1,1} = 0) = 1 - p_1(0), \qquad \mathrm{P}(G^{\mathrm{c}}_{1,1} = 1) = p_1(0),$$

$$\mathrm{P}(G^{\mathrm{c}}_{n,1} = x) = \sum_{y=1}^{n-x} p_n(y) \frac{n}{y} \binom{y}{x} \binom{n - y - 1}{n - y - x} + p_n(0)\delta_{x,1}, \qquad x = 0, 1, \dots, \left[\frac{n}{2}\right], \ n \geq 2.$$

Corollary 3.2. *The conditional PDF* $\mathrm{P}(G^{\mathrm{c}}_{n,k} = x \mid S_{n,1} = n - y)$ *is given by*

$$\mathrm{P}(G^{\mathrm{c}}_{n,k} = x \mid S_{n,1} = n) = \delta_{x,1}, \qquad x \geq 0,$$

$$\begin{aligned} &\mathrm{P}(G^{\mathrm{c}}_{n,k} = x \mid S_{n,1} = n - y) \\ &\quad = \binom{n-1}{y-1}^{-1} \binom{y}{x} \sum_{i=0}^{[(n-y-kx)/k]} (-1)^i \binom{y - x}{i} \binom{n - k(x+i) - 1}{n - k(x+i) - y}, \end{aligned} \qquad (3.10)$$

$y = 1, 2, \dots, n,\ x = 0, 1, \dots, [(n - y)/k]$.

Proof. It follows directly from Proposition 2.1.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Corollary 3.2 yields the following result, by means of Corollary 2.1.

The conditional PDF $\mathrm{P}(G^{\mathrm{c}}_{n,1} = x \mid S_{n,1} = n - y)$ is given by

$$\mathrm{P}(G^{\mathrm{c}}_{n,1} = x \mid S_{n,1} = n) = \delta_{x,1}, \qquad x \geq 0,$$
$$\mathrm{P}(G^{\mathrm{c}}_{n,1} = x \mid S_{n,1} = n - y) = \binom{n-1}{y-1}^{-1}\binom{y}{x}\binom{n-y-1}{n-y-x},$$

$y = 1, 2, \ldots, n,\ x = 0, 1, \ldots, n - y.$

Proposition 3.4. *Let $\mu_{G^{\mathrm{c}}_{n,k}}$ denote the mean of $G^{\mathrm{c}}_{n,k}$. Then,*

(a) *for $s = 0$,*

$$\mu_{G^{\mathrm{c}}_{n,k}} = \begin{cases} p^n + nqp^k, & k = 1, 2, \ldots, n-1, \\ p^n, & k = n; \end{cases} \tag{3.11}$$

(b) *for $s > 0$, $\alpha = w/s$, and $\beta = b/s$,*

$$\mu_{G^{\mathrm{c}}_{n,k}} = \begin{cases} \dfrac{\mathrm{B}(\alpha+n, \beta) + n\mathrm{B}(\alpha+k, \beta+1)}{\mathrm{B}(\alpha, \beta)}, & k = 1, 2, \ldots, n-1, \\ \dfrac{\mathrm{B}(\alpha+n, \beta)}{\mathrm{B}(\alpha, \beta)}, & k = n. \end{cases}$$

Proof. (a) The representation $G^{\mathrm{c}}_{n,k} = \sum_{i=k}^{n} E^{\mathrm{c}}_{n,i}$ and (3.4) imply that

$$\mu_{G^{\mathrm{c}}_{n,k}} = \sum_{i=k}^{n} \mu_{E^{\mathrm{c}}_{n,i}} = \begin{cases} nq^2 \displaystyle\sum_{i=k}^{n-2} p^i + nqp^{n-1} + p^n, & k = 1, 2 \ldots, n-2, \\ nqp^{n-1} + p^n, & k = n-1, \\ p^n, & k = n, \end{cases}$$

from which we obtain (3.11).

(b) It follows from part (a) and Proposition 2.2.

4. Statistics referring to the sum of the lengths of success runs

In this section we deal with the statistics related to the sum of the lengths of success runs of length greater than or equal to a threshold length k and to the associated waiting time until the sum of lengths equals or exceeds a predetermined value r for the first time. The exact PDF of the bivariate RVs $\boldsymbol{M}_{n,k}$ and $\boldsymbol{M}^{\mathrm{c}}_{n,k}$ are also studied. Besides their own independent merit, they provide a useful means for obtaining the PDFs of $S_{n,k}$ and $S^{\mathrm{c}}_{n,k}$. The $S_{n,k}$, $\boldsymbol{M}_{n,k}$ and $S^{\mathrm{c}}_{n,k}$, $\boldsymbol{M}^{\mathrm{c}}_{n,k}$ statistics are studied in Section 4.1, and the $T_{r,k}$ statistic is examined in Section 4.2. The results referring to $S_{n,k}$, $\boldsymbol{M}_{n,k}$ and $S^{\mathrm{c}}_{n,k}$, $\boldsymbol{M}^{\mathrm{c}}_{n,k}$ are derived by means of Propositions 2.1 and 2.2, whereas the study of $T_{r,k}$ is carried out following a minor modification of Proposition 2.1.

4.1. The sum of lengths of all success runs of length at least k

The $S_{n,k}$, $\boldsymbol{M}_{n,k} = (S_{n,k}, G_{n,k})$ and $S^{\mathrm{c}}_{n,k}$, $\boldsymbol{M}^{\mathrm{c}}_{n,k} = (S^{\mathrm{c}}_{n,k}, G^{\mathrm{c}}_{n,k})$ statistics are considered for trials arranged on a line and on a circle in Sections 4.1.1 and 4.1.2, respectively. In addition, we relate the RVs $G_{n,k}$, $S_{n,k}$ and $G^{\mathrm{c}}_{n,k}$, $S^{\mathrm{c}}_{n,k}$ to the RVs L_n and L^{c}_n, respectively, and W_k.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

4.1.1. *Trials arranged on a line.* In this section we initially derive the exact joint PDF of $S_{n,k}$ and $G_{n,k}$ (Theorem 4.1). Then we obtain the marginal PDF of $S_{n,k}$ (Theorem 4.2). We then give a computationally tractable formula for the exact conditional PDF of $S_{n,k}$, given $S_{n,1}$ (Corollary 4.1), followed by an exact formula for the mean of $S_{n,k}$ for $s \geq 0$ (Proposition 4.1). An efficient scheme for the recursive computation of $S_{n,k}$ is provided by Theorem 4.3. Finally, Proposition 4.2 relates the specific probabilities of $G_{n,k}$, $S_{n,k}$, L_n, and W_k. For $s = 0$, Proposition 4.2 gives a formula for the reliability of a linear consecutive system.

Theorem 4.1. *The exact joint PDF of $S_{n,k}$ and $G_{n,k}$ is given as follows:*

(a)

$$\mathrm{P}(S_{n,k} = 0,\ G_{n,k} = 0) = \sum_{y=0}^{n} p_n(y) C(n-y, y+1, k-1);$$

(b) *for $x = k, k+1, \dots, n$ and $m = 1, 2, \dots, [x/k]$,*

$$\begin{aligned}\mathrm{P}(S_{n,k} = x,\ G_{n,k} = m) \\ = \binom{x-(k-1)m-1}{m-1} \sum_{y=0}^{n-x} p_n(y) \binom{y+1}{m} C(n-y-x, y+1-m, k-1).\end{aligned} \tag{4.1}$$

Proof. For $1 \leq m \leq [x/k]$, $k \leq x \leq n$, m cells are chosen from the $y+1$ ones in $\binom{y+1}{m}$ ways. For each of these combinations, the number of ways x Ss are placed in the m chosen distinguishable cells so that each cell receives at least k balls is equal to $J(x, m, k)$ by Lemma 2.3. The number of ways the remaining $n-y-x$ Ss can be placed in the remaining $y+1-m$ cells with no cell receiving more than $k-1$ Ss is equal to $C(n-y-x, y+1-m, k-1)$ by Corollary 2.2. Therefore, the total number of allocations of $n-y$ Ss in the $y+1$ cells yielding m success runs of length greater than or equal to k, with total number of successes equal to x, is given by

$$N_{(S_{n,k}, G_{n,k})}(x, m, y) = \binom{y+1}{m} J(x, m, k) C(n-y-x, y+1-m, k-1).$$

For $x = 0$ and $m = 0$, the $n-y$ Ss are placed in the $y+1$ distinguishable cells with no cell receiving more than $k-1$ Ss in $C(n-y, y+1, k-1)$ ways. Hence, $N_{(S_{n,k}, G_{n,k})}(0, 0, y) = C(n-y, y+1, k-1)$. The result then follows from Proposition 2.1.

The next result is a direct consequence of Theorem 4.1.

Theorem 4.2. *The PDF of $S_{n,k}$ is given as follows:*

(a)

$$\mathrm{P}(S_{n,k} = 0) = \sum_{y=[n/k]}^{n} p_n(y) C(n-y, y+1, k-1);$$

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

(b) *for* $x = k, k+1, \ldots, n$,

$$\mathrm{P}(S_{n,k} = x) = \sum_{y=0}^{n-x} p_n(y) \sum_{m=1}^{[x/k]} \binom{y+1}{m} \binom{x-(k-1)m-1}{m-1} C(n-y-x, y+1-m, k-1). \tag{4.2}$$

For $s = 0$ and $k = 1$, (4.2) reduces to the ordinary binomial distribution with PDF

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

For $s = 0$ and $k > 1$, (4.2) gives a new exact formula for the PDF of $S_{n,k}$. For $s \neq 0$, Theorem 4.2 provides new distributions.

Corollary 4.1. *The conditional PDF* $\mathrm{P}(S_{n,k} = x \mid S_{n,1} = n-y)$ *for* $0 \le y \le n$ *is given as follows:*

(a)

$$\mathrm{P}(S_{n,k} = 0 \mid S_{n,1} = n-y) = \binom{n}{y}^{-1} \sum_{j=0}^{[(n-y)/k]} (-1)^j \binom{y+1}{j} \binom{n-kj}{n-kj-y};$$

(b) *for* $x = k, k+1, \ldots, n-y$,

$$\mathrm{P}(S_{n,k} = x \mid S_{n,1} = n-y) = \binom{n}{y}^{-1} \sum_{m=1}^{[x/k]} \binom{y+1}{m} \binom{x-(k-1)m-1}{m-1} \times \sum_{j=0}^{[(n-y-x)/k]} (-1)^j \binom{y+1-m}{j} \binom{n-kj-x-m}{n-kj-x-y}. \tag{4.3}$$

Proof. It follows directly from Proposition 2.1 and Theorem 4.2.

Antzoulakos *et al.* (2003a) have also obtained (in their Theorem 4.2) the conditional PDF of $S_{n,k}$, given $S_{n,1}$, and they used it in nonparametric tests of randomness. Their expressions, in comparison to (4.3), involve two additional consecutive summations of binomial coefficients. Thus, the formulae of Corollary 4.1 may be evaluated faster computationally.

Proposition 4.1. *Let* $\mu_{S_{n,k}}$ *denote the mean of* $S_{n,k}$. *Then,*

(a) *for* $s = 0$,

$$\mu_{S_{n,k}} = p^k (k + (n-k)(kq+p)), \qquad n \ge k; \tag{4.4}$$

(b) *for* $s > 0$,

$$\mu_{S_{n,k}} = \frac{1}{\mathrm{B}(\alpha, \beta)} (k\mathrm{B}(\alpha+k, \beta) + k(n-k)\mathrm{B}(\alpha+k, \beta+1) + (n-k)\mathrm{B}(\alpha+k+1, \beta)).$$

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Proof. (a) Let Z_i, $1 \le i \le n$, be as given in the proof of Proposition 3.1. We define the RV U_j, $k \le j \le n$, as follows:

$$U_j = \begin{cases} 1 & \text{if } \prod_{i=j-k}^{j} Z_i = 1, \\ k & \text{if } Z_{j-k} = 0,\ \prod_{i=j-k+1}^{j} Z_i = 1, \\ 0 & \text{otherwise} \end{cases}$$

(convention: $Z_0 = 0$). Then, $S_{n,k} = \sum_{j=k}^{n} U_j$, so that

$$\mu_{S_{n,k}} = \mathrm{E}(U_k) + \sum_{j=k+1}^{n} \mathrm{E}(U_j) = kp^k + (n-k)(p^{k+1} + kqp^k).$$

(b) It follows from part (a) and Proposition 2.2.

Antzoulakos *et al.* (2003a) derived (4.4) using a different approach based on recursive relations.

Theorem 4.3. *Let* $S_{n,k,w,b,s}$ *denote (explicitly) the statistic* $S_{n,k}$ *defined on a* PE(w, b, s) *sampling scheme. Then its PDF satisfies the following relations.*

(a) *For* $x < 0$ *or* $x > n$, $\mathrm{P}(S_{n,k,w,b,s} = x) = 0$.

(b) *For* $0 \le n < k$, $\mathrm{P}(S_{n,k,w,b,s} = 0) = 1$ *and* $\mathrm{P}(S_{n,k,w,b,s} = x) = 0$, $x \ge 1$.

(c) *For* $n \ge k$,

$$\mathrm{P}(S_{n,k,w,b,s} = x)$$
$$= \begin{cases} \displaystyle\sum_{i=0}^{k-1} p_{i+1,w,b,s}(1)\,\mathrm{P}(S_{n-1-i,k,w+is,b+s,s} = 0) & \textit{for } x = 0, \\ \displaystyle\sum_{i=0}^{k-1} p_{i+1,w,b,s}(1)\,\mathrm{P}(S_{n-1-i,k,w+is,b+s,s} = x) & \\ \quad \displaystyle + \sum_{i=k}^{x-k} p_{i+1,w,b,s}(1)\,\mathrm{P}(S_{n-1-i,k,w+is,b+s,s} = x - i) & \\ \quad + p_{x+1,w,b,s}(1)\,\mathrm{P}(S_{n-1-x,k,w+xs,b+s,s} = 0) & \textit{for } x = k, \dots, n-1, \\ p_{n,w,b,s}(0) & \textit{for } x = n, \end{cases}$$

where $p_{n,w,b,s}(y) \equiv p_n(y)$ *is given by (2.1).*

Proof. Obviously, for $x < 0$ or $x > n$ and for $0 \le n < k$, parts (a) and (b) hold. Given a PE(w, b, s) sampling scheme let Z_i, $1 \le i \le n$, be a binary sequence such that $Z_i = 1$ if the outcome of the ith draw is a success and $Z_i = 0$ otherwise. For $n \ge k$, we define the events $A_0 = \{Z_1 = 0\}$, $A_i = \{Z_1 = \cdots = Z_i = 1,\ Z_{i+1} = 0\}$, $i = 1, \dots, n-1$, and $A_n = \{Z_1 = \cdots = Z_n = 1\}$. From the definitions of A_i and the Pólya–Eggenberger sampling scheme, we see that $\mathrm{P}(A_i) = p_{i+1,w,b,s}(1)$, $i = 0, 1, \dots, n-1$, and $\mathrm{P}(A_n) = p_{n,w,b,s}(0)$. Then, for $x = 0, k, k+1, \dots, n$,

$$\mathrm{P}(S_{n,k,w,b,s} = x) = \mathrm{P}\bigg(\bigcup_{i=0}^{n} [(S_{n,k,w,b,s} = x) \cap A_i]\bigg) = \sum_{i=0}^{n} \mathrm{P}(S_{n,k,w,b,s} = x \mid A_i)\,\mathrm{P}(A_i).$$

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

We first observe that, for $i = 0, 1, \dots, k-1$,

$$\mathrm{P}(S_{n,k,w,b,s} = 0 \mid A_i) = \mathrm{P}(S_{n-1-i,k,w+is,b+s,s} = 0)$$

and that, for $i = k, k+1, \dots, n$, $\mathrm{P}(S_{n,k,w,b,s} = 0 \mid A_i) = 0$. Also, we note that, for $x = k, k+1, \dots, n-1$,

$$\mathrm{P}(S_{n,k,w,b,s} = x \mid A_i) = \mathrm{P}(S_{n-1-i,k,w+is,b+s,s} = x)$$

for $i = 0, 1, \dots, k-1$;

$$\mathrm{P}(S_{n,k,w,b,s} = x \mid A_i) = \mathrm{P}(S_{n-i-1,k,w+is,b+s,s} = x - i)$$

for $i = k, k+1, \dots, x-k$; $\mathrm{P}(S_{n,k,w,b,s} = x \mid A_i) = 0$ for $i = x-k+1, \dots, x-1$ or $i = x+1, \dots, n$; and

$$\mathrm{P}(S_{n,k,w,b,s} = x \mid A_x) = \mathrm{P}(S_{n-x-1,k,w+xs,b+s,s} = 0).$$

Moreover,

$$\mathrm{P}(S_{n,k,w,b,s} = n) = \mathrm{P}(A_n) = p_{n,w,b,s}(0).$$

The result then follows.

Theorem 4.3 gives a new recursive scheme for the PDF of $S_{n,k}$. Numerical investigations indicate that, for $s = 0$, this scheme may be more efficient than the corresponding one given by Theorem 3.2 of Antzoulakos *et al.* (2003a).

Proposition 4.2, below, relates the probability $\mathrm{P}(G_{n,k} = 0) = \mathrm{P}(S_{n,k} = 0)$ to the cumulative distribution function of L_n, the tail probabilities of W_k, and the reliability of a linear consecutive system. In Section 4.2 the PDF of W_k is obtained using a different approach.

Proposition 4.2. *Let L_n and W_k respectively denote the length of the longest success run and the number of draws according to a* $\mathrm{PE}(w, b, s)$ *scheme until the occurrence of the first success run of length k. Then,*

$$\begin{aligned}\mathrm{P}(L_n < k) &= \mathrm{P}(W_k > n)\\ &= \mathrm{P}(G_{n,k} = 0)\\ &= \mathrm{P}(S_{n,k} = 0)\\ &= \sum_{y=0}^{n} p_n(y) \sum_{i=0}^{[(n-y)/k]} (-1)^i \binom{y+1}{i}\binom{n-ki}{n-ki-y}.\end{aligned} \tag{4.5}$$

Taking $s = 0$ we obtain a formula for the reliability $R(k, n; 1-p) \equiv \mathrm{P}(L_n < k)$ of a linear consecutive-k-out-of $n{:}F$ system with common component reliability $1 - p = b/(w+b)$. For alternative formulae for the distributions of L_n and W_k, see Makri *et al.* (2007a), (2007b).

Next, by conditioning on $S_{n,1}$ we obtain

$$\mathrm{P}(L_n < k \mid S_{n,1} = n - y) = \binom{n}{y}^{-1} \sum_{i=0}^{[(n-y)/k]} (-1)^i \binom{y+1}{i}\binom{n-ki}{n-ki-y}. \tag{4.6}$$

Equation (4.6) is equivalent to Equation (2.59) of Balakrishnan and Koutras (2002), which was derived by Burr and Cane (1961).

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

4.1.2. *Trials arranged on a circle.* In the following we assume that the outcomes of the n draws, according to a $\mathrm{PE}(w, b, s)$ sampling scheme, are arranged on a circle. First we derive the exact joint PDF of the RVs $S_{n,k}^{\mathrm{c}}$ and $G_{n,k}^{\mathrm{c}}$ (Theorem 4.4), and then we obtain the PDF and the conditional PDF, given $S_{n,1}$, of $S_{n,k}^{\mathrm{c}}$ (Theorem 4.5 and Corollary 4.2). Next, we consider the mean of $S_{n,k}^{\mathrm{c}}$ for $s \geq 0$ (Proposition 4.3) and, for $s = 0$, we relate the PDF of $S_{n,k}^{\mathrm{c}}$ to the PDF of $S_{n,k}$ (Theorem 4.6). Finally, Proposition 4.4 relates the specific probabilities of $G_{n,k}^{\mathrm{c}}$, $S_{n,k}^{\mathrm{c}}$, and L_n^{c}. For $s = 0$, we obtain a formula for the reliability of a circular consecutive system.

Theorem 4.4. *The exact joint PDF of the RVs $S_{n,k}^{\mathrm{c}}$ and $G_{n,k}^{\mathrm{c}}$ is given as follows:*

(a)

$$\mathrm{P}(S_{n,k}^{\mathrm{c}} = 0,\ G_{n,k}^{\mathrm{c}} = 0) = \sum_{y=[n/k]}^{n} p_n(y)\frac{n}{y}C(n-y, y, k-1);$$

(b) *for $x = k, k+1, \dots, n-1$ and $m = 1, 2, \dots, [x/k]$,*

$$\begin{aligned}\mathrm{P}(S_{n,k}^{\mathrm{c}} = x,\ G_{n,k}^{\mathrm{c}} = m)\\ &= \binom{x-(k-1)m-1}{m-1}\sum_{y=1}^{n-x} p_n(y)\frac{n}{y}\binom{y}{m}C(n-y-x, y-m, k-1); \qquad (4.7)\end{aligned}$$

(c) $\mathrm{P}(S_{n,k}^{\mathrm{c}} = n, G_{n,k}^{\mathrm{c}} = 1) = p_n(0)$.

Proof. For $1 \leq m \leq [x/k]$, $x = k, k+1, \dots, n$, m cells can be chosen from the y ones in $\binom{y}{m}$ ways. For each of these combinations, the number of ways x Ss are placed in the m cells so that each cell receives at least k Ss is equal to $J(x, m, k)$ by Lemma 2.3. The number of ways the remaining $n-y-x$ Ss can be placed in the remaining $y-m$ cells with no cell receiving more than $k-1$ Ss is equal to $C(n-y-x, y-m, k-1)$ by Corollary 2.2. Continuing along the lines of the proof of Theorem 3.2, we obtain

$$N_{(S_{n,k}^{\mathrm{c}}, G_{n,k}^{\mathrm{c}})}(x, m, y) = \frac{n}{y}\binom{y}{m}J(x, m, k)C(n-y-x, y-m, k-1).$$

For $x = 0$ and $m = 0$, the $n-y$ Ss are placed in the y distinguishable cells so that each cell receives no more than $k-1$ Ss in $C(n-y; y; k-1)$ ways. Hence,

$$N_{(S_{n,k}, G_{n,k})}(0, 0, y) = \frac{n}{y}C(n-y; y; k-1).$$

Finally, observing that $N_{(S_{n,k}, G_{n,k})}(n, 1, 0) = 1$, the proof is completed by means of Proposition 2.1.

We next note the following direct consequence of Theorem 4.4.

Theorem 4.5. *The PDF of the RV $S_{n,k}^{\mathrm{c}}$ is given as follows:*

(a)

$$\mathrm{P}(S_{n,k}^{\mathrm{c}} = 0) = \sum_{y=[n/k]}^{n} p_n(y)\frac{n}{y}C(n-y, y, k-1);$$

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

(b) *for* $x = k, k+1, \dots, n-1$,

$$\mathrm{P}(S^{\mathrm{c}}_{n,k} = x) = \sum_{y=1}^{n-x} p_n(y) \frac{n}{y} \sum_{m=1}^{[x/k]} \binom{y}{m} \binom{x-(k-1)m-1}{m-1} C(n-y-x, y-m, k-1); \tag{4.8}$$

(c) $\mathrm{P}(S^{\mathrm{c}}_{n,k} = n) = p_n(0)$.

For $s = 0$ and $k = 1$, (4.8) reduces to the ordinary binomial distribution. Under a $\mathrm{PE}(w, b, s)$ sampling scheme, for $(s, k) \neq (0, 1)$, Theorem 4.5 provides new distributions.

Corollary 4.2. *The conditional PDF* $\mathrm{P}(S^{\mathrm{c}}_{n,k} = x \mid S^{\mathrm{c}}_{n,1} = n - y)$ *for* $n \geq k$ *is given as follows:*

(a)

$$\mathrm{P}(S^{\mathrm{c}}_{n,k} = x \mid S^{\mathrm{c}}_{n,1} = n) = \delta_{x,n}, \qquad x \geq 0;$$

(b) *for* $y = 1, 2, \dots, n$,

$$\mathrm{P}(S^{\mathrm{c}}_{n,k} = x \mid S^{\mathrm{c}}_{n,1} = n - y) = \begin{cases} \binom{n-1}{y-1}^{-1} \sum_{j=0}^{[(n-y)/k]} (-1)^j \binom{y}{j} \binom{n-kj-1}{n-kj-y}, & x = 0, \\ \binom{n-1}{y-1}^{-1} \sum_{m=1}^{[x/k]} \binom{y}{m} \binom{x-(k-1)m-1}{m-1} & \\ \times \sum_{j=0}^{[(n-y-x)/k]} (-1)^j \binom{y-m}{j} \binom{n-x-kj-m-1}{n-x-kj-y}, & \\ & x = k, k+1, \dots, n-y. \end{cases} \tag{4.9}$$

Proof. It follows directly from Proposition 2.1 and Theorem 4.5.

Proposition 4.3. *Let* $\mu_{S^{\mathrm{c}}_{n,k}}$ *denote the mean of* $S^{\mathrm{c}}_{n,k}$. *Then,*

(a) *for* $s = 0$,

$$\mu_{S^{\mathrm{c}}_{n,k}} = np^k(kq + p), \qquad n \geq k;$$

(b) *for* $s > 0$, $\alpha = w/s$, *and* $\beta = b/s$,

$$\mu_{S^{\mathrm{c}}_{n,k}} = \frac{nk\mathrm{B}(\alpha + k, \beta + 1) + n\mathrm{B}(\alpha + k + 1, \beta)}{\mathrm{B}(\alpha, \beta)}, \qquad n \geq k.$$

Proof. (a) Let Z_i, $1 \leq i \leq n$, be as given in Proposition 3.1. We define an RV U_j, $1 \leq j \leq n$, as follows. For $j = 1, \dots, k$,

$$U_j = \begin{cases} 1 & \text{if } \prod_{i=1}^{j} Z_i \prod_{i=n-k+j}^{n} Z_i = 1, \\ k & \text{if } \prod_{i=1}^{j} Z_i \prod_{i=n-k+j+1}^{n} Z_i = 1,\ Z_{n-k+j} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

For $j = k+1, \dots, n$,

$$U_j = \begin{cases} 1 & \text{if } \prod_{i=j-k}^{j} Z_i = 1, \\ k & \text{if } \prod_{i=j-k+1}^{j} Z_i = 1,\ Z_{j-k} = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $S^{\mathrm{c}}_{n,k} = \sum_{j=1}^{n} U_j$, so that

$$\mu_{S^{\mathrm{c}}_{n,k}} = \sum_{j=1}^{n} \mathrm{E}(U_j) = nkqp^k + np^{k+1}.$$

(b) It follows from part (a) and Proposition 2.2.

For $s = 0$, the RV $S^{\mathrm{c}}_{n,k}$ is related to the RV $S_{n,k}$ by the following theorem.

Theorem 4.6. *For* $s = 0$, $n \geq k$, *and* $0 < p < 1$,

(a)

$$\mathrm{P}(S^{\mathrm{c}}_{n,k} = 0) = q^2 \sum_{i=0}^{k-1} (i+1) p^i \, \mathrm{P}(S_{n-2-i,k} = 0);$$

(b) *for* $x = k, k+1, \dots, n-2$,

$$\mathrm{P}(S^{\mathrm{c}}_{n,k} = x) = q^2 \sum_{i=0}^{k-1} (i+1) p^i \, \mathrm{P}(S_{n-2-i,k} = x) + q^2 \sum_{i=k}^{x} (i+1) p^i \, \mathrm{P}(S_{n-2-i,k} = x - i);$$

(c)

$$\mathrm{P}(S^{\mathrm{c}}_{n,k} = n-1) = \begin{cases} nqp^{n-1} & \textit{if } n \geq k+1, \\ 0 & \textit{otherwise;} \end{cases}$$

(d) $\mathrm{P}(S^{\mathrm{c}}_{n,k} = n) = p^n$.

Proof. The proofs of parts (a), (c), and (d) are straightforward. Thus, we proceed to prove part (b). Consider the sequence $Z_1, Z_2, \dots, Z_n$, as given in the proof of Proposition 3.1. We consider the events $A_0 = \{Z_1 = 0\}$, $B_0 = \{Z_n = 0\}$, $A_j = \{Z_1 = \cdots = Z_j = 1,\ Z_{j+1} = 0\}$, $B_l = \{Z_{n-l} = 0,\ Z_{n-l+1} = \cdots = Z_n = 1\}$, $j, l = 1, \dots, n-2$. Then, for $x = k, k+1, \dots,$ we find that

$$\{S^{\mathrm{c}}_{n,k} = x\} = \bigcup_j \bigcup_l [\{S^{\mathrm{c}}_{n,k} = x\} \cap A_j \cap B_l],$$

where the double union is over the integers j, l satisfying the conditions $j, l = 0, \dots, x$, $0 \leq j + l \leq x$. So, we obtain

$$\begin{aligned} \mathrm{P}(S^{\mathrm{c}}_{n,k} = x) &= \sum_j \sum_l \mathrm{P}(S^{\mathrm{c}}_{n,k} = x \mid A_j \cap B_l) \, \mathrm{P}(A_j \cap B_l) \\ &= \sum_{j+l=0}^{k-1} (j+l+1) q^2 p^{j+l} \, \mathrm{P}(S_{n-(j+l)-2,k} = x) \\ &\quad + \sum_{j+l=k}^{x} (j+l+1) q^2 p^{j+l} \, \mathrm{P}(S_{n-(j+l)-2,k} = x - (j+l)), \end{aligned}$$

and this establishes the result.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Proposition 4.4, below, relates the probability $\mathrm{P}(G^{\mathrm{c}}_{n,k} = 0) = \mathrm{P}(S^{\mathrm{c}}_{n,k} = 0)$ to the cumulative distribution function of L^{c}_n, the length of the longest (circular) success run, and the reliability of a circular consecutive system.

Proposition 4.4. *It holds true that*

$$\mathrm{P}(L^{\mathrm{c}}_n < k) = \mathrm{P}(G^{\mathrm{c}}_{n,k} = 0) = \mathrm{P}(S^{\mathrm{c}}_{n,k} = 0) = \sum_{y=1}^{n} p_n(y) \frac{n}{y} \sum_{i=0}^{[(n-y)/k]} (-1)^i \binom{y}{i} \binom{n-ki-1}{n-ki-y}. \tag{4.10}$$

Taking $s = 0$ we obtain a formula for the reliability $R^{\mathrm{c}}(k, n; 1-p) \equiv \mathrm{P}(L^{\mathrm{c}}_n < k)$ of a circular consecutive-k-out-of n:F system with common component reliability $1 - p = b/(w + b)$. For an alternative derivation of the distribution of L^{c}_n, see Makri *et al.* (2007b).

By conditioning on $S_{n,1}$ we obtain

$$\mathrm{P}(L^{\mathrm{c}}_n < k \mid S_{n,1} = n - y) = \binom{n-1}{y-1}^{-1} \sum_{i=0}^{[(n-y)/k]} (-1)^i \binom{y}{i} \binom{n-ki-1}{n-ki-y}, \tag{4.11}$$

which is equivalent to Equation (2.80) of Balakrishnan and Koutras (2002).

4.2. The waiting time associated with the sum of success run lengths

In this section we consider the statistic $T_{r,k}$ for $r \geq k > 0$. By a slight modification of Proposition 2.1 the exact PDF of $T_{r,k}$ is established in Theorem 4.7 for $s \geq 0$ and in Proposition 4.5 for $s = -1$. In the sequel, $\mathbf{1}_A(\cdot)$ stands for the indicator function of the set $A = \{1, 2, \dots, k - 1\}$.

Theorem 4.7. *Let $T_{r,k}$ ($s \geq 0$) denote the number of draws according to the Pólya–Eggenberger sampling scheme until the total number of successes in runs of length greater than or equal to k is equal to or exceeds the value r for the first time. Then,*

(a) $\mathrm{P}(T_{r,k} = r) = p_r(0)$;

(b) *for* $n \geq r + 1$,

$$\mathrm{P}(T_{r,k} = n) = \sum_{\alpha=0}^{k-1} \sum_{y=1}^{n-r} p_n(y) \sum_{i=0}^{y} \binom{y}{i} \binom{r - k - (k-1)i + \alpha - \mathbf{1}_A(\alpha)}{r - (i+1)k + \alpha} \times C(n - y - r - \alpha, y - i, k - 1). \tag{4.12}$$

Proof. Obviously part (a) holds. We proceed to prove part (b). Set $B_\alpha = \{$the total number of successes in runs of length greater than or equal to k in n trials is equal to $r + \alpha\}$, $\alpha = 0, 1, \dots, k - 1$, and denote by Y_n the total number of failures in the sequence of n draws. Then, a typical element of the event $\{T_{r,k} = n\} \cap B_\alpha \cap \{Y_n = y\}$ is a permutation of $n - y$ Ss and y Fs such that a success run of length at least k (for $\alpha = 0$) or exactly equal to k (for $\alpha = 1, 2, \dots, k - 1$) follows the last F, and the total number of successes in runs of length greater than or equal to k is equal to $r + \alpha$. The calculation of the number $N(n, \alpha, y)$ of these permutations can be carried out as follows. Initially i cells, $U_{j_1}, \dots, U_{j_i}$, are chosen from the cells $U_1, \dots, U_y$ in $\binom{y}{i}$ ways, $i = 0, 1, \dots, y$. Fix i and $j_1, \dots, j_i$. We note that any allocation of $n - y$ Ss in $y + 1$ cells under the restrictions of the problem is carried out in two stages for $\alpha = 0$ and in three consecutive stages for $\alpha = 1, 2, \dots, k - 1$. Namely, for $\alpha = 0$, at the first

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

stage r Ss are placed in the cells $U_{j_1}, \dots, U_{j_i}, U_{y+1}$, with each cell receiving at least k Ss in $J(r, i+1, k)$ ways by Lemma 2.3 and, for $\alpha = 1, 2, \dots, k-1$, k Ss are placed in the cell U_{y+1}, and in the sequel $r + \alpha - k$ Ss are allocated to the cells $U_{j_1}, \dots, U_{j_i}$, with each cell receiving at least k Ss in $J(r + \alpha - k, i, k)$ ways. Next, the remaining $n - y - r - \alpha$ Ss are placed in the remaining $y - i$ cells with no cell receiving more than $k - 1$ Ss in $C(n - y - r - \alpha, y - i, k - 1)$ ways by Corollary 2.2. Summing with respect to i and applying the multiplicative principle, we find that

$$N(n, 0, y) = \sum_{i=0}^{y} \binom{y}{i} J(r, i+1, k) C(n - y - r, y - i, k - 1)$$

and, for $\alpha = 1, 2, \dots, k-1$, we find that

$$N(n, \alpha, y) = \sum_{i=0}^{y} \binom{y}{i} J(r + \alpha - k, i, k) C(n - y - r - \alpha, y - i, k - 1).$$

Each of these permutations has probability $p_n(y)$, and hence

$$\mathrm{P}(\{T_{r,k} = n\} \cap B_\alpha \cap \{Y_n = y\}) = N(r, \alpha, y) p_n(y).$$

Finally, summing with respect to α and y, $y = 1, \dots, n - r$, the probability $\mathrm{P}(T_{r,k} = n)$, $n \geq r + 1$, follows.

For $s = 0$ and $k = 1$, (4.12) reduces to the ordinary negative binomial distribution with PDF

$$f(n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \qquad n = r, r+1, \dots.$$

For $s > 0$ and $r > k$, Theorem 4.7 provides new distributions. For $s = 0$ and $r \geq k$, it gives new exact formulae for the PDF of $T_{r,k}$ which, for $r = k$, becomes a new expression of the geometric distribution of order k (Philippou *et al.* (1983)) as $W_k = T_{k,k}$. A recursive evaluation of the PDF of $T_{r,k}$ for $s = 0$ and $r \geq k$ was given by Antzoulakos *et al.* (2003b). Makri *et al.* (2007a) provided an alternative formula for the PDF of $T_{k,k}$ for $s \geq 0$.

The case in which $s = -1$ requires special attention because the random variable $T_{r,k}$ might take on the value ∞ with positive probability. We give in the sequel its probability distribution.

Proposition 4.5. *Let $T_{r,k}$ be the balls drawn from an urn containing w white balls and b black balls, without replacement ($s = -1$), until the total number of successes in runs of length greater than or equal to k is equal to or exceeds the value r for the first time. Then,*

(a) $\mathrm{P}(T_{r,k} = r) = w^{(r)}/(w+b)^{(r)}$, $r \leq w$;

(b) *for* $n = r+1, \dots, w+b$,

$$\mathrm{P}(T_{r,k} = n) = \sum_{\alpha=0}^{k-1} \sum_{y=1}^{n-r} \frac{w^{(n-y)} b^{(y)}}{(w+b)^{(n)}} \sum_{i=0}^{y} \binom{y}{i} \binom{r - k - (k-1)i + \alpha - \mathbf{1}_A(\alpha)}{r - (i+1)k + \alpha} \times C(n - y - r - \alpha, y - i, k - 1);$$

(c) $\mathrm{P}(T_{r,k} = \infty) = \sum_{x=0}^{r-1} \mathrm{P}(S_{w+b,k,w,b,-1} = x)$, *where* $S_{w+b,k,w,b,-1} \equiv S_{w+b,k}$ *for* $s = -1$.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Proof. It is straightforward.

For $r > k$, Proposition 4.5 provides a new distribution. For $r = k$, it gives new formulae for the PDF of $W_k = T_{k,k}$, which are alternative to the corresponding ones given by Panaretos and Xekalaki (1986), Ling (1988), Godbole (1990), and Makri *et al.* (2007a).

5. Numerical examples

In this section we illustrate the distributions derived in Sections 3 and 4 by means of numerical examples. In Example 5.1 we calculate the exact joint and marginal PDFs of the RVs $S_{n,k}$, $G_{n,k}$ and $S^{\mathrm{c}}_{n,k}$, $G^{\mathrm{c}}_{n,k}$. In Examples 5.2 and 5.3 we provide the means and variances of the RVs $G_{n,k}$, $G^{\mathrm{c}}_{n,k}$, $S_{n,k}$, $S^{\mathrm{c}}_{n,k}$, and $T_{r,k}$. Finally, in Example 5.4 we give some critical values of the conditional distributions of the RVs $G_{n,k}$, $S_{n,k}$, L_n, and $G^{\mathrm{c}}_{n,k}$, $S^{\mathrm{c}}_{n,k}$, L^{c}_n, given the number of successes $S_{n,1}$ in a linear and a circular binary sequence, respectively.

Example 5.1. For $s = 0$ and $k = 2$, we calculate the exact joint and marginal distributions of $S_{n,k}$, $G_{n,k}$ (linear case) and $S^{\mathrm{c}}_{n,k}$, $G^{\mathrm{c}}_{n,k}$ (circular case) as functions of p and q. The value $n = 5$ was chosen so that the required computations can also be carried out by hand and, thus, it is possible to gain insight into the formulae presented in the theorems.

Linear case: Let

$$p_{i,j} = \mathrm{P}(\boldsymbol{M}_{5,2} = (i, j)) = \mathrm{P}(S_{5,2} = i, G_{5,2} = j), \qquad i \in \{0, 2, 3, 4, 5\},\ j \in \{0, 1, 2\}.$$

Then we have

$$\begin{aligned} p_{0,0} &= p^3q^2 + 6p^2q^3 + 5pq^4 + q^5, & p_{2,1} &= 6p^3q^2 + 4p^2q^3, \\ p_{3,1} &= 2p^4q + 3p^3q^2, & p_{4,1} &= 2p^4q, \\ p_{4,2} &= p^4q, & p_{5,1} &= p^5, \end{aligned}$$

and $p_{i,j} = 0$ otherwise, by (4.1). Let

$$f_i = \mathrm{P}(S_{5,2} = i), \qquad i = 0, 2, 3, 4, 5,$$

and

$$g_j = \mathrm{P}(G_{5,2} = j), \qquad j = 0, 1, 2.$$

Then we have

(a)

$$\begin{aligned} f_0 &= p^3q^2 + 6p^2q^3 + 5pq^4 + q^5, & f_2 &= 6p^3q^2 + 4p^2q^3, \\ f_3 &= 2p^4q + 3p^3q^2, & f_4 &= 3p^4q, \\ f_5 &= p^5, \end{aligned}$$

by (4.2);

(b)

$$\begin{aligned} g_0 &= p^3q^2 + 6p^2q^3 + 5pq^4 + q^5, & g_1 &= p^5 + 4p^4q + 9p^3q^2 + 4p^2q^3, \\ g_2 &= p^4q, \end{aligned}$$

by (3.5).

The marginal distributions f_i and g_j can also be derived using the joint distribution $p_{i,j}$.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Circular case: Let

$$p_{i,j} = \mathrm{P}(\boldsymbol{M}^{\mathrm{c}}_{5,2} = (i, j)) = \mathrm{P}(S^{\mathrm{c}}_{5,2} = i, G^{\mathrm{c}}_{5,2} = j), \qquad i \in \{0, 2, 3, 4, 5\},\ j \in \{0, 1\}.$$

Then we have

$$\begin{aligned} p_{0,0} &= 5p^2q^3 + 5pq^4 + q^5, & p_{2,1} &= 5p^3q^2 + 5p^2q^3, \\ p_{3,1} &= 5p^3q^2, & p_{4,1} &= 5p^4q, \\ p_{5,1} &= p^5, \end{aligned}$$

and $p_{i,j} = 0$ otherwise, by (4.7). Let

$$f_i = \mathrm{P}(S^{\mathrm{c}}_{5,2} = i), \quad i = 0, 2, 3, 4, 5, \qquad \text{and} \qquad g_j = \mathrm{P}(G^{\mathrm{c}}_{5,2} = j), \quad j = 0, 1.$$

Then we have

(a)

$$\begin{aligned} f_0 &= 5p^2q^3 + 5pq^4 + q^5, & f_2 &= 5p^3q^2 + 5p^2q^3, \\ f_3 &= 5p^3q^2, & f_4 &= 5p^4q, \\ f_5 &= p^5, \end{aligned}$$

by (4.8);

TABLE 1: The means and variances of $G_{n,k}$ and $S_{n,k}$.

s	w	b	n	k	$\mathrm{E}(G_{n,k})$	$V(G_{n,k})$	$\mathrm{E}(S_{n,k})$	$V(S_{n,k})$
−1	5	5	5	1	1.61	0.32	2.50	0.69
				2*	0.64	0.27	1.53	1.73
			10	1	3.00	0.67	5.00	0.00
				3*	0.50	0.25	1.67	2.94
	9	1	10	1	1.80	0.16	9.00	0.00
				7*	0.60	0.24	4.80	15.76
0	5	5	10	1	2.75	0.69	5.00	2.50
				3*	0.56	0.36	2.13	5.52
			100	3	6.19	3.52	24.63	61.77
				6*	0.75	0.68	5.23	34.24
	9	1	10	1	1.71	0.54	9.00	0.90
				7*	0.62	0.24	5.64	20.23
			100	3	7.80	3.41	87.04	23.50
				27*	0.48	0.35	16.85	448.76
1	5	5	10	1	2.55	0.82	5.00	4.55
				3*	0.59	0.40	2.45	7.80
			100	7*	0.82	1.65	7.62	171.33
	9	1	100	39*	0.42	0.31	27.62	1396.26
5	5	5	10	1	2.00	1.20	5.00	10.00
				3*	0.60	0.43	3.20	13.44
			100	11*	0.65	1.38	14.04	739.03

TABLE 2: The means and variances of $G^{\text{c}}_{n,k}$ and $S^{\text{c}}_{n,k}$.

s	w	b	n	k	$\text{E}(G^{\text{c}}_{n,k})$	$V(G^{\text{c}}_{n,k})$	$\text{E}(S^{\text{c}}_{n,k})$	$V(S^{\text{c}}_{n,k})$
−1	5	5	5	1	1.39	0.25	2.50	0.69
				2*	0.70	0.21	1.81	1.80
			10	1	2.78	0.62	5.00	0.00
				3*	0.60	0.24	2.02	3.01
	9	1	10	1	1.00	0.00	9.00	0.00
				9*	1.00	0.00	9.00	0.00
0	5	5	10	1	2.50	0.62	5.00	2.50
				3*	0.63	0.35	2.50	6.25
			100	3	6.25	3.52	25.00	62.50
				6*	0.78	0.70	5.47	35.78
	9	1	10	1	1.25	0.26	9.00	0.90
				8*	0.78	0.17	7.32	15.45
			100	3	7.29	3.57	87.48	22.25
				29*	0.47	0.34	17.90	518.35
1	5	5	10	1	2.28	0.74	5.00	4.55
				3*	0.62	0.37	2.81	8.77
			100	7*	0.85	1.72	7.98	183.33
	9	1	100	44*	0.40	0.26	30.50	1649.38
5	5	5	10	1	1.76	1.00	5.00	10.00
				4*	0.42	0.26	3.00	14.87
			100	12*	0.56	1.10	13.74	754.54

(b)

$$g_0 = 5p^2q^3 + 5pq^4 + q^5, \quad g_1 = p^5 + 5p^4q + 10p^3q^2 + 5p^2q^3,$$

by (3.9).

The marginal distributions f_i and g_j can also be derived using the joint distribution $p_{i,j}$.

Example 5.2. The means and variances of $G_{n,k}$, $S_{n,k}$, and $G^{\text{c}}_{n,k}$, $S^{\text{c}}_{n,k}$, using several sampling schemes PE(w, b, s), are presented in Tables 1 and 2, respectively, for various values of n and k. The results of the tables show a variety of possible configurations and highlight similarities and discrepancies between the means and variances of linear and circular sequences.

The selected values of the parameters s, w, b, n, and k illustrate sampling schemes of special interest; commonly used initial urn contents, especially for the case in which $s = 0$; values of n ranging from small to large; and values of k of special meaning. For instance, the value $k = 1$ is related to the number of success runs—via $G_{n,1}$ $(G^{\text{c}}_{n,1})$—as well as to the number of successes in the sequence—via $S_{n,1} \equiv S^{\text{c}}_{n,1}$. The values of k with stars (*) equal the nearest integer to the mean length of the longest success run, a characteristic number of consecutive successes in every linear or circular sequence. To obtain the results of Tables 1 and 2 we have used (3.5), (3.9), (4.2), (4.5), (4.8), and (4.10).

Example 5.3. Here we calculate the mean and variance of the RV $T_{r,k}$ for $t = \min\{n : \sum_{x=r}^{n} \text{P}(T_{r,k} = x) \geq 0.995\}$ or for $t = 125$ trials, whichever stopping time t comes first.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

TABLE 3: The stopping times, cumulative sums, means, and variances of $T_{r,k}$.

s	w	b	k	r	t	$\sum_{x=r}^{t} \mathrm{P}(T_{r,k} = x)$	$\mathrm{E}(T_{r,k})$	$V(T_{r,k})$
0	5	5	2	2	26	0.9953	5.85	18.87
				4	38	0.9958	11.82	39.63
	9	1	1	1	3	0.9990	1.11	0.11
				2	4	0.9963	2.20	0.22
			2	2	6	0.9962	2.32	0.58
				4	10	0.9978	4.67	1.27
				8	15	0.9969	9.13	1.94
1	5	5	1	1	13	0.9952	2.16	3.22
				2	21	0.9954	4.37	8.14
			2	2	79	0.9950	7.55	78.12
				4	125	0.9940	15.19	217.91
	9	1	1	1	3	0.9955	1.11	0.12
				2	5	0.9950	2.22	0.29
			2	2	9	0.9962	2.36	0.92
				4	14	0.9952	4.73	2.22
				8	24	0.9957	9.30	5.02
5	5	5	1	1	125	0.9921	4.42	100.91
				2	125	0.9841	7.84	177.82
			2	2	125	0.9170	10.68	346.93
				4	125	0.8735	15.83	454.89

Table 3 presents t, $\sum_{x=r}^{t} \mathrm{P}(T_{r,k} = x)$, $\mathrm{E}(T_{r,k})$, and $V(T_{r,k})$ for various PE(w, b, s) sampling schemes with $s \geq 0$ and several values of k and r. To obtain the results of Table 3 we have used (4.12).

Example 5.4. We illustrate the conditional distributions of the statistics $G_{n,k}$, $S_{n,k}$, and L_n, and $G^{\mathrm{c}}_{n,k}$, $S^{\mathrm{c}}_{n,k}$, and L^{c}_n, given the number of successes $S_{n,1} = n - y$ in binary sequences ordered linearly and circularly, respectively. Results similar to the ones presented here may be useful in nonparametric tests of randomness in which the null distributions are given by (3.6), (3.10), (4.3), (4.6), (4.9), and (4.11).

Let U_n denote either the RV L_n or the RV L^{c}_n, and let $V_{n,k}$ denote any of the RVs $G_{n,k}$, $G^{\mathrm{c}}_{n,k}$, $S_{n,k}$, $S^{\mathrm{c}}_{n,k}$. For $0 \leq y \leq n$, $1 \leq k \leq n - y$, and a given nominal probability α (the significant level), let u_β and v_γ be integers (critical values) in the support of the corresponding RV such that

$$\mathrm{P}(U_n \geq u_\beta \mid S_{n,1} = n - y) = \beta, \qquad \mathrm{P}(V_{n,k} \geq v_\gamma \mid S_{n,1} = n - y) = \gamma.$$

The probabilities β $(0 < \beta \leq \alpha < 1)$ and γ $(0 < \gamma \leq \alpha < 1)$, the exact α values, or the natural significant levels are the largest real numbers which do not exceed α. They may not be equal to the assigned nominal probability α, as they refer to discrete random variables. In Table 4 we give such upper-tailed critical values of u_β and v_γ for $\alpha = 0.05$ and $n = 10$ linearly and circularly ordered trials. The values of β and γ are shown in brackets. For the linear case, the fourth column refers to $G_{n,k}$ and the fifth column refers to $S_{n,k}$. For the circular case, the

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

TABLE 4: u_β and v_γ for $\alpha = 0.05$ and $n = 10$. The values of β and γ are shown in brackets.

$n - y$	u_β	k	v_γ	
		Linear case		
4	4 (0.0333)	1	–	–
		2	–	–
		3	–	4 (0.0333)
		4	1 (0.0333)	4 (0.0333)
		5	–	–
		6	–	–
		7	–	–
5	5 (0.0238)	1	5 (0.0238)	–
		2	–	–
		3	–	5 (0.0238)
		4	–	5 (0.0238)
		5	1 (0.0238)	5 (0.0238)
		6	–	–
		7	–	–
6	6 (0.0238)	1	5 (0.0238)	–
		2	3 (0.0476)	–
		3	2 (0.0476)	–
		4	–	6 (0.0238)
		5	–	6 (0.0238)
		6	1 (0.0238)	6 (0.0238)
		7	–	–
7	7 (0.0333)	1	–	–
		2	–	–
		3	–	–
		4	–	7 (0.0333)
		5	–	7 (0.0333)
		6	–	7 (0.0333)
		7	1 (0.0333)	7 (0.0333)
		Circular case		
4	4 (0.0476)	1	–	–
		2	–	–
		3	–	4 (0.0476)
		4	1 (0.0476)	4 (0.0476)
		5	–	–
		6	–	–
		7	–	–
5	5 (0.0397)	1	5 (0.0079)	–
		2	–	–
		3	–	5 (0.0397)
		4	–	5 (0.0397)
		5	1 (0.0397)	5 (0.0397)
		6	–	–
		7	–	–
6	6 (0.0476)	1	–	–
		2	3 (0.0476)	–
		3	–	–
		4	–	6 (0.0476)
		5	–	6 (0.0476)
		6	1 (0.0476)	6 (0.0476)
		7	–	–

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

fourth column refers to $G^{c}_{n,k}$ and the fifth column refers to $S^{c}_{n,k}$. Similar tables can be provided for larger values of n and for any α.

The data of Table 4 admits the following interpretation for a possible application in nonparametric tests of randomness. For instance, suppose that we want to test the null hypothesis of randomness, H_0, versus the alternative hypothesis of nonrandomness, H_1, for a linearly ordered binary sequence with $S_{10,1} = 5$. Then, by Table 4, at a significant level of at most $\alpha = 0.05$, the null hypothesis, H_0, is rejected when

(a) the length of the observed longest success run is at least 5 (because $u_\beta = 5$, with $\beta = 0.0238$);

(b) the number of success runs of length at least 1 observed is at least 5 ($v_\gamma = 5$, with $\gamma = 0.0238$), or the number of success runs of length at least 5 observed is (at least) equal to 1 ($v_\gamma = 1$, with $\gamma = 0.0238$);

(c) the sum of the lengths of all the success runs of length at least k (with $k = 3$, 4, or 5) observed is at least 5 ($v_\gamma = 5$, with $\gamma = 0.0238$ for $k = 3$, 4, or 5).

A similar interpretation holds for circularly ordered sequences. Hence, based on analogous arguments, we do not reject the null hypothesis of randomness, at a significant level of at most 0.05, of the sequence SSSFSFSSFS arranged on a line or on a circle (given in Section 2.1).

Acknowledgement

We wish to thank the anonymous referee for his/her insightful comment, which led to the addition of Remark 2.2.

References

Antzoulakos, D. L., Bersimis, S. and Koutras, M. V. (2003a). On the distribution of the total number of run lengths. *Ann. Inst. Statist. Math.* **55,** 865–884.

Antzoulakos, D. L., Bersimis, S. and Koutras, M. V. (2003b). Waiting times associated with the sum of success run lengths. In *Mathematical and Statistical Methods in Reliability*, eds B. Lindqvist and K. Doksum, World Scientific, River Edge, NJ, pp. 141–157.

Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. John Wiley, New York.

Barton, D. A. and David, F. N. (1958). Runs in a ring. *Biometrika* **45,** 572–578.

Burr, E. J. and Cane, G. (1961). Longest run of consecutive observations having a special attribute. *Biometrica* **48,** 461–465.

Charalambides, C. A. (2002). *Enumerative Combinatorics.* Chapman and Hall/CRC, Boca Raton, FL.

Cochran, W. G. (1938). An extension of Gold's method of examining the apparent persistence of one type of weather. *Quart. J. R. Meteor. Soc.* **64,** 631–634.

Eryilmaz, S. and Demir, S. (2007). Success runs in a sequence of exchangeable binary trials. *J. Statist. Planning Inference* **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. and Koutras, M. V. (1994). Distribution theory of runs: a Markov chain approach. *J. Amer. Statist. Assoc.* **89,** 1050–1058.

Fu, J. C. and Lou, W. Y. W. (2003). *Distribution Theory of Runs and Patterns and Its Applications: A Finite Markov Chain Imbedding Approach*. World Scientific, River Edge, NJ.

Fu, J. C., Lou, W. Y. W., Bai, Z. and Li, G. (2002). The exact and limiting distributions for the number of successes in success runs within a sequence of Markov-dependent two-state trials. *Ann. Inst. Statist. Math.* **54,** 719–730.

George, E. O. and Bowman, D. (1995). A full likelihood procedure for analysing exchangeable binary data. *Biometrics* **51,** 512–523.

Gibbons, J. D. and Chakraborti, S. (2003). *Nonparametric Statistical Inference,* 4th edn. Marcel Dekker, New York.

Godbole, A. P. (1990). On hypergeometric and related distributions of order k. *Commun. Statist. Theory Meth.* **21,** 1291–1301.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press

Goldstein, L. (1990). Poisson approximation in DNA sequence matching. *Commun. Statist. Theory Meth.* **19,** 4167–4179.

Hirano, K. and Aki, S. (1993). On number of occurrences of success runs of specified length in a two-state Markov chain. *Statistica Sinica* **3,** 313–320.

Johnson, N. and Kotz, S. (1977). *Urn models and Their Application.* John Wiley, New York.

Koutras, M. V. and Alexandrou, V. A. (1997). Non-parametric randomness tests based on success runs of fixed length. *Statist. Prob. Lett.* **32,** 393–404.

Koutras, M. V., Papadopoulos, G. K. and Papastavridis, S. G. (1995). Runs on a circle. *J. Appl. Prob.* **32,** 396–404.

Ling, K. D. (1988). On discrete distributions of order k defined on Pólya–Eggenberger urn model. *Soochow. J. Math.* **2,** 199–210.

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. (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., Philippou, A. N. and Psillakis, Z. M. (2007a). Pólya, inverse Pólya, and circular Pólya distributions of order k for l-overlapping success runs. *Commun. Statist. Theory Meth.* **36,** 657–668.

Makri, F. S., Philippou, A. N. and Psillakis, Z. M. (2007b). Shortest and longest length of success runs in binary sequences. *J. Statist. Planning Inference* **137,** 2226–2239.

Mood, A. M. (1940). The distribution theory of runs. *Ann. Math. Statist.* **11,** 367–392.

Móri, T. F. (1991). On the waiting time till each of some given patterns occurs as a run. *Prob. Theory Relat. Fields* **87,** 313–323.

Mosteller, F. (1941). Note on an application of runs to quality control charts. *Ann. Math. Statist.* **12,** 228–232.

Muselli, M. (1996). Simple expressions for success run distributions in Bernoulli trials. *Statist. Prob. Lett.* **31,** 121–128.

Panaretos, J. and Xekalaki, E. (1986). On some distributions arising from certain generalized sampling schemes. *Commun. Statist. Theory Meth.* **15,** 873–891.

Philippou, A. N. (1986). Distributions and Fibonacci polynomials of order k, longest runs, and reliability of consecutive-k-out-of-n:F systems. In *Fibonacci Numbers and Their Applications*, eds A. N. Philippou *et al.*, Reidel, Dordrecht, pp. 203–227.

Philippou, A. N., Georgiou, C. and Philippou, G. N. (1983). A generalized geometric distribution and some of its properties. *Statist. Prob. Lett.* **1,** 171–175.

Riordan, J. (1964). *An Introduction to Combinatorial Analysis*, 2nd edn. John Wiley, New York.

Schuster, E. F. (1991). Distribution theory of runs via exchangeable random variables. *Statist. Prob. Lett.* **11,** 379–386.

Schwager, S. J. (1983). Run probabilities in sequences of Markov-dependence trials. *J. Amer. Statist. Assoc.* **78,** 168–175.

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.

Sen, K., Agarwal, M. and Chakraborty, S. (2002). Lengths of runs and waiting time distributions by using Pólya–Eggenberger sampling scheme. *Studia Sci. Math. Hung.* **39,** 309–332.

Tripsiannis, G. A. and Philippou, A. N. (1997). A new multivariate inverse Pólya distribution of order k. *Commun. Statist. Theory Meth.* **26,** 149–158.

Wolfowitz, J. (1943). On the theory of runs with some applications to quality control. *Ann. Math. Statist.* **14,** 280–288.

https://doi.org/10.1239/aap/1198177236 Published online by Cambridge University Press