Ann. Inst. Statist. Math.
Vol. 49, No. 1, 141–153 (1997)

FORMULAE AND RECURSIONS FOR THE JOINT DISTRIBUTION OF SUCCESS RUNS OF SEVERAL LENGTHS

ANANT P. GODBOLE[1*], STAVROS G. PAPASTAVRIDIS[2] AND ROBERT S. WEISHAAR[3*]

[1] *Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, U.S.A.*
[2] *Department of Mathematics, University of Athens, Panepistemiopolis, 15710 Athens, Greece*
[3] *Department of Mathematics, The Ohio State University, Columbus, OH 43210, U.S.A.*

(Received November 15, 1993; revised March 18, 1996)

Abstract. Consider a sequence of n independent Bernoulli trials with the j-th trial having probability p_j of success, $1 \le j \le n$. Let $M(n, \boldsymbol{K})$ and $N(n, \boldsymbol{K})$ denote, respectively, the r-dimensional random variables $(M(n, k_1), \dots, M(n, k_r))$ and $(N(n, k_1), \dots, N(n, k_r))$, where $\boldsymbol{K} = (k_1, k_2, \dots, k_r)$ and $M(n, s)$ $[N(n, s)]$ represents the number of overlapping [non-overlapping] success runs of length s. We obtain exact formulae and recursions for the probability distributions of $M(n, \boldsymbol{K})$ and $N(n, \boldsymbol{K})$. The techniques of proof employed include the inclusion-exclusion principle and generating function methodology. Our results have potential applications to statistical tests for randomness.

Key words and phrases: Overlapping and non-overlapping success runs, distributions of order k, generating functions, tests for randomness, inclusion-exclusion.

1. Introduction

Let $X_1, X_2, \dots, X_n$ be n independent Bernoulli random variables such that $\boldsymbol{P}(X_j = 1) = p_j$, $1 \le j \le n$, and let $M = M(n, k)$ and $N = N(n, k)$ denote, respectively, the number of overlapping and non-overlapping success runs of fixed length $k \ge 1$. For example, the sequence $FSSSSSF$ corresponds to $M(7, 2) = 4$ and $N(7, 2) = 2$. The exact distribution theory of M has been studied (just to mention some of the more recent literature) by Ling (1988), Godbole (1992*b*) and Chryssaphinou *et al.* (1993), while asymptotic results may be found in Chryssaphinou and Papastavridis (1988), Godbole (1991, 1992*b*) and Barbour *et al.* (1992). Previous work on the exact and approximate distribution theory of N can be found in Hirano (1986), Philippou and Makri (1986), Godbole (1990,

*The research of the first and third-named authors was partially supported by National Science Foundation Grants DMS-9100829 and DMS-9200409.

1991, 1992*a*), Papastavridis (1990) and Barbour *et al.* (1992). Different but allied random variables have been studied by Panaretos and Xekalaki (1986, 1989), Philippou *et al.* (1989) and Xekalaki *et al.* (1987).

Agin and Godbole (1992) exhibited the fact that a non-parametric statistical test based on N can be significantly more powerful than standard criteria in detecting certain kinds of clustering; also, a preliminary check (in the same paper) indicated that a multivariate test based on *several* values of k gave an *even higher* empirical power than the univariate test. This is the primary motivation for our studying and laying down, in this paper, the distribution-theoretic foundation of multivariate random quantities based on the variables M and N: For any r integers $k_1, k_2, \ldots, k_r$, let $\boldsymbol{K} = (k_1, k_2, \ldots, k_r)$, $M(n, \boldsymbol{K}) = (M(n, k_1), \ldots, M(n, k_r))$ and $N(n, \boldsymbol{K}) = (N(n, k_1), \ldots, N(n, k_r))$, where $M(n, k_j)$ and $N(n, k_j)$ are each defined to equal zero if $n < k_j$ for some j; we shall call the distribution of the variables $M(n, \boldsymbol{K})$ and $N(n, \boldsymbol{K})$ the *overlapping binomial distribution of order* $\boldsymbol{K}$ and the *binomial distribution of order* $\boldsymbol{K}$ respectively. These distributions reduce to the well-known and similarly named univariate quantities on setting $r = 1$. Exact formulae and recursions for $M(n, \boldsymbol{K})$ and $N(n, \boldsymbol{K})$ will be provided in Sections 2 and 3 respectively. The inclusion-exclusion principle will be the method of choice in Section 2, while generating function methodology will yield several recursions (notably for unequal p's) and exact formulae in Section 3. The latter technique has been used, e.g., by Koutras and Papastavridis (1992).

Since $\boldsymbol{K}$ does not change through the paper, we will often denote the variables $M(n, \boldsymbol{K})$ and $N(n, \boldsymbol{K})$ simply by $M(n)$ and $N(n)$ respectively.

2. Exact formulae

We start by proving a result that extends the univariate Theorem 2.1 of Philippou and Makri (1986). We note, however, that our proof employs geometric variables truncated at the level $(n-1)$ as in Ling (1988); the original method used by Philippou and Makri (1986) truncates these variables at the value $(k-1)$, where k is the success run length. However, this procedure does not permit an efficient multivariate generalization.

THEOREM 2.1. *Let $X_1, X_2, \ldots, X_n$, $n \geq 1$, be i.i.d. Bernoulli (p) random variables. Then for any r positive integers $k_1, k_2, \ldots, k_r$, and with $X = (x_1, x_2, \ldots, x_r)$, we have:*

$$\boldsymbol{P}(N(n) = X) = \sum_{y=0}^{n} q^y p^{n-y} \sum_{i=0}^{n} \sum_{3} \binom{y}{y_1, y_2, \ldots, y_n}, \tag{2.1}$$

where $\sum_3$ is over all non-negative integers $\{y_j\}_{j=1}^n$ satisfying the conditions

$$\sum_{j=1}^{n} j y_j + i = n; \tag{2.2}$$

$$\sum_{j=1}^{n} y_j = y; \tag{2.3}$$

and

$$\left[\frac{i}{k_l}\right]+\sum_{j=k_l+1}^{n}\left[\frac{j-1}{k_l}\right]y_j=x_l \qquad (1\le l\le r). \tag{2.4}$$

PROOF. Following the combinatorial method established by Philippou *et al.* (1983), we note that any elementary event in $\{N(n)=X\}$ can be described as follows: Let y_j ($j=1,2,\dots,n$) and i ($0\le i\le n$) represent, respectively, the number of arrangements $SS\cdots SSF$ ($j-1$ S's) and the number of successes after the last failure in the sequence. Then it is easy to see that $N(n)=X$ if and only if (2.2) and (2.4) hold. Thus

$$\boldsymbol{P}(N(n)=X)=\sum_{i=0}^{n}\sum_{2}\binom{y_1+y_2+\cdots+y_n}{y_1,y_2,\dots,y_n}q^{\sum_{j=1}^{n}y_j}p^{n-\sum_{j=1}^{n}y_j}, \tag{2.5}$$

where $\sum_2$ is over all non-negative integers $\{y_j\}_{j=1}^n$ satisfying (2.2) and (2.4). (2.1) now follows on conditioning on the number y of failures, which clearly equals $\sum_{j=1}^{n}y_j$. □

The next result follows on using methods similar to those above; it extends the result of Ling (1988). We mention that the recursive techniques of Section 3 will provide an alternative formula that involves just a double summation. Theorem 2.2 is important in its own right, however, and will be used, furthermore, as a starting point for the proof of the exact formula in Theorem 2.3.

THEOREM 2.2. *Let $X_1,X_2,\dots,X_n$, $n\ge 1$, be i.i.d. Bernoulli (p) random variables. Then for any r positive integers $k_1,k_2,\dots,k_r$, and with $X=(x_1,x_2,\dots,x_r)$, we have*:

$$\boldsymbol{P}(M(n)=X)=\sum_{y=0}^{n}q^yp^{n-y}\sum_{i=0}^{n}\sum_{3}\binom{y}{y_1,y_2,\dots,y_n}, \tag{2.6}$$

where $\sum_3$ is over all non-negative integers $\{y_j\}_{j=1}^n$ satisfying (2.2), (2.3) *and*

$$\max\{0,i-k_l+1\}+\sum_{j=k_l+1}^{n}(j-k_l)y_j=x_l \qquad (1\le l\le r). \tag{2.7}$$

Representations (2.1) and (2.6) *can* be used as the basis of a computational procedure, but searching for the solutions of the integer equations (2.2), (2.3), (2.4) and (2.7) exhausts a large amount of computer time, even for moderately large values of n. For this reason, as well as to explore different methodologies, we seek alternative formulae for the probability density functions of $M(n,\boldsymbol{K})$ and $N(n,\boldsymbol{K})$. One of the combinatorial procedures that can be employed towards such an end, and which results in a representation entirely in terms of binomial coefficients, is illustrated in the next result. For simplicity we state and prove Theorem 2.3 below only for the case of overlapping runs and only for $r=2$.

Although the statement of Theorem 2.3 below looks complicated, a careful glance will reveal that from the standpoint of both mathematical depth and numerical efficiency, it is far above the simpler looking Theorem 2.2.

THEOREM 2.3. *The joint p.d.f. of the numbers of overlapping runs of two different lengths $k_1 = k$ and $k_2 = l$ has an alternative representation entirely in terms of binomial coefficients, as follows:*

$$
\begin{aligned}
&\boldsymbol{P}(M(k) = X, M(l) = Y) \\
&\quad = \sum_y q^y p^{n-y} \sum_v \binom{y}{v} \sum_t \binom{y-v}{t} \\
&\qquad \bullet \Bigg\{ \sum_j (-1)^j \binom{v}{j} \binom{X - Y - (w+j)(l-k) - 1}{v-1} \\
&\qquad\quad \cdot \sum_m (-1)^m \binom{w}{m} \binom{Y - m(n-l) - 1}{w-1} \\
&\qquad\quad \cdot \sum_r (-1)^r \binom{t+1}{r} \binom{n - X - k(y-t) - kr}{t} \\
&\qquad\quad + \sum_j (-1)^j \binom{t}{j} \binom{n - X - k(y-t+1) - jk}{t-1} \\
&\qquad\quad \cdot \sum_m (-1)^m \binom{w}{m} \binom{Y - m(n-l) - 1}{w-1} \\
&\qquad\quad \cdot \sum_r (-1)^r \binom{v+1}{r} \binom{X - (l-k)w - Y - 1 - r(l-k)}{v} \\
&\qquad\quad + \sum_j (-1)^j \binom{t}{j} \binom{n - X - k(y-t+1) - jk}{t-1} \\
&\qquad\quad \cdot \sum_m (-1)^m \binom{v}{m} \binom{X - Y - (w+m)(l-k) - 1}{v-1} \\
&\qquad\quad \cdot \sum_r (-1)^r \binom{w+1}{r} \binom{Y - r(n-l) - 1}{w} \Bigg\}.
\end{aligned}
$$

PROOF. Assume, without loss of generality, that $k < l$. We start with (2.6) and note that (2.2) and (2.7) can be rewritten, with $X = x_1$ and $Y = x_2$, as

$$
\sum_{j=1}^{k} j y_j = n - i - X - k(y-t) + \max\{0, i-k+1\} \tag{2.8}
$$

$$
\begin{aligned}
\sum_{j=k+1}^{l} j y_j = {} & X + k(y-t) - l(y-v-t) - Y \\
& - \max\{0, i-k+1\} + \max\{0, i-l+1\}
\end{aligned} \tag{2.9}
$$

and

$$\sum_{j=l+1}^{n} jy_j = Y + l(y-v-t) - \max\{0, i-l+1\} \tag{2.10}$$

where

$$t = \sum_{j=1}^{k} y_j, \tag{2.11}$$

$$v = \sum_{j=k+1}^{l} y_j, \tag{2.12}$$

and

$$w = y - v - t = \sum_{j=l+1}^{n} y_j. \tag{2.13}$$

Conditioning on v and t thus yields

$$\begin{aligned} &\boldsymbol{P}(M(k)=X, M(l)=Y) \\ &\quad = \sum_y q^v p^{n-v} \sum_v \binom{y}{v} \sum_t \binom{y-v}{t} \\ &\quad \bullet \sum_{i=0}^{n} \sum_5 \binom{t}{y_1, y_2, \ldots, y_k} \sum_6 \binom{v}{y_{k+1}, \ldots, y_l} \sum_7 \binom{w}{y_{l+1}, \ldots, y_n} \end{aligned} \tag{2.14}$$

where $\sum_5$, $\sum_6$ and $\sum_7$ are over non-negative integers satisfying (2.8) and (2.11); (2.9) and (2.12); and (2.10) and (2.13) respectively. On splitting the sum over i into three components, (2.14) reduces to

$$\begin{aligned} &\boldsymbol{P}(M(k)=X, M(l)=Y) \\ &\quad = \sum_y q^y p^{n-y} \sum_v \binom{y}{v} \sum_t \binom{y-v}{t} \\ &\quad \bullet \Bigg\{ \sum_{i=0}^{k-1} \sum_{5_1} \binom{t}{y_1, y_2, \ldots, y_k} \sum_{6_1} \binom{v}{y_{k+1}, \ldots, y_l} \\ &\qquad \cdot \sum_{7_1} \binom{w}{y_{l+1}, \ldots, y_n} \\ &\qquad + \sum_{i=k}^{l-1} \sum_{5_2} \binom{t}{y_1, y_2, \ldots, y_k} \sum_{6_2} \binom{v}{y_{k+1}, \ldots, y_l} \\ &\qquad \cdot \sum_{7_2} \binom{w}{y_{l+1}, \ldots, y_n} \\ &\qquad + \sum_{i=l}^{n} \sum_{5_3} \binom{t}{y_1, y_2, \ldots, y_k} \sum_{6_3} \binom{v}{y_{k+1}, \ldots, y_l} \\ &\qquad \cdot \sum_{7_3} \binom{w}{y_{l+1}, \ldots, y_n} \Bigg\}. \end{aligned} \tag{2.15}$$

The truncation employed above enables us to consider sums over sets of integers that are related to those in (2.8) through (2.13), but which do not involve equations with terms of the form $\max(s,t)$. For example, the sum 5_1 is over integers $y_1, \ldots, y_k$ satisfying the conditions (2.11) and

$$\sum_{j=1}^{k} jy_j - n - i - X - k(y-t) \tag{2.16}$$

whereas the sum 6_3 is over $y_{k+1}, \ldots, y_l$ satisfying (2.12) and

$$\sum_{j=k+1}^{l} jy_j = X + k(y-t+1) - l(y-v-t+1) - Y. \tag{2.17}$$

We shall now need to use the following

LEMMA 2.1.

$$\sum_{x_1+\cdots+x_r=s;x_1+2x_2+\cdots rx_r=m} \binom{s}{x_1, x_2, \ldots, x_r} = \sum_j (-1)^j \binom{s}{j}\binom{m-jr-1}{s-1}. \tag{2.18}$$

PROOF. Omitted. See, e.g., the proofs of Lemmas 2 and 3 in Godbole (1990). □

Using Lemma 2.1 and the easily verified fact that the six sums 5_2, 5_3, 6_1, 6_3, 7_1 and 7_2 do not depend on i, (2.15) can be reexpressed as

$$\begin{aligned}
&\boldsymbol{P}(M(k)=X, M(l)=Y) \\
&= \sum_y q^y p^{n-y} \sum_v \binom{y}{v} \sum_t \binom{y-v}{t} \\
&\quad \bullet \left\{ \sum_j (-1)^j \binom{v}{j}\binom{X-Y-(w+j)(l-k)-1}{v-1} \right. \\
&\quad \cdot \sum_m (-1)^m \binom{w}{m}\binom{Y-m(n-l)-1}{w-1} \\
&\quad \cdot \sum_{i=0}^{k-1}\sum_r (-1)^r \binom{t}{r}\binom{n-X-k(y-t)-i-1-kr}{t-1} \\
&\quad + \sum_j (-1)^j \binom{t}{j}\binom{n-X-k(y-t+1)-jk}{t-1} \\
&\quad \cdot \sum_m (-1)^m \binom{w}{m}\binom{Y-m(n-l)-1}{m-1}
\end{aligned} \tag{2.19}$$

$$\cdot \sum_{i=k}^{l-1} \sum_{r} (-1)^r \binom{v}{r} \binom{X-(l-k)w-Y+k-2-i-r(l-k)}{v-1}$$

$$+ \sum_{j} (-1)^j \binom{t}{j} \binom{n-X-k(y-t+1)-jk}{t-1}$$

$$\cdot \sum_{m} (-1)^m \binom{v}{m} \binom{X-Y-(w+m)(l-k)-1}{v-1}$$

$$\cdot \sum_{i=l}^{n} \sum_{r} (-1)^r \binom{w}{r} \binom{Y-2-r(n-l)-i}{w-1} \Bigg\}.$$

As our final step, we simplify the summations that depend on i; towards this end, we state

LEMMA 2.2.

$$\sum_{i=0}^{k-1} \sum_{m} (\ 1)^m \binom{v}{m} \binom{n-x-k(y-v)-mk-i-1}{v-1} \tag{2.20}$$

$$= \sum_{m} (-1)^m \binom{v+1}{m} \binom{n-x-k(y-v)-mk}{v}$$

PROOF. Omitted; contained in the proof of Theorem 1 in Godbole (1990). □

Assuming, without loss of generality, that $i \leq n-1$, (2.19) and Lemma 2.2 now yield the following representation for the joint p.d.f. of $M(n,k)$ and $M(n,l)$:

$$\boldsymbol{P}(M(k)=X, M(l)=Y) \tag{2.21}$$

$$= \sum_{y} q^y p^{n-y} \sum_{v} \binom{y}{v} \sum_{t} \binom{y-v}{t}$$

$$\bullet \Bigg\{ \sum_{j} (-1)^j \binom{v}{j} \binom{X-Y-(w+j)(l-k)-1}{v-1}$$

$$\cdot \sum_{m} (-1)^m \binom{w}{m} \binom{Y-m(n-l)-1}{w-1}$$

$$\cdot \sum_{r} (-1)^r \binom{t+1}{r} \binom{n-X-k(y-t)-kr}{t}$$

$$+ \sum_{j} (-1)^j \binom{t}{j} \binom{n-X-k(y-t+1)-jk}{t-1}$$

$$\cdot \sum_{m} (-1)^m \binom{w}{m} \binom{Y-m(n-l)-1}{w-1}$$

$$\cdot \sum_{r} (-1)^r \binom{v+1}{r} \binom{X-(l-k)w-Y-1-r(l-k)}{v}$$

$$+\sum_j (-1)^j \binom{t}{j} \binom{n-X-k(y-t+1)-jk}{t-1}$$
$$\cdot \sum_m (-1)^m \binom{v}{m} \binom{X-Y-(w+m)(l-k)-1}{v-1}$$
$$\cdot \sum_r (-1)^r \binom{w+1}{r} \binom{Y-r(n-l)-1}{w} \Bigg\}.$$

This completes the proof of Theorem 2.3. □

3. Recursions

We begin by deriving a recursive formula for the distribution of $M(n) = M(n, \boldsymbol{K})$ that is valid even in the non-i.i.d. case. Assume, then, that the probability of success at the j-th trial is p_j and set $q_j = 1 - p_j$ $(1 \le j \le n)$. By convention, we let $p_0 = 0$. For ease of exposition, we write $\boldsymbol{P}(M(n) = X)$ as $P(n, x_1, x_2, \ldots, x_r)$, and set, for $n \le 0$,

$$\boldsymbol{P}(n, x_1, \ldots, x_r) = \delta_{|x_1|+|x_2|+\cdots+|x_r|,0}, \tag{3.1}$$

where the Kronecker delta $\delta_{i,j}$ equals one if $i = j$ and zero otherwise. For $x \in \boldsymbol{Z}$, let $x^+ := \max\{0, x\}$, where ':=' means "equal by definition"; throughout this paper we will use the standard notation $\boldsymbol{N} = \{0, 1, \ldots\}$, $\boldsymbol{Z}^+ = \{1, 2, \ldots\}$ and $\boldsymbol{Z} = \{0, \pm 1, \pm 2, \ldots\}$.

THEOREM 3.1. *Given $n \ge 1$ independent Bernoulli trials $X_1, X_2, \ldots, X_n$, with $\boldsymbol{P}(X_j = 1) = p_j$, we have*

$$\begin{aligned} \boldsymbol{P}(n, x_1, x_2, \ldots, x_r) &= q_n \boldsymbol{P}(n-1, x_1, \ldots, x_r) \\ &\quad + \sum_{i=1}^{n} q_{n-i} p_{n-i+1} \cdots p_n \\ &\quad \cdot \boldsymbol{P}(n-i-1, x_1 - (i-k_1+1)^+, \ldots, x_r - (i-k_r+1)^+). \end{aligned} \tag{3.2}$$

PROOF. The event under consideration splits into two subevents, according as the last trial is a failure or a success. If $X_n = F$, we get the first term on the right hand side of (3.2). On the other hand, the event $X_n = S$ can be further subdivided into n subevents, depending on the length i of the (maximum) final run of successes, $1 \le i \le n$. This accounts for the sum on the right hand side of (3.2). □

Remark. Theorem 3.1 provides a recursion of order n in the non-i.i.d. case, and can easily be extended to provide a (direct) recursive formula for

$\boldsymbol{P}(M(n,\boldsymbol{K}) \geq X) := \boldsymbol{P}(M(n,k_1) \geq x_1, \ldots, M(n,k_r) \geq x_r)$. Such a formula would be of greater interest from the viewpoint of statistical applications.

We next consider the i.i.d. case and derive a simpler recursive formula than that given by Theorem 3.1; this recursion will be seen to be of (smaller) order $\min_i(x_i + k_i - 1)$. If $X = (x_1, \ldots, x_r)$, where $x_i \in \boldsymbol{N}$, $(i = 1, \ldots, r)$, then we shall say that $X \geq 0$ if $x_i \geq 0$ $\forall i$. Let $r \geq 1$ and consider the r-tuple $(k_1, \ldots, k_r)$ of positive integers. If $k \in \boldsymbol{Z}^+$, then $\mu_k : \boldsymbol{N} \to \boldsymbol{N}$ is defined by

$$\mu_k(n) = (n-k+1)^+. \tag{3.3}$$

Next, define $\mu : \boldsymbol{N} \to \boldsymbol{N}^r$ by

$$\mu(n) = (\mu_{k_1}(n), \ldots, \mu_{k_r}(n)). \tag{3.4}$$

Finally, let $z_1, \ldots, z_r$ be formal variables independent over the ring $\boldsymbol{Z}$, and, with $Z = (z_1, \ldots, z_r)$, $X = (x_1, \ldots, x_r)$, let Z^X denote the quantity $z_1^{x_1} \cdots z_r^{x_r}$.

Definition. Let $n \in \boldsymbol{N}$, $m \in \boldsymbol{Z}^+$, $X \in \boldsymbol{N}^r$. Then $Q(n,m,X)$ is the number of m-tuples $(n_1, \ldots, n_m)$ in $\boldsymbol{N}^m$ such that $n_1 + \cdots + n_m = n$ and $\mu(n_1) + \cdots + \mu(n_m) = X$, where addition in $\boldsymbol{N}^r$ is defined componentwise. By convention, let $Q(n,0,X) = 0$, and $Q(n,m,X) = 0$ if $n < 0$ or $m < 0$ or $x_i < 0$ for any i.

Defining the formal series $A(t,Z)$ by

$$A(t,Z) = \sum_{n \geq 0} t^n Z^{\mu(n)} \tag{3.5}$$

we have

Lemma 3.1.

$$\text{(a)} \quad \sum_{n,X \geq 0} Q(n,m,X) t^n Z^X = \{A(t,Z)\}^m \tag{3.6}$$

$$\text{(b)} \quad \sum_{n,X \geq 0; m \geq 1} Q(n,m,X) t^n y^m Z^X = \frac{1}{1 - yA(t,Z)} - 1. \tag{3.7}$$

Proof. (3.7) follows immediately from (3.6) and the fact that $Q(n,0,X) = 0$, while (3.6) can easily be seen to be true from the relevant definitions. □

Lemma 3.2.

$$\boldsymbol{P}(M(n) = X) = \sum_{m \geq 0} Q(m, n-m+1, X) p^m q^{n-m}. \tag{3.8}$$

Proof. Similar in spirit (but not in substance) to the proof of Theorem 2.2: we condition on the number of *successes* (not the number of failures), and note that $n - m + 1$ spaces are generated between the $n - m$ failures. We let n_i be the number of successes in the i-th space ($1 \leq i \leq n-m+1$) to obtain the required formula. □

By Lemma 3.2, we have

$$\begin{aligned}
&\sum_{n,X} \boldsymbol{P}(M(n)=X)t^n Z^X \\
&\quad = \sum_{n,m,X>0} Q(m,n-m+1,X)p^m q^{n-m} t^n Z^X \\
&\quad = \sum_{n,m,X\geq 0} Q(m,n+1,X)p^m q^n t^{n+m} Z^X \\
&\qquad \text{(on replacing } n-m \text{ by } n) \\
&\quad = \frac{1}{qt} \sum_{m,n,X\geq 0} Q(m,n+1,X)(pt)^m (qt)^{n+1} Z^X \\
&\quad = \frac{1}{qt}\left\{\frac{1}{1-qtA(pt,Z)} - 1\right\} \\
&\qquad \text{(on using Lemma 3.1 with the roles of } m \text{ and } n \text{ interchanged)} \\
&\quad = \frac{A(pt,Z)}{1-qtA(pt,Z)}
\end{aligned}$$

so that

$$(1-qtA(pt,Z)) \sum_{n,X\geq 0} \boldsymbol{P}(M(n)=X)t^n Z^X = A(pt,Z). \tag{3.9}$$

On equating coefficients of $t^n Z^X$ on both sides of (3.9) and recalling the definition of $A(t,Z)$ we obtain

THEOREM 3.2. *Given n i.i.d. Bernoulli (p) random variables, we have the recursion*

$$\begin{aligned}
\boldsymbol{P}(M(n,K)=X) - \sum_{s\geq 0} qp^s \boldsymbol{P}(M(n-1-s,K) = X-\mu(s)) &= p^n \qquad [X=\mu(n)] \\
&= 0 \qquad [X \neq \mu(n)].
\end{aligned}$$

The recursion in Theorem 3.2 is *not* just a special case of that in Theorem 3.1. As pointed out earlier, it yields a deeper result for the i.i.d. case. We now derive an exact formula for $P(M(n)=X)$, alternative to that given by Theorem 2.2, also expressed in terms of multinomial coefficients, but containing only a double summation:

THEOREM 3.3. *Under the conditions of Theorem 3.2,*

$$\boldsymbol{P}(M(n)=X) = \delta_{0,X}q^n + \sum_{m=1}^{n}\sum_{2} \binom{n-m+1}{t_0, t_1, \ldots, t_m} p^m q^{n-m}, \tag{3.10}$$

where δ is the Kronecker delta and the sum $\sum_2$ varies over all $m+1$-tuples $(t_0, \ldots, t_m)$ satisfying the conditions

$$t_0 + t_1 + \cdots + t_m = n - m + 1 \tag{3.11}$$

$$t_1 + 2t_2 + \cdots + mt_m = m \tag{3.12}$$

and

$$t_1\mu(1) + t_2\mu(2) + \cdots + t_m\mu(m) = X. \tag{3.13}$$

PROOF. Using Lemma 3.1(a), for $m > 0$,

$$\sum_{n,X\geq 0} Q(n,m,X)t^n Z^X = (A(t,Z))^m = \left(\sum_{n\geq 0} t^n Z^{\mu(n)}\right)^m \tag{3.14}$$

so that for $m, n > 0$, $Q(n,m,X)$ is the coefficient of $t^n Z^X$ in $(A(t,Z))^m$; in other words,

$$Q(n,m,X) = \sum_3 \binom{m}{m_0, m_1, \ldots, m_n} \tag{3.15}$$

where $\sum_3$ varies over all $(n+1)$-tuples $(m_0, \ldots, m_n)$ satisfying the conditions

$$\sum_{i=0}^{n} m_i = m \tag{3.16}$$

$$\sum_{i=1}^{n} im_i = n \tag{3.17}$$

and

$$\sum_{i=1}^{n} m_i\mu(i) = X. \tag{3.18}$$

It follows from Lemma 3.2 that

$$\begin{aligned} \boldsymbol{P}(M(n) = X) &= Q(0, n+1, X)q^n + \sum_{m=1}^{n} Q(m, n-m+1, X)p^m q^{n-m} \\ &= \delta_{0,X} q^n + \sum_{m=1}^{n} \sum_2 \binom{n-m+1}{t_0, t_1, \ldots, t_m} p^m q^{n-m}, \end{aligned}$$

as asserted. $\square$

If $r = 1$, then Theorem 3.3 provides a double summation formula for $\boldsymbol{P}(M(n) = X)$ which is essentially equivalent to the triple summation expression of Ling (1988). Theorem 2.3 provides an exact formula, but only for $r = 2$. Theorem 3.3 is a general result, therefore, that is valid for all values of r. We now turn our attention to developing recursions and exact formulae for $P(N(n, K) = X)$. It turns out, however, that all the hard work in this direction has already been done; the approach used to analyze overlapping runs carries through with few changes. Specifically, if we define

$$\mu(n) = \left(\left[\frac{n}{k_1}\right], \left[\frac{n}{k_2}\right], \dots, \left[\frac{n}{k_r}\right]\right) \tag{3.19}$$

and suitable modify the definitions of $Q(n, m, X)$ and $A(t, Z)$, then it easy to see that the following results hold:

THEOREM 3.4. *Given $n \geq 1$ independent Bernoulli trials $X_1, X_2, \dots, X_n$, with $\boldsymbol{P}(X_j = 1) = p_j$, we have*

$$\begin{aligned} \text{(3.20)}\quad \boldsymbol{P}(N(n) = X) &:= P(n, x_1, x_2, \dots, x_r) \\ &= q_n P(n-1, x_1, x_2, \dots, x_r) + p_1 p_2 \cdots p_n \delta_{([n/k_1],[n/k_2],\dots,[n/k_r]),X} \\ &\quad + \sum_{i=1}^{n-1} q_{n-i} p_{n-i+1} \cdots p_n P\left(n - i - 1, x_1 - \left[\frac{i}{k_1}\right], \dots, x_r - \left[\frac{i}{k_r}\right]\right). \end{aligned}$$

PROOF. Exactly the same as that of Theorem 3.1; for variety, we split the required probability into *three* subevents, however, according as the last trial is a failure, all the trials are successes, or the last i are successes $(1 \leq i \leq n-1)$. □

THEOREM 3.5. *Equations* (3.6) *and* (3.7) *hold without any changes, using the modified definitions of $Q(n, m, X)$ and $A(t, Z)$. Also, $P(N(n) = X)$ is given, in the i.i.d. case, by an equation identical to* (3.8)*, from which one can deduce a recursion that is the same as that given by Theorem* 3.2. *Finally, the exact formula* (3.10) *of Theorem* 3.3 *translates verbatim into one for $P(N(n) = X)$, but with $\mu(n)$ being defined by* (3.19)*. More specifically, we have*

$$\boldsymbol{P}(N(n) = X) = \delta_{0,X} q^n + \sum_{m=1}^{n} \sum_{2} \binom{n-m+1}{t_0, t_1, \dots, t_m} p^m q^{n-m}, \tag{3.21}$$

where δ is the Kronecker delta and the sum $\sum_2$ varies over all $m+1$ tuples $(t_0, \dots, t_m)$ satisfying the conditions

$$t_0 + t_1 + \cdots + t_m = n - m + 1 \tag{3.22}$$

$$t_1 + 2t_2 + \cdots + mt_m = m \tag{3.23}$$

and

$$t_1[1/k_l] + t_2[2/k_l] + \cdots + t_m[m/k_l] = x_l \qquad (1 \leq l \leq r). \tag{3.24}$$

References

Agin, M. and Godbole, A. P. (1992). A new exact runs test for randomness, *Computing Science and Statistics: Proceedings of the 22nd Symposium on the Interface* (eds. C. Page and R. LePage), 281–285, Springer, New York.

Barbour, A., Holst, L. and Janson, S. (1992). *Poisson Approximation*, Oxford University Press, Oxford.

Chryssaphinou, O. and Papastavridis, S. (1988). A limit theorem on the number of overlapping occurrences of a pattern in a sequence of independent trials, *Probab. Theory Related Fields*, **79**, 129–143.

Chryssaphinou, O., Papastavridis, S. and Tsapelas, T. (1993). On the number of overlapping success runs in a sequence of independent Bernoulli trials, *Applications of Fibonacci Numbers, Vol. 5: Proceedings of the 5th International Conference on Fibonacci Numbers* (eds. G. E. Bergum, A. Horadam and A. Philippou), 103–112, Kluwer, Boston.

Godbole, A. P. (1990). Specific formulæ for some success run distributions, *Statist. Probab. Lett.*, **10**, 119–124.

Godbole, A. P. (1991). Poisson approximations for runs and patterns of rare events, *Adv. in Appl. Probab.*, **23**, 851–865.

Godbole, A. P. (1992*a*). Approximate reliabilities of m-consecutive-k-out-of-n: failure systems, *Statist. Sinica*, **3**, 321–328.

Godbole, A. P. (1992*b*). The exact and asymptotic distribution of overlapping success runs, *Comm. Statist. Theory Methods*, **21**, 953–967.

Hirano, K. (1986). Some properties of the distributions of order k, *Fibonacci Numbers with Applications: Proceedings of the 1st International Conference on Fibonacci Numbers and their Applications* (eds. A. Philippou, G. Bergum and A. Horadam), Reidel, Dordrecht.

Koutras, M. and Papastavridis, S. (1992). On the number of runs and related statistics, *Statist. Sinica*, **3**, 277–294.

Ling, K. D. (1988). On binomial distributions of order k, *Statist. Probab. Lett.*, **6**, 247–250.

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

Panaretos, J. and Xekalaki, E. (1989). A probability distribution associated with events with multiple occurrences, *Statist. Probab. Lett.*, **8**, 389–395.

Papastavridis, S. (1990). m-consecutive-k-out-of-n : F systems, *IEEE Translation on Reliability*, **R39**, 386–388.

Philippou, A. N. and Makri, F. (1986). Successes, runs and longest runs, *Statist. Probab. Lett.*, **4**, 211–215.

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

Philippou A. N., Tripsianis, G., and Antzoulakos, D. (1989). New Polya and Inverse Polya Distributions of order k, *Comm. Statist. Theory Methods*, **18**, 2125–2137.

Xekalaki, E., Panaretos, J. and Philippou, A. (1987). On some mixtures of distributions of order k, *Fibonacci Quarterly*, **25**, 151–160.