Journal of the Korean Statistical Society (2004), **33**: 1, pp 35–58

DISTRIBUTIONS OF PATTERNS OF TWO FAILURES SEPARATED BY SUCCESS RUNS OF LENGTH k

KANWAR SEN[1] AND BABITA GOYAL[1]

ABSTRACT

For fixed positive integers n and k ($n \geq k + 2$), the exact probability distributions of non-overlapping and overlapping patterns of two failures separated by (i) exactly k successes, (ii) at least k successes and (iii) at most k successes have been obtained for Bernoulli independent and Markov dependent trials by using combinatorial technique. The waiting time distributions for the first occurrence and the r^{th} ($r > 1$) occurrence of the patterns have also been obtained.

AMS 2000 subject classifications. Primary 60C05; Secondary 62E15.
Keywords. Runs, patterns, probability functions, waiting time distributions.

1. INTRODUCTION

In literature, many authors have obtained various probability distributions related to runs, in independent Bernoulli as well as in Markov dependent trials. In the classical literature, "a success run of length k" meant an uninterrupted sequence of exactly k successes (Mood, 1940). For example, the sequence $SSSSSSFFSSSSFS$ contains three success runs, one each of length 6, 4 and 1. Feller (1968) proposed the counting from scratch as soon as a run of length k occurs, *i.e.*, in the above sequence, there are three runs of length 3. Ling (1988) proposed the concept of overlapping runs where an individual success can contribute to at most k runs. In the above sequence, there are six success runs of length 3 by Ling's way of counting. Goldstein (1990) counted the number of runs of length k or more. The distributions of success runs of length k or more have been termed as distributions of order k.

Received November 2002; accepted September 2003.
[1]Department of Statistics, University of Delhi, Delhi-110007, India (e-mail : ksen@ndb.vsnl.net.in)

The study of the distributions of order k in independent Bernoulli and Markov dependent trials has been undertaken by various authors, Feller (1968), Rajarshi (1974), Aki (1985, 1992), Hirano (1986), Philippou and Makri (1986), Ling (1988), Koutras and Papastavridis (1993), Fu (1996), Koutras (1996), to name a few. The study of runs has been found useful in various fields, particularly in reliability, statistical quality control, molecular biology and DNA detection, pattern recognition and matching, computer science, *etc.*

The method of run counting in various fields depends on the problem under consideration. Feller's method of counting is useful in renewal theory and reliability theory, for example, in m-Consecutive-k-out-of-n: F systems (Papastavridis, 1990) whereas in DNA sequence matching identical runs of length at least k (Goldstein, 1990) have been found useful.

A run is a particular pattern in which there are all identical elements. A pattern is a specific string of outcomes and we say that a pattern has occurred when a sub string of a sequence of outcomes exactly matches with the pattern. There can be different elements and sub patterns in a pattern. For example, Feller (1968) considered the distribution of "a success run of length r or a failure run of length ρ".

In this paper we have defined three patterns as

(i) Two failures separated by exactly k successes : $F\underbrace{SS\cdots S}_{k}F$;

(ii) Two failures separated by at least k successes : $F\underbrace{SS\cdots S}_{\geq k}F$;

(iii) Two failures separated by at most k successes : $F\underbrace{SS\cdots S}_{\leq k}F$.

Hirano and Aki (1993) considered the pattern in (ii) in Markov dependent trials and derived the exact probability distribution and probability generating function. Their result for *p.m.f.* involves solution of simultaneous system of linear equations in n variables. Later in 1997, Koutras considered the waiting time distribution of the same pattern for Bernoulli trials and obtained the *p.g.f.* and moments.

These patterns are special cases of scan statistics. A scan statistic can be defined as follows: Let $X_1, X_2, \ldots, X_N$ be a sequence of integer valued random variables. For $2 \leq m \leq N-1$, consider the moving sums of m consecutive observations. Then the linear unconditional scan statistics are defined as

$$S_m = \max_{1\leq t\leq N-m+1} Y_t.$$

where $Y_t = \sum_{i=t}^{t+m-1} X_i$ (Glaz and Balakrishnan, 1999). These statistics are widely used in various fields, including DNA analysis and protein detection (Altschul and Erickson, 1988; Arratia *et al.*, 1989; Leung and Yamashita, 1999), epidemiology (Krauth, 1992a, b), mine field detection (Glaz, 1996), quality control and reliability theory (Balakishnan *et al.*, 1993), radar detection and sociology, to name a few.

Most of the work in the field of scan statistics has been focussed on the case when $X_1, X_2, \ldots, X_N$ are *iid* non-negative random variables, in particular 0-1 *iid* Bernoulli variables. The dependent case for 0-1 Markov dependent trials was discussed by Glaz (1983) and Koutras and Alexandrou (1995). Krauth (1992a, b) and Wallenstein *et al.* (1989) discussed the case of multinomial random vector. In this paper we have obtained the exact probability distributions for r occurrences of patterns (overlapping and non-overlapping) in n Bernoulli and Markov dependent trials. Waiting time distributions for 1^{st} and r^{th} $(r > 1)$ occurrences of patterns have also been considered. In contrast to results of Hirano and Aki (1993), we have obtained the results in terms of Binomial coefficients.

2. NON-OVERLAPPING PATTERNS FOR INDEPENDENT BERNOULLI TRIALS

Let n and k be fixed positive integers such that $n \geq k + 2$. In a sequence of n Bernoulli trials, let the possible outcomes be a success (S) and a failure (F). Consider the pattern of length $k + 2$ $(k \geq 1)$ of two failures separated by a run of k successes. We define the following random variables:

$X_n^{(k)} \equiv$ the number of times a pattern of length exactly $k + 2$ of two failures separated by a success run of exact length k occurs in n trials;

$Y_n^{(k)} \equiv$ the number of times a pattern of length at least $k + 2$ of two failures separated by a success run of length at least k occurs in n trials;

$Z_n^{(k)} \equiv$ the number of times a pattern of length at most $k + 2$ of two failures separated by a non-void success run of length at most k occurs in n trials.

We state and prove the following results, which will be used in the sequel.

RESULT 2.1. *The number of ways of placing n indistinguishable balls into m identical cells such that no cell receives exactly r balls is*

$$N(n,m;\sim r)=\sum_{j=0}^{\min([n/r],m)}(-1)^j\binom{m}{j}\binom{m+n-(r+1)j-1}{n-rj}. \tag{2.1}$$

PROOF. It is obvious that

$$\begin{aligned}
N(n,m;\sim r) &= \text{coefficient of } z^n \text{ in the generating function}\\
&\quad (1+z+z^2+\cdots+z^{r-1}+z^{r+1}+\cdots)^m\\
&= \text{coefficient of } z^n \text{ in } \left(\frac{1}{1-z}-z^r\right)^m\\
&= \text{coefficient of } z^n \text{ in } (1-z)^{-m}\{1-z^r(1-z)\}^m\\
&= \text{coefficient of } z^n \text{ in } (1-z)^{-m}\sum_j\binom{m}{j}\{-z^r(1-z)\}^j\\
&= \text{coefficient of } z^n \text{ in } \sum_j(-1)^j\binom{m}{j}z^{jr}\sum_i(-1)^i\binom{-m+j}{i}z^i\\
&= \text{R.H.S. of (2.1).}
\end{aligned}$$

□

RESULT 2.2. *The number of ways of placing n indistinguishable balls into m identical cells such that no cell receives more than r balls is*

$$N(n,m;r)=\sum_{j=0}^{\min\{[n/(r+1)],m\}}\binom{m}{j}\binom{m+n-(r+1)j-1}{m-1} \tag{2.2}$$

(See Rosenstock and Maradudin, 1961; Papastavridis, 1990).

RESULT 2.3. *The number of ways of placing n indistinguishable balls into m identical cells such that no cell receives less than r balls is*

$$M(n,m;r)=\binom{m+n-mr-1}{n-mr}. \tag{2.3}$$

RESULT 2.4. *The number of ways of placing n indistinguishable balls into m identical cells such that no cell receives more than r balls and less than l balls ($r \geq l$) is*

$$N(n,m;r,l)=N(n-ml,m;r-l). \tag{2.4}$$

We, then, have the following theorems.

THEOREM 2.1. *For $n \geq r(k+2)$, $k \geq 1$,*

$$
\begin{aligned}
P\{X_n^{(k)} = r\} \\
&= \sum_{s=rk}^{n-2r} \sum_{d=0}^{s-rk} \left[\binom{n-s-r-1}{r} \binom{r+d+1}{d} \right. \\
&\quad \times N(s-rk-d, n-s-2r-1; \sim k) \\
&\quad \left. + \binom{n-s-r-1}{r-1} \binom{r+d}{d} N(s-rk-d, n-s-2r; \sim k) \right] p^s q^{n-s},
\end{aligned} \tag{2.5}
$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$, $p = P[S]$, $q = 1-p$.

PROOF. Let s be the number of successes in n trials. Then, out of n trials, $r(k+2)$ trials are exhausted in r occurrences of the pattern. Considering one pattern as one outcome, the sequence now has $n - r(k+2) + r = n - r(k+1)$ outcomes, of which $s - rk$ are successes, r patterns and $n - s - 2r$ failures, to be arranged so that $\{X_n^{(k)} = r\}$ remains true.

Now, between two failures, a success run of length k cannot appear. However, a success run of length k can appear between two patterns or a pattern and a failure, *i.e.*, after a pattern, without adding to the number of patterns. If S, F, and E respectively denote a success, a failure and a pattern, then

$$FE\underbrace{S\cdots S}_{k}E\underbrace{S\cdots S}_{k}FFEFF$$

is one of the possible arrangements.

Two cases arise:

(i) The last failure happens to be an independent failure.

(ii) The last failure belongs to a pattern.

Case (i): If the last failure is an independent failure then r places (after r events) are available where a success run of length k can occur. The $n-s-2r-1$ failures and r patterns can be arranged in $\binom{n-s-r-1}{r}$ ways.

Let d be the number of successes placed in r places after patterns ($0 \leq d \leq s-rk$) and at the beginning and at the end of the sequence. Since these successes can be placed unconditionally so the number of possible arrangements is $\binom{r+d+1}{d}$.

Remaining $s-rk-d$ successes are to be placed in $n-s-2r-1$ places between individual failures such that no place gets exactly k successes and the number of ways in which this can be done is $N(s - rk - d, n - s - 2r - 1; \sim k)$. *e.g.*, let $n = 24$; $r = k = 2$; $p = q = 1/2$; $s = 14 \Rightarrow n - s = 10$.

Let E stand for the pattern "$FSSF$". For two occurrences of the pattern, the numbers of exhausted successes and failures are 4 each and the number of remaining successes and failures are 10 and 6 respectively. Then the following arrangement is one of the possible arrangements where the places marked by $*$ denote the places where d successes can be placed unconditionally and $\diamond$ denotes the places where a success run of length 2 cannot occur as the occurrence of a success run of length 2 at one or more of these places will either increase the number of occurrences of the patterns or will result in shifting of the occurrence of the patterns:

$$*F \diamond E * F \diamond F \diamond F \diamond E * F \diamond F*.$$

Case (ii): If the last failure belongs to a pattern then $r - 1$ places (between r patterns) are available where a success run of length k can be placed. The number of ways in which $n - s - 2r$ failures and $r - 1$ patterns can be arranged is $\binom{n-s-r-1}{r-1}$.

The number of ways of placing d $(0 \le d \le s - rk)$ successes can be arranged unconditionally in $r + 1$ places ($r - 1$ between patterns, one at the beginning of the sequence and one at the end) is $\binom{r+d}{d}$.

The remaining $s - rk - d$ successes can be placed in $n - s - 2r$ places such that at no place, number of consecutive successes is equal to k. The number of ways this can be done is $N(s - rk - d, n - s - 2r; \sim k)$.

In the above example, one of the possible arrangements is

$$*F \diamond F \diamond E * F \diamond F \diamond F \diamond E*.$$

Combining the two cases and $p^s q^{n-s}$ being the probability of a sequence, we get (2.5). □

THEOREM 2.2. *For $n \ge r(k + 2)$, $k \ge 1$,*

$$\begin{aligned} P\{Y_n^{(k)} = r\} &= \sum_{s=rk}^{n-2r} \sum_{d=0}^{s-rk} \left[\binom{n-s-r-1}{r} \binom{2r+d+1}{d} N(s-d, n-s-2r-1; k-1) \right. \\ &\quad \left. + \binom{n-s-r-1}{r-1} \binom{2r+d}{d} N(s-d, n-s-2r; k-1) \right] p^s q^{n-s}, \end{aligned} \tag{2.6}$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$.

PROOF. In this case, r pairs of failures are separated by at least k successes. In these pairs, we place rk successes. Now, we have $s - rk$ remaining successes, r patterns and $n - s - 2r$ failures, the $s - rk$ successes are to be arranged in such a way that $\{Y_n^{(k)} = r\}$ remains true. Then, as in earlier case, considering two cases regarding the status of the last failure and observing that after a failure, a success run of length k or more cannot appear (except possibly at the last trial if it happens to be an individual failure), we get (2.6). □

THEOREM 2.3. *For* $n \geq r(k+2)$, $k \geq 1$,

$$\begin{aligned} P\{Z_n^{(k)} = r\} &= \sum_{s=r}^{n-2r} \sum_{s_0=r}^{\min(s,rk)} N(s_0, r; k, 1) \Bigg[\binom{n-s-r-1}{r} \binom{r+s-s_0+1}{s-s_0} \\ &\times M(s-s_0, n-s-2r-1; k+1) + \binom{n-s-r-1}{r-1} \binom{r+s-s_0}{s-s_0} \\ &\times M(s-s_0, n-s-2r; k+1) \Bigg] p^s q^{n-s}, \end{aligned} \tag{2.7}$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$.

PROOF. The result can be proved on similar lines on observing that in r pairs of failures, s_0 successes are to be placed in such a way that no pair gets more than k successes and each pair gets at least one. The remaining successes are to be arranged in remaining places in such a way that $\{Z_n^{(k)} = r\}$ remains true. □

In the following theorems, we obtain the joint probability distributions of above patterns.

THEOREM 2.4. *For* $0 \leq r_1 \leq r_2 \leq [n/(k+2)]$,

$$\begin{aligned} &P\{X_n^{(k)} = r_1, Y_n^{(k)} = r_2\} \\ &= \binom{r_2}{r_1} \sum_{s=r_2k}^{n-2r_2} \sum_{d=0}^{s-r_2k} \Bigg[\binom{n-s-r_2-1}{r_2} \binom{2r_2+d+1-r_1}{d} \\ &\quad \times N(s-r_2k-d, n-s-2r_2-1; k-1) \\ &\quad + \binom{n-s-r_2-1}{r_2-1} \binom{2r_2+d-r_1}{d} \\ &\quad \times N(s-r_2k-d, n-s-2r_2; k-1) \Bigg] p^s q^{n-s}. \end{aligned} \tag{2.8}$$

PROOF. The event defined in L.H.S. of (2.8) is equivalent to the event

$$\{X_n^{(k)} = r_1, Y_n^{(k+1)} = r_2 - r_1\}. \tag{2.9}$$

For this, $2r_1$ failures are exhausted in occurrence of $\{X_n^{(k)} = r_1\}$ and $2(r_2 - r_1)$ in occurrence of $\{Y_n^{(k+1)} = r_2 - r_1\}$. The number of ways in which r_1 patterns can be selected out of r_2 is $\binom{r_2}{r_1}$. We are left with $n - s - r_2$ $(= n - s - 2r_1 - 2(r_2 - r_1))$ failures and r_2 $(= r_1 + r_2 - r_1)$ patterns.

For placement of successes, we put k successes each in r_1 pairs of failures and remaining $s - r_2k$ successes are placed in $r_2 - r_1$ pairs of events, r_2 places after events, between individual failures and at the beginning and the end of the sequence in such a manner that (2.9) remains satisfied. Then, (2.8) can be obtained on similar lines as earlier results. □

THEOREM 2.5. *For* $0 \le r_1 \le r_2 \le [n/(k+2)]$,

$$\begin{aligned} P\{X_n^{(k)} &= r_1, Z_n^{(k)} = r_2\} \\ &= \binom{r_2}{r_1} \sum_{s=r_1k+(r_2-r_1)}^{n-2r_2} \sum_{s_0=r_2-r_1}^{\min\{s,(r_2-r_1)k\}} \sum_{d=0}^{s-s_0-r_1k} N(s_0, r_2 - r_1; k, 1) \\ &\quad \times \left[\binom{n-s-r_2-1}{r_2}\binom{r_2+d+1}{d}\right. \\ &\quad \times M(s - s_0 - r_1k - d, n - s - 2r_2 - 1; k+1) \\ &\quad + \binom{n-s-r_2-1}{r_2-1}\binom{r_2+d}{d} \\ &\quad \left. \times M(s - s_0 - r_1k - d, n - s - 2r_2; k+1)\right] p^s q^{n-s}. \end{aligned} \tag{2.10}$$

THEOREM 2.6. *For* $0 \le r_1 \le r_2, r_3 \le [n/(k+2)]$, $k \ge 1$,

$$\begin{aligned} P\{X_n^{(k)} &= r_1, Y_n^{(k)} = r_2, Z_n^{(k)} = r_3\} \\ &= \binom{r_2 + r_3 - r_1}{r_1, r_2 - r_1, r_3 - r_1} \\ &\quad \times \sum_{s=r_2k+r_3-r_1}^{n-2r_2-2(r_3-r_1)} \sum_{s_0=r_3-r_1}^{\min\{s-r_2k,(r_3-r_1)k\}} N(s_0, r_3 - r_1; k-1, 1) \\ &\quad \times \sum_{d=0}^{s-s_0-r_2k} \left[\binom{n-s-r_2-r_3+r_1-1}{r_2+r_3-r_1}\binom{2r_2+r_3-2r_1+d+1}{d}\right. \end{aligned} \tag{2.11}$$

$$+\binom{n-s-r_2-r_3+r_1-1}{r_2+r_3-r_1-1}\binom{2r_2+r_3-2r_1+d}{d}\Bigg]p^s q^{n-s}.$$

PROOF. The event $\{X_n^{(k)}=r_1, Y_n^{(k)}=r_2, Z_n^{(k)}=r_3\}$ is equivalent to the event

$$\{X_n^{(k)}=r_1, Y_n^{(k+1)}=r_2-r_1, Z_n^{(k-1)}=r_3-r_1\}. \tag{2.12}$$

Here, $2r_1+2(r_2-r_1)+2(r_3-r_1)=2(r_2+r_3-r_1)$ failures are exhausted in $r_2+r_3-r_1$ occurrences of the events. The number of ways in which patterns can occur is

$$\binom{r_2+r_3-r_1}{r_1, r_2-r_1, r_3-r_1}.$$

The remaining number of failures is $n-s-2(r_2+r_3-r_1)$ and $r_2+r_3-r_1$ patterns are there, among which successes are to be placed.

In r_2 patterns, put k successes in each pattern thus exhausting r_2k successes. In r_3-r_1 patterns put a total of s_0 successes in such a manner that at no point, more than $k-1$ successes occur and each pattern gets at least one. The number of ways in which this can be done is $N(s_0, r_3-r_1; k-1, 1)$. Then, we are left with $s-s_0-r_2k$ successes, which are to be placed in between patterns and failures such that (2.12) remains satisfied.

Now, between individual failures, no successes can appear. Also, no more successes can be placed in r_1 patterns of type two failures separated by a success run of length k and r_3-r_1 patterns of type two failures separated by a non void success run of length at most k so the only places left are r_2-r_1 patterns of type two failures separated by a success run of length $k+1$ or more and after patterns, *i.e.*, a total of $r_2-r_1+r_2+r_3-r_1=2r_2+r_3-2r_1$ places. Then, proceeding as in earlier cases, we get (2.11). □

REMARK 2.1. From Feller (1968, p. 64),

$$\sum_{\nu=0}^{r}\binom{\nu+k-1}{k-1}=\binom{r+k}{k}.$$

This implies

$$\sum_{d=0}^{s-s_0-r_2k}\binom{2r_2+r_3-2r_1+d+1}{d}$$
$$=\sum_{d=0}^{s-s_0-r_2k}\binom{2r_2+r_3-2r_1+d+1}{2r_2+r_3-2r_1+1}$$

$$= \binom{2r_2 + r_3 - 2r_1 + 2 + s - s_0 - r_2k}{2r_2 + r_3 - 2r_1 + 2}$$

$$= \binom{s - s_0 - r_2k + 2(r_2 - r_1 + 1) + r_3}{2(r_2 + r_1 - 1) + r_3}$$

and

$$\sum_{d=0}^{s-s_0-r_2k} \binom{2r_2 + r_3 - 2r_1 + d}{d} = \binom{s - s_0 - r_2k + 2(r_2 - r_1 + 1) + r_3 + 1}{2(r_2 + r_1 - 1) + r_3 + 1}.$$

Thus, (2.11) becomes

$$\begin{aligned} &P\{X_n^{(k)} = r_1, Y_n^{(k)} = r_2, Z_n^{(k)} = r_3\} \\ &= \binom{r_2 + r_3 - r_1}{r_1, r_2 - r_1, r_3 - r_1} \\ &\times \sum_{s=r_2k+r_3-r_1}^{n-2r_2-2(r_3-r_1)} \sum_{s_0=r_3-r_1}^{\min\{s-r_2k,(r_3-r_1)k\}} N(s_0, r_3 - r_1; k-1, 1) p^s q^{n-s} \qquad (2.13) \\ &\times \left[\binom{n-s-r_2-r_3+r_1-1}{r_2 + r_3 - r_1} \binom{s - s_0 - r_2k + 2(r_2 - r_1 + 1) + r_3}{2(r_2 + r_1 - 1) + r_3} \right. \\ &\left. + \binom{n-s-r_2-r_3+r_1-1}{r_2 + r_3 - r_1 - 1} \binom{s - s_0 - r_2k + 2(r_2 - r_1 + 1) + r_3 + 1}{2(r_2 + r_1 - 1) + r_3 + 1} \right]. \end{aligned}$$

3. WAITING TIME DISTRIBUTIONS OF NON-OVERLAPPING PATTERNS FOR INDEPENDENT BERNOULLI TRIALS

To obtain waiting time distributions of the patterns considered in Section 2, we define the following random variables:

$X_n^{(k,r)} \equiv$ the number of trials needed for the r^{th} occurrence of a pattern of two failures separated by a success run of exact length k;

$Y_n^{(k,r)} \equiv$ the number of trials needed for the r^{th} occurrence of a pattern of two failures separated by a success run of at least length k;

$Z_n^{(k,r)} \equiv$ the number of trials needed for the r^{th} occurrence of a pattern of two failures separated by a success run of at most length k.

$$X^{(k,1)} = X^{(k)},\ Y^{(k,1)} = Y^{(k)},\ Z^{(k,1)} = Z^{(k)}.$$

Then we have the following results.

THEOREM 3.1. *For* $k \geq 1$,

$$P\{X^{(k)} = n\} = \sum_{s=k}^{n-2} \sum_{s_0=0}^{s-k} N(s - s_0 - k, n - s - 2; \sim k) p^s q^{n-s}, \tag{3.1}$$

where $n = k + 2, k + 3, \ldots$

PROOF. Partition the sequence of outcomes at a point where the event has occurred

$$\cdots \mid F \underbrace{S \cdots S}_{k \text{ successes}} F$$

$$\text{I} \qquad \text{II}$$

Then in sequence I, there is no success run of length k except possibly in the beginning and last trial is always a failure. Further, partitioning subsequence I at the points where first success run ends, if it happens to be the initial run and let s_0 be the number of successes in this success run. Then $s - s_0 - k$ successes are to be distributed in $n - s - 2$ places such that at no place, a success run of length k occurs. The number of ways in which this can be done is $N(s - s_0 - k, n - s - 2; \sim k)$. Hence (3.1). □

THEOREM 3.2. *For* $k \geq 1$,

$$P\{X^{(k,r)} = n\} = \sum_{s=rk}^{n-2r} \sum_{d=0}^{s-rk} \binom{n-s-r-1}{r-1} \binom{r+d-1}{r-1} \times N(s - rk - d, n - s - 2r; \sim k) p^s q^{n-s}, \tag{3.2}$$

where $n \geq r(k + 2)$.

PROOF. In this case also, the sequence always ends with a pattern. After forming r patterns, we have $n - s - 2r$ failures, $s - rk$ successes and r patterns. The number of ways in which $n-s-2r$ failures and $r-1$ patterns can be arranged is $\binom{n-s-r-1}{r-1}$. Put d successes after patterns and at the beginning of the sequence unconditionally. Then remaining successes are to be placed in remaining places such that at no place, a success run of length k occurs. Hence (3.2). □

THEOREM 3.3. *For* $k \geq 1$,

$$P\{Y^{(k)} = n\} = \sum_{s=k}^{n-2} \sum_{s_0=0}^{s-k} N(s - s_0 - k, n - s - 2; k - 1) p^s q^{n-s}, \tag{3.3}$$

where $n = k + 2, k + 3, \ldots$

REMARK 3.1. The result can be obtained on partitioning the sequence as in earlier case and then on observing that in remaining trials, no success run of length k or more, except possibly in the beginning, is possible.

THEOREM 3.4. *For* $k \geq 1$,

$$\begin{aligned} P\{Y^{(k,r)} = n\} = \sum_{s=rk}^{n-2r} \sum_{d=0}^{s-rk} & \binom{n-s-r-1}{r-1} \binom{2r+d-1}{2r-1} \\ & \times N(s - rk - d, n - s - 2r; k - 1) p^s q^{n-s}, \end{aligned} \tag{3.4}$$

where $n = k + 2, k + 3, \ldots$

Similarly the distributions of $Z^{(k)}$ and $Z^{(k,r)}$ can be obtained.

4. OVERLAPPING PATTERNS FOR INDEPENDENT BERNOULLI TRIALS

For overlapping patterns, we observe that a failure, which is not at either end of the sequence can contribute at most to two patterns. For this case, we define the following random variables:

$\overline{X}_n^{(k)}$ $\equiv$ the number of times a pattern of length $k+2$ of two failures separated by exactly k successes occurs when the patterns may overlap;

$\overline{Y}_n^{(k)}$ $\equiv$ the number of times a pattern of length at least $k+2$ of two failures separated by at least k successes occurs when the patterns may overlap;

$\overline{Z}_n^{(k)}$ $\equiv$ the number of times a pattern of length at most $k+2$ of two failures separated by at most k successes occurs when the patterns may overlap;

$\overline{X}_n^{(k,r)}$ $\equiv$ the number of trials required for r^{th} occurrence of a pattern of length $k+2$ of two failures separated by exactly k successes in overlapping patterns;

$\overline{Y}_n^{(k,r)} \equiv$ the number of trials required for r^{th} occurrence of a pattern of length at least $k+2$ of two failures separated by at least k successes in overlapping patterns,

$$\overline{X}_n^{(k,1)} = \overline{X}^{(k)},\ \overline{Y}_n^{(k,1)} = \overline{Y}^{(k)}.$$

Then, we have the following results.

THEOREM 4.1. *For* $n \geq r(k+1)+2$, $k \geq 1$,

$$P\left\{\overline{X}_n^{(k)} = r\right\} = \sum_{s=rk}^{n-r-1} \sum_{d=0}^{s-rk} (d+1)\binom{n-s-1}{r} \times N(s-rk-d, n-s-r-1; \sim k)p^s q^{n-s}, \tag{4.1}$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$.

PROOF. In this case, $(r+1)$ failures can be used to form r patterns. Then, the remaining $n-s-r-1$ failures are to be arranged in $r+1$ places generated by r patterns unconditionally and the number of ways in which this can be done is

$$\binom{n-s-r-1+r+1-1}{r+1-1} = \binom{n-s-1}{r}.$$

For placement of successes, we observe that after placing rk successes in r patterns, the remaining successes, *i.e.*, $s-rk$ in number, are to be placed in places generated by $n-s-r-1$ failures in such a way that at no place, except possibly in the beginning and the end, a success run can be of exact length k. If d is the number of successes put in the beginning and the end together, then (4.1) can be obtained using the same argument as in non-overlapping patterns. □

THEOREM 4.2. *For* $n \geq r(k+1)+2$, $k \geq 1$,

$$P\left\{\overline{Y}_n^{(k)} = r\right\} = \sum_{s=rk}^{n-r-1} \binom{n-s-1}{r} \sum_{d=0}^{s-rk} \sum_{s_1=0}^{s-rk-d} N(s-rk-d, n-s-1; k-1)p^s q^{n-s}, \tag{4.2}$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$.

Similarly, the distributions of $\overline{Z}_n^{(k)}$, $\overline{X}_n^{(k)}$ and $\overline{Y}_n^{(k)}$ can be obtained from the distributions of $Z_n^{(k)}$, $X_n^{(k,r)}$ and $Y_n^{(k,r)}$ by replacing the number of ways in which patterns can be arranged and then rearranging the limits.

5. MARKOV DEPENDENT TRIALS

Now, instead of independent Bernoulli trials, we consider the case of two-state, homogeneous $\{0,1\}$ valued Markov chain $X_1, X_2, \ldots, X_n$ with the possible outcomes a success (S) and a failure (F) and with transition probability matrix

$$\text{From}\quad \begin{matrix} \\ S \\ F \end{matrix} \quad \overset{\text{To}}{\underset{}{\begin{matrix} S & F \end{matrix}}} \atop \begin{pmatrix} p_1 & q_1 \\ p_2 & q_2 \end{pmatrix}; \quad p_t + q_t = 1;\ t = 1, 2. \tag{5.1}$$

Define $S_n \equiv$ the number of successes in n trials. We then have the following theorems.

THEOREM 5.1. *For initial trial a success and for $n \geq r(k+2)$, $k > 1$, $s = rk$,*

$$P\{X_n^{(k)} = r, S_n = s\} = \binom{n - rk - r}{r} \left(p_2 p_1^{k-1} q_1\right)^r q_2^{n-rk-r}, \tag{5.2}$$

and for $s = rk+1, \ldots, n-2r$,

$$\begin{aligned} &P\{X_n^{(k)} = r, S_n = s\} \\ &= \binom{n-s-r}{r} \sum_{s_1=0}^{s-rk} \sum_{s_2=0}^{s-rk-s_1} \sum_{s_3=1}^{\min\{r(k-1),s-rk-s_2-s_1\}} \\ &\times \sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min(n-s-2r,s_3)} N(s_3, r_1; 1, \sim k) \\ &\times \sum_{r_2=1}^{\min(r,s-rk-s_1-s_2-s_3)} \binom{r}{r_2} M(s - s_1 - s_2 - s_3 - rk, r_2; 1) \\ &\times I'_{[i,0]}\left(\frac{p_2}{p_1}\right) p_1^{(k-1)r+s_1} p_2^{s-(k-1)r-s_1} q_1^{r+r_1+r_2-1} q_2^{n-r-r_1-r_2+1}, \end{aligned} \tag{5.3}$$

where

$$I'_{[i,0]}(u) = \begin{cases} 1, & \textit{if} \quad i = 0, \\ u, & \textit{otherwise}, \end{cases}$$

and

$$N(s_3, r_1; 1, \sim k) = \textit{number of ways of putting } s_3 \textit{ successes into } r_1 \textit{ cells such that each cell has at least one success but not exactly } k \textit{ successes,}$$

$$\textit{i.e.,}\quad N(s_3, r_1; 1, \sim k) = N(s_3 - 1, r_1; \sim k - 1),$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$.

PROOF. The four possible outcomes of a sequence are:

(i) $S \cdots S$, *i.e.*, a sequence starting with a success and ending with a success;

(ii) $S \cdots F$, *i.e.*, a sequence starting with a success and ending with a failure;

(iii) $F \cdots S$, *i.e.*, a sequence starting with a failure and ending with a success;

(iv) $F \cdots F$, *i.e.*, a sequence starting with a failure and ending with a failure.

Partition the sequence of outcomes at the points where the first success run (followed by the first failure) ends and where the last success run (preceded by the last failure) begins.

Let s_1 be the number of successes in the first subsequence and s_2 be the number of successes in the last subsequence.

$$\underbrace{\cdots\cdots}_{s_1} \mid F \cdots\cdots F \mid \underbrace{\cdots\cdots}_{s_2}$$

$$\text{I} \qquad\qquad \text{II} \qquad\qquad \text{III}$$

If initial trial is a success, then for a sequence of type (i), $1 \le s_1 \le s - rk$, $1 \le s_2 \le s - rk - s_1$ with probability contribution $p_2 p_1^{s_1+s_2-1}$ to the total probability; for type (ii), $1 \le s_1 \le s - rk, s_2 = 0$ with probability contribution $p_1^{s_1}$; for type (iii), $s_1 = 0$, $1 \le s_2 \le s - rk$ with probability contribution $p_2 p_1^{s_2-1}$; and for type (iv), $s_1 = s_2 = 0$.

Now, the subsequence II contains $n - s_1 - s_2$ trials, in which there are r patterns, exhausting rk successes and $2r$ failures. In these patterns, rk successes and r failures (after successes) contribute $(p_2 p_1^{k-1} q_1)^r$ to the total probability.

Out of remaining $n - s_1 - s_2 - r(k+2)$ trials, $s - s_1 - s_2 - rk$ are successes and $n - s - 2r$ are independent failures (not belonging to any pattern). The number of ways in which r patterns and $n - s - 2r$ failures can be arranged unconditionally is $\binom{n-s-r}{r}$. Then $n-s-r-2$ places are available for placing $s-s_1-s_2-rk$ successes in such a manner that after a failure, no success run of length k appears.

Let s_3 be the number of successes, to be placed in r_1 places after failures such that at least one success is placed and no place gets equal to k successes. The number of ways in which this can be done is $N(s_3-1, r_1; \sim k-1)\binom{n-s-2r}{r_1}$ with probability $(p_2q_1)^r p_1^{s_3-r_1}$.

Remaining $s-s_1-s_2-s_3-rk$ successes are arranged in r_2 places after patterns in such a way that at least one success is placed in each cell. The number of ways in which this can be done is $M(s-s_1-s_2-s_3-rk, r_2; 1)\binom{r}{r_2}$ with probability $(p_2q_1)^{r_2}p_1^{s-s_1-s_2-s_3-rk}$. Remaining failures are placed with probability q_2 each, *i.e.*, $q_2^{n-s-r-r_1-r_2}$.

Combining these components we get (5.3). (5.2) is obvious. Similarly results can be obtained when initial trial is a failure. □

THEOREM 5.2. *For initial trial a success and for* $n \geq r(k+2)$, $k > 1$,

$$
\begin{aligned}
P\{Y_n^{(k)} = r, S_n = s\}
&= \binom{n-s-r}{r} \sum_{s_0=0}^{rk} \sum_{s_1=0}^{s-s_0} \sum_{s_2=0}^{s-s_0-s_1} \sum_{s_3=1}^{\min\{r(k-1), s-s_1-s_2\}} \\
&\quad \times \sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min\{(n-s-2r), s_3\}} \binom{n-s-2r}{r_1} \sum_{r_2=1}^{\min(r, s-s_0-s_1-s_2-s_3)} \binom{r}{r_2} \\
&\quad \times N(s_3, r_1; 1, k) M(s-s_0-s_1-s_2-s_3, r_2; 1) I'_{[i,0]}\left(\frac{p_2}{p_1}\right) \\
&\quad \times p_1^{s-r-r_1-r_2} p_2^{r+r_1+r_2} q_1^{r+r_1+r_2} q_2^{n-s-r-r_1-r_2},
\end{aligned}
\tag{5.4}
$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$, $s = rk, \ldots, n-2r$.

PROOF. In this case, r pairs of failures are separated by at least k successes. In these pairs, we place s_0 successes in such a way that each pair gets at least k successes.

Now, we have $s-s_0$ remaining successes, r patterns and $n-s-2r$ failures, the $s-s_0$ successes are to be arranged in such a way that $\{Y_n^{(k)} = r\}$ remains true. The probability contribution of these $s_0+s_1+s_2$ successes and r failures (after successes) to the total probability is $p_1^{s_0+s_1+s_2-r}(p_2q_1)^r$ for sequences of type (ii) and (iv) and $p_1^{s_0+s_1+s_2-r-1}p_2^{r+1}q_1^r$ for sequences of type (i) and (iii).

The number of ways in which r patterns and $n-s-2r$ failures can be arranged unconditionally is $\binom{n-s-r}{r}$. Then $n-s-r-2$ places are available for placing $s-s_0-s_1-s_2$ successes in such a manner that after a failure, no success run of length k or more appears.

Let s_3 be the number of successes, to be placed in r_1 places after failures such that each cell gets at least one success and no cell gets more than $k-1$ successes. The number of ways in which these could be done is $N(s_3, r_1; 1, k-1)\binom{n-s-2r}{r_1}$ with probability $(p_2q_1)^{r_1}p_1^{s_3-r_1}q_2^{n-s-2r-r_1}$.

Remaining $s-s_0-s_1-s_2-s_3$ successes are arranged in r_2 places (if any) after patterns in such a way that at least one success is placed in each cell. The number of ways in which this could be done is $M(s-s_0-s_1-s_2-s_3, r_2; 1, k-1)\binom{r}{r_2}$ with probability $(p_2q_1)^{r_2}p_1^{s-s_0-s_1-s_2-s_3}q_2^{r-r_2}$. Remaining failures are placed with probability q_2 each, *i.e.*, $q_2^{n-s-r-r_1-r_2}$.

Combining these components, we get (5.4). □

THEOREM 5.3. *For initial trial a success and for $n \geq r(k+2)$, $k > 1$,*

$$\begin{aligned}
&P\{Z_n^{(k)} = r, S_n = s\} \\
&= \binom{n-s-r}{r} \sum_{s_0=0}^{rk} \sum_{s_1=0}^{s-s_0} \sum_{s_2=0}^{s-s_0-s_1} \sum_{r_1=1}^{\min\{(n-s-2r), s-s_0-s_1-s_2\}} \\
&\quad \times \sum_{s_3=(k+1)r_1}^{\min\{r(k-1), s-s_0-s_1-s_2\}} \sum_{r_2=1}^{\min(r, s-s_0-s_1-s_2-s_3)} \binom{r}{r_2} M(s_3, r_1; k+1) \\
&\quad \times M(s-s_0-s_1-s_2-s_3, r_2; 1)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right) p_1^{s-r-r_1-r_2} p_2^{r+r_1+r_2} \\
&\quad \times q_1^{r+r_1+r_2} q_2^{n-s-r-r_1-r_2},
\end{aligned} \tag{5.5}$$

where $r = 0, 1, 2, \ldots, [n/(k+2)]$, $s = k, \ldots, n-2r$.

In the following theorems, we shall obtain the joint probability distributions of these patterns.

THEOREM 5.4. *For initial trial a success and for $0 \leq r_1 \leq r_2 \leq [n/(k+2)]$, $k > 1$,*

$$\begin{aligned}
&P\{X_n^{(k)} = r_1, Y_n^{(k)} = r_2, S_n = s\} \\
&= \binom{n-s-r_2}{r_2}\binom{r_2}{r_1} \sum_{s_0=r_2k}^{s} \sum_{s_1=0}^{s-s_0} \sum_{s_2=0}^{s-s_0-s_1} \sum_{s_3=1}^{\min\{r_2(k-1), s-s_0-s_3\}}
\end{aligned}$$

$$\times \sum_{r_3=[s_3/(k-1)]}^{\min(r_2,s-s_0-s_1-s_2-s_3)} \binom{n-s-2r_2}{r_3} \sum_{r_4=1}^{\min(n-s-2r_2,s_3)} \binom{r_2}{r_4} \tag{5.6}$$
$$\times M(s-s_0-s_1-s_2-s_3,r_4;1)M(s_0-r_1k,r_2-r_1;k+1)$$
$$\times N(s_3,r_3;1,k)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right) p_1^{s-r_2-r_3-r_4} p_2^{r_2+r_3+r_4}$$
$$\times q_1^{r_2+r_3+r_4} q_2^{n-s-r_2-r_3-r_4},$$

where $r = 0,1,2,\ldots,[n/(k+2)], s = rk,\ldots,n-2r_2$.

PROOF. The event defined in the L.H.S. of (5.6) is equivalent to the event

$$\{X_n^{(k)} = r_1, Y_n^{(k+1)} = r_2 - r_1\}. \tag{5.7}$$

For this event, $2r_1$ failures are exhausted in occurrence of $\{X_n^{(k)} = r_1\}$ and $2(r_2 - r_1)$ are exhausted in occurrence of $\{Y_n^{(k+1)} = r_2 - r_1\}$. The number of ways in which r_1 patterns can be selected out of r_2 is $\binom{r_2}{r_1}$. We are left with $n - s - 2r_1 - 2(r_2 - r_1) = n - s - 2r_2$ failures and r_2 $(= r_1 + r_2 - r_1)$ patterns. Then, (5.6) can be obtained on similar lines as earlier results. □

THEOREM 5.5. *For initial trial a success and for* $0 \le r_1 \le r_2 \le [n/(k+2)]$, $k > 1$,

$$P\{X_n^{(k)} = r_1, Z_n^{(k)} = r_2, S_n = s\}$$
$$= \binom{n-s-r_2}{r_2}\binom{r_2}{r_1} \sum_{s_0=r_2k}^{s} \sum_{s_1=0}^{s-s_0} \sum_{s_2=0}^{s-s_0-s_1} \sum_{r_3=1}^{\min(n-s-2r_2,s-s_0-s_1-s_2)}$$
$$\times \sum_{s_3=(k+1)r_2}^{\min(s-s_0-s_1-s_2,r_2)} \binom{n-s-2r_2}{r_3} \sum_{r_4=1}^{\min(r_2,s-s_0-s_1-s_2-s_3)} \binom{r_2}{r_4} \tag{5.8}$$
$$\times M(s-s_0-s_1-s_2-s_3,r_4;1)M(s_3,r_3;k+1)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right)$$
$$\times p_1^{s-r_2-r_3-r_4} p_2^{r_2+r_3+r_4} q_1^{r_2+r_3+r_4} q_2^{n-s-r_2-r_3-r_4},$$

where $r = 0,1,2,\ldots,[n/(k+2)]$, $s = rk,\ldots,n-2r_2$.

NOTE. In above results, if any of the factors is equal to zero, then the corresponding components contribute unity.

6. WAITING TIME DISTRIBUTIONS FOR MARKOV DEPENDENT TRIALS

Now, we obtain waiting time distributions of the patterns defined in Section 5.

THEOREM 6.1. *For initial trial a success and $k > 1$,*

$$\begin{aligned} &P\{X^{(k)} = n\} \\ &= \sum_{s=k}^{n-2} \sum_{s_1=0}^{s-k} \sum_{\substack{s_2=0 \\ (s_2 \neq k)}}^{s-s_1-k} \sum_{r_1=0}^{\min(s-s_1-s_2-k, n-s-3)} \binom{n-s-3}{r_1} \\ &\times N(s-s_1-s_2-k-r_1, r_1; \sim k-1)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right) \\ &\times p_1^{s-r_1-1}(p_2 q_1)^{r_1+1} q_2^{n-s-r_1-1}, \end{aligned} \tag{6.1}$$

where $n = k+2, k+3, \ldots$

PROOF. Partition the sequence of outcomes in two subsequences, at the point where the event has occurred.

$$\cdots \mid F\underbrace{S\cdots S}_{k}F$$
$$\text{I} \qquad \text{II}$$

Then in subsequence I, there is no success run of length k except possibly in the beginning. Further let s_1 be the number of successes in the beginning of subsequence I and s_2 ($\neq k$) be the number of successes in the end. Then $s - s_1 - s_2 - k$ successes are to be distributed in $n - s - 3$ places (generated by $n - s - 2$ failures) such that at no place, a success run of length k occurs.

Let at r_1 places after failures, $s - s_1 - s_2 - k$ successes be placed in such a manner that at least one success is placed at each cell and no cell gets k successes. The number of ways of choosing r_1 places is $\binom{n-s-3}{r_1}$ and the number of ways of placing successes is $N(s - s_1 - s_2 - k - r_1, r_1; \sim (k-1))$. Hence (6.1). □

THEOREM 6.2. *For initial trial a success and $k > 1$,*

$$\begin{aligned} &P\{X^{(k,r)} = n\} \\ &= \sum_{s=(r-1)k}^{n-2r} \binom{n-s-r-1-\mathrm{I}[s_2,0]}{r-2} \sum_{s_1=0}^{s-(r-1)k-1} \sum_{s_2=0}^{s-s_1-(r-1)k} \\ &\times \sum_{s_3=1}^{\min\{(r-1)(k-1), s-(r-1)k-s_1-s_2\}} \sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min\{n-s-2(r-1), s_3\}} \binom{n-s-2r-1}{r-1} \end{aligned}$$

$$\times \sum_{r_2=1}^{\min(r-1,s-r-s_1-s_2-s_3)} \binom{r-2}{r_2} N(s_3 - r_1, r_1; \sim k-1)$$

$$\times M\big(s - s_1 - s_2 - s_3 - (r-1)k, r_2; 1\big)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right)$$

$$\times p_1^{s_1+s_2+(k-1)(r-1)} p_2^{s-s_1-s_2-(k-1)(r-1)} q_1^{r+r_1+r_2-2} q_2^{n-s-r-r_1-r_2+2}$$

$$+ \sum_{s=(r-1)k}^{n-2r} \binom{n-s-r-2}{r-1} \sum_{s_1=0}^{s-(r-1)k-1} \sum_{\substack{s_2=1\\(s_2\neq k)}}^{s-s_1-(r-1)k} \tag{6.2}$$

$$\times \sum_{s_3=1}^{\min\{(r-1)(k-1),s-(r-1)k-s_1-s_2\}} \sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min\{n-s-2(r-1),s_3\}} \binom{n-s-2r-1}{r_1}$$

$$\times \sum_{r_2=1}^{\min(r-1,s-r-s_1-s_2-s_3)} \binom{r-2}{r_2} N(s_3 - r_1, r_1; \sim k-1)$$

$$\times M\big(s - s_1 - s_2 - s_3 - (r-1)k, r_2; 1\big)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right)$$

$$\times p_1^{s_1+s_2+(k-1)(r-1)-1} p_2^{s-s_1-s_2-(k-1)(r-1)+1}$$

$$\times q_1^{r+r_1+r_2-2} q_2^{n-s-r-r_1-r_2+2},$$

where $\mathrm{I}_{[s_2,0]} = 1$ *if* $s_2 = 0$; 0 *otherwise*, $n \geq r(k+2)$.

PROOF. Again, in this case also, the sequence always ends with a pattern. After forming r patterns, we have $n-s-2r$ failures, $s-rk$ successes and r patterns. Partition the sequence of outcomes at the point where the last event has occurred.

In the remaining $n-k-2$ trials, we have $n-s-2r$ failures, $s-rk$ successes and $r-1$ patterns to be arranged in such a manner that after an independent failure, a success run of length r should not appear. The number of ways in which $n-s-2r$ failures and r patterns can be arranged is $\binom{n-s-r-1}{r-1}$ as in case (ii) of Theorem 5.1.

Consider subsequence I. If the last trial is a failure, then the required probability is given by

$$\sum_{s=(r-1)k}^{n-2r} \binom{n-s-r-1}{r-1} \sum_{s_1=0}^{s-(r-1)k} \sum_{s_3=1}^{s-(r-1)k-s_1} \sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min\{n-s-2(r-1),s_3\}}$$

$$\times\binom{n-s-2r}{r_1}\sum_{r_2=1}^{\min(r-1,s-r-s_1-s_3)} N(s_3-r_1,r_1;\sim k-1)$$
$$\times M(s-s_1-s_3-(r-1)k,r_2;1)p_1^{s_1+(k-1)(r-1)}p_2^{s-s_1-(k-1)(r-1)} \tag{6.3}$$
$$\times q_1^{r+r_1+r_2-2}q_2^{n-s-r-r_1-r_2+2}.$$

If the last trial is a success, then two cases arise:

(i) If the failure preceding the last run of successes is independent failure, then $s_2 \neq k$ so the required probability is given by

$$\sum_{s=(r-1)k}^{n-2r}\binom{n-s-r-2}{r-1}\sum_{s_1=0}^{s-(r-1)k-1}\sum_{\substack{s_2=1\\(s_2\neq k)}}^{s-s_1-(r-1)k}$$
$$\times\sum_{s_3=1}^{\min\{(r-1)(k-1),s-(r-1)k-s_1-s_2\}}\sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min\{n-s-2(r-1),s_3\}}\binom{n-s-2r}{r_1}$$
$$\times\sum_{r_2=1}^{\min(r-1,s-r-s_1-s_2-s_3)}\binom{r-2}{r_2}N(s_3-r_1,r_1;\sim k-1) \tag{6.4}$$
$$\times M(s-s_1-s_2-s_3-(r-1)k,r_2;1)\,I'_{[i,0]}\left(\frac{p_2}{p_1}\right)$$
$$\times p_1^{s_1+s_2+(k-1)(r-1)-1}p_2^{s-s_1-s_2-(k-1)(r-1)+1}$$
$$\times q_1^{r+r_1+r_2-2}q_2^{n-s-r-r_1-r_2+2}.$$

(ii) If the failure preceding the last run of successes belongs to a pattern, then the required probability is given by

$$\sum_{s=(r-1)k}^{n-2r}\binom{n-s-r-2}{r-2}\sum_{s_1=0}^{s-(r-1)k-1}\sum_{s_2=1}^{s-s_1-(r-1)k}$$
$$\times\sum_{s_3=1}^{\min\{(r-1)(k-1),s-(r-1)k-s_1-s_2\}}\sum_{r_1=[s_3/(k-1)]}^{\min\{n-s-2(r-1),s_3\}}\binom{n-s-2r-1}{r_1}$$
$$\times\sum_{r_2=1}^{\min(r-1,s-r-s_1-s_2-s_3)}\binom{r-2}{r_2}N(s_3-r_1,r_1;\sim k-1) \tag{6.5}$$
$$\times M(s-s_1-s_2-s_3-(r-1)k,r_2;1)I'_{[i,0]}\left(\frac{p_2}{p_1}\right)$$

$$\times p_1^{s_1+s_2+(k-1)(r-1)-1} p_2^{s-s_1-s_2-(k-1)(r-1)+1}$$
$$\times q_1^{r+r_1+r_2-2} q_2^{n-s-r-r_1-r_2+2}.$$

Combining (6.3), (6.4) and (6.5), we get (6.2). □

THEOREM 6.3. *For initial trial a success and $k > 1$,*

$$\begin{aligned} P\{Y^{(k)} = n\} &= \sum_{s=k}^{n-2} \sum_{s_0=k}^{s} \sum_{s_1=0}^{s-s_0} \sum_{\substack{s_2=0 \\ (s_2 \le k)}}^{s-s_0-s_1} \sum_{r_1=0}^{\min(s-s_0-s_1-s_2,n-s-3)} \binom{n-s-3}{r_1} \\ &\times N(s-s_0-s_1-1, r_1; 1, k-1)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right) p_1^{s-r_1-1} \\ &\times (p_2 q_1)^{r_1+1} q_2^{n-s-r_1-1}, \end{aligned} \tag{6.6}$$

where $n = k+2, k+3, \ldots$

THEOREM 6.4. *For initial trial a success and $k > 1$,*

$$\begin{aligned} P\{Y^{(k,r)} = n\} &= \sum_{s=(r-1)k}^{n-2r} \sum_{s_0=k}^{s} \binom{n-s-r-1-\mathrm{I}_{[s_2,0]}}{r-2} \sum_{s_1=0}^{s-s_0} \sum_{s_2=0}^{s-s_0-s_1} \\ &\times \sum_{s_3=1}^{\min\{(r-1)(k-1),s-s_0-s_1-s_2\}} \sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min\{n-s-2(r-1),s_3\}} \binom{n-s-2r-1}{r_1} \\ &\times \sum_{r_2=1}^{\min(r-1,s-s_0-s_1-s_2-s_3)} \binom{r-2}{r_2} N(s_3, r_1; 1, k-1) \\ &\times M(s-s_0-s_1-s_3, r_2; 1)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right) p_1^{s_0+s_1+s_2} p_2^{s-s_0-s_1-s_2} \\ &\times q_1^{r+r_1+r_2-2} q_2^{n-s-r-r_1-r_2+2} \\ &+ \sum_{s=(r-1)k}^{n-2r} \sum_{s_0=k}^{s} \binom{n-s-r-2}{r-1} \sum_{s_1=0}^{s-s_0-s_1} \sum_{\substack{s_2=1 \\ (s_2 \le k)}}^{s-s_0-1} \end{aligned} \tag{6.7}$$

$$\times \sum_{s_3=1}^{\min\{(r-1)(k-1),s-s_0-s_1-s_2\}} \sum_{r_1=\lceil s_3/(k-1)\rceil}^{\min\{n-s-2(r-1),s_3\}} \binom{n-s-2r-1}{r_1}$$

$$\times \sum_{r_2=1}^{\min\{r-1,s-s_0-s_1-s_2-s_3\}} \binom{r-2}{r_2} N(s_3, r_1; 1, k-1)$$

$$\times M(s-s_0-s_1-s_2-s_3, r_2; 1)\, I'_{[i,0]}\left(\frac{p_2}{p_1}\right)$$

$$\times p_1^{s_0+s_1+s_2-1} p_2^{s-s_0-s_1-s_2+1} q_1^{r+r_1+r_2-2} q_2^{n-s-r-r_1-r_2+2},$$

where $n \geq r(k+2)$.

Similarly the distributions of $Z^{(k)}$ and $Z^{(k,r)}$ can be derived.

ACKNOWLEDGEMENTS

The authors are grateful to the referees for their valuable comments.

REFERENCES

Altshul, S. F. and Erickson, B. W. (1988). "Significance levels of biological sequence comparisons using non-linear similarity function", *Bulletin of Mathematical Biology,* **50**, 77–92.

Aki, S. (1985). "Discrete distributions of order k on a binary sequence", *Annals of the Institute of Statistical Mathematics,* **37**, 205–224.

Aki, S. (1992). "Waiting time problems for a sequence of discrete random variables", *Annals of the Institute of Statistical Mathematics.* **44**, 363–378.

Arratia, R., Goldstein, L. and Gordon, L. (1989). "Two moments suffice for Poisson approximations : The Chen-Stein method". *The Annals of Probability,* **17**, 9–25.

Balakrishnan, N., Balasubramanian, K. and Viveros, R. (1993). "On sampling inspection plans based on the theory of runs", *The Mathematical Scientist,* **18**, 113–126.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications,* Vol. I, 3rd ed., John Wiley & Sons, New York-London-Sydney.

Fu, J. C. (1996). "Distribution theory of runs and patterns associated with a sequence of multi-state trials", *Statistica Sinica,* **6**, 957–974.

Glaz, J. (1983). "Moving window detection for discrete data". *IEEE Transactions on Information Theory,* **IT-29**, 457–462.

Glaz, J. (1996). "Discrete scan statistics with applications to minefield detection". *Proceedings of the Conference of SPEI,* **2765**, 420–429.

Glaz, J. and Balakrishnan, N. (1999). *Scan Statistics and Applications.* Birkhäuser Verlag.

Goldstein, L. (1990). "Poisson approximations and DNA sequence matching", *Communications in Statistics-Theory & Methods,* **19**. 4167–4179.

Hirano, K. (1986). "Some properties of the distributions of order k", In *Fibonacci Numbers and Their Applications* (A. N. Philippou, G. E. Bergum and A. F. Horadam, eds.), 43–53, Reidel, Dordrecht.

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.

Koutras, M. V. (1996). "On a waiting time distribution in a sequence of Bernoulli trials", *Annals of the Institute of Statistical Mathematics,* **48**, 789–806.

Koutras, M. V. and Alexandrou, V. A. (1995). "Runs, scans, and urn model distributions : A unified Markov chain approach", *Annals of the Institute of Statistical Mathematics,* **47**, 743–766.

Koutras, M. V. and Papastavridis, S. G. (1993). "On the number of runs and related statistics", *Statistica Sinica,* **3**, 277–294.

Krauth, J. (1992a). "Bounds for the upper-tail probabilities of the circular ratchet scan statistics", *Biometrics,* **48**, 1177–1185.

Krauth, J. (1992b). "Bounds for the upper tail probabilities of the linear ratchet scan statistics", In *Analyzing and Modelling Data and Knowledge* (M. Schader, ed.), 51–61, Springer-Verlag, Berlin.

Ling, K. D. (1988). "On binomial distributions of order k", *Statistics & Probability Letters,* **6**, 247–250.

Leung, M. Y. and Yamashita, T. E. (1999). "Applications of the scan statistic in DNA sequence analysis", In *Scan Statistics and Applications* (J. Glaz and N. Balakrishnan, eds.), 269–286, Birkhäuser Boston, Boston.

Mood, A. M. (1940). "The distribution theory of runs", *Annals of Mathematical Statistics,* 367–392.

Philippou, A. N. and Makri, F. S. (1986). "Successes, runs and longest runs", *Statistics & Probability Letters,* **4**, 211–215.

Papastavridis, S. (1990). "m-Consecutive-k-out-of-n : F systems", *IEEE Transactions on Reliability,* **39**, 386–388.

Rajarshi, M. B. (1974). "Success runs in a two-state Markov chain", *Journal of Applied Probability,* **11**, 190–192.

Rosenstock, H. B. and Maradudin, A. A. (1962). "On certain combinatorial problems", *American Journal of Physics,* **30**, 330–332.

Wallenstein, S., Weinberg, C. R. and Gould, M. (1989). "Testing for a pulse in seasonal event data", *Biometrics,* **45**, 817–830.