Distribution Theory of Runs: A Markov Chain Approach

J. C. Fu; M. V. Koutras

Journal of the American Statistical Association, Vol. 89, No. 427. (Sep., 1994), pp. 1050-1058.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199409%2989%3A427%3C1050%3ADTORAM%3E2.0.CO%3B2-L

Journal of the American Statistical Association is currently published by American Statistical Association.

Distribution Theory of Runs: A Markov Chain Approach

J. C. Fu and M. V. Koutras*

The statistics of the number of success runs in a sequence of Bernoulli trials have been used in many statistical areas. For almost a century, even in the simplest case of independent and identically distributed Bernoulli trials, the exact distributions of many run statistics still remain unknown. Departing from the traditional combinatorial approach, in this article we present a simple unified approach for the distribution theory of runs based on a finite Markov chain imbedding technique. Our results cover not only the identical Bernoulli trials, but also the nonidentical Bernoulli trials. As a byproduct, our results also yield the exact distribution of the waiting time for the mth occurrence of a specific run.

KEY WORDS: Bernoulli random variables; Patterns; Reliability; Transition probability matrix.

1. INTRODUCTION

The concept of runs has been used in various areas. For example, in the early 1940s it was used in the area of hypothesis testing (run-test) by Wald and Wolfowitz (1940) and Wolfowitz (1943) and in the area of statistical quality control by Mosteller (1941) and Wolfowitz (1943). Recently, it has been successfully used in many other areas, such as reliability of engineering systems, quality control, DNA sequencing, psychology, ecology, and radar astronomy (see Schwager 1983).

There are various definitions of runs (see Schwager 1983). We shall not give a general definition here, because new advances and application of new criteria to new problems will probably soon render most definitions obsolete. We shall only define the statistics of runs when the statistical model is clearly specified.

In this article we consider only the statistics of success runs of both identical and nonidentical independent Bernoulli trials. In this special model a success run is defined as a sequence of consecutive successes (S) preceded and succeeded by failures (F). The number of successes in a run will be referred to as its length.

Let us first define the five most important and frequently used statistics of success runs associated with nonidentical independent Bernoulli trials $X_1, \ldots, X_n$ with success probabilities p_t and failure probabilities $q_t = 1 - p_t$, for $t = 1, \ldots, n$. Given n and k ($1 \le k \le n$), let

(a) $E_{n,k}$ be the number of success runs of size exactly k, in the sense of Mood's (1940) counting;
(b) $G_{n,k}$ be the number of success runs of size greater than or equal to k;
(c) $N_{n,k}$ be the number of nonoverlapping consecutive k successes, in the sense of Feller's (1968) counting;
(d) $M_{n,k}$ be the number of overlapping consecutive k successes, in the sense of Ling's (1988) counting; and
(e) L_n be the size of the longest success run (see Gibbons 1971).

To make these definitions clear and transparent, we give the following example: suppose that 10 different coins are tossed, with outcomes $SSFSSSSFFF$. Then $L_{10} = 4$, and for $k = 2$, we have $E_{10,2} = 1$, $G_{10,2} = 2$, $N_{10,2} = 3$, and $M_{10,2} = 4$. From the definitions (a)–(f) it is clear that the following relationships are always true:

$$E_{n,k} \le G_{n,k} \le N_{n,k} \le M_{n,k},$$

$$E_{n,k} = G_{n,k} - G_{n,k+1},$$

and

$$L_n < k \quad \text{iff } N_{n,k} = 0.$$

Currently, except for a few special cases, the exact distributions of the five run statistics (a)–(f) are mainly unknown, especially for nonidentical Bernoulli trials.

The distribution theory of run has had a long and stormy history. The theory seems to have started at the end of the nineteenth century. There were considerable research works around 1940, including those of Wishart and Hirsheld (1936), Cochran (1938), Mood (1940), Wald and Wolfowitz (1940), and Wolfowitz (1943). Most of the publications during this period focused on studying the conditional distributions of success runs given n_1 successes and $n_2 = n - n_1$ failures. Let γ_{1i} (γ_{2i}) denote the number of success (failure) runs of size i given n_1 successes and n_2 failures. Also let $\gamma_1 = \sum \gamma_{1i}$ ($\gamma_2 = \sum \gamma_{2i}$) be the total number of success (failure) runs. Stevens (1939) and Wald and Wolfowitz (1940) proved that

$$P(\gamma_1 = x) = \frac{\binom{n_1 - 1}{x - 1}\binom{n_2 + 1}{x}}{\binom{n}{r_1}}. \tag{1}$$

Mood (1940) obtained the exact distributions of γ_{ji}, $j = 1, 2$ and their first two moments, but his formulas are rather complex. In the case of unknown n_1 and n_2, as far as we are aware there is no explicit formula for $E_{n,k}$.

There was very little published during the period 1950–1970. Recently, during the late 1980s and early 1990s, this

* J. C. Fu is Professor, Department of Statistics, The University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2. M. V. Koutras is Associate Professor, Department of Mathematics, University of Athens, Athens, Greece 15784. This work was supported in part by Natural Science and Engineering Research Council of Canada Grant NSERC A-9216. The authors thank the referees for helpful comments and suggestions.

Journal of the American Statistical Association
September 1994, Vol. 89, No. 427, Theory and Methods

area became very active again. For iid Bernoulli trials, Philippou and Makri (1986) and Hirano (1986) independently obtained the exact distribution for the random variable $N_{n,k}$, namely,

$$P(N_{n,k}=x) = \sum_{m=0}^{k-1} \sum_{x_1+2x_2+\cdots+kx_k=n-m-kx} \binom{x_1+\cdots+x_k+x}{x_1,\ldots,x_k,x} \times p^n \left(\frac{q}{p}\right)^{x_1+\cdots+x_k} \quad (2)$$

for $x=0,1,\ldots,[n/k]$. Godbole (1990) gave an alternative formula for $N_{n,k}$,

$$P(N_{n,k}=x) = \sum_{[[n-kx]/k]\le y\le n-kx} q^y p^{n-y} \binom{y+x}{x} \times \sum_{0\le j\le[(n-kx-y)/k]} (-1)^j \binom{y+1}{j} \times \binom{n-kx-jk}{y} \quad (3)$$

for $x=0,1,\ldots,[n/k]$. The main advantage of Godbole's formula over Philippou and Makri's and Hirano's formula is that it is easier to evaluate by computer. The distribution of $N_{n,k}$ was called Binomial distribution of order k. It was studied by, for example, Aki (1985), Aki and Hirano (1984, 1986, 1988), Aki, Kuboki, and Hirano (1984), Hirano (1986), and Philippou (1986). Ling (1988) gave recurrence relation formulas for the distribution and the probability generating function (pgf) of the random variable $M_{n,k}$ (iid model only). Hirano, Aki, Kashiwagi, and Kuboki (1991) and Hirano and Aki (1992) obtained an explicit form of the pgf. Chrysaphinou, Papastavridis, and Tsapelas (1993), extending Ling's recursive formulas to the non-iid model, proved that

$$\begin{aligned} P(M_{n,k}=x) &= \prod_{i=1}^{n} p_i, \quad \text{if } x=n-k+1, \\ &= (q_1p_n+q_np_1)\prod_{i=2}^{n-1} p_i, \quad \text{if } x=n-k, \\ &= \sum_{i=0}^{x+k-1}\left(\prod_{m=n-i+1}^{n+1} p_m\right) q_{n-1} \\ &\quad \times P(M_{n-i-1,k}=x-\max\{0,i-k+1\}), \\ &\qquad \text{if } 0\le x<n-k. \quad (4) \end{aligned}$$

The foregoing formulas of $N_{n,k}$ and $M_{n,k}$ are incredibly complex and hard to compute. There are many approximation formulas for those run statistics that can be found in Walsh's (1965) *Handbook of Nonparametric Statistics.* As Mood (1940) wrote, "the distribution problem is, of course, a combinatorial one, and the whole development depends on some identities in combinatory analysis." Mood's point of view had a great influence in current research works in this area. If one adopted Mood's suggested combinatorial method to obtain the formulas for the exact distributions of other nontrivial run statistics, it would be a very tedious if not impossible, task. This is the main reason why the exact distributions of so many run statistics still remain unknown.

In this manuscript we systematically develop formulas for the exact distributions of all five run statistics that are intuitively and conceptually easy to understand and are computationally simple.

To achieve these goals, we shall take a completely different approach to solving this problem. Basically, we adopt an approach similar to the one used by Fu (1986, 6.7a), Fu and Hu (1987), and Chao and Fu (1989, 1990), for the study of certain reliability problems. More specifically, we imbed the problem into a finite Markov chain. Then the distribution of the random variable of interest can be expressed in terms of transition probabilities of the Markov chain.

In Section 2 we introduce the concept of finite Markov chain imbedding and give our basic results. In Section 3 we obtain the distributions for all five random variables, $E_{n,k}$, $G_{n,k}$, $N_{n,k}$, $M_{n,k}$, and L_n, by applying the results of Section 2. We provide numerical examples of exact distributions of those run statistics in Section 4, and study the exact distribution of waiting time of first (or mth) consecutive success run of size k in Section 5. Finally, in section 6 we discuss this approach and its possible extensions.

2. FINITE MARKOV CHAIN

For given n, let $\Gamma_n = \{0, 1, \ldots, n\}$ be an index set and $\Omega = \{a_1, \ldots, a_m\}$ be a finite state space. For convenience, we use the random variable $X_{n,k}$ to represent any of the five random variables $E_{n,k}$, $G_{n,k}$, $N_{n,k}$, $M_{n,k}$, and L_n.

Definition 2.1. A nonnegative integer random variable $X_{n,k}$ can be imbedded into a finite Markov chain if:

(a) there exists a finite Markov chain $\{Y_t: t\in\Gamma_n\}$ defined on the finite state space Ω,
(b) there exists a finite partition $\{C_x, x=0,1,\ldots,l\}$ on the state space Ω, and
(c) for every $x=0,1,\ldots,l$, we have $P(X_{n,k}=x) = P(Y_n\in C_x)$.

Let Λ_t be the $m\times m$ transition probability matrix of the finite Markov chain $(\{Y_t: t\in\Gamma_n\}, \Omega)$. Let U_r be a $1\times m$ unit vector having 1 at the rth coordinate and 0 elsewhere, and let U_r' be the transpose of U_r. Finally, for every C_x, define the $1\times m$ vector

$$U(C_x) = \sum_{r: a_r\in C_x} U_r.$$

Theorem 2.1. If $X_{n,k}$ can be imbedded into a finite Markov chain, then

$$P(X_{n,k}=x) = \pi_0\left(\prod_{t=1}^{n}\Lambda_t\right)U'(C_x), \quad (5)$$

where $\pi_0 = (P(Y_0=a_1), P(Y_0=a_2), \ldots, P(Y_0=a_m))$ is the initial probability of the Markov chain.

Proof. For any state $a_r \in \Omega$, it follows from the Chapman–Kolmogorov equations that

$$P(Y_n = a_r) = \pi_0\left(\prod_{t=1}^{n} \Lambda_t\right) U_r'.$$

Because $X_{n,k}$ can be imbedded into the finite Markov chain, it is evident (from the definition) that for every x,

$$P(X_{n,k} = x) = P(Y_n \in C_x) = \sum_{a_r \in C_x} P(Y_n = a_r)$$

$$= \sum_{a_r \in C_x} \pi_0\left(\prod_{t=1}^{n} \Lambda_t\right) U_r' = \pi_0\left(\prod_{t=1}^{n} \Lambda_t\right) U'(C_x).$$

This completes the proof.

The moments $E(X_{n,k}^r)$, $r = 1, 2, \ldots$ can be written as

$$E(X_{n,k}^r) = \pi_0\left(\prod_{t=1}^{n} \Lambda_t\right) V_r', \tag{6}$$

where

$$V_r = \sum_{x=0}^{l} x^r U(C_x).$$

Similarly, the probability-generating function of $X_{n,k}$ can be written as

$$\varphi_{n,k}(s) = \pi_0\left(\prod_{t=1}^{n} \Lambda_t\right) W'(s), \tag{7}$$

where

$$W(s) = \sum_{x=0}^{l} s^x U(C_x).$$

If the Markov chain is homogeneous (iid case), that is, $\Lambda_t = \Lambda$ for all $t \in \Gamma_n$, then the exact distribution of the random variable $X_{n,k}$ can simply be expressed by

$$P(X_{n,k} = x) = \pi_0 \Lambda^n U'(C_x), \quad \forall\, x = 0, \ldots, l.$$

In view of these results, to find the distribution, moments, and probability-generating function for any imbeddable random variable, we need to construct only three things: (a) a proper state space Ω, (b) a proper partition $\{C_x\}$ for the state space, and (c) the transition probability matrix Λ_t associated with the imbedded Markov chain.

3. DISTRIBUTION OF RUNS

In a series of reliability papers by Fu (1986), Fu and Hu (1987), and Chao and Fu (1989, 1991), it was proven that the probability $P(N_{n,k} = 0)$ can be evaluated by the formula

$$P(N_{n,k} = 0) = \pi_0\left(\prod_{t=1}^{n} \Lambda_t\right) U',$$

where $\pi_0 = (1, 0, \ldots, 0)$, and $U = (1, 1, \ldots, 1, 0)$ are $1 \times (k+1)$ vectors and Λ_t is a $(k+1) \times (k+1)$ transition probability matrix given by

$$\Lambda_t = \left[\begin{array}{cccccccc|c} q_t & p_t & 0 & 0 & \cdot & \cdot & \cdot & & 0 \\ q_t & 0 & p_t & 0 & \cdot & \cdot & \cdot & & 0 \\ q_t & 0 & 0 & p_t & \cdot & \cdot & \cdot & & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ q_t & 0 & 0 & 0 & \cdot & \cdot & \cdot & & p_t \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 1 \end{array}\right] = \left[\begin{array}{c|c} A_t & p_t e_k' \\ \hline 0 & 1 \end{array}\right], \tag{8}$$

where $e_k = (0, \ldots, 0, 1)$ is a $1 \times k$ unit vector.

The matrix A_t will play an important role in our approach of establishing the transition matrices associated with the various ways of counting runs. The fundamental idea of finding the exact distribution of a run statistic is to transform the way we count the success runs in a sequence of Bernoulli trials into a finite Markov chain with a proper state space and transition probability matrix. To illustrate the entire procedure, we start with the following example.

Example. Assume that $FSSSFSSS$ is a sequence of outcomes of eight Bernoulli trials. Given the sequence, let x be the value of the random variable $X_{n,k}$ and let m the number of trailing successes (i.e., the number of last consecutive successes counting backward). In this case, $m = 3$ and the random variables $E_{8,3} = G_{8,3} = N_{8,3} = M_{8,3} = x = 2$. If the ninth trial is F (with probability q_9), then the sequence of outcomes becomes $FSSSFSSSF$, and thus now we have $m = 0$ and all four random variables have the same value $x = 2$. But if the ninth Bernoulli trial is S (with probability p_9), then the sequence of outcomes becomes $FSSSFSSSS$, resulting in $m = 4$ and $E_{9,3} = 1$, $G_{9,3} = 2$, $N_{9,3} = 2$, and $M_{9,3} = 3$. Roughly speaking, x stands for the number of occurrences of specific patterns in the sequence, and the auxiliary variable m provides information about the stage of formation of the next pattern.

The typical element of the state space Ω will be represented by a 2-tuple (x, i), and the transition probability matrix will be of the form

$$\Lambda_t = (p_t(y, j; x, i)),$$

where the states of 2-tuples (x, i) are lexicographically ordered and

$$p_t(y, j; x, i) = P(Y_t = (y, j) \mid Y_{t-1} = (x, i)).$$

For $i = 0$, the event $Y_t = (x, 0)$ means that, in the sequence of outcomes up to the tth trial, x appearances of the pattern (i.e., consecutive successes) have occurred and the last outcome is F, or we start counting for the next pattern. A vital important common property of the five random variables of interest is that if the system is in the state (x, i) after $(t-1)$th trial, and an F occurs in the tth trial, then the system enters the state $(x, 0)$. This is because whenever the F occurs, we must start counting the next pattern from the beginning. Hence the transition probability is

$$p_t(x, 0; x, i) = q_t. \tag{9}$$

This can easily be seen in the foregoing example. Now to complete the whole transition matrix Λ_t, we need only specify the transition probabilities when a success (S) occurs at the tth Bernoulli trial. Of course, those transition probabilities depend heavily on the structure of the specified run statistic; the details are given in the following subsections. For convenience and simplicity, throughout this article we assume that $P(Y_0 = (0, 0)) = 1$ and that the last state of the Markov chain is an absorbing state; that is, the last row of Λ_t is equal to $(0, 0, \ldots, 0, 1)$.

All the results given in this section refer to independent but not necessarily identical Bernoulli trials. The iid model is obviously covered as special case.

3.1 The Distribution of $N_{n,k}$

Let us consider the state space

$$\Omega = \Omega(N_{n,k}) = \{(x, i): x = 0, 1, \ldots, l \text{ and } i = 0, 1, \ldots, k-1\}, l = [n/k]$$

and the finite Markov chain $\{Y_t: t \in \Gamma_n\}$ defined on Ω as follows. For any sequence of outcomes of length t, $SFS \ldots F\overbrace{SS \ldots S}^{m}$, let m be the number of trailing successive S's *counting backwards* ($m = 0$ if the tth outcome is F). We define $Y_t = (x, i)$ if there exist x nonoverlapping consecutive k successes and $m \equiv i(\text{mod } k)$. For given $x(0 \le x \le l)$, we introduce the subset $C_x = \{(x, i): i = 0, 1, \ldots, k-1\}$. Clearly, $\{C_x: x = 0, 1, \ldots, l\}$ forms a partition of state space $\Omega(N_{n,k})$ and $P(N_{n,k} = x) = P(Y_n \in C_x)$. As mentioned before, the description of the model will be completed if we specify the transition probabilities from $t-1$ to t, when a success occurs at the tth Bernoulli trial. Because we are counting nonoverlapping appearances of consecutive success runs, it is evident that if a success occurs at the tth trial, then we have

(a) transition from state (x, i) to state $(x, i+1)$ for $0 \le x \le l$ and $0 \le i \le k-2$, and
(b) transition from state $(x, k-1)$ to state $(x+1, 0)$ for $0 \le x \le l-1$.

Therefore,

$$p_t(x, i+1; x, i) = p_t \quad \text{for} \quad 0 \le x \le l, \quad 0 \le i \le k-2$$

and

$$p_t(x+1, 0; x, k-1) = p_t \quad \text{for} \quad 0 \le x \le l-1. \tag{10}$$

All the nonzero elements of the transition probability matrices $\Lambda_t = \Lambda_t(N) = (p_t(y, j; x, i))$ are given by (9) and (10), with the last entry of their last row being 1.

As an illustration, the transition matrix of $N_{5,2}$ is given by

$$\begin{array}{c} \qquad (0,0)\ (0,1)\ (1,0)\ (1,1)\ (2,0)\ (2,1) \\ \Lambda_t(N) = \left[\begin{array}{cc|cc|cc} q_t & p_t & 0 & 0 & & \\ q_t & 0 & p_t & 0 & & \\ \hline & & q_t & p_t & 0 & \\ & & q_t & 0 & p_t & 0 \\ \hline & & & & q_t & p_t \\ & & & & 0 & 1 \end{array}\right]. \end{array} \tag{11}$$

For the general case, let B_t be a $k \times k$ matrix having p_t at the entry $(k, 1)$ and 0s elsewhere, and/or A_t^* be a matrix of the same form as A_t, defined in (8), except from its last row, which is replaced by $(0, \ldots, 0, 1)$. In view of (9) and (10), the transition probability matrix $\Lambda_t(N)$ can be written as a bidiagonal blocked matrix of the form

$$\Lambda_t(N) = \begin{bmatrix} A_t & B_t & & & \\ & A_t & B_t & & \\ & & \ddots & \ddots & \\ & & & A_t & B_t \\ & & & & A_t^* \end{bmatrix}, \tag{12}$$

and its dimension is $\sum_{x=0}^{l} |C_x| = (l+1)k$. Hence, by virtue of Theorem 2.1, we may state that

$$P(N_{n,k} = x) = \pi_0 \left(\prod_{t=1}^{n} \Lambda_t(N) \right) U'(C_x)$$
$$x = 0, 1, \ldots l, \tag{13}$$

which yields our basic result for $N_{n,k}$.

3.2 The Distribution of $G_{n,k}$

For a sequence of n Bernoulli trials, the random variable $G_{n,k}$ enumerates the number of success runs of length greater than or equal to k in the sequence. Let us consider the state space

$$\Omega = \Omega(G_{n,k}) = \{(x, i): x = 0, 1, \ldots, l \text{ and } i = -1, 0, 0, \ldots, k-1\} - \{(0, -1)\},$$

$$l = \left[\frac{n+1}{k+1}\right]$$

and the finite Markov chain $\{Y_t: t \in \Gamma_n\}$, defined on Ω as follows. For any sequence of outcomes of length t, say $SFS \ldots F\overbrace{SS \ldots S}^{m}$, we define $Y_t = (x, i)$, $0 \le i \le k-1$ if and only if there exist exactly $x \ge 0$ success runs of size at least k before the last $m+1$ outcomes $F\overbrace{SS \ldots S}^{m}$, and $i = m \le k-1$. If $m \ge k$ and $x \ge 1$ and there are $x-1$ success runs before the last $m+1$ outcomes, then we define $Y_t = (x, -1)$. The state $(x, -1)$ can be viewed as a *waiting state,* in the sense that the last m outcomes ($m \ge k$) are S's and we are looking for an F so that they become a success run of length greater than or equal to k. Note that the condition $m \ge k$ implies that x must be equal or greater than 1; that is, the state $(0, -1)$ does not make any sense. This is why that case was ruled out from the state space $\Omega(G_{n,k})$.

Regarding the partitioning of $\Omega(G_{n,k})$, we define

$$C_0 = \{(0, i): i = 0, 1, \ldots, k-1\}$$
$$C_x = \{(x, i): i = -1, 0, 1, \ldots, k-1\},$$
$$x = 1, 2, \ldots, l, \tag{14}$$

and the equality $P(G_{n,k} = x) = P(Y_n \in C_x)$, $x = 0, 1, \ldots, l$ is rather obvious. The one-step transition probabilities in this model are specified by (9) and

$$p_t(x, i+1; x, i) = p_t \quad \text{for} \quad 0 \le x \le l, \quad 0 \le i \le \max(0, k-2)$$

$$p_t(x+1, -1; x, k-1) = p_t \quad \text{for} \quad 0 \le x \le l-1,$$

and

$$p_t(x, -1; x, -1) = p_t \quad \text{for} \quad 1 \le x \le l.$$

For example, in the special case $n = 5$, $k = 2$, the transition matrix $\Lambda_t = \Lambda_t(G)$ is given by

$$\begin{array}{c}
\begin{array}{cccccccc} (0,0) & (0,1) & (1,-1) & (1,0) & (1,1) & (2,-1) & (2,0) & (2,1) \end{array} \\
\Lambda_t(G) = \left[\begin{array}{cc|c|cc|c|cc}
q_t & p_t & 0 & & & & & \\
q_t & 0 & p_t & & & & & \\
\hline
 & & p_t & q_t & 0 & 0 & & \\
\hline
 & & & q_t & p_t & 0 & & \\
 & & & q_t & 0 & p_t & & \\
\hline
 & & & & & p_t & q_t & 0 \\
\hline
 & & & & & & q_t & p_t \\
 & & & & & & 0 & 1
\end{array}\right].
\end{array}$$

In general, $\Lambda_t(G)$ is a bidiagonal blocked matrix of the form

$$\Lambda_t(G) = \begin{bmatrix}
A_t & p_t e_k' & & & & & \\
 & p_t & q_t e_1 & & & & \\
 & & A_t & p_t e_k' & & & \\
 & & & \cdot & \cdot & & \\
 & & & & \cdot & \cdot & \\
 & & & & & p_t & q_t e_1 \\
 & & & & & & A_k^*
\end{bmatrix}, \tag{15}$$

where $e_1 = (1, 0, \ldots, 0)$, and $e_k = (0, \ldots, 0, 1)$ are unit $1 \times k$ vectors. It is clear that its dimension equals $\sum_{x=0}^{l} |C_x| = (l + 1)(k + 1) - 1$. The distribution function, moments and probability generating function of the random variable $G_{n,k}$, can now be easily derived through formulas (5), (6), and (7).

3.3 The Distribution of $M_{n,k}$

The random variable $M_{n,k}$ gives the number of overlapping consecutive k successes—in the sense of Ling's (1988) counting—in a sequence of n Bernoulli trials. The Markov chain $\{Y_t: t \in \Gamma_n\}$ associated with $M_{n,k}$ is now defined over the state space

$$\begin{aligned}
\Omega &= \Omega(M_{n,k}) \\
&= \{(x, i): x = 0, 1, \ldots, l - 1 \quad \text{and} \\
&\qquad i = -1, 0, 1, \ldots, k - 1\} \cup \{(l, -1)\} - \{(0, -1)\}, \\
&\qquad\qquad l = n - k + 1,
\end{aligned}$$

as follows. Let m again be the number of trailing successes counting backwards, and let x be the number of overlapping success runs (Ling's counting) of length k up to tth trial. We define $Y_t = (x, m)$ if $m \le k - 1$ and $Y_t = (x, -1)$ if $m \ge k$. It is easy to verify that the transition probabilities of $\Lambda_t = \Lambda_t(M)$ can be obtained from (9) and

$$p_t(x, i + 1; x, i) = p_t \quad \text{for} \quad 0 \le x \le l - 1 \quad \text{and} \quad 0 \le i \le k - 2,$$

$$p_t(x + 1, -1; x, k - 1) = p_t \quad \text{for} \quad 0 \le x \le l - 1,$$

and

$$p_t(x + 1, -1; x, -1) = p_t \quad \text{for} \quad 1 \le x \le l - 1.$$

The corresponding partition of the state space $\Omega(M_{n,k})$ is specified by

$$C_0 = \{(0, i): i = 0, 1, \ldots, k - 1\}, \qquad C_l = \{(l, -1)\}$$

and

$$C_x = \{(x, i): i = -1, 0, 1, \ldots, k - 1\} \qquad x = 1, \ldots, l - 1.$$

For $n = 4$, $k = 2$, the transition matrix $\Lambda_t = \Lambda_t(M)$ is

$$\begin{array}{c}
\begin{array}{ccccccccc} (0,0) & (0,1) & (1,-1) & (1,0) & (1,1) & (2,-1) & (2,0) & (2,1) & (3,-1) \end{array} \\
\Lambda_t(M) = \left[\begin{array}{cc|c|cc|c|cc|c}
q_t & p_t & 0 & & & & & & \\
q_t & 0 & p_t & & & & & & \\
\hline
 & & 0 & q_t & 0 & p_t & & & \\
\hline
 & & & q_t & p_t & 0 & & & \\
 & & & q_t & 0 & p_t & & & \\
\hline
 & & & & & 0 & q_t & 0 & p_t \\
\hline
 & & & & & & q_t & p_t & 0 \\
 & & & & & & q_t & 0 & p_t \\
\hline
 & & & & & & 0 & 0 & 1
\end{array}\right].
\end{array} \tag{16}$$

In general, the transition matrix associated with the random variable $M_{n,k}$ has same form as matrix (16) with dimension $\sum_{x=0}^{l} |C_x| = l(k + 1)$. We leave the details to the reader.

3.4 The Distribution of $E_{n,k}$

For the study of the distribution of the number $E_{n,k}$ of success runs of size exactly k, we define the finite Markov chain $\{Y_t: t \in \Gamma_n\}$ with state space

$$\Omega = \Omega(E_{n,k}) = \{(x, i): x = 0, 1, \ldots l \text{ and } i = -2, -1, 0, \ldots, k - 1\} - \{(0, -2)\},$$

$$l = \left[\frac{n + 1}{k + 1}\right].$$

The states (x, i) for $x = 0, 1, \ldots, l$ and $i = 0, 1, \ldots, k - 1$ are defined similarly to those in the previous paragraphs. The additional states $(x, -1)$, $(x, -2)$ are used for managing the fact that a success run whose length becomes greater than k does not count (for $E_{n,k}$) as a success run any more. More specifically, in a sequence of outcomes $SFS \ldots F\overbrace{SS \ldots S}^{m}$, we define the following two states:

- Overflow states: $(x, -1)$, $x = 0, 1, \ldots, l$. $Y_t = (x, -1)$, $x = 0, 1, \ldots, l$ means that $m > k$ and exactly x success runs of size k appear before the last $m + 1$ outcomes (overflow state).
- Waiting states: $(x, -2)$, $x = 1, \ldots, l$. $Y_t = (x, -2)$, means that $m = k$ and exactly x success runs of size k have occurred.

With these in mind, we can easily specify the following partition of $\Omega(E_{n,k})$:

$$C_0 = \{(0, i): i = -1, 0, \ldots, k-1\},$$
$$C_x = \{(x, i): i = -2, -1, 0, \ldots, k-1\}, \qquad x = 1, 2, \ldots, l.$$

We also can specify the transition probabilities of $\Lambda_t = \Lambda_t(E)$:

$$p_t = (x, i+1; x, i) = p_t \qquad \text{for} \quad 0 \le x \le l, \quad 0 \le i \le k-2,$$
$$p_t = (x+1, -2; x, k-1) = p_t \quad \text{for} \quad 1 \le x \le l-1,$$
$$p_t = (x-1, -1; x, -2) = p_t \quad \text{for} \quad 1 \le x \le l,$$

and

$$p_t = (x, -1; x, -1) = p_t \quad \text{for} \quad 0 \le x \le l.$$

As an illustration, the special case $n = 5$, $k = 2$ yields

$$\Lambda_t(E) = \left[\begin{array}{c|cc|cc|cc|cc|cc}
p_t & q_t & 0 & & & & & & & & \\
\hline
 & q_t & p_t & 0 & & & & & & & \\
 & q_t & 0 & p_t & & & & & & & \\
\hline
p_t & & & 0 & 0 & q_t & 0 & & & & \\
 & & & 0 & p_t & q_t & 0 & & & & \\
\hline
 & & & & & q_t & p_t & 0 & & & \\
 & & & & & q_t & 0 & p_t & & & \\
\hline
 & & & & p_t & & & 0 & 0 & q_t & \\
 & & & & & & & 0 & p_t & q_t & \\
\hline
 & & & & & & & & & q_t & p_t \\
 & & & & & & & & & 0 & 1
\end{array}\right]. \tag{17}$$

In general, the transition probability matrix of $E_{n,k}$ has the same form as matrix (17) with dimension $\sum_{x=0}^{l} |C_x| = (l+1)(k+2) - 1$.

3.5 The Distribution of the Longest Success Run

The longest success run in a sequence of Bernoulli trials is related to the random variables $N_{n,k}$, $G_{n,k}$, and $M_{n,k}$ in the following simple way:

$$L_n < k \quad \text{iff } G_{n,k} = N_{n,k} = M_{n,k} = 0. \tag{18}$$

Thus the distribution of L_n can easily be obtained through the following theorem.

Theorem 3.1. For given n, we have

$$P(L_n = k) = \pi_{0,k+1}\left(\prod_{t=1}^{n} \Lambda_t(k+1)\right)U'(k+1) - \pi_{0,k}\left(\prod_{t=1}^{n} \Lambda_t(k)\right)U'(k), \quad k = 0, 1, \ldots, n,$$

where $\pi_{0,j} = (1, 0, \ldots, 0)$, $U(j) = (1, 1, \ldots, 1, 0)$ $(j = k, k+1)$ are $1 \times (j+1)$ vectors, and $\Lambda_t = \Lambda_t(k)$ is given by (8).

Proof. It follows from (18) that

$$P(L_n < k) = P(N_{n,k} = 0)$$

for every $k = 0, 1, \ldots, n$. Because

$$P(L_n = k) = P(L_n < k+1) - P(L_n < k) = P(N_{n,k+1} = 0) - P(N_{n,k} = 0),$$

the desired result is easily verified.

It is worth mentioning that the distribution of L_n also could be obtained directly by using Theorem 2.1. For example, the special case $n = 3$ can be effortlessly transferred to the setup of Theorem 2.1 by introducing the state space

$$\Omega = \Omega(L_3) = \{(0, 0), (1, -1), (1, 0), (1, 1), (2, -1), (2, 0), (2, 1), (3, -1)\}$$

and the transition probability matrix

$$\Lambda_t(L) = \left[\begin{array}{c|c|cc|c|ccc|c}
q_t & p_t & 0 & 0 & & & & & \\
\hline
 & 0 & q_t & 0 & p_t & & & & \\
 & 0 & q_t & p_t & 0 & & & & \\
 & 0 & q_t & 0 & p_t & & & & \\
\hline
 & & & & 0 & q_t & 0 & 0 & p_t \\
 & & & & 0 & q_t & p_t & 0 & 0 \\
 & & & & 0 & q_t & 0 & p_t & 0 \\
 & & & & 0 & q_t & 0 & 0 & p_t \\
\hline
 & & & & & & & & 1
\end{array}\right].$$

The probabilities $P(L_3 = x)$, $x = 0, 1, 2, 3$ now can be derived by a direct application of Theorem 2.1 on each one of the sets:

$$C_0 = \{(0, 0)\}, C_1 = \{(1, -1), (1, 0), (1, 1)\},$$
$$C_2 = \{(2, -1), (2, 0), (2, 1)\}, \quad \text{and} \quad C_3 = \{(3, -1)\}.$$

The general form of $\Lambda_t(L)$ is rather simple; we leave it to the reader.

4. NUMERICAL EXAMPLE

Table 1 gives the exact distributions and means of the random variables $E_{5,2}$, $G_{5,2}$, $N_{5,2}$, $M_{5,2}$, and $L_{5,2}$ for a sequence of five Bernoulli trials, in which the tth trials' probability of success is $p_t = 1/(t+1)$, $t = 1, 2, \ldots, 5$ (non-iid case). The values of $n = 5$ and $k = 2$ were purposely chosen small, so that the reader can repeat the computations by hand and thus gain insight into the mechanism that generates them.

In addition to the foregoing numerical example, Table 2 is given to show that the combinatorial methods (even in the simple case of $n = 15$, $k = 2$) for finding the exact distributions of run statistics $E_{15,2}$, $G_{15,2}$, $N_{15,2}$, $M_{15,2}$, and L_{15} can be erroneous, and that our approach succeeds easily.

The numerical results in Table 2 were obtained by a computer program based on Theorem 2.1, with the transition probability matrices and $P_t = 1/(t+1)$ developed in Section 3; see (12), (15), (16), and (17). The computing was done on a computer; CPU time for each case was less than 1 minute. The computer program is available from the authors on request.

5. WAITING TIME DISTRIBUTION OF SUCCESS RUN

The Markov chain imbedding technique for the exact distribution of $N_{n,k}$ developed in Section 3.1 can also be used

Table 1. The Exact Distributions and Means of Five Runs Statistics With n = 5

x	0	1	2	3	4	5	Mean
$E_{5,2}$	.7931	.2028	.0042				.2112
$G_{5,2}$	.7375	.2583	.0042				.2667
$N_{5,2}$	.7375	.2486	.0139				.2764
$M_{5,2}$	.7375	.2028	.0500	.0083	.0014		.3333
L_5	.1667	.5708	.2069	.0459	.0083	.0014	1.1625

to find the waiting time distribution of the first success run of length k or mth ($m = 1, 2, \ldots$) success run of length k (see Aki et al. 1984; Feller 1968; Philippou and Muwafi 1982; Philippou et al. 1983, and references therein). For given m and k ($m = 1, 2, \ldots$ and $k = 1, 2, \ldots$), let us consider the random variable $T_{m,k}$ to be the smallest number of independent trials in which the mth success run of length k (in Feller's counting) occurred.

Theorem 5.1. For given m and k, the distribution of waiting time $T_{m,k}$ can be specified by

$$P(T_{m,k} = mk) = \prod_{t=1}^{mk} p_t$$

and

$$P(T_{m,k} = n) = \pi_0\left(\prod_{t=1}^{n-k} \Lambda_t(N_{n-k,k})\right)U'(m-1, 0) \prod_{t=n-k+1}^{n} p_t, \quad n > mk, \quad (19)$$

where π_0 and $\Lambda_t(N_{n,k})$ are given in Section 3.1 and $U(m-1, 0) = (0, \ldots, 0, 1, 0, \ldots, 0)$ is a unit vector having "1" at the coordinate associated with the state $(m - 1, 0)$.

Proof. For $n > mk$, the random variable $T_{m,k}$ equals n if, and only if that (i) the last k trials are all S's and the Markov chain at time $(n - k)$ has to be at the state $(m - 1, 0)$. Hence, we have, by applying Theorem 2.1,

$$\begin{aligned} P(T_{m,k} = n) &= P(N_{n-k,k} = (m-1, 0) \text{ and the last } k \text{ trials are successes}) \\ &= P(N_{n-k,k} = (m-1, 0))P(\text{last } k \text{ trials are successes}) \\ &= \pi_0\left(\prod_{t=1}^{n-k} \Lambda_t(N_{n-k,k})\right)U'(m-1, 0) \prod_{t=n-k+1}^{n} p_t. \end{aligned}$$

This completes the proof.

It is worth mentioning that Formula (19) can easily be modified too capture the waiting time distributions for alternative counting methods (see, for example, Ling 1989, which introduced the waiting time for the mth overlapping success run). To illustrate Theorem 5.1, we give the following numerical results in Table 3.

6. DISCUSSION

Our approach of finding the distribution of run statistics departs from the traditional combinatorial approach. The simple results of Section 2 provide a unified procedure for the evaluation of their distributions. In view of the special cases treated in Section 3, one could state that if a specific run (or pattern) is given, then it is not too difficult to create the proper state space, partition, and Markov chain for Theorem 2.1 to be applied. This provides a great potential for extending our results to any enumerating method.

The results presented in Section 3 not only are of theoretical interest, but also are very useful for many practical applications. Our numerical evaluation of the probabilities of a number of occurrences of various runs are rather straightforward. Thus the critical values for nonparametric tests based on these runs could easily be obtained (see, for example, Cochran 1938, Gibbons 1971, Mood 1940, and Walsh 1965). It is worth mentioning that Agin and Godbole (1992) proved that many

Table 2. The Exact Distributions and Means of Five Runs Statistics With n = 15

x \ r.v.	$E_{15,2}$	$G_{15,2}$	$N_{15,2}$	$M_{15,2}$	L_{15}
0	.73200	.67163	.67163	.67163	.06250
1	.24771	.30120	.29046	.24125	.60913
2	.01976	.02646	.03602	.06881	.26135
3	.00051	.00070	.00184	.01516	.05531
4	3.99×10^{-6}	5.27×10^{-6}	.00004	.00270	.00991
5	5.08×10^{-9}	5.69×10^{-9}	4.86×10^{-7}	.00040	.00156
6			2.28×10^{-9}	.00005	.00021
7			3.10×10^{-12}	5.68×10^{-6}	.00003
8				5.54×10^{-7}	2.88×10^{-6}
9				4.81×10^{-8}	2.85×10^{-7}
10				3.71×10^{-9}	2.58×10^{-8}
11				2.55×10^{-10}	2.14×10^{-9}
12				1.58×10^{-11}	1.64×10^{-10}
13				7.64×10^{-13}	1.16×10^{-11}
14				4.78×10^{-14}	7.65×10^{-13}
15					4.78×10^{-14}
Mean	.28879	.35625	.36821	.43750	1.34668

Table 3. The Waiting Time Distributions of the First Success Run of Length k.

	p = .90		p(i) = i/i + 1		p(i) = 1 − 1/2^i		
n	k = 2	k = 3	k = 2	k = 3	k = 2	k = 3	k = 5
2	.8100		.3333		.3750		
3	.0810	.7290	.2500	.2500	.3281	.3281	
4	.0810	.0729	.2000	.2000	.2051	.3076	
5	.0154	.0729	.1111	.1667	.0710	.1987	.2980
6	.0088	.0729	.0595	.1429	.0177	.1118	.2933
7	.0023	.0198	.0271	.0937	.0028	.0397	.1940
8	.0010	.0144	.0117	.0611	.0003	.0111	.1104
9	.0003	.0091	.0046	.0383	.0000	.0026	.0588
10	.0001	.0038	.0017	.0219		.0004	.0303
11	.0000	.0024	.0006	.0122		.0001	.0108
12		.0013	.0002	.0066			.0032
13		.0007	.0001	.0034			.0008
14		.0004		.0017			.0002
15		.0002		.0008			.0000
16		.0001		.0004			
17		.0001		.0002			
18				.0001			

test procedures based on fixed length runs are significantly more powerful than the standard ones.

Many recent results on distribution theory, probability of patterns, and reliability theory can be derived as special cases of Theorem 2.1 by manipulating the general expression (5). For example, the recurrence relations given by Hirano (1986) and Aki and Hirano (1988) for the binomial distribution of order k are effortlessly deduced through (5), (6), and (7) by using the transition probability matrix (12). Similarly, the recurrence relations provided by Ling (1988) (iid case) and Chrysaphinou et al. (1993) (non-iid case) can be easily generated through the transition matrix (16). Many results on the reliability of linearly connected systems can also be considered as direct consequences of (5) with transition probability matrix (12).

Another interesting feature of the approach presented in this article is that the general results of Section 2 are also valid for the case of a two-state Markov chain (instead of independent Bernoulli trials). Some obvious modifications on the transition matrices (12), (15), and (16), combined by straightforward algebraic manipulations on expression (5), capture quite a few of the results given by Hirano and Aki (1992), and Aki and Hirano (1993).

Finally, we mention that our approach gives a very efficient computational tool for tabulation purposes. Should one be interested in constructing tables of the probabilities $P(X_{r,k} = x)$ for $r = k, k + 1, \ldots, n$, we could start with the transition probability matrix of the variable $X_{n,k}$ and proceed from the distribution of $X_{r-1,k}$ to the distribution of $X_{r,k}$ by a single matrix multiplication (between $\prod_{t=1}^{r-1} \Lambda_t$ and Λ_r), $r = k + 1, \ldots, n$.

Note that the transition matrices given in Section 3 are in their full form, in the sense that in every state, all possible substates are listed, even if they are never reached by the system. For example, one can easily check that the random variable $E_{5,2}$ cannot reach the states (2, −2), (2, 0), and (2, 1). Therefore, the transition matrix given in Section 3.4 can be reduced to an 8×8 matrix (instead of 11×11) by using (0, . . . , 0, 1) as its eighth row and column. These kinds of shortcuts are very useful, especially when the calculations are done by hand.

[Received January 1993. Revised August 1993.]

REFERENCES

Agin, M. A., and Godbole, A. P. (1992), "A New Exact Runs Test for Randomness," in *Computing Science and Statistics,* eds. C. Page and R. LePage, New York: Springer-Verlag, pp. 281–285.

Aki, S. (1985), "Discrete Distributions of Order k on a Binary Sequence," *Annals of the Institute of Statistical Mathematics* Part A, 37, 205–224.

Aki, S., and Hirano, K. (1986), "Some Characteristics of The Binomial and Some Other Distributions of Order k," Research Memorandum 316, The Institute of Statistical Mathematics.

——— (1988), "Some Characteristics of the Binomial Distribution of Order k and Related Distributions, *Statistical Theory and Data Analysis* II, ed. K. Matusita, Amsterdam: North-Holland pp. 211–222.

——— (1993), "Discrete Distributions Related to Succession Events in a Two-State Markov Chain," *Statistical Science & Data Anal.,* 467–476.

Aki, S., Kuboki, H., and Hirano, K. (1984), "On Discrete Distributions of Order k," *Annals of the Institute of Statistical Mathematics,* 36, 431–440.

Chao, M. T., and Fu, J. C. (1989), "A Limit Theorem of Certain Repairable Systems," *Annals of the Institute of Statistical Mathematics,* 41, 809–818.

——— (1991), "The Reliability of Large Series Systems Under Markov Structure," *Advances in Applied Probability,* 23, 894–908.

Chrysaphinou, 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* 5, 103–112.

Cochran, W. G. (1938), "An Extension of Gold's Method for Examining the Apparent Persistence of One Type of Weather," *Royal Meteorological Society Quarterly Journal,* 64, 631–634.

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

Fu, J. C. (1986), "Reliability of Consecutive-k-out-of-n: F Systems With (k − 1)-Step Markov Dependence," *IEEE Transactions on Reliability,* R35, 602–606.

Fu, J. C., and Hu, B. (1987), "On Reliability of a Large Consecutive-k-out-of-n: F System With (k − 1)-Step Markov Dependence," *IEEE Transactions on Reliability,* R 36, 75–77.

Gibbons, J. D. (1971), *Nonparametric Statistical Inference,* New York: McGraw-Hill.

Godbole, A. P. (1990), "Specific Formula for Some Success Run Distributions," *Statistics and Probability Letters,* 10, 119–124.

Hirano, K. (1986), "Some Properties of the Distributions of Order k," *Fibonacci Numbers and Their Applications,* eds. A. N. Philippou, G. E. Bergum, and A. F. Horadam, Dordrecht: Reidel, pp. 43–53.

Hirano, K., and Aki, S. (1987), "Properties of the Extended Distributions of Order k," *Statistics and Probability Letters,* 6, 67–69.

——— (1992), "On the Number of Occurrences of Success Runs of Length k in a Two-State Markov Chain," *Statistica Sinica,* 3, 313–319.

Hirano, K., Aki, S., Kashiwagi, N., and Kuboki, H. (1991), "On Ling's Binomial and Negative Binomial Distributions of Order k," *Statistics and Probability Letters,* 11, 503–509.

Ling, K. D. (1988), "On Binomial Distributions of Order k," *Statistics and Probability Letters,* 6, 247–250.

——— (1989), "A New Class of Negative Binomial Distributions of Order k," *Statistics and Probability Letters,* 7, 371–376.

Mood, A. M. (1940), "The Distribution Theory of Runs," *Annals of Mathematical Statistics,* 11, 367–392.

Mosteller, F. (1941), "Note on an Application of Runs to Quality Control Charts," *Annals of Mathematical Statistics,* 12, 228–232.

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, G. E. Bergum, and A. F. Horadam, Dordrecht: Reidel, 203–227.

Philippou, A. N., Georgiou, C., and Phililppou, G. M. (1983), "A Generalized Geometric Distribution and Some of Its Properties," *Statistics and Probability Letters,* 1, 171–175.

Philippou, A. N., and Makri, F. S. (1985), "Longest Success Runs and Fibonacci-Type Polynomials," *Fibonacci Quarterly,* 23, 338–346.

——— (1986), "Successes, Runs, and Longest Runs," *Statistics and Probability Letters,* 4, 101–105.

Philippou, A. N., and Muwafi, A. A. (1982), "Waiting for the kth Consecutive Success and the Fibonacci Sequence of Order k," *The Fibonacci Quarterly,* 20, 28–32.

Schwager, S. J. (1983), "Run Probabilities in Sequences of Markov-Dependent Trials," *Journal of the American Statistical Association,* 78, 168–175.

Stevens, W. L. (1939), "Distributions of Groups in a Sequence of Alternatives," *Annals of Eugenics,* 9, 10–17.

Wald, A., and Wolfowitz, J. (1940), "On a Test Whether Two Populations are From the Same Population," *Annals of Mathematical Statistics,* 11, 147–162.

Walsh, J. E. (1965), *Handbook of Nonparametric Statistics,* New York: D. Van Nostrand.

Wishart, J., and Hirshfeld, H. O. (1936), "On a Test Whether Two Samples are From the Same Population," *London Mathematical Society Journal,* 11, 227–235.

Wolfowitz, J. (1943), "On the Theory of Runs With Some Applications to Quality Control," *Annals of Mathematical Statistics,* 14, 280–288.