Statistica Sinica **6**(1996), 957-974

DISTRIBUTION THEORY OF RUNS AND PATTERNS ASSOCIATED WITH A SEQUENCE OF MULTI-STATE TRIALS

James C. Fu

National Donghwa University

Abstract: Runs and patterns in a sequence of Bernoulli trials and multi-state trials have broadly been used for various purposes and in many areas of statistics and applied probability. Recently, Fu and Koutras (1994) developed a new method based on finite Markov chain imbedding technique to study the exact distributions for the number of specified runs and patterns in a sequence of Bernoulli trials. In this manuscript, a "forward and backward principle" for the method of finite Markov chain imbedding is introduced to study the exact and joint distributions for the numbers of runs and patterns in a sequence of multi-state trials. Waiting time distribution for a pattern is also obtained.

Key words and phrases: Multi-state trials, distributions of runs and patterns, Markov chain, transition probability matrix.

1. Introduction

Run statistics and patterns in a sequence of Bernoulli trials have been successfully used in various areas, such as hypothesis testing (run-test, Wald and Wolfowitz (1940) and Walsh (1965)), statistical quality control (Mosteller (1941) and Wolfowitz (1943)), DNA sequencing, psychology, and ecology (Schwager (1983)). The distribution theory of runs seems to have started at the end of the nineteenth century (Stevens (1939)). There were a considerable amount of research works on the distribution theory of runs around 1940, for example, Wishart and Hirsheld (1936), Cochran (1938), Mood (1940), Wald and Wolfowitz (1940) and Wolfowitz (1943). Most of the research during this period was focused on the study of the conditional distributions of runs given the total number of successes. There was very little research on the exact and limiting distributions of runs during the period between 1950 and 1970. However, there were some very interesting approximate formulae for the distributions of runs and patterns developed during this period (Walsh (1965) and Gibbons (1985)). During the late 1980's and early 1990's , this area became very active again. Recent publications in this area are mainly focused on studying the exact and limiting distributions for certain patterns and runs in a sequence of independent identically distributed (i.i.d.) Bernoulli trials.

The number $N_{n,k}$ of non-overlapping consecutive k successes (in the sense of Feller's counting (1968, Ch. XIII)) in a sequence of Bernoulli trials is probably one of the most studied run statistic. In the context of distribution theory, the distribution of $N_{n,k}$ is called a binomial distribution of order k. The run statistic $N_{n,k}$ also plays an even more important role in finding the distributions for other runs. Its distribution has been applied successfully in various areas, especially to the reliabilities of some engineering systems. For example the probability of $N_{n,k} = 0$ is the reliability of a consecutive-k-out-of-n : F system and the tail probability of $N_{n,k} < m$ is the reliability of a m-consecutive-k-out-of-n : F system. It is heavily studied by many researchers in the area of reliability theory.

Traditionally, combinatorial methods were used to find the exact distributions for the numbers of runs and patterns. Even in the case of i.i.d. Bernoulli trials, the formulae for the distributions of runs and patterns are often complex and tedious. It is usually difficult to extend those i.i.d. results by combinatorial methods to a sequence of independent but non-identical Bernoulli trials, or to extend to a more complex case such as a sequence of m-step Markov dependent multi-state trials. Recently, Fu and Koutras (1994) took a Markov chain approach to study the exact distributions for the numbers of runs and patterns in a sequence of Bernoulli trials. Their ideas were largely borrowed from the sequence of papers by Fu (1986) and Chao and Fu (1989, 1991) for studying the reliabilities of linearly connected engineering systems. The exact distribution of a specified run statistic is studied under the framework of a finite Markov chain, and is expressed in terms of its transition probability matrices. In that paper, they have obtained the exact distributions of five often used run statistics in a sequence of i.i.d. Bernoulli trials (viz., the number $E_{n,k}$ of success runs of size exactly k, the number $G_{n,k}$ of success runs of size greater than or equal to k, the number $N_{n,k}$ of non-overlapping consecutive k successes, the number $M_{n,k}$ of overlapping consecutive k successes, and the length L_n of the longest success run).

Until recently, there were not many general results on the exact and limiting distributions for runs and patterns associated with a sequence of multi-state trials. This was probably due to the complexity of combinatorial analysis in the case of multi-state trials. Schwager (1983) gave a very interesting paper that studied the probability of a simple run (run containing one or two symbols) in a sequence of n multi-state trials each of which has s ($s \geq 2$) possible outcomes (symbols) by using the recursive method of renewal equations. He argued that the recursive approach is superior to the generating function approach (see Feller (1968), Ch. XIII) to treat the runs in the latter can only be used for the multiple runs (viz., runs containing more than two symbols) and the generalized multiple runs (viz., a collection of runs) for the i.i.d. multi-state trials. For finding the exact and joint distributions of runs under independent but non-identical multi-state trials or Markov dependent multi-state trials, both methods become tedious and intractable. There have been no general theorems to deal with those dependent multi-state trials.

There were also considerable applications of waiting time (sooner or later) distribution of a specified run to various areas of statistics, especially biostatistics and quality control. Most of these results were focused on i.i.d. Bernoulli trials (see Hahn and Gage (1983), Benevento (1984), Ebneshahrashoob and Sobel (1990), and Balasubramanian, Viveros and Balakrishnan (1993)). There were very few general results regarding the waiting time distribution for the case of multi-state trials and there was no result in Markov dependent multi-state trials.

This article mainly introduces a "forward and backward principle" for the method of finite Markov chain imbedding developed by Fu and Koutras (1994) to study the exact and joint distributions for the runs and patterns in four different cases; (1) i.i.d., (2) independent but non-identical, (3) homogeneous Markov dependent, and (4) non-homogeneous Markov dependent sequences of multi-state trials. It also demonstrates that this method could be used to obtain the waiting time distribution of a specified run.

The rest of this manuscript is organized in the following ways. In Section 2, the method of Markov chain imbedding for finding the exact distribution of runs is introduced. In Section 3, the "forward and backward principle" for constructing a Markov chain associated with a specified run in a sequence of multi-state trials is systematically developed. Section 4 studies mainly the marginal and joint distributions for success runs and failure runs in a sequence of Bernoulli trails. Section 5 gives a general result regarding the waiting time distribution of the mth occurrence of a specified pattern. In Section 6, several numerical examples are given to illustrate the main results. In Section 7, several technical remarks, discussions and possible extensions are given.

2. Finite Markov Chain Imbedding

Let $Z_1, \ldots, Z_n$ be a sequence of n i.i.d. multi-state trials, each of which has s ($s \geq 2$) states (or symbols), labeled $b_1, \ldots, b_s$ and their corresponding probabilities $p_1, \ldots, p_s$ of occurring. Throughout this section, we denote the random variable $X_n(X_n = \phi(Z_1, \ldots, Z_n))$ by the number of runs (or patterns) in a sequence of n multi-state trials. For a given positive integer n, let $\Gamma_n = \{0, 1, \ldots, n\}$ be a finite index set and $\Omega = \{a_1, \ldots, a_m\}$ be a state space having m states.

Definition 2.1. The random variable $X_n = \phi(Z_1, \ldots, Z_n)$, the number of a specified runs occurring in a sequence of n multi-state trials $Z_1, \ldots, Z_n$, is finite Markov chain imbeddable if

(i) there exists a finite Markov chain $\{Y_t, t \in \Gamma_n\}$ defined on a finite state space Ω with transition matrices $M_t, t \in \Gamma_n$, and
(ii) there exists a partition $\{C_x, x = 0, 1, \ldots, l\}$ on the state space Ω (where C_x and l may depend on n), such that, for every $x = 0, 1, \ldots, l$,

$$P(X_n = x) = P(Y_n \in C_x). \tag{2.1}$$

It is well known that the probabilistic behavior of a Markov chain is uniquely characterized by its own transition matrices. The following specially constructed formula (2.2) for computing the probability of $Y_n \in C_x$ introduced by Fu and Koutras (1994) is a matrix version of the Chapman-Kolmogorov Theorem. It has been successfully used to obtain the exact distribution (also moments and generating function) for the number of runs (or patterns) in Bernoulli trials. It will play an even more important role for developing the distribution of the number of runs in a sequence of multi-state trials. For this reason, we repeat the formula here (see Fu and Koutras (1994) for details).

Let $\{Y_t, \Omega, M_t : t \in \Gamma_n\}$ be a Markov chain with initial probabilities $\pi_0 = (p_1, \ldots, p_m)$, where $p_j = P(Y_0 = a_j)$, for $j = 1, \ldots, m$, $U(a_i) = (0, \ldots, 0, 1, 0, \ldots, 0)$ be a unit vector with 1 at ith place and 0 elsewhere. For given n, if X_n is finite Markov chain imbeddable, then

$$P(X_n = x) = \pi_o\Big(\prod_{t=1}^{n} M_t\Big)U'(C_x), \tag{2.2}$$

where $U(C_x) = \sum_{a_i \in C_x} U(a_i)$. Similarly, the moments and the generating function of the random variable X_n can be obtained by replacing the vector $U'(C_x)$ in (2.2) with vectors V'_r and $W'(s)$ respectively, where $V_r = \sum_{x=0}^{l} x^r U(C_x)$ and $W(s) = \sum_{x=0}^{l} s^x U(C_x)$.

Remark. The random variable X_n is finite Markov chain imbeddable for every n that does not imply the sequence $\{X_n\}$ is a Markov chain.

3. Forward and Backward Principle

The Markov chain imbedding technique for finding the distribution of a specified run introduced by Fu and Koutras (1994) for Bernoulli trials involves basically three essential steps: (i) the construction of a proper state space Ω based on the structure of the specified run, (ii) the construction of a finite Markov chain and its transition probability matrices, and (iii) the construction of a partition $\{C_x\}$ on the state space Ω which is one to one corresponding to the random variables X_n in the sense that $P(X_n = x) = P(Y_n \in C_x)$ for all x. Even in the case of a sequence of i.i.d. Bernoulli trials, the task of finding the exact distribution of a specified pattern (run) by using the Markov chain imbedding technique may not always be trivial. There is no good definition about what is a pattern (or run). Wolfowitz (1943) wrote " $\cdots$ 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". In this manuscript we consider mainly two types of often used patterns:

Definition 3.1. Λ is a simple pattern if Λ is composed of a specified sequence of k symbols, i.e. $\Lambda = b_{i_1}, \ldots, b_{i_k}$ (the length of the pattern k is fixed, and the symbols in the pattern are allowed to be repeated).

Define $\Lambda_1 \cup \Lambda_2$ to be either the pattern Λ_1 occurred or the pattern Λ_2 occurred.

Definition 3.2. Λ is a compound pattern if it is a union of m distinct simple patterns, i.e. $\Lambda = \Lambda_1 \cup \ldots \cup \Lambda_m$ (m is fixed).

To find the exact distribution for a specified pattern not only requires a deep understanding of the structure of the specified pattern but also its counting procedure throughout the sequence of n multi-state trials. In order to facilitate our study of the general finite Markov chain imbedding technique for simple and compound patterns in multi-state trials without heavy mathematics, let us consider a sequence of n i.i.d. three-state trials $Z_1, \ldots, Z_n$, each of which has three possible outcomes (symbols), labeled b_1, b_2, and b_3 with corresponding probabilities p_1, p_2, and p_3 of occurring. Suppose we are interested in finding the exact distribution of the random variable X_n, the number of non-overlapping specified simple patterns $\Lambda = b_1b_1b_2$ in a sequence of n three-state trials.

In the following we introduce the *forward and backward principle* for the finite Markov chain imbedding technique to find the exact distribution of a given simple pattern Λ.

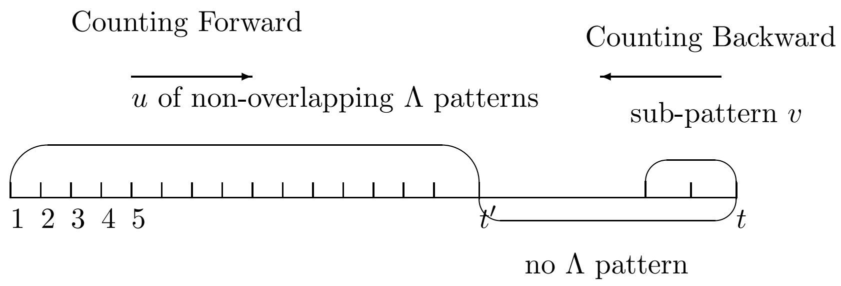
(i) *Decompose* the pattern $\Lambda = b_1b_1b_2$ *forward* into three sub-patterns labeled 0, $1 = b_1$, and $2 = b_1b_1$, where the label "0" stands for neither sub-pattern "1" nor sub-pattern "2". We shall refer to these three sub-patterns' 0, 1, and 2 as *ending blocks.*

(ii) Let $\omega = (z_1, \ldots, z_n)$ be a realization of a sequence of n three-state trials where z_i is the outcome of the ith trials. We define the Markov chain $\{Y_t : t = 1, \ldots, n\}$ operating on ω as $Y_t(\omega) = (u, v)$ for every $t = 1, \ldots, n$, where

$u =$ the total number of non-overlapping patterns Λ that occurred in the first t trials (counting *forward* from the first trials to the tth trials),

$v =$ the sub-pattern (ending block), counting *backward* from t.

The following diagram illustrates the definitions of u and v in a sequence of t trials:



For example, let us consider a realization $\omega = (b_3b_1b_1b_2b_1b_1b_3)$ of a sequence of seven three-state trials. Applying the forward and backward principle, then the realization of the imbedded Markov chain Y_t on ω is given by $\{Y_1(\omega) = (0,0)$, $Y_2(\omega) = (0,1)$, $Y_3(\omega) = (0,2)$, $Y_4(\omega) = (1,0)$, $Y_5(\omega) = (1,1)$, $Y_6(\omega) = (1,2)$ and $Y_7(\omega) = (1,0)\}$. Note that for every given ω, the realization of the imbedded Markov chain $Y_t(\omega) = (u,v)$ is uniquely by determined by (i) and (ii) under non-overlapping counting.

Based on (i) and (ii), the state space Ω associated with the imbedded Markov chain is defined by

$$\Omega = \{(u,v) : u = 0,1,\dots,l, \text{ and } v = 0,1,\dots,k-1\} \tag{3.1}$$

with $m = (l+1)k$ states, where k is the length of the simple pattern Λ and $l = [n/k]$ is the maximum number of patterns Λ possible in the sequence of n multi-state trials.

(iii) For $t = 1,\dots,n$, the transition probabilities of the transition matrix M_t is determined by the following two equations : for $u = 0,1,\dots,l-1$,

$$P(Y_t = (u+1,0)|Y_{t-1} = (u,k-1)) = p_j, \tag{3.2}$$

where p_j is the probability of b_j the last symbol of the specified pattern Λ, for $u = 0,1,\dots,l$ and $v,v' = 0,1,\dots,k-1$,

$$P(Y_t = (u,v')|Y_{t-1} = (u,v)) = \sum_{v\to v'} p_i, \tag{3.3}$$

where $\sum_{v\to v'}$ sums over all p_i corresponding to b_i of which the ending block v is changed to the ending block v', and zero otherwise.

(iv) The partition $\{C_x = [(x,v) : (x,v) \in \Omega, v = 0,1,\dots,k-1]$, for $x = 0,1,\dots,l\}$ on Ω is one-to-one corresponding to the random variable X_n in the sense that $P(X_n = x) = P(Y_n \in C_x)$ for every $x = 0,1,\dots,l$.

It follows from our construction that the transition probabilities given by (3.2) and (3.3) depend only on $Y_{t-1} = (u,v)$; hence the sequence $\{Y_t\}$ is a Markov chain. The four steps (i), (ii), (iii), and (iv) of construction mentioned above establish that the random variable X_n, the number of simple patterns Λ in a sequence of n independent multi-state trials, is finite Markov chain imbeddable.

For our example of $n = 7$ and $\Lambda = b_1b_1b_2$, it follows from the forward and backward principle that the imbedded Markov chain $\{Y_t : t = 1, \ldots, 7\}$ defined on the state space $\Omega = \{(0,0),(0,1),(0,2),(1,0),(1,1),(1,2),(2,0),(2,1),(2,2)\}$ has transition probability matrices

$$M_t = \left[\begin{array}{ccc|ccc|ccc}
p_2+p_3 & p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
p_2+p_3 & 0 & p_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
p_3 & 0 & p_1 & p_2 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & p_2+p_3 & p_1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p_2+p_3 & 0 & p_1 & 0 & 0 & 0 \\
0 & 0 & 0 & p_3 & 0 & p_1 & p_2 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & p_2+p_3 & p_1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p_2+p_3 & 0 & p_1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right] \qquad (3.4)$$

for $t = 1, 2, \ldots, 7$. The partition associated with the random variable X_7 is generated by the subsets $C_x = \{(x, v) : v = 0, 1, 2\}$, $x = 0, 1, 2$. The distribution of X_7 could be easily computed by using the equation (2.2). The numerical result is given in Section 6.

To demonstrate the usefulness of this approach for the compound pattern, let us consider an example of $n = 5$ and the specified compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$ being a union of two simple patterns $\Lambda_1 = b_1b_2$ and $\Lambda_2 = b_3b_1$. We are interested in the distribution of the random variable X_5, the number of patterns either Λ_1 or Λ_2 occurring in a sequence of five i.i.d. three-state trials. Note that the compound pattern $\Lambda_1 \cup \Lambda_2$ can be decomposed into 0, $1 = b_1$, and $2 = b_3$ three ending blocks, where the ending block "0" stands for neither the ending block "1" nor the ending block "2". The imbedded Markov chain associated with the compound pattern has a state space $\Omega = \{(u, v) : u = 0, 1, 2 \text{ and } v = 0, 1, 2\}$ and transition probability matrices

$$M_t = \left[\begin{array}{ccc|ccc|ccc}
p_2 & p_1 & p_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p_1 & p_3 & p_2 & 0 & 0 & 0 & 0 & 0 \\
p_2 & 0 & p_3 & p_1 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & p_2 & p_1 & p_3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_1 & p_3 & p_2 & 0 & 0 \\
0 & 0 & 0 & p_2 & 0 & p_3 & p_1 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & p_2 & p_1 & p_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right] \qquad (3.5)$$

for $t = 1, \ldots, 5$. The numerical results are given in Section 6.

With some modifications of the forward and backward principle, the method could also be extended to a sequence of homogeneous (or non-homogeneous) Markov dependent multi-state trials. We provide the following example. Let $\{Z_i; i = 1, \ldots, n\}$ be a three-state $\{a, b, c\}$ homogeneous Markov chain with a 3×3 transition probability matrix (p_{ij}), where $i, j = a, b$, and c. Let $\Lambda = aa$ be the pattern of interest and the random variable $X_n(\Lambda)$ be the number of patterns Λ in a sequence of n Markov dependent trials. Define the state space $\Omega = \{(x, y) : x = 0, 1, \ldots, [n/2]$, and $y = a^*, a, b, c\}$ and partition $\{C_x = [(x, y) : y = a^*, a, b, c]$, for $x = 0, 1, \ldots, [n/2]\}$. For every t and $\omega = (z_1, \ldots, z_n)$, define $Y_t(\omega) = (x, y)$, where

$$x = \text{ number of a full pattern } \Lambda \text{ in the sequence } z_1, \ldots, z_t,$$

$$y = \begin{cases} a^*, & \text{if a full pattern } \Lambda \text{ occurs at the } t\text{th trial in non-overlapping,} \\ a, & \text{if a half pattern occurs at the } t\text{th trial in non-overlapping,} \\ b \text{ or } c, & \text{if the } t\text{th trial } Z_t = b \text{ or } c \text{ respectively.} \end{cases}$$

We give the three key cases of transition probabilities of $Y_{t+1} = (u, v)$ given $Y_t = (x, y)$ below; for $x = 0, 1, \ldots, ([n/2] - 1)$,

(i) $P(Y_{t+1} = (x + 1, a^*)|Y_t = (x, a)) = p_{aa}$,

(ii) $P(Y_{t+1} = (x, a)|Y_t = (x, a^*)) = p_{aa}$,

and

(iii) $P(Y_{t+1} = (x, v)|Y_t = (x, y)) = p_{yv}$, for $y = b, c$, and $v = a, b$, and c,

and leave the remaining cases to the readers. The exact distribution of $X_n(\Lambda)$ can be obtained by equation (2.2) with ease.

For n s-state (s ≥ 2) trials, the total number of possible realizations $\{\omega = (z_1, \ldots, z_n)\}$ is s^n which tends to infinity exponentially fast. When n the number of trials is relatively large, say $n \geq 50$, the total number of possible realizations becomes extremely large and the computer cannot handle the case. For the case of a simple pattern, the forward and backward principle classifies a realization ω according to the number of patterns occurring in the realization and its ending block. It reduces the number of states to a linear function of n (at most $m = (l + 1)k$ states). This reduction is significant and makes the computation manageable.

4. Exact and Joint Distributions of Runs in Bernoulli Trials

In this section, the “forward and backward principle” and the Markov chain imbedding technique are used to study the marginal and joint distributions of the random variable $R(n, S)$, the number of success runs, and the random variable $R(n, F)$, the number of failure runs, under three different types of Bernoulli trials; (i) i.i.d., (ii) independent but not identically distributed, and (iii) one-step homogeneous or non-homogeneous Markov chain dependency.

4.1. Exact distributions

Let $Z_1, \dots, Z_n$ be n i.i.d. Bernoulli trial and for each trials with probability $p(0 < p < 1)$ for S (Success) and probability q $(q = 1 - p)$ for F (Failure).

(i) In the sequential counting procedure, the pattern "success run" can be decomposed into two ending blocks (or sub-patterns) F and S respectively (it means that a sequence ends with either F or S).

(ii) For $1 \le t \le n$, define the Markov chain $Y_t(\omega) = (u, v)$, with the understanding that the sequence ω has u success runs with an ending block v $(v = F$ or $S)$ in the first t trials. Therefore the state space Ω can be defined as

$$\Omega = \{(0, F)\} \cap \{(u, v) : u = 1, \dots, [(n + 1)/2] \text{ and } v = S, F\}, \tag{4.1}$$

where the state $(0, F)$ indicates that either all the outcomes are failures or when the system is at the initial time, $t = 0$. The partition of the state space Ω associated with the random variable $R(n, S)$ is given by

$$\{C_0, C_1, \dots, C_l; l = [(n + 1)/2]\}, \tag{4.2}$$

where $C_0 = \{(0, F)\}$ and $C_u = \{(u, S) \text{ and } (u, F)\}$, for $u = 1, \dots, l$.

(iii) The transition probability matrix associated with the Markov chain is

$$M(S) = \begin{array}{c} \\ (0,F) \\ (1,S) \\ (1,F) \\ \vdots \\ \vdots \\ (l,S) \\ (l,F) \end{array} \begin{array}{c} \begin{array}{ccccccc} (0,F) & (1,S) & (1,F) & \cdots & \cdots & (l,S) & (l,F) \end{array} \\ \left[\begin{array}{ccccccc} q & p & 0 & & \cdots & & 0 \\ 0 & p & q & 0 & & \cdots & 0 \\ 0 & \ddots & q & p & 0 & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & & 0 & q & p & 0 \\ 0 & & \cdots & & 0 & p & q \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{array}\right] \end{array}, \tag{4.3}$$

where the states (u, v) are arranged lexicographically. It is a $(2l + 1) \times (2l + 1)$ stochastic matrix. Given $\pi_0 = (1, 0, \dots, 0)$, it follows from our construction and equation (2.2) that the distribution of success runs is specified by the following equation, for $r = 0, 1, \dots, l$,

$$P(R(n, S) = r) = \pi_0 M^n(S) U'(C_r). \tag{4.4}$$

For independent but non-identically distributed Bernoulli trials with p_i and q_i being success and failure probabilities respectively, the exact distribution of a success run can be represented by,

$$P(R(n,S)=r)=\pi_0\Big(\prod_{t=1}^{n} M^n(S)\Big)U'(C_r) \text{ for } r=0,1,\dots,l, \tag{4.5}$$

where the transition probability matrix $M_t(S)$ has the same form as $M(S)$ except that p is replaced by p_t and q is replaced by q_t.

Further the result can also be easily extended to a sequence of one-step homogeneous Markov dependent Bernoulli trials $\{Z_t : t = 1,\dots,n\}$ having a transition probability matrix

$$A=\begin{pmatrix} p_{SS} & p_{SF} \\ p_{FS} & p_{FF} \end{pmatrix}. \tag{4.6}$$

Given the initial condition that $P(Z_1 = S) = p$ and $P(Z_1 = F) = q$, the exact distribution of success runs can also be represented by the equation (4.5) with $M_1 = M(S)$ and

$$M_t=\left[\begin{array}{cccccc} p_{FF} & p_{FS} & & & & \\ & p_{SS} & p_{SF} & & 0 & \\ & & p_{FF} & p_{FS} & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & 0 & & & p_{SS} & p_{SF} \\ & & & & & 1 \end{array}\right] \quad \text{for } t=2,3,\dots,n. \tag{4.7}$$

The above results (4.4), (4.5) and (4.7) can be easily extended to the distribution of a failure run $R(n,F)$ by interchanging the S with F. Note that all the results derived from this method are simple and tractable whether they are in a sequence of i.i.d., or non-identical, or one-step homogeneous Markov dependent Bernoulli trials. They differ only in the transition matrices which are rather minor.

4.2. Joint distribution

Let $R(n)$ be the total number of success runs and failure runs in a sequence of n Bernoulli trials. It follows that for every n

$$R(n)=R(n,S)+R(n,F). \tag{4.8}$$

Note that the random variables $R(n,S)$ and $R(n,F)$ are highly related in the following ways. If there are x of success runs, then there can only be $x+1$, x, or $x{+}1$ of failure runs. If $R(n)=2x$ is an even number, then $R(n,S)=R(n,F)=x$, and if $R(n)=2x-1$ is an odd number, then either $R(n,S)+1=R(n,F)=x$ or in the situation $R(n,S)=R(n,F)+1=x$. Hence, for any sequence of Bernoulli trials it can only be classified into one for the following four types of states:
(i) $(x,x-1)$ being a sequence which has x success runs and $(x-1)$ failure runs with an ending block S.
(ii) $(x,x+1)$ being a sequence which has x success runs and $(x+1)$ failure runs with an ending block F.
(iii) (x,x,F) being a sequence which has x success runs and x failure runs with an ending block F.
(iv) (x,x,S) being a sequence which has x success runs and x failure runs with an ending block S.

For finding the distribution of $R(n)$, we could first find the joint distribution of $R(n,S)$ and $R(n,F)$, then project the joint distribution of $(R(n,S),R(n,F))$ on the partition generated by $R(n,S)+R(n,F)$.

Let $\{Y_t\}$ be a Markov chain defined on the state space specified by (i) to (iv),

$$\Omega=\{(1,0),(0,1),(1,1,S),(1,1,F),\ldots,(l,1,S),(l,1,F)\}. \tag{4.9}$$

For each $\omega=(SSFS\ldots FF)$, we define $Y_t(\omega)=$ (the number of success runs, the number of failure runs, ending sub-pattern) as an element of Ω. In the case of independent but non-identical Bernoulli trials, the definition of Y_t, $t=1,\ldots,n$ yields the initial distribution $\pi_1=(p_1,q_1,0,\ldots,0)$ of Y_1 and the transition probability matrices for $t=2,3,\ldots,n$

$$M_t=\begin{array}{c}(1,0)\\(0,1)\\(1,1,S)\\(1,1,F)\\\cdot\\\cdot\\\cdot\\(l,l-1)\\(l-1,l)\\(l,l,S)\\(l,l,F)\end{array}\begin{bmatrix}
p_t & 0 & 0 & q_t & & & & & & & \\
 & q_t & p_t & 0 & 0 & & & & 0 & & \\
 & & p_t & 0 & 0 & q_t & & & & & \\
 & & & q_t & p_t & & & & & & \\
 & & & & \cdot & \cdot & \cdot & \cdot & & & \\
 & & & & & \cdot & \cdot & \cdot & \cdot & & \\
 & & & & & & \cdot & \cdot & \cdot & \cdot & \\
 & & & & & & & p_t & 0 & 0 & q_t \\
 & & & & & & & & q_t & p_t & 0 \\
 & & 0 & & & & & & & 1 & 0 \\
 & & & & & & & & & & 1
\end{bmatrix}. \tag{4.10}$$

The exact distribution of the total number of runs can be represented by

$$P(R(n)=r)=(p_1,q_1,0,\ldots,0)\Big(\prod_{t=2}^{n}M_t\Big)U'(C_r),\ \text{for } r=1,\ldots,l, \tag{4.11}$$

where $C_r = \{(r/2, r/2, S), (r/2, r/2, F)\}$, if r is an even number, and $C_r = \{((r+1)/2, (r-1)/2), ((r-1)/2, (r+1)/2)\}$, if r is an odd number.

The marginal distributions of $R(n, S)$ and $R(n, F)$ can also be obtained by projecting the joint distribution to the partitions generated by random variables $R(n, S)$ and $R(n, F)$ respectively. Again, with some simple modifications, the result (4.11) also holds for the both homogeneous and non-homogeneous Markov dependent Bernoulli trials.

5. Waiting Time Distribution

Consider a simple pattern $\Lambda = b_{i_1}, \ldots, b_{i_k}$ composed of a sequence of k specified symbols, $b_{i_1}, \ldots,$ and b_{i_k}. The pattern Λ can always be decomposed forward as $k+1$ ending blocks $0{,}1 = b_{i_1}, \ldots, k-1 = b_{i_1}, \ldots, b_{i_k}$. We define the state space $\Omega_{m,k}$ associated with the pattern Λ as

$$\Omega_{m,k} = \{(x, y) : x = 0, 1, \ldots, m-1 \text{ and } y = 0, 1, \ldots, k-1\} \cup \{\alpha\}, \qquad (5.1)$$

where the α is an absorbing state. Construct a Markov chain $\{Y_t : t = 1, \ldots, n\}$ on the state space $\Omega_{m,k}$ with

$$Y_t(\omega) = \begin{cases} (x, y), & \text{if there are } x \ \Lambda \text{ patterns with an ending block } y \text{ in the} \\ & \text{first } t \text{ trials,} \\ \alpha, & \text{if there are } m \text{ or more than } m \ \Lambda \text{ patterns in the first } t \\ & \text{trials,} \end{cases}$$

and the transition probability matrix

$$M_t(m) = \begin{array}{c} (0,0) \\ (0,1) \\ \vdots \\ (m-1, k-1) \\ -------- \\ \alpha \end{array} \left[\begin{array}{ccc|c} & & & \\ & & & \\ & p_{(x,y),(u,v)}(t) & & p_{(x,y),\alpha}(t) \\ & & & \\ -- & ----- & - \ - & ---- \\ 0 & \cdots & 0 & 1 \end{array} \right], \qquad (5.2)$$

where α is an absorbing state and $M_t(m)$ is a $(mk+1) \times (mk+1)$ matrix whose transition probabilities $p_{(x,y),(u,v)}(t)$ and $p_{(x,y),\alpha}(t)$ are determined by the equation (3.2) and (3.3). Let $X_n(\Lambda)$ be the number of patterns occurring in a sequence of n multi-state trials and $W(m, \Lambda)$ be the waiting time (the number of trials required) for the mth ($m = 1, 2, \ldots$) pattern to have occurred. For example, $W(1, \Lambda) = 5$ means the first time the pattern Λ occurred at the 5th trial.

Given Λ, m and $n (mk \le n)$, it follows from the definitions that the two random variables $X_n(\Lambda)$ and $W(m,\Lambda)$ are one-to-one related in the following way:

$$W(m,\Lambda) \le n \text{ if and only if } X_n(\Lambda) \ge m. \tag{5.3}$$

Since the exact distribution of $X_n(\Lambda)$ can be obtained by using the method of Markov chain imbedding, hence the relationship (5.3) yields the distribution for the waiting time $W(m,\Lambda)$:

$$P(W(m,\Lambda)=n) = \pi_0 \Big(\prod_{t=1}^{n-1} M_t(m) \Big) (I - M_n(m)) U'(m) \tag{5.4}$$

for $n \ge m \times k$, where $\pi_0 = (1,0,\dots,0)$ is a $1 \times (km+1)$ vector, $M_t(m)$ is given by (5.2), k is the length of the pattern Λ, and $U'(m) = \sum_{x=0}^{m-1} U'(C_x)$.

By properly defining the transition matrix, this general theorem covers many well-known results about the waiting time distributions given by Hahn and Gage (1983), Benevento (1984), Ebneshahrashoob and Sobel (1990), Balassubramanian, Viveros and Balakrishnan (1993), and Fu and Koutras (1994). For example, let $\Lambda = SS\dots SS$ be a pattern of consecutive k successes in a sequence of i.i.d. Bernoulli trials having probabilities p and q for success and failure respectively. The simple pattern $\Lambda = SS\dots SS$ of consecutive k successes can be decomposed into k ending blocks $0 = F$, $1 = S$, $2 = SS, \dots, k-1$ consecutive successes. The state space $\Omega = \{(0,y) : y = 0,1,\dots,k-1\} \cup \{\alpha\}$ contains $k+1$ states having α as absorbing state. For the i.i.d. Bernoulli trials, the transition matrices defined by (5.2) become,

$$M = \begin{array}{cc} \begin{array}{c} (0,0) \\ (0,1) \\ \cdot \\ \cdot \\ (0,k-1) \\ \alpha \end{array} & \left[\begin{array}{ccccccc} q & p & \cdot & \cdot & \cdot & 0 & 0 \\ q & 0 & p & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ q & 0 & 0 & \cdot & \cdot & 0 & p \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 1 \end{array} \right] \end{array} \quad \text{for all } t = 1,\dots,n. \tag{5.5}$$

It follows immediately from (5.4) that the distribution of the waiting time for the first occurrence of the pattern $\Lambda = SS\dots SS$, is given by

$$P(W(1,\Lambda)=n) = \pi_0 M^{n-k} U' p^k, \tag{5.6}$$

where $\pi_0 = (1,0,\dots,0), U = (1,0,\dots,0)$, and M is given by (5.5). By the same token, the above result (5.4) is also true for the compound pattern.

6. Numerical Examples

Several numerical examples are given here to illustrate our method and results developed in previous sections.

Example 1. Consider 10 i.i.d. Bernoulli trials each with $p = 0.4$ and $q = 1 - p = 0.6$. By (2.2) and transition matrix (4.10), the following Table 1 gives the joint distribution and marginal distributions of success runs and failure runs respectively.

Table 1. Joint and marginal distributions of runs $R(10, S)$ and $R(10, F)$

$R(10,F)$ $R(10,S)$	1	2	3	4	5		$R(10,S)$
0		.006047					.006047
1	.000105	.023557	.073503				.097165
2		.020726	.169656	.175211			.365593
3			.079272	.225210	.096082		.400564
4				.054057	.059985	.009157	.123199
5					.005839	.000796	.006635
$R(10,F)$	.000105	.050330	.322431	.454478	.161907	.009953	1.0000

Projecting the joint distribution of $R(10, S)$ and $R(10, F)$ on the partition generated by $R(10) = R(10, S) + R(10, F)$, this yields the distribution of the total number of success and failure runs shown by the following Table 2.

Table 2. Distribution of the total number $R(10)$ of runs in ten trials.

$R(10)$	1	2	3	4	5	6	7	8	9	10
Prob.	.006152	.023557	.094229	.169656	.254483	.225210	.150139	.059985	.001505	.000796

Example 2. Consider a sequence of i.i.d. trials each of them having three states b_1, b_2, and b_3 and corresponding probabilities of occurring $p_1 = p_2 = 0.3$ and $p_3 = 0.4$. Suppose we are interested in finding the *exact* distribution for the number of a specified compound pattern $\Lambda = b_1b_2 \cup b_3b_1$ in n trials. Applying the forward and backward principle, this yields a Markov chain Y_t defined on the state space $\Omega = \{(u, v) : u = 0, 1, \ldots, l, l = [n/2] \text{ and } v = 0, 1, 2\}$, having transition matrices

$$M_t = \begin{bmatrix} A & B & 0 & 0 & 0 & 0 \\ 0 & A & B & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & A & B \\ 0 & 0 & 0 & 0 & 0 & A^* \end{bmatrix}, \quad \text{for } t = 1, \ldots, n,$$

where

$$A = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0 & 0.3 & 0.4 \\ 0.3 & 0 & 0.4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 0.3 & 0 & 0 \\ 0.3 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad A^* = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and the partition $\{C_x = [(x, v) : v = 0, 1, 2]; x = 0, 1, \ldots, l.\}$. The following Table 3 provides numerical results for the distributions of the number of patterns Λ for $n = 7, 11,$ and 21.

Table 3. Exact distributions for the number of patterns Λ in $n = 7, 11$ and 21 three-state trials.

n	x	$p(x)$	n	x	$p(x)$
7	0	.241887	21	0	.0092905
	1	.468508		1	.0600785
	2	.257324		2	.1680750
	3	.032281		3	.2654140
				4	.2587810
11	0	.095315		5	.1600370
	1	.307973		6	.0620535
	2	.365549		7	.0143712
	3	.189825		8	.0017987
	4	.039238		9	.0000999
	5	.002100		10	$.155\times10^{7}$

The above numerical results were done with Mathematica on an IBM 386 PC. For each case, the CPU time is rather minimum usually it takes no more than a few seconds even in the case of $n = 21$.

7. Discussions

The results about the exact distributions of number of patterns obtained by combinatorial, generating function, and recursive equation methods for i.i.d. multi-state trials are usually difficult to extend to Markov chain dependent trials. The recursive method of the renewal equation (see Schwager (1983)) could not be used for searching the joint distribution of runs. On the contrary, from examples in the previous sections, the results for i.i.d. trials obtained by using the Markov chain imbedding technique can be easily extended to trials with one-step (or m-step) Markov chain dependency. The basic idea of handling the one step (or m-step) Markov chain dependence is to have an additional coordinate to indicate the outcome of the last trial (or last m trials).

Suppose $X_n(\Lambda)$, the number of patterns Λ in a sequence of n multi-state trials, could be imbedded into a finite Markov chain $\{Y_t : t = 1, \ldots, n\}$ defined on the state space

$$\Omega = \{(x,y) : x = 0,\ldots,l_n \text{ and } y = 0,\ldots,m_x - 1\} = \{C_x : x = 0,\ldots,l_n\}, \tag{7.1}$$

where $l_n = [n/k]$, k is the length of the pattern and m_x and $\mathrm{card}(C_x)$, and having transition matrices of the following form

$$M_t = \left[\begin{array}{cccccc}
A_t(0) & B_t(0) & & & & \\
 & A_t(1) & B_t(1) & & 0 & \\
 & \cdot & \cdot & & & \\
 & & \cdot & \cdot & & \\
 & & & \cdot & \cdot & \\
 & & & \cdot & \cdot & \\
 & 0 & & & \cdot & B_t(l_{n-1}) \\
 & & & & & A_t(l_n)
\end{array}\right] \tag{7.2}$$

for $t = 1,\ldots,n$. There are many runs and patterns of interest whose transition matrices do have the above form. For instance, all the transition matrices associated with the run statistics $N_{n,k}, M_{n,k}, G_{n,k}$, (see Fu and Koutras (1994)) and patterns $\Lambda = b_1b_1b_2$ and $\Lambda = b_1b_2 \cup b_3b_1$ (see Section 3) all have the form (7.2).

Denote vectors, $\alpha_t(x) = (P(Y_t = (x,0)),\ldots,P(Y_t = (x,m-1)))$ for $t = 1,\ldots,n$. The probability of $X_n(\Lambda) = x$ can be represented as

$$P(X_n(\Lambda) = x) = \alpha_n(x)\mathbf{1}', \text{ for all } x = 0,1,\ldots,l_n, \tag{7.3}$$

where $\mathbf{1}' = (1,\ldots,1)'$. Since $M_t = K_t + H_t$, where K_t is a diagonal matrix with components $A_t(x)$, for $x = 0,1,\ldots,l_n$ and H_t is an upper diagonal matrix with components $B_t(x)$, for $x = 0,1,\ldots,l_n - 1$, it follows from the fact $\pi_0(\prod_{j=1}^{t} M_j) = \pi_0(\prod_{j=1}^{t-1} M_j)M_t$ for every $t = 1,\ldots,n$ that the following recursive equations hold

$$\begin{cases} \alpha_t(0) = \alpha_{t-1}(0)A_t(0) \\ \\ \alpha_t(x) = \alpha_{t-1}(x-1)B_t(x-1) + \alpha_{t-1}(x)A_t(x), \ x = 1,\ldots,l_n, \end{cases} \tag{7.4}$$

for every $t = 1,\ldots,n$. The above recursive formula (7.4) provides an efficient way to compute the probabilities $P(X_n(\Lambda) = x) = \alpha_n(x)\mathbf{1}'$, for all $x = 0,1,\ldots,l_n$, especially when the dimension of the transition matrix M_t is so large that the formula $\pi_0(\prod_{t=1}^{n} M_t)U'(C_x)$ takes longer time to do the computation. From backward multiplication, the finite Markov chain imbedding technique always provides a recursive equation like (7.4) automatically for the distribution of $X_n(\Lambda)$ which could not be easily obtained through the combinatorial or renewal equation methods. The recursive equation (7.4) provides a very useful tool to study the limiting distribution for the random variable $X_n(\Lambda)$. In spite of their potential interests, we will not pursuit these here.

The Markov chain imbedding method can also be modified to study the distributions of certain runs and patterns on permutations of n symbols or permutations of several symbols alike. The extensions of the method for these cases are often non-trivial. Let $\{\pi_n = (3, 2, n, \dots, 5)\}$ be all permutation of n integers 1,2,$\dots$,n. Define the index functions of successions for $i = 1, \dots, n-1$, $I_i(\pi_n) = 1$ if $\pi_n(i) + 1 = \pi_n(i+1)$, and 0 otherwise, where $\pi_n(i)$ is the number of ith coordinate of the permutation π_n and the number of successions in a random permutation, $X_n(\pi_n) = \sum_{i=1}^{n} I_i(\pi_n)$. By extending the Markov chain imbedding technique to random permutation, Fu (1995) obtained the exact distribution for the random variable $X_n(\pi_n)$, the number of successes, and also showed that the limiting distribution is a Poisson distribution with parameter $\lambda = 1$.

For the compound pattern $\Lambda = \cup_{i=1}^{s-1}\Lambda_i$, where $\Lambda_i = b_i b_{i+1}$, $i = 1, \dots, s-1$, if $s = n$ and $p_i = 1/n$ for all $i = 1, \dots, n$ we believe that the random variable $X_n(\Lambda)$ also has a Poisson distribution as $n \to \infty$ for the same reason that the number of successions in a random permutation of n integers has a Poisson distribution as $n \to \infty$. In practice, this means that when the number of states s in each trial is very large, then the distribution of the number of patterns Λ can be approximated by a Poisson distribution.

The finite Markov chain imbedding may not be unique. Often there are several different ways to imbed a random variable. To find the best imbedded Markov chain it requires experience and understanding the structure of the counting process associated with the random variable. Sometimes in order to obtain the distribution of a pattern Λ_1, it may be more efficient and simpler to find the joint distribution of pattern Λ_1 and Λ_2 first, then project the joint distribution on the partition $\{C_x\}$ associated with the pattern Λ_1. This can be seen from the procedure for finding the distribution of the total number of runs given in the Section 4.

With today's high speed computers, the exact distributions (also means and variances) of runs and patterns can be obtained by the method of finite Markov chain imbedding with ease. All our numerical results in Section 6 were carried out on an IBM 386 PC with less then one minute in CPU time in each case. For very large n, if the probability of pattern Λ occurring is very small, say having an order of λ/n, then the distribution of the pattern Λ occurring can be approximated by a Poisson distribution with parameter λ. This fact had been broadly used in studying reliabilities of certain engineering systems when the failure probabilities of components are very small, for instance the reliability of consecutive-k-out-of-n : F system.

Acknowledgement

The author would like to thank the referee for his useful suggestions and comments. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant NSERC A-9216, and National Science Council of Republic of China under Grant 85-2121-M-259-003.

References

Balasubramanian, K., Viveros, R. and Balakrishnan, N. (1993). Sooner and later waiting time problems for Markovian Bernoulli trials. *Statist. Probab. Lett.* **18**, 153-161.

Benevento, R. V. (1984). The occurrence of sequence patterns in ergodic Markov chains. *Stochastic Process. Appl.* **17**, 369-373.

Chao, M. T. and Fu, J. C. (1989). A limit theorem of certain repairable systems. *Ann. Inst. Statist. Math.* **41**, 809-818.

Chao, M. T. and Fu, J. C. (1991). The reliability of large series systems under Markov structure. *Adv. Appl. Probability* **23**, 894-908.

Cochran, W. G. (1938). An extension of Gold's method for examining the apparent persistence of one type of weather. *Roy. Meteorol. Soc. Quart. Journal* **64**, 631-634.

Ebneshahrashoob, M. and Sobel, M. (1990). Sooner and later waiting time problems for Bernoulli trials : frequency and run quotas. *Statist. Probab. Lett.* **9**, 5-11.

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

Fu, J. C. (1986). Reliability of consecutive-k-out-of-n : F systems with $(k-1)$-step Markov dependence. *IEEE Trans. Reliability* **R35**, 602-606.

Fu, J. C. (1995). Exact and limiting distributions of the number of successions in a random permutation. *Ann. Inst. Statist. Math.* **47**, 435-446.

Fu, J. C. and Koutras, M. V. (1994). Distribution theory of runs : A Markov chain approach. *J. Amer. Statist. Assoc.* **89**, 1050-1058.

Gibbons, J. D. (1985). *Non-parametric Statistical Inference.* M. Dekker, New York.

Hahn, G. J. and Gage, J. B. (1983). Evaluation of a start-up demonstration test. *J. Qual. Technol.* **15**, 103-106.

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

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

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

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

Stevens, W. L. (1939). Distributions of groups in a sequence of alternatives. *Ann. of Eugenics* **9**, 10-17.

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Ann. Math. Statist.* **11**, 147-162.

Walsh, J. E. (1965). *Handbook of Non-Parametric Statistics.* D. Van Nostrand Company, Inc. New York.

Wishart, J. and Hirshfeld, H. O. (1936). A theorem concerning the distribution of joins between line segments. *London Math. Soc. J.* **11**, 227-235.

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

Institute of Applied Mathematics, National Donghwa University, Hualien, Taiwan.

(Received August 1994; accepted January 1996)