ELSEVIER

Statistics & Probability Letters 32 (1997) 393–404

STATISTICS & PROBABILITY LETTERS

Non-parametric randomness tests based on success runs of fixed length

M.V. Koutras *, V.A. Alexandrou [1]

Department of Statistics, University of Athens, Athens 15784, Greece

Received June 1994; revised March 1996

Abstract

Let X_n be a random variable enumerating the number of appearances of a specific pattern in a sequence of n Bernoulli trials. A new method is presented for obtaining the conditional distribution of X_n given the number of successes in the n trials. The method is applied to three fixed-length run statistics and the results are used for establishing and investigating certain non-parametric tests of randomness.

Keywords: Bernoulli trials; Success runs; Non-parametric tests of randomness; Markov chain imbeddable variables; Generating functions

1. Introduction

Tests of randomness are an important addition to statistical theory. One of the best known and easiest to apply procedure, for testing random versus non-random ordering in a sequence of two types of symbols is the classical runs test (Gibbons, 1971). A run is in general defined as a succession of one or more identical symbols which are followed and preceded by a different symbol or no symbol at all. Clues to lack of randomness are provided by any tendency of the symbols to exhibit a definite pattern in the sequence. Since a run which is unusually long reflects a tendency for like objects to cluster and therefore a trend, Mosteller (1941) suggested a test for randomness based on the length of the longest run. Mood (1940) gave a valuable collection of formulae related to run tests.

When dealing with the distribution theory of runs one has to face two classes of different but interrelated problems. The first one, which is combinatorial in nature, is concerned with random arrangements of a fixed number of elements of each of two kinds. The second, instead of having the number of elements of each kind fixed, assumes that they are randomly drawn from a binomial population. As an application of the first problem we mention the well-known Wald and Wolfowitz (1940) test for checking the hypothesis that two samples come from the same continuous distribution. The second set-up is very useful in quality control of

* Corresponding author.

[1] Supported by the National Scholarship Foundation of Greece.

PII S0167-7152(96)00099-5

manufactured items where one of the fundamental problems is to decide whether the observations are random or in the language employed in this field, whether statistical control exists.

In this paper we propose a certain procedure for developing non-parametric tests based on run-related statistics. Though our target is to study random arrangements of a fixed number of elements of two kinds, we start off with the study of binomial populations and develop a general method for passing to the non-parametric set-up.

Let X_n denote a statistic related to a sequence of independent binary variables. In Section 2 we provide a means for computing the double generating function of the conditional distribution of X_n given the number S_n of successes in the n trials by making use of the generating function $\varphi_x(w; p)$ of the unconditional probabilities $\{P(X_n = x)\}_{n\geq 0}$. In Section 3 we discuss in brief the Markov chain imbedding technique introduced by Fu and Koutras (1994) and subsequently refined by Koutras and Alexandrou (1995), and derive a matrix formula for the evaluation of $\varphi_x(w; p)$ for Markov chain imbeddable variables. As an application, the generating functions of certain fixed-length run statistics are obtained. Moreover, we derive simple formulae (involving binomial coefficients) for the exact conditional distribution of the run statistics. Finally, in Section 4 we use Monte Carlo simulation techniques to compare certain non-parametric randomness tests based on fixed-length run statistics to the classical run and longest run test.

2. Generating functions for enumerating variables: connection between the conditional and unconditional distribution

Consider a sequence $Z_1, Z_2, \ldots, Z_n$ of n Bernoulli trials with constant success (S) probabilities $p = P(Z_i = 1)$ and failure (F) probabilities $q = P(Z_i = 0), i = 1, 2, \ldots, n$. Let also X_n denote a random variable counting the number of appearances of a specific pattern in the sequence $Z_1, Z_2, \ldots, Z_n$. As examples of such variables we mention

(a) the total number R_n, of runs of either type (S or F) (Mood, 1940; Gibbons, 1971),
(b) the number $G_{n,k}$ of success runs of length at least k (k is a positive integer) (Hirano and Aki, 1993),
(c) the number $M_{n,k}$ of overlapping success runs of length k (Ling, 1988),
(d) the number $N_{n,k}$ of non-overlapping and recurrent success runs of length k (Feller, 1968).

For the evaluation of the probability mass function $P(X_n = x)$ of the statistics mentioned above, several methods have been proposed, even for the case of non-identical or Markov dependent Bernoulli trials.

A detailed study of R_n can be found in Gibbons (1971). Recently Lou (1996a, b) used a Markov chain imbedding technique to study R_n for iid or Markov dependent trials. The distributions of $N_{n,k}, M_{n,k}$ have been called *binomial distributions of order k* (because of their apparent similarity to the binomial distribution, which is obtained for $k = 1$), and have been extensively investigated during the last 10 years by Ling (1988), Hirano et al. (1991), etc. For $G_{n,k}$ and the case of Markov dependent trials we refer to Hirano and Aki (1993), Fu and Koutras (1994) and Koutras and Alexandrou (1995).

In this section, motivated by the problem of establishing non-parametric randomness tests based on enumerating variables X_n as the ones described above, we provide a method for obtaining the conditional distribution of X_n, given the number S_n of successes observed in the sequence of outcomes $Z_1, Z_2, \ldots, Z_n$. Our method is based on generating function techniques.

Theorem 1. *Let*

$$\varphi_x(w; p) = \sum_{n=0}^{\infty} P(X_n = x) w^n \tag{1}$$

be the generating function of the probabilities $\{P(X_n = x)\}_{n\geq 0}$ (*Convention*: $P(X_0 = 0) = 1$). *Then the double generating function of the quantitites*

$$\alpha(x, y; n) = \binom{n}{y} P(X_n = x \mid S_n = n - y) \tag{2}$$

is given by

$$\psi_x(t, w) = \sum_{n=0}^{\infty} \sum_{y=0}^{n} \alpha(x, y; n) t^y w^n = \varphi_x\left((1 + t)w; \frac{1}{1 + t}\right).$$

Moreover

$$P(X_n = x \mid S_n = n - y) = \frac{(n - y)!}{(n!)^2} \left[\frac{\partial^{n-y}}{\partial t^y \partial w^n} \psi_x(t, w)\right]_{t=w=0}.$$

Proof. Since S_n is a sufficient statistic for p, we conclude that the conditional distribution of X_n given that $S_n = n - y$ does not depend on p. Hence the quantity

$$\binom{n}{y} P(X_n = x \mid S_n = n - y)$$

is a function of x, y and n only, a fact justifying the notation introduced in (2). By virtue of the total probability theorem we may write

$$P(X_n = x) = \sum_{y=0}^{n} P(X_n = x \mid S_n = n - y) \binom{n}{y} p^{n-y} q^y = \sum_{y=0}^{n} \alpha(x, y; n) p^n (q/p)^y$$

and imputing $P(X_n = x)$ in formula (1) we deduce

$$\varphi_x(w; p) = \sum_{n=0}^{\infty} \sum_{y=0}^{n} \alpha(x, y; n)(q/p)^y (pw)^n$$

which, on letting $t = q/p$, implies

$$\varphi_x\left(w; \frac{1}{1 + t}\right) = \sum_{n=0}^{\infty} \sum_{y=0}^{n} \alpha(x, y; n) t^y \left(\frac{w}{1 + t}\right)^n.$$

The required result follows immediately. □

For the sake of illustration, let us consider the case $X_n = R_n$ (total number of runs of either type). To evaluate the generating function $\varphi_x(w; p)$ we can think of x cells where the even numbered ones contain at least one success (i.e. their enumerating generating function is $\sum_{i\geq 1}(pw)^i = (pw)/(1 - pw)$) and the odd numbered at least one failure (i.e. their enumerating generating function is $\sum_{i\geq 1}(qw)^i = (qw)/(1 - qw)$). Therefore

$$\varphi_x(w; p) = \begin{cases} 2\left(\dfrac{pw}{1 - pw}\right)^{x/2} \left(\dfrac{qw}{1 - qw}\right)^{x/2} & \text{if } x \text{ even,} \\ \left(\dfrac{pw}{1 - pw}\right)^{(x+1)/2} \left(\dfrac{qw}{1 - qw}\right)^{(x-1)/2} + \left(\dfrac{pw}{1 - pw}\right)^{(x-1)/2} \left(\dfrac{qw}{1 - qw}\right)^{(x+1)/2} & \text{if } x \text{ odd,} \end{cases}$$

and making use of Theorem 1 we deduce an explicit formula for $\psi_x(t, w)$.

Expanding now $\psi_x(t,w)$ with respect to t and w we readily obtain the well-known formulae for $P(R_n = x \mid S_n = n - y)$ (cf. Gibbons, 1971, p. 53).

3. Generating functions for Markov chain imbeddable variables

As a result of Theorem 1, should one be able to compute the generating function $\varphi_x(w; p)$ of the sequence $\{P(X_n = x)\}_n$, he would easily obtain the conditional distribution $P(X_n = x \mid S_n = n - y)$ by using the Taylor series expansion of $\psi_x(t,w) = \varphi_x((1+t)w;(1+t)^{-1})$ with respect to t and w. We now turn our attention to the problem of evaluating the generating function $\varphi_x(w; p)$.

In two recent papers by Fu and Koutras (1994) and Koutras and Alexandrou (1995) it was shown that random variables X_n enumerating certain patterns in a sequence $Z_1, Z_2, \ldots, Z_n$ of Bernoulli trials can be studied efficiently by considering a proper Markov chain imbedding technique. Since this approach provides a universal tool for the analysis of such variables, it would be very useful to derive a formula for evaluating $\varphi_x(w; p)$ when the Markov chain description of X_n is known. For the sake of completeness of our exposition, before stating the main result of this paragraph, we present in brief the main terminology and results of the Markov chain approach; more details can be found in Fu and Koutras (1994), Koutras and Alexandrou (1995), Fu (1995) and Lou (1996a, b).

A random variable X_n taking values in $\{0, 1, \ldots, \ell_n\}$ (n is a non-negative integer) will be called a *Markov chain imbeddable Variable of Binomial type* (MVB) if

(i) there exists a Markov chain $\{Y_t : t \geqslant 0\}$ defined on the state space Ω with

$$\Omega = \bigcup_{x \geqslant 0} C_x, \qquad C_x = \{c_{x0}, c_{x1}, \ldots, c_{x,s-1}\},$$

(ii) $P(Y_t = c_{yj} \mid Y_{t-1} = c_{xi}) = 0$ for all $y \neq x, x+1$,

(iii) for every $x = 0, 1, \ldots, \ell_n$ the probabilities $P(X_n = x)$ are given by

$$P(X_n = x) = P(Y_n \in C_x).$$

It is sufficient for our purposes and also of greater simplicity to consider the special case of *homogeneous Markov chains* $\{Y_t : t \geqslant 0\}$ described by

(a) the initial probabilities $\pi_x = (P(Y_0 = c_{x0}),\ P(Y_0 = c_{x1}), \ldots, P(Y_0 = c_{x,s-1}))$, $x \geqslant 0$,
(b) the *within states* one step transition matrix $A = (a_{ij})_{s \times s}$, $a_{ij} = P(Y_t = c_{xj} \mid Y_{t-1} = c_{xi})$,
(c) the *between states* one step transition matrix

$$B = (b_{ij})_{s \times s}, \qquad b_{ij} = P(Y_t = c_{x,j} \mid Y_{t-1} = c_{x-1}).$$

Let $\boldsymbol{f}_t(x)$, $0 \leqslant t \leqslant n$ denote the *probability* (row) *vectors*

$$\boldsymbol{f}_t(x) = (P(Y_t = c_{x0}),\quad P(Y_t = c_{x1}), \ldots, P(Y_t = c_{x,s-1}))$$

and $\mathbf{1} = (1, 1, \ldots, 1)$ the row vector of R^s with all its entries being 1. As Koutras and Alexandrou (1995) point out, the probability mass function of X_n can be expressed as

$$P(X_n = x) = \boldsymbol{f}_n(x)\mathbf{1}', \quad x = 0, 1, \ldots, \ell_n$$

with $\boldsymbol{f}_t(x)$, $0 \leqslant x \leqslant \ell_n$, $1 \leqslant t \leqslant n$ being a double sequence of vectors which satisfies the recurrence relations

$$\boldsymbol{f}_t(0) = \boldsymbol{f}_{t-1}(0)A, \quad \boldsymbol{f}_t(x) = \boldsymbol{f}_{t-1}(x)A + \boldsymbol{f}_{t-1}(x-1)B, \qquad 1 \leqslant x \leqslant \ell_n,\ t = 1, 2, \ldots, n \tag{3}$$

with initial conditions $\boldsymbol{f}_0(x) = \pi_x$, $0 \leqslant x \leqslant \ell_n$.

We are now ready to prove the main result of this paragraph, which expresses the generating function $\varphi_x(w; p)$ in terms of A, B and π_0.

Theorem 2. *If X_n is a MVB then the generating function $\varphi_x(w; p) = \sum_{\{n: x \leqslant \ell_n\}} P(X_n = x)w^n$ is given by*

$$\varphi_x(w; p) = w^x \pi_0((I - wA)^{-1}B)^x (I - wA)^{-1}\mathbf{1}' = w^x \pi_0 (I - wA)^{-1}(B(I - wA)^{-1})^x \mathbf{1}'. \tag{4}$$

Proof. Condition (ii) of the MVB definition implies that $\ell_{n+1} - \ell_n \in \{0, 1\}$. Let $\{s_x\}_{x \geqslant 0}$ be a sequence of integers such that $n \geqslant s_x$ if and only if $x \leqslant \ell_n$ (convention: $\ell_0 = s_0 = 0$). Then $s_x > s_{x-1}$ for all $x = 1, 2, \ldots$ and $\boldsymbol{f}_n(x) = \mathbf{0}$ for all $n < s_x$. We consider the vector generating function

$$\boldsymbol{\varphi}_x(w; p) = \sum_{\{n: x \leqslant \ell_n\}} \boldsymbol{f}_n(x) w^n = \sum_{n \geqslant s_x} \boldsymbol{f}_n(x) w^n, \quad x \geqslant 0.$$

For $x = 0$ we get, by virtue of (3),

$$\boldsymbol{\varphi}_0(w; p) = \sum_{n \geqslant 0} \boldsymbol{f}_n(0) w^n = \boldsymbol{f}_0(0) + wA \sum_{n \geqslant 1} \boldsymbol{f}_{n-1}(0) w^{n-1}$$

or equivalently $\boldsymbol{\varphi}_0(w; p) = \pi_0 + wA\boldsymbol{\varphi}_0(w; p)$, which implies

$$\boldsymbol{\varphi}_0(w; p) = \pi_0 (I - wA)^{-1}. \tag{5}$$

For $x > 0$ we have, because of (3),

$$\boldsymbol{\varphi}_x(w; p) = w \left\{ \left(\sum_{n \geqslant s_x - 1} \boldsymbol{f}_n(x) w^n \right) A + \left(\sum_{n \geqslant s_x} \boldsymbol{f}_{n-1}(x-1) w^{n-1} \right) B \right\},$$

which can be alternatively written as

$$\boldsymbol{\varphi}_x(w; p) = w \left(\boldsymbol{\varphi}_x(w; p)A + \boldsymbol{\varphi}_{x-1}(w; p)B + \boldsymbol{f}_{s_x - 1}(x) w^{s_x - 1} A - \left(\sum_{s_{x-1} \leqslant n < s_x - 1} \boldsymbol{f}_n(x-1) w^n \right) B \right).$$

It is not difficult to verify that the last two terms of the above formula's RHS vanish; thus $\boldsymbol{\varphi}_x(w) = w\boldsymbol{\varphi}_x(w)A + w\boldsymbol{\varphi}_{x-1}(w; p)B$, yielding

$$\boldsymbol{\varphi}_x(w; p) = w\boldsymbol{\varphi}_{x-1}(w; p)B(I - wA)^{-1}, \quad x \geqslant 1. \tag{6}$$

Relations (4) are now readily obtained by combining formulae (5) and (6). □

It should be mentioned that usually the vector π_0 of initial probabilities equals $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$. Moreover, in most of the applications, matrix B contains only one or two non-zero entries; as a consequence, the evaluation of the power $((I - wA)^{-1}B)^x$ in formula (4) turns out to be reasonably easy.

Corollary 1. *The generating function $\varphi_x(w; p)$ of the number $G_{n,k}$ of success runs of length at least k is given by*

$$\varphi_x(w; p) = \frac{(pw)^{kx}(qw)^{x-1}}{(1 - pw)^x \{1 - qw(1 - (pw)^k)(1 - pw)^{-1}\}^{x+1}}, \quad x \geqslant 1. \tag{7}$$

Proof. In this case, the within states and between states transition matrices are given by (see Koutras and Alexandrou, 1995)

$$A=\begin{bmatrix} q & p & 0 & \cdot & 0 & 0\\ q & 0 & p & \cdot & 0 & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ q & 0 & 0 & \cdot & p & 0\\ q & 0 & 0 & \cdot & 0 & 0\\ q & 0 & 0 & \cdot & 0 & p \end{bmatrix}_{(k+1)\times(k+1)} \qquad B=\begin{bmatrix} & & & & 0\\ & & \mathbf{0}_{k\times k} & & 0\\ & & & & \cdot\\ & & & & 0\\ & & & & p\\ 0 & 0 & 0 & \cdot\ 0 & 0 \end{bmatrix}_{(k+1)\times(k+1)}$$

If $(I-wA)^{-1}=(\gamma_{ij})$, it is immediate that

$$\Delta=(I-wA)^{-1}B=p\begin{bmatrix} & & & \gamma_{1k}\\ & \mathbf{0}_{k\times k} & & \gamma_{2k}\\ & & & \cdot\\ & & & \gamma_{kk}\\ 0 & 0 & \cdot\ 0 & \gamma_{k+1,k} \end{bmatrix}$$

and making use of $\Delta^x=(p\gamma_{k+1,k})^{x-1}\Delta$, $x\geqslant 1$ we may write (4) as

$$\varphi_x(w;p)=w^x(p\gamma_{k+1,k})^{x-1}(e_1\Delta)(I-wA)^{-1}\mathbf{1}', \quad x\geqslant 1. \tag{8}$$

Routine calculations show that

$$\gamma_{k+1,k}=qw(pw)^{k-1}\,|I-wA|^{-1}, \qquad |I-wA|=(1-pw)\{1-qw(1-(pw)^k)(1-pw)^{-1}\},$$

$$e_1\Delta=p\gamma_{1k}\cdot e_{k+1}=p(pw)^{k-1}(1-pw)\,|I-wA|^{-1}\cdot e_{k+1}, \qquad e_{k+1}(I-wA)^{-1}\mathbf{1}'=|I-wA|^{-1}$$

(where $e_{k+1}=(0,0,\ldots,0,1)$), and substituting in (8) we readily deduce (7). □

Corollary 2. *The generating function of the number $M_{n,k}$ of overlapping success runs of length k is given by*

$$\varphi_x(w;p)=(pw)^{k-x-1}\frac{\{1-qw(1-(pw)^{k-1})(1-pw)^{-1}\}^{x-1}}{\{1-qw(1-(pw)^k)(1-pw)^{-1}\}^{x+1}}, \quad x\geqslant 1. \tag{9}$$

Proof. Now the transition matrices are of the same form as for $G_{n,k}$, with the $(k+1,k+1)$ entry of A being 0 and the same entry of B being p. If $(I-wA)^{-1}=(\gamma_{ij})$, $\Delta=(I-wA)^{-1}B$ we get

$$\Delta^x=\{p(\gamma_{k+1,k}+\gamma_{k+1,k+1})\}^{x-1}\Lambda,\ \gamma_{k+1,k}=\frac{qw(pw)^{k-1}}{|I-wA|},\ \gamma_{k+1,k-1}=1,\ |I-wA|=1-qw\frac{1-(pw)^k}{1-pw}$$

and (4) simplifies to

$$\varphi_x(w;p)=w^x(p(\gamma_{k+1,k}+\gamma_{k-1,k+1}))^{x-1}e_1\Delta(I-wA)^{-1}\mathbf{1}', \quad x\geqslant 1,$$

which on using the obvious relations

$$e_1\Lambda=p(\gamma_{1,k}+\gamma_{1,k+1})e_{k+1}=p^kw^{k-1}\,|I-wA|^{-1}e_{k+1}, \qquad e_{k+1}(I-wA)^{-1}\mathbf{1}'=|I-wA|^{-1}$$

proves (9). □

Corollary 3. *The generating function $\varphi_x(w; p)$ of the number $N_{n,k}$ of non-overlapping success runs of length k is given by*

$$\varphi_x(w; p) = (pw)^{kx} \frac{(1-(pw)^k)(1-pw)^{-1}}{\{1-qw(1-(pw)^k)(1-pw)^{-1}\}^{x+1}}, \quad x \geq 0. \tag{10}$$

Proof. Recalling the transition matrices A and B for $N_{n,k}$ from Koutras and Alexandrou (1995) and letting again $(I-wA)^{-1} = (\gamma_{ij}), \Lambda = (I-wA)^{-1}B$ we may easily verify that

$$\Lambda^x = (p\gamma_{1k})^{x-1}\Lambda, \quad e_1\Lambda = p\gamma_{1k}e_1, \ \gamma_{1k} = (pw)^{k-1}|I-wA|^{-1},$$

$$\sum_{j=1}^{k}\gamma_{1j} = (1-(pw)^k)(1-pw)^{-1}|I-wA|^{-1}, \quad |I-wA| = 1-qw(1-(pw)^k)(1-pw)^{-1}.$$

Substituting now in (4) we deduce

$$\varphi_x(w; p) = w^x(p\gamma_{1k})^x e_1(I-wA)^{-1}\mathbf{1}' = w^x(p\gamma_{1k})^x \sum_{j=1}^{k}\gamma_{1j}, \quad x \geq 1,$$

which effortlessly leads to formula (10). It is not difficult to verify that the same formula holds true for $x = 0$ since $\varphi_0(w; p) = e_1(I-wA)^{-1}\mathbf{1}' = \sum_{j=1}^{k}\gamma_{1j}$. □

We mention that the result of Corollary 3 could also be derived from the double generating function (see Feller, 1968, p. 341)

$$\sum_{x=0}^{\infty}\sum_{n=kx}^{\infty} P(N_{n,k} = x)w^n z^x = \frac{1-(pw)^k}{1-w+qw(pw)^k-(1-pw)(pw)^k z}$$

by expanding the RHS with respect to z.

In the next section we are going to discuss tests of randomness based on the number of runs of specific length k, i.e. $N_{n,k}, G_{n,k}, M_{n,k}$. In order to apply such tests we need the conditional probability distribution of each of the three random variables given the number S_n of successes in the n trials. The evaluation of these distributions can be easily achieved by Theorem 1 in conjunction with the outcomes of Corollaries 1–3.

For the sake of illustration let us consider the conditional distribution of the number $M_{n,k}$ of overlapping success runs of length k, given $S_n = n - y$. If $x \geq 1$ then using formula (9) and Theorem 1 we get

$$\psi_x(t, w) = \frac{w^{k-x-1}(1+w^k t - twd(w))^{x-1}}{(1-twd(w))^{x+1}}, \quad d(w) = \frac{1-w^k}{1-w},$$

and expanding $\psi_x(t, w)$ in power series in t yields

$$\sum_{n=0}^{\infty} a(x, y; n)w^n = \sum_{j=0}^{\min(x-1,y)} \binom{x-1}{j}\binom{y+1}{y-j} d^{y-j}(w)w^{x+y+(k-1)(j+1)}. \tag{11}$$

If $x - 1 < y$, it is plain that the RHS simplifies to $\sum_{n=0}^{\infty}\sum_{j=0}^{x-1}\sum_{i=0}^{\infty} A_n(x, y; i, j)w^n$ where

$$A_n(x, y; i, j) = (-1)^i \binom{x-1}{j}\binom{y+1}{y-j}\binom{y-j}{i}\binom{n-k(i+j+1)-x}{y-j-1}. \tag{12}$$

Thus, formula (11) yields

$$P(M_{n,k} = x \mid S_n = n - y) = \binom{n}{y}^{-1} \alpha(x, y; n) = \binom{n}{y}^{-1} \sum_{j=0}^{x-1} \sum_{i=0}^{\infty} A_n(x, y; i, j). \tag{13}$$

If $x - 1 \geqslant y \geqslant 1$, the RHS of (11) reads

$$\sum_{j=0}^{y-1} \binom{x-1}{y} \binom{y+1}{y-j} d^{y-j}(w) w^{x-y+(k-1)(j+1)} + \binom{x-1}{y} w^{k+x-1-ky}$$

and a similar procedure as before shows that

$$P(M_{n,k} = x \mid S_n = n - y) = \binom{n}{y}^{-1} \left\{ \sum_{j=0}^{y} \sum_{i=0}^{\infty} A_n(x, y; i, j) + \binom{n - k(y+1)}{y} \delta_{n,x+(y+1)k-1} \right\} \tag{14}$$

where δ_{ij} denotes Kronecker's delta (i.e. $\delta_{ij} = 1$ if and only if $i = j$). Formulae (13), (14) can be written in a single expression for all $x \geqslant 1$, as

$$P(M_{n,k} = x \mid S_n = n - y) = \binom{n}{y}^{-1} \left\{ \sum_{j=0}^{\min(x-1,y)} \sum_{i=0}^{\infty} A_n(x, y; i, j) + \binom{n - k(y+1)}{y} \delta_{n,x+(y+1)k-1} \right\}. \tag{15}$$

It is worth mentioning that one could effortlessly derive the unconditional distribution of $M_{n,k}$ as a triple summation by making use of the last formula and the total probability theorem. An explicit expression for the probability function $P(M_{n,k} = x)$, $x \geqslant 0$ was given for the first time by Ling (1988). His formula involved a sum of multiple sums of probabilities with multinomial coefficients. Recently, Godbole (1992) obtained a more attractive formula by manipulating over Ling's expression and reducing it to a sum of two terms, each of them being a quadruple sum. Obviously, our representation is much more manageable and computationally tractable.

Making use of the outcome of Corollary 3 one can work in a similar fashion to derive Godbole's (1990) formula for the conditional distribution of $N_{n,k}$. Finally, Corollary 1 leads to the following expression for the conditional distribution of $G_{n,k}$,

$$P(G_{n,k} = x \mid S_n = n - y) = \binom{y+1}{x} \binom{n}{y}^{-1} \sum_{j=0}^{\lfloor (n-y-kx)/k \rfloor} (-1)^j \binom{y+1-x}{j} \binom{n - k(x+j)}{y}, \tag{16}$$

for $n \geqslant kx + y$ and $x \geqslant 0$. To the best of our knowledge this last formula has not appeared before.

Note also that for $x = 0$ we have $P(M_{n,k} = 0 \mid S_n = n - y) = P(N_{n,k} = 0 \mid S_n = n - y) = P(G_{n,k} = 0 \mid S_n = n - y)$ and (16) offers an easy to apply formula for the evaluation of all these quantities.

4. Non-parametric tests of randomness

In the 1940s, when interest in the theory of runs was quite high (cf. Mood, 1940; Wald and Wolfowitz, 1940; etc.), two different randomness tests were proposed: the classical run test which was based on the total number R_n of runs of either type and the longest-run test which utilises the length L_n of the longest success run. Larsen et al. (1973) proposed a rank-based procedure whereas O'Brien and Dyck (1985) derived a test for randomness which makes use of the variances of the run lengths. Recently, Agin and Godbole (1992) using the classical runs test as a model, developed a new exact test based on a conditional version of

$N_{n,k}$. Since this last test was found to be significantly more powerful in detecting certain types of clustering (non-randomness) we decided to explore the performance of tests based on the (similar in nature) quantities $G_{n,k}, M_{n,k}$. The results were quite encouraging, at least for $M_{n,k}$.

Our tests were upper tailed and the critical values for rejection were determined via formulae (15), (16) and Godbole's (1990) formula (for $N_{n,k}$) which provide the exact null distributions.

In order to evaluate the operational characteristics (power) of the test, we used Monte Carlo techniques, i.e. we specified particular alternative hypotheses, artificially generated (through a random number generator) 1000 sequences obeying the specific rule, and then evaluated the ability of each test to correctly reject the null hypothesis.

The parametric configurations upon which the comparisons were performed are the following (p_i, $i = 1,2,\ldots,n$, denotes the success probability of the ith trial):

1. First-order Markov dependence: $p_1 = 0.5$ and

$$p_i = \begin{cases} p & \text{if the } (i-1)\text{th trial is a success} \\ 0.5 & \text{if the } (i-1)\text{th trial is a failure} \end{cases} \quad \text{where } p = 0.90, 0.99.$$

2. Non-systematic unimodal and bimodal clustering:

$$p_i = \begin{cases} 0.9 & \text{if } |i - [n/2]| \leqslant [n/20] \\ 0.3 & \text{otherwise} \end{cases}$$

and

$$p_i = \begin{cases} 0.9 & \text{if } |i - [n/4]| \leqslant [n/20] \text{ or } |i - [3n/4]| \leqslant [n/20] \\ 0.3 & \text{otherwise.} \end{cases}$$

3. Cyclical clustering (with cycle length equal to 10):

$$p_i = \begin{cases} p & \text{if } 10r + 1 \leqslant i \leqslant 10r + c,\ r = 0,1,2,\ldots \\ 0.5 & \text{otherwise,} \end{cases}$$

where $c \leqslant 10$ is a fixed integer and $p = 0.95, 1.0$.

It is worth mentioning that these sequences are indicative of real situations and have appeared before in certain practical applications. For example, the interest in the non-systematic clustering configuration originates from studying the occurrence of patterns of demyelination of internodes in nerve fibres. Specifically if demyelination could be said to have occurred at random, this would be suggestive of Schwann cell disease. For more details, we refer to Dyck et al. (1984). On the other hand, the Markov dependence alternative is of special significance in psychological achievement testing, animal learning studies, athletical competition, etc., where the researcher is usually called to decide whether a specific outcome justifies the "*success breeds success*" assumption (i.e. attaining a positive outcome makes it more probable that a positive outcome will also be attained on the next trial). For other fields of application of such configurations see Schwager (1983).

For the purpose of making comparisons of the tests based on $N_{n,k}, G_{n,k}$ and $M_{n,k}$ to the classical tests based on R_n and L_n, we evaluated the empirical power of them for $n = 20$, 50, 100, 200 and level of significance $\alpha = 0.05$, 0.10. Since the null distribution of the first three statistics depends on k, it is obvious that the critical values and the power of the tests will be affected by the choice of k. For the selection of k one could follow the empirical rule suggested by Agin and Godbole (1992), i.e. $k - 1$ is taken equal to the mean value of the longest success run in a random sequence of n Bernoulli trials ($p = 0.5$). An alternative empirical rule for determining k is the following: choose k so that $P(L_n = k)$ is maximised. Our numerical experimentation showed that both choices are in general very close to the value of k yielding the maximum power (which is shown in our tables).

Table 1
Empirical power at level $\alpha = 0.10$ and $\alpha = 0.05$

		$\alpha = 0.10$					$\alpha = 0.05$				
n	Alternative	$G_{n,k}$	$M_{n,k}$	$N_{n,k}$	R_n	L_n	$G_{n,k}$	$M_{n,k}$	$N_{n,k}$	R_n	L_n
20	NSU	0.102	0.153	0.102	0.175	0.176	0.062	0.12	0.061	0.082	0.121
	NSB	0.124	0.181	0.169	0.226	0.096	0.046	0.104	0.052	0.097	0.042
	MD1	0.042	0.424	0.123	0.438	0.443	0.033	0.325	0.070	0.315	0.326
	MD2	0.007	0.483	0.143	0.271	0.907	0.002	0.252	0.044	0.256	0.672
50	NSU	0.276	0.421	0.274	0.237	0.425	0.151	0.264	0.151	0.151	0.259
	NSB	0.451	0.599	0.435	0.432	0.407	0.335	0.456	0.341	0.317	0.233
	MD1	0.072	0.733	0.341	0.786	0.422	0.042	0.648	0.238	0.709	0.295
	MD2	0.0	0.520	0.201	0.538	0.791	0.0	0.508	0.166	0.537	0.754
100	NSU	0.650	0.794	0.657	0.517	0.765	0.567	0.716	0.567	0.367	0.700
	NSB	0.754	0.952	0.759	0.795	0.868	0.611	0.912	0.639	0.671	0.800
	MD1	0.140	0.922	0.649	0.953	0.449	0.091	0.862	0.508	0.912	0.325
	MD2	0.0	0.664	0.283	0.674	0.676	0.0	0.577	0.218	0.599	0.612
200	NSU	0.865	0.957	0.874	0.724	0.933	0.816	0.926	0.819	0.574	0.890
	NSB	0.961	0.997	0.972	0.945	0.984	0.894	0.994	0.943	0.920	0.963
	MD1	0.234	0.990	0.877	0.996	0.559	0.163	0.975	0.799	0.992	0.431
	MD2	0.0	0.806	0.421	0.812	0.568	0.0	0.756	0.287	0.774	0.473

Table 2
Empirical power at level $\alpha = 0.10$ for several choices of k

		NSU			NSB			MD1			MD2		
n	k	$G_{n,k}$	$M_{n,k}$	$N_{n,k}$	$G_{n,k}$	$M_{n,k}$	$N_{n,k}$	$G_{n,k}$	$M_{n,k}$	$N_{n,k}$	$G_{n,k}$	$M_{n,k}$	$N_{n,k}$
100	5	0.164	0.791	0.479	0.444	0.931	0.759	0.029	0.896	0.576	0.0	0.658	0.228
	6	0.356	0.767	0.477	0.649	0.906	0.740	0.068	0.862	0.620	0.0	0.662	0.249
	7	0.650	0.714	0.657	0.728	0.879	0.739	0.103	0.823	0.565	0.0	0.664	0.283
	8	0.600	0.608	0.600	0.754	0.813	0.755	0.140	0.794	0.544	0.0	0.663	0.235
200	5	0.290	0.957	0.786	0.499	0.997	0.970	0.006	0.990	0.877	0.0	0.799	0.421
	6	0.444	0.945	0.820	0.755	0.996	0.972	0.035	0.981	0.863	0.0	0.802	0.406
	7	0.700	0.932	0.855	0.896	0.991	0.950	0.083	0.968	0.865	0.0	0.803	0.355
	8	0.865	0.885	0.874	0.927	0.985	0.957	0.154	0.954	0.842	0.0	0.804	0.319

NSU: non-systematic unimodal; NSB: non-systematic bimodal; MD1: Markov dependence $p = 0.90$; MD2: Markov dependence $p = 0.99$.

It is clear from Table 1 that the test based on the number $M_{n,k}$ of overlapping success runs of length k usually ranks first in empirical power for both unimodal and bimodal non-systematic clustering alternatives. If the sequence of outcomes exhibits a first-order Markov dependence, $M_{n,k}$ is still better from the other two fixed-length run statistics $G_{n,k}, N_{n,k}$, but it becomes slightly less sensitive than L_n or R_n as p approaches 1. An additional practical advantage of both $M_{n,k}$ and $N_{n,k}$ over $G_{n,k}$ is that an improper choice of k by the researcher does not result in significant loss of power. This is clearly illustrated in Table 2, where the empirical power of the three fixed-length run tests is compared for several values of k.

Table 3
Empirical power at level $\alpha = 0.05$ for cyclical clustering

n	p	c	$G_{n,k}$	$M_{n,k}$	$N_{n,k}$	R_n	L_n
100	1.0	5	0.883	0.202	0.770	0.619	0.009
	1.0	6	0.931	0.084	0.882	0.677	0.009
	1.0	7	0.912	0.028	0.939	0.672	0.012
	1.0	8	0.834	0.004	0.884	0.468	0.011
	0.95	5	0.622	0.140	0.499	0.433	0.023
	0.95	6	0.696	0.063	0.547	0.453	0.014
	0.95	7	0.580	0.021	0.499	0.407	0.012
	0.95	8	0.377	0.013	0.300	0.223	0.017
200	1.0	5	1.0	0.444	0.971	0.856	0.017
	1.0	6	1.0	0.133	0.993	0.882	0.014
	1.0	7	0.995	0.025	0.997	0.869	0.017
	1.0	8	0.976	0.001	0.993	0.688	0.010
	0.95	5	0.927	0.265	0.795	0.690	0.020
	0.95	6	0.927	0.090	0.849	0.703	0.018
	0.95	7	0.851	0.017	0.780	0.639	0.011
	0.95	8	0.651	0.007	0.547	0.350	0.017

In order to assess the performance of the tests in the case of cyclical clustering, we considered $c = 5, 6, 7, 8$, $p = 0.95, 1$ and $n = 100, 200$. As Table 3 indicates $M_{n,k}$ and L_n are very unreliable in detecting the existence of such alternatives. On the contrary, the tests based on $G_{n,k}, N_{n,k}$ and R_n were proved to be very sensitive (especially for $n = 200$), with $G_{n,k}$ displaying the highest power in most of the cases and R_n the lowest.

References

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

Dyck, P., H. Nukada and J. Karnes (1984), Permanent axotomy: a model of chronic neuronal degeneration preceded by axonal atrophy, myelin remodelling and degeneration, in: P. Dyck et al., eds., *Peripheral Neuropathy*, Vol. 1 (W.B. Saunders, Philadelphia, 2nd ed.) pp. 1103–1138.

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

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

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

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

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

Godbole, A. (1992), The exact and asymptotic distribution of overlapping success runs, *Commun. Statist. Theory Methods* **21A**, 953–996.

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

Hirano, K., S. Aki, N. Kashiwagi and H. Kuboki (1991), On Ling's binomial and negative binomial of order k, *Statist. Probab. Lett.* **11**, 503–509.

Koutras, M.V. and V.A. Alexandrou (1995), Runs, scans and urn model distributions: a unified Markov chain approach, *Ann. Inst. Statist. Math.* **47**, 743–766.

Larsen, R.J., C.L. Holmes and C.W. Heath (1973), A statistical test for measuring unimodal clustering: a description of the test and of its application to cases of acute leukaemia in metropolitan Atlanta, Georgia, *Biometrics* **29**, 301–309.

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

Lou, W.W. (1996a), On runs and longest runs test: method of finite Markov chain imbedding, *J. Amer. Statist. Assoc.*, to appear.

Lou, W.W. (1996b), An application of the method of finite Markov chain imbedding to runs tests, *Statist. Probab. Lett.*, to appear.
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.
O'Brien, P. and P. Dyck (1985), A runs test based on run lengths, *Biometrics* **41**, 237–244.
Schwager, S. (1983), Run probabilities in sequences of Markov dependent trials, *J. Amer. Statist. Assoc.* **78**, 168–175.
Wald, A. and J. Wolfowitz (1940), On a test whether two samples are from the same population, *Ann. Math. Statist.* **11**, 147–162.