A MARKOV CHAIN APPROACH TO THE PROBLEM OF RUNS AND PATTERNS

Zhiying Liang

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science at
Concordia University
Montreal, Quebec, Canada

March 1993

ISBN 0-612-01261-1

ABSTRACT

A Markov Chain Approach To
The Problem Of Runs And Patterns

Zhiying Liang

The analysis of exact distribution for randomly arranged repetitive runs and patterns in repeated Bernoulli, decimal, alphabetical or other types of sequences have been studied since 1940's. They can be classified into four types.

In this thesis, we propose a different approach to the problem of runs and patterns than the approaches taken by all the other authors on this topic by converting it into the problem of Markov chain with discrete state space. We derive methods to compute the exact probabilities and to derive the asymptotic distributions of the number of occurrence of runs and patterns of any kinds of length.

ACKNOWLEDGEMENTS

First I would like to thank Dr. Y. H. Wang for sharing his statistical knowledge with me as a teacher and as my thesis supervisor. I thank him for suggesting the topic and providing the assistance essential to the completion of this research. I am also grateful to him for his precious help getting a teaching assistantship and his patience.

TABLE OF CONTENTS

INTRODUCTION

The purpose of this thesis is to propose a approach to the problem of runs and patterns different from the approaches considerations of all the other authors on this topic by a conversion into the problem of Markov chain with discrete state space is used. We shall find ways to compute the exact distributions and to derive the asymptotic distributions of the number of occurrence of runs and patterns of length k for all $k \geq 1$ are found.

To be more precise, we explain in details the case with a binary sequence. Let $n > k \geq 1$ be two fixed positive integers. For a given sequence of n randomly arranged S (success) and F (failure), we are interested how to count the number of occurrences of the pattern SS$\cdots$S (a run of k S) in a total of n trials.

There are many ways of counting the number of *runs of S of length k*. They can be classified into four types as follows:

Type I: A run of S of length k means a string of exactly k S followed by an F.

Type II: A run of S of length k means a string of uninterrupted k or more S.

Type III: A run of S of length k means a string of exactly k S with recounting starting immediately after a run occurs.

Type IV: A run of S of length k means a string of exactly k S allowing overlapping runs.

For example, consider a realization of a sequence S and F such as

S-S-S-S-F-S-S-S-F-F-F-S-S-S-S-S-S-F-F-F-F-S-S.

In this sequence $n = 23$. If we take $k = 3$, then there is only one run of three S of type I; there are three runs of three S of type II; four runs of three S of type III and seven runs of three S of type IV.

Runs of type I are the most restricted one. In a narrow sense, the word "pattern" means this type of runs. In linguistics, a literary text can be viewed as sample sequences drawn from a population of possible texts from an author. (See Yule (1944).) Counting the number of occurrences of a particular pa,tern (a cluster of letters or words) in a randomly selected text of an author is equivalent to counting the number of "runs" of type I of such a pattern. (See Brainerd and Chang (1982).)

Runs of type II are a natural way of counting runs and are the ones most commonly accepted in the classical literature before Feller (1968) came up with the definition of type III runs. In the literature, runs of type II have often been referred to as " the classical way of counting" runs. In statistics, one often wants to know whether a set of observed data available for some statistical analysis is random. To test the randomness in this situation one method is to use the total number of runs above and below the median in the set of data. The "runs" in the runs test are runs of type II for $k = 1$. (See Mood (1940).)

In reliability, a "consecutive-k-out-of-n:F system" consists of n linearly ordered components. The failure times of the components are assumed to be independent and identically distributed. The system fails if and only if at least k out of its n components fail. (See Chiang and Niu (1981).) The study of the reliability of such a system is reduced to the study of the number of type II runs with "failure" substituting for "success". The reliability of the consecutive-k-out-of-n:F system is the probability that a run of "failures" of length k of type II has never occurred.

An extension of the consecutive-k-out-of n:F system is a system with $m-1$, $m \geq 2$, identical back-up systems. Such a system is known as "m-consecutive-k-out-of-n:F system". (See Griffith (1986) and Papastavridis (1991).) For such a system to fail, it is necessary to have m or more repeated runs of failure of length k of type II. Therefore, the reliability of the m-consecutive-k-out-of-n:F system is the probability that there are at most $m-1$ runs of failures of length k of type II.

Feller (1968) proposed runs of type III from the point of view of renewal process. Thus runs of type III have been called "Feller's way of counting" runs in the literature. As Feller noted (1968, p.279) that "If we are to use the theory of recurrent events, then the notion of runs of length k must be defined so that we start from scratch every time a run is completed. This means adopting the following definition. A sequence of n letters S and F contains as many runs of length k as there are non-overlapping uninterrupted successions of exactly k letters S. In a sequence of Bernoulli trials a run of length k occurs at the n-th trial if the n-th trial adds a new run to the sequence."

We believe that Feller is the first person to consider the problem of "runs of length k". Before him, people were only concerned with the problem of "runs", i.e. "runs of length 1 or more", as the "runs" defined in the runs test based on the total number of runs. His definition appeared in the first edition of his book (1968) which was published in 1950.

Three interesting examples of type III runs are:

Example A. (See Aki (1985).) An urn contains w white and r red balls. Let k be a fixed integer such that $k \leq r$. A ball is drawn at random. If it is a white ball, it is replaced into the urn, if red it is laid beside the urn. Another random drawing is made from the urn. If the ball is red it is laid beside the urn and the drawing continues. But when a white ball is drawn, the white ball and all the red balls which have been accumulated beside the urn are replaced into the urn at the same time. The procedure is repeated in identical manner as long as the red balls accumulated outside the urn is less than k. If the number of red balls outside the urn reaches k, all the k red balls outside the urn are replaced into the urn and the process starts anew. A binary sequence is obtained by recording S or F for each random drawing according to whether it is a red or a white ball. In this example, the occurrence of consecutive k successes means that the number of red balls outside the urn has reached the value k.

Example B. (See Aki (1985).) An electric bulb is lighted. It is checked whether it has failed or not at the end of each day. If it is found to be burnt out, then a new one is replaced immediately. If a bulb has been lighted for k consecutive days, it is replaced with a new one even if it has not failed. Define a binary sequence by recording S and F every day, according to whether the electric bulb is in working condition or has failed. In this example, the occurrence of consecutive k successes means that an electric bulb which has not failed will be replaced by a new one.

Example C. (*Counters of Type I*). A sequence of Bernoulli trials is performed. A counter is designed to register successes, but the mechanism is locked for exactly $k-1$ trials following each registration. In other words, a success at the n-th trial is registered if, and only if, no registration has occurred in the preceding $k-1$ trials. The counter is then locked at the conclusion of trials number $n, \cdots, n+k-1$, and is freed at the conclusion of the $(n+k)$-th trial provided that this trial is a failure. However, whenever the counter is free (not locked) the situation is exactly the same, and the trials start from scratch. In this example, the occurrence of consecutive k successes means that the counter is locked for a period of k trials, including the initial trial which locked the counter. (See Feller (1968).)

Type IV runs was recently defined by Ling (1988, 1989) in conjunction with binomial and negative binomial distributions of order k. The following example is a natural one for type IV runs.

Example D. (*Counters of Type II*). Same as Example C except that *each success* locks the counter for k time units ($k-1$ trials following the success) so that a success during a locked period prolongs that period. For example, take $k \geq 2$, if a success at the n-th trial is registered Bwhich locks the counter to the $(n+k-1)$-th trial, and another sucess at the $(n+1)$-th trial is again registered, then the locking period of the counter is prolonged to the $(n+k)$-th trial. In this example, the occurrence of consecutive k successes means exactly the same as in the previous example, but allowing overlapping in counting of runs of length k. (Feller (1968).)

In our view, "pattern" is a special case of "runs of length k" for some $k \geq 2$. Therefore in the sequel we treat the whole problem of runs and patterns simply as the problem of runs. Further illustrative examples of our view will be given in Chapter One.

Throughout this paper, we adopt the usual convention of denoting "1" for "success" and "0" for "failure", and vice versa. Let $X_1, X_2, \cdots, X_n$ be a sequence of Markov Bernoulli random variables with the following stationary transition probabilities

$$\begin{matrix} & 0 & 1 \\ 0 & 1-\alpha & \alpha \\ 1 & \beta & 1-\beta \end{matrix}, \qquad 0 < \alpha, \beta < 1, \qquad (1)$$

and the initial probability

$$P(X_1 = 1) = p = 1 - P(X_1 = 0), \qquad 0 \leq p \leq 1.$$

It is well known that the above model contains the following two special cases:

1). If $p = \alpha/(\alpha + \beta)$, then $\boldsymbol{X}$ is completely stationary in the sense that $P(X_i = 1) = p$ for all $i = 1, 2, \cdots$. It is known that the transition matrix for this model can be expressed as

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} \quad\quad 0 & \quad\quad\quad\quad 1 \end{array} \\ \begin{bmatrix} 1-(1-\pi)p & (1-\pi)p \\ (1-\pi)(1-p) & (1-\pi)p+\pi \end{bmatrix} \end{array},$$

where π is the correlation coefficient of X_i and X_{i+1}. (See Edwards (1960).)

2) If $\alpha = p$ and $\beta = 1 - p$, then $\boldsymbol{X}$ is the ordinary sequence of i.i.d. Bernoulli variables.

Now, we define $\mathbb{N}_I$ to be the number of occurrences of consecutive k successes of type I. The other three counting variables N_{II}, N_{III} and N_{IV} are defined accordingly. (For brevity we suppress the dependence of all four counting variables N on k and n.) Evidently we have the stochastic ordering of $N_I \leq N_{II} \leq N_{III} \leq N_{IV}$, i.e. for fixed ω in the sample space, $N_I(\omega) \leq N_{II}(\omega) \leq N_{III}(\omega) \leq N_{IV}(\omega)$.

If $k = 1$, then N_{II} is the number of transitions from F to S and $N_{III} = N_{IV} = S_n = X_1 + \cdots + X_n$ is the occupation time of S. The exact distributions of the four counting variables and related topics have been a subject of study in recent history, especially in the last decade. Some relevant references on this topic, in addition to the references cited so far, are von Mises (1921), Rajarshi (1974), Aki (1985), Hirano (1986), Philippou and Makri (1986) and Hirano et al (1990). The literature on this topic is very numerous. Some of the recent articles are listed in Godbole (1991).

CHAPTER ONE

MARKOVIAN STRUCTURE

Let us consider the consecutive k-out-of-n:F system. It has been shown that such a system forms a Markov chain with $k+1$ states. (See Chao and Fu (1989).) The state 0 is interpreted as that the system is functioning flawlessly with all the n components in perfectly working condition. The states 1 to $k-1$ denote levels of deterioration of the system, such that state i, for $1 \leq i \leq k-1$, indicates that the system is functioning with i of its n components out of order. The state k indicates that the entire system breaks down.

We shall extend this fact a step further for all the four types of runs. In what follows, we shall let $\{Y_i : i = 1, 2, \cdots\}$ be the induced Markov chain of the Markov Bernoulli sequence $\{X_i : i = 1, 2, \cdots\}$. Strictly speaking, there are four induced Markov chain $\{Y_i : i = 1, 2, \cdots\}$ for the four types of runs, but we shall not distinguish them at the moment. Which of the induced Markov chain belongs to which type of the runs shall be clear in the text.

§1.1 For type I runs

We denote the state space $\boldsymbol{S}$ by {0', 0, 1, $\cdots, k, k'$ } where state j for $0 \leq j \leq k$ indicates that there are j consecutive S, preceded by an F. State k' denotes that a run of type I has been completed. That is to say a sequence of consecutive k S followed by an F. The state 0' is for the sequence of $k+1$ or more consecutive S. Therefore, initially

$$P(Y_1 = 1) = P(X_1 = 1) = p = 1 - P(Y_1 = 0)\,. \tag{2}$$

And the transition probabilities are, for $i \geq 2$,

$$P(Y_i = j | Y_{i-1} = j-1) \quad = \begin{cases} P(X_i = 1 | X_{i-1} = 1), & \text{for } 2 \leq j \leq k. \\ \\ P(X_i = 1 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(Y_i = 0 | Y_{i-1} = j-1) \quad = \begin{cases} P(X_i = 0 | X_{i-1} = 1), & \text{for } 2 \leq j \leq k. \\ \\ P(X_i = 0 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(\ Y_i = k' \mid Y_{i-1} = k\ ) \quad = \quad P(X_i = 0 | X_{i-1} = 1).$$

$$P(\ Y_i = 0' \mid Y_{i-1} = k\ ) \quad = \quad P(X_i = 1 | X_{i-1} = 1).$$

$$
\begin{aligned}
P(\ Y_i = 0' \mid Y_{i-1} = 0'\ ) &= P(X_i = 1 | X_{i-1} = 1). \\
P(\ Y_i = 0 \mid Y_{i-1} = 0'\ ) &= P(X_i = 0 | X_{i-1} = 1). \\
P(\ Y_i = 0 \mid Y_{i-1} = k'\ ) &= P(X_i = 0 | X_{i-1} = 0). \\
P(\ Y_i = 1 \mid Y_{i-1} = k'\ ) &= P(X_i = 1 | X_{i-1} = 0).
\end{aligned}
$$

For $k = 3$, (a run means exactly three S followed by an F), the transition matrix can be expressed as follows:

$$
\begin{array}{c} \\ 0' \\ 0 \\ 1 \\ 2 \\ 3 \\ 3' \end{array}
\begin{array}{c}
\begin{array}{cccccc} 0' & 0 & 1 & 2 & 3 & 3' \end{array} \\
\begin{bmatrix}
1-\beta & \beta & 0 & 0 & 0 & 0 \\
0 & 1-\alpha & \alpha & 0 & 0 & 0 \\
0 & \beta & 0 & 1-\beta & 0 & 0 \\
0 & \beta & 0 & 0 & 1-\beta & 0 \\
1-\beta & 0 & 0 & 0 & 0 & \beta \\
0 & 1-\alpha & \alpha & 0 & 0 & 0
\end{bmatrix}
\end{array}. \tag{3}
$$

§1.2 For type II runs

The state space $\boldsymbol{S}$ for the type II runs is $\{\ 0,\ 1,\ \cdots, k, k'\ \}$ where state j for $0 \le j \le k$ indicates an F followed by j consecutive S. State k' means there are at least $k+1$ consecutive S. The initial probability of $\{Y_i\}$ is the same as (2). The transition probabilities can be seen to be, for $i \ge 2$,

$$P(Y_i = j | Y_{i-1} = j-1) = \begin{cases} P(X_i = 1 | X_{i-1} = 1), & \text{for } 2 \le j \le k. \\ P(X_i = 1 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(Y_i = 0 | Y_{i-1} = j-1) = \begin{cases} P(X_i = 0 | X_{i-1} = 1), & \text{for } 2 \le j \le k. \\ P(X_i = 0 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(Y_i = k' \,|\, Y_{i-1} = k) = P(X_i = 1 | X_{i-1} = 1).$$

$$P(Y_i = 0 \,|\, Y_{i-1} = k) = P(X_i = 0 | X_{i-1} = 1).$$

$$P(Y_i = k' \,|\, Y_{i-1} = k') = P(X_i = 1 | X_{i-1} = 1).$$

$$P(Y_i = 0 \,|\, Y_{i-1} = k') = P(X_i = 0 | X_{i-1} = 1).$$

For $k = 3$, (a run means a string of three or more consecutive S), the transition matrix is as follows:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 3' \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 3' \end{array} \\ \begin{bmatrix} 1-\alpha & \alpha & 0 & 0 & 0 \\ \beta & 0 & 1-\beta & 0 & 0 \\ \beta & 0 & 0 & 1-\beta & 0 \\ \beta & 0 & 0 & 0 & 1-\beta \\ \beta & 0 & 0 & 0 & 1-\beta \end{bmatrix} \end{array}. \qquad (4)$$

§1.3 For type III runs

The state space $\boldsymbol{S}$ for the type III runs is $\{\ 0,\ 1,\ \cdots, k\ \}$

where state j, for $0 \leq j \leq k$, it means that an **F** is followed by m consecutive S with $m = i \text{ mod}(k+1)$. The initial probability of $\{Y_i\}$ is the same as (2). And the transition probabilities are, for $i \geq 2$,

$$P(Y_i = j | Y_{i-1} = j-1) = \begin{cases} P(X_i = 1 | X_{i-1} = 1), & \text{for } 2 \leq j \leq k. \\ P(X_i = 1 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(Y_i = 0 | Y_{i-1} = j-1) = \begin{cases} P(X_i = 0 | X_{i-1} = 1), & \text{for } 2 \leq j \leq k+1. \\ P(X_i = 0 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(\ Y_i = 1 \mid Y_{i-1} = k\ ) = P(X_i = 1 | X_{i-1} = 1).$$

For $k = 3$, (a run means a string of three consecutive S with recounting starts right after a run occurs), the transition matrix is as follows:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1-\alpha & \alpha & 0 & 0 \\ \beta & 0 & 1-\beta & 0 \\ \beta & 0 & 0 & 1-\beta \\ \beta & 1-\beta & 0 & 0 \end{bmatrix} \end{array}. \qquad (5)$$

§1.4 For type IV runs

The state space $\boldsymbol{S}$ for the type IV runs is $\{ 0, 1, \cdots, k \}$ where state i, for $1 \leq i < k$, indicates there are i consecutive S which is preceded by an F. State k denotes there are at least k uninterrupted S preceded by an F. The initial probability of $\{Y_i\}$ is the same as (2). And the transition probabilities are, for $i \geq 2$,

$$P(Y_i = j | Y_{i-1} = j-1) = \begin{cases} P(X_i = 1 | X_{i-1} = 1), & \text{for } 2 \leq j \leq k. \\ P(X_i = 1 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(Y_i = 0 | Y_{i-1} = j-1) = \begin{cases} P(X_i = 0 | X_{i-1} = 1), & \text{for } 2 \leq j \leq k. \\ P(X_i = 0 | X_{i-1} = 0), & \text{for } j = 1. \end{cases}$$

$$P(Y_i = k \mid Y_{i-1} = k) = P(X_i = 1 | X_{i-1} = 1).$$

$$P(Y_i = 0 \mid Y_{i-1} = k) = P(X_i = 0 | X_{i-1} = 1).$$

For $k = 3$, (a run means a string of three consecutive S allowing overlapping runs), the transition matrix is as follows:

$$
\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}
\begin{array}{c}
\begin{array}{cccc} \quad 0 & \quad 1 & \quad 2 & \quad 3 \end{array} \\
\begin{bmatrix}
1-\alpha & \alpha & 0 & 0 \\
\beta & 0 & 1-\beta & 0 \\
\beta & 0 & 0 & 1-\beta \\
\beta & 0 & 0 & 1-\beta
\end{bmatrix}
\end{array} . \tag{6}
$$

It is easy to verify that for $k = 1$, the transition matrices of types III and IV runs are identical.

In passing we would like to make two remarks. One is that in the completely stationary and the i.i.d. cases, all four induced Markov chains have much simpler structures, because it invclves less parameters. Most of the results in this area deal with these two special cases.

The other is that the binary sequence we considered here can be easily extended to multivalent sequences, such as the decimal sequence with numbers $0, 1, \cdots, 9$, or the alphabetical sequence with 26 letters; or to mixtures of several types of runs, such as a run of k_1 S of type I followed immediately by another run of k_2 F of type II, etc.

CHAPTER TWO

THE PROBABILITY DISTRIBUTIONS

OF THE COUNTING VARIABLES

We begin this chapter by considering a general homogeneous Markov chain $\boldsymbol{Y}=\{Y_i : i = 1, 2, \cdots\}$ with transition matrix $\boldsymbol{P} = (p_{ij})$ and countable state space $\boldsymbol{S} = \{0, 1, \cdots, s.\}$. Some of the procedures to be described next can be found in many books on Markov chain, e.g. Chung (1967) or Cinlar (1975).

Let $k \in \boldsymbol{S}$ be a fixed state and $N^{(k)}$ be the total number of visits to state k by the process $\boldsymbol{Y}$. Denote $\{V_n\}$ to be the sequence of the arrival times at the state k, so that if $V_n = k_1$ and $V_{n+1} = k_2$, then $k_1 < k_2$ and $Y_{k_1} = k = Y_{k_2}$, but $Y_j \neq k$ for all $k_1 < j < k_2$. If $\mathbb{N}^{(k)} = m < \infty$, then $V_{m+1}(\omega) - V_m(\omega) = V_{m+2}(\omega) - V_{m+1}(\omega) = \cdots = \infty$. Denote $T_i = V_i - V_{i-1}$ (take $V_0 = 1$) to be the sequence of the interarrival times into state k. The Markovian property of the process $\boldsymbol{Y}$ guarantees that the interarrival times T_i are mutually independent. Furthermore because the transition probabilities are stationary, T_i, for $i \geq 2$, are identically distributed.

Denote

$$f_{ak}(x) = P(T_1 = x | Y_1 = a)$$

to be the probability that the process $\boldsymbol{Y}$ starts at the state a and reaches the state k for the first time in x steps. Given the transition matrix $\boldsymbol{P}$, the computation of the probability $f_{ak}(x)$ can be done as follows:

For $x = 1$,

$$f_{ak}(1) = P(Y_2 = k | Y_1 = a) = p_{ak}. \tag{7}$$

For $x \geq 2$,

$$\begin{aligned} f_{ak}(x) & \\ &= \sum_{b \neq k} P(Y_2 = b | Y_1 = a) P(Y_3 \neq k, \cdots, Y_x \neq k, Y_{x+1} = k | Y_2 = b), \\ &= \sum_{b \neq k} p_{ab} f_{bk}(x-1). \end{aligned} \tag{8}$$

Thus given two states a and k and the transition matrix $\boldsymbol{P}$, $f_{ak}(1)$ is the a-th entry of the k-th column of $\boldsymbol{P}$. Denote the column vector $v_k(x)$ by

$$v_k(x) = (f_{0k}(x), f_{1k}(x), \cdots, f_{sk}(x))',$$

so that $f_{ak}(x)$ is the a-th entry of $v_k(x)$, and $v_k(1)$ is the k-th column of $\boldsymbol{P}$. It follows from (8) that

$$v_k(x) = \boldsymbol{Q}_k v_k(x-1), \qquad \text{for all } x \geq 2, \text{ and } k \in \boldsymbol{S}, \tag{9}$$

where $\boldsymbol{Q}_k$ is the matrix obtained from the transition matrix $\boldsymbol{P}$ by deleting its k-th column and replacing it with a zero vector.

Let $n > 1$ be a fixed finite integer, consider the first n steps $\{Y_1, Y_2, \cdots, Y_n\}$ of the process $\boldsymbol{Y}$. Define

$$\mathbb{N} = \text{the total number of visits to state } k \text{ by } \{Y_1, Y_2, \cdots, Y_n\}.$$

For a and k in $\boldsymbol{S}$, and let $f_{ak}(x)$ and $f_{kk}(x)$ be defined by (7) and (8) for all $x \geq 1$, except that for computational convenience later, we shall let $f_{ak}(0) = 0$ for all a and k, even for $a = k$, which is often taken to be 1.

Denote

$$g^1_{ak}(x) = f_{ak}(x), \quad \text{for } x = 1, 2, \cdots, n-1$$

and for $m \geq 2$

$$g_{ak}^{(m)} = g_{ak}^{1} \; * \; g_{kk}^{(m-1)},$$

where $g_{kk}^{(m-1)}$ is the $(m-1)$-th convolution of f_{kk}. Thus $g_{ak}^{(m)}$ is the convolution of g_{ak}^{1} and $g_{kk}^{(m-1)}$.

Define

$$h_{ak}(y) = \sum_{x=0}^{n-1} g_{ak}^{(y)}(x), \qquad \text{for } y = 1, 2, \cdots, K,$$

where K is the maximum possible run of length k in first n steps of the process $\boldsymbol{Y}$ which depends on the initial state a, the duration of the time n and the type of runs in question. For the type I runs $K = \left[\frac{n-a}{k+1}\right]$, for type II runs $K = \left[\frac{n}{k+1}\right]$, for type III runs $K = \left[\frac{n-1+a}{k}\right]$, and for type IV runs $K = n - k + a$ It is evident that

$$P(I\!\!N \geq y | X_1 = a) = h_{ak}(y), \;\; \text{for } y = 1, 2, \cdots, K. \qquad (10)$$

Thus (10) is the conditional probability that starting from state a at time 1, the process $\boldsymbol{Y}$ has visited state k at least y times in its first n steps. From (10) it follows that

$$P(N = y | X_1 = a) = \begin{cases} h_{ak}(y) - h_{ak}(y+1), & \text{for } y = 1, 2, \cdots, K-1, \\ h_{ak}(K), & \text{for } y = K, \\ 1 - h_{ak}(1), & \text{for } y = 0. \end{cases} \tag{11}$$

Theorem 1. The probabilities of N_i, the number of occurrences of success runs of length k of type i, for i =I, II, III, IV, are

$$P(N_i = j)$$
$$= \begin{cases} pP(N_i = j | X_1 = 1) + (1-p)P(N_i = j | X_1 = 0), & \text{for } 1 \le j \le K, \\ 0, & \text{for } j > K, \end{cases} \tag{12}$$

where $P(N_i = j | X_1 = a)$ is obtained by (11) with the transition probabilities for each type of runs as defined in Sections 1.3.a through 1.3.d.

Now for each a and k in $\boldsymbol{S}$, we define

$$\phi_{ak} = \sum_{x=1}^{\infty} f_{ak}(x). \tag{13}$$

So that ϕ_{ak} is the probability that starting at state a the process

$\boldsymbol{Y}$ will ever visit state k. Summing over x in (8), we get

$$\phi_{ak} = p_{ak} + \sum_{b \neq k} p_{ab}\phi_{bk}. \tag{14}$$

Equation (14) defines a relation between the ever-reaching probability ϕ_{ak} and the transition matrix $\boldsymbol{P}$.

Denote the column vector $v_k = (\phi_{0k}, \phi_{1k}, \cdots, \phi_{sk})'$,

$$v_k = v_k(1) + Q_k v_k, \qquad \text{for all } k \in \boldsymbol{S}, \tag{15}$$

where Q_k is defined as in (9) and $v_k(1)$ is the k-th column of $\boldsymbol{P}$. The linear equation (15) can be used to solve for v_k, for all $k \in \boldsymbol{S}$.

Note that $\mathbb{N}^{(k)}$ is the total number of visits to state k by the process $\boldsymbol{Y}$ so that $\mathbb{N}^{(k)} = m$ if and only if the interarrival times $T_i < \infty$ for $i \leq m$ and $T_i = \infty$ for $i > m$. As mentioned before the interarrival times $T_1, T_2, \cdots, T_{m+1}$ are independent and

$$P(T_1 < \infty | Y_1 = a) = \phi_{ak},$$

$$P(T_j < \infty) = \phi_{kk}, \quad \text{for } j = 2, \cdots, m,$$

and

$$P(T_{m+1} = \infty) = 1 - \phi_{kk}.$$

We conclude that, if $\phi_{kk} < 1$, then

$$P(\mathbb{N}^{(k)} = m | Y_1 = k) = (1 - \phi_{kk})\phi_{kk}^{(m-1)}, \quad \text{for } m = 1, 2, \cdots, \tag{16}$$

and for $a \neq k$,

$$P(\mathbb{N}^{(k)} = m | Y_1 = a) = \begin{cases} 1 - \phi_{ak}, & \text{for } m = 0, \\ \phi_{ak}(1 - \phi_{kk})\phi_{kk}^{(m-1)}, & \text{for } m = 1, 2, \cdots. \end{cases} \tag{17}$$

If $\phi_{kk} = 1$, then the process $\boldsymbol{Y}$ will return to state k after leaving it with probability one. By the Markovian property, it will return over and over again, so that $P(N^{(k)} = \infty) = 1$ and the probabilities in (16) and (17) are all zeros.

The numbers obtained by solving the linear equation (14) are of great significance. It determines whether or not a state k is visited infinitely often. If $\phi_{kk} = 1$, then the state k is *recurrent.* If $\phi_{kk} < 1$, then almost surely the state will be visited only finitely many times, and the state k is *transient.*

In all the four cases, if $\alpha = 0$ and $\beta \neq 0$ in the Markov Bernoulli model (1), state 0 of the induced Markov chain $\boldsymbol{Y}$ is an absorbing state and is recurrent. Since, in all the four types of runs, state 0 is accessible from any other states, all other states are transient. Therefore, looking at the induced process $\boldsymbol{Y}$ from a long term point of view, i.e. $n = \infty$, state k will be visited only finitely many times, and all the probabilities of the four counting variables are well defined.

To compute their probabilities, we need to compute ϕ_{ak} ($\phi_{ak'}$, in case of type I runs) for $a = 0, 1$, and k (k' for type I runs) and then apply (16) and (17). Their values for $\alpha = 0$ and $\beta \neq 0$ are summarized as follows:

For type I runs:

$$\phi_{00} = 1, \quad \phi_{1k'} = (1-\beta)^{k-1}\beta, \quad \text{and} \quad \phi_{k'k'} = 0.$$

For type II runs:

$$\phi_{00} = 1, \quad \phi_{1k} = (1-\beta)^{k-1}, \quad \text{and} \quad \phi_{kk} = 0.$$

For type III runs:

$$\phi_{00} = 1, \quad \phi_{1k} = (1-\beta)^{k-1}, \qquad \text{and} \qquad \phi_{kk} = (1-\beta)^k.$$

For type IV runs:

$$\phi_{00} = 1, \quad \phi_{1k} = (1-\beta)^{k-1}, \qquad \text{and} \qquad \phi_{kk} = (1-\beta).$$

The computations of ϕ_{00} are relatively straight forward. We shall briefly show the derivations of ϕ_{kk} and ϕ_{1k} for type II and type III runs, the values of ϕ_{kk} and ϕ_{1k} for other two types can be derived in an identical manner.

For type II runs; we see that $\phi_{00} = 1$ and $\phi_{0i} = 0$ for all $i \neq 0$. From (14) it follows that

$$\phi_{k'k} = 0 + \beta\phi_{0k} + (1-\beta)\phi_{k'k} = (1-\beta)\phi_{k'k},$$

which implies that $\phi_{k'k} = 0$. Thus

$$\phi_{kk} = 0 + \beta\phi_{0k} + (1-\beta)\phi_{k'k} = (1-\beta)\phi_{k'k} = 0.$$

And we also have

$$\phi_{12} \;\; = \;\; (1-\beta) + \beta\phi_{02} = (1-\beta),$$

$$
\begin{aligned}
\phi_{23} &= (1-\beta) + \beta\phi_{03} = (1-\beta), \\
\phi_{13} &= 0 + \beta\phi_{03} + (1-\beta)\phi_{23} = (1-\beta)\phi_{23} = (1-\beta)^2, \\
\cdots & \cdots \cdots\cdots\cdots\cdots \\
\phi_{k-1,k} &= (1-\beta) + \beta\phi_{0k} = (1-\beta), \\
\phi_{k-2,k} &= 0 + \beta\phi_{0k} + (1-\beta)\phi_{k-1,k} = (1-\beta)\phi_{k-1,k} = (1-\beta)^2, \\
\cdots & \cdots \cdots\cdots\cdots\cdots \\
\phi_{2k} &= 0 + \beta\phi_{0k} + (1-\beta)\phi_{3k} = (1-\beta)\phi_{3k} = (1-\beta)^{k-2}, \\
\phi_{1k} &= 0 + \beta\phi_{0k} + (1-\beta)\phi_{2k} = (1-\beta)\phi_{2k} = (1-\beta)^{k-1},
\end{aligned}
$$

For type III runs; by repeatedly applying (14) and using $\phi_{00} = 1$ and $\phi_{0i} = 0$ for all $i \neq 0$, we have

$$
\begin{aligned}
\phi_{12} &= (1-\beta) + \beta\phi_{02} = (1-\beta), \\
\phi_{23} &= (1-\beta) + \beta\phi_{03} = (1-\beta), \\
\phi_{13} &= 0 + \beta\phi_{03} + (1-\beta)\phi_{23} = (1-\beta)\phi_{23} = (1-\beta)^2, \\
\cdots & \cdots \cdots\cdots\cdots\cdots
\end{aligned}
$$

$$
\begin{aligned}
\phi_{k-1,k} &= (1-\beta)+\beta\phi_{0k} = (1-\beta), \\
\phi_{k-2,k} &= 0+\beta\phi_{0k}+(1-\beta)\phi_{k-1,k} = (1-\beta)\phi_{k-1,k} = (1-\beta)^2, \\
\cdots & \cdots \cdots\cdots\cdots \\
\phi_{2k} &= 0+\beta\phi_{0k}+(1-\beta)\phi_{3k} = (1-\beta)\phi_{3k} = (1-\beta)^{k-2}, \\
\phi_{1k} &= 0+\beta\phi_{0k}+(1-\beta)\phi_{2k} = (1-\beta)\phi_{2k} = (1-\beta)^{k-1}, \\
\phi_{kk} &= 0+\beta\phi_{0k}+(1-\beta)\phi_{1k} = (1-\beta)\phi_{1k} = (1-\beta)^{k}.
\end{aligned}
$$

Conditioning on X_1, we have the following theorem.

Theorem 2. If $\alpha = 0$, $\beta \neq 0$ and $n = \infty$, the distributions of the four counting variables are:

$$
P(N_I = m) = \begin{cases} p(1-\beta)^{k-1}\beta, & \text{for } m = 1, \\ 1 - p(1-\beta)^{k-1}\beta, & \text{for } m = 0. \end{cases}
$$

$$
P(N_{II} = m) = \begin{cases} p(1-\beta)^{k-1}, & \text{for } m = 1, \\ 1 - p(1-\beta)^{k-1}, & \text{for } m = 0. \end{cases}
$$

$$
P(N_{III} = m) = \begin{cases} 1 - p(1-\beta)^{k-1}, & \text{for } m = 0, \\ p(1-\beta)^{k-1}[1-(1-\beta)^{k}](1-\beta)^{(m-1)k}, & \text{for } m = 1, 2, \cdots. \end{cases}
$$

and

$$P(N_{IV}=m)=\begin{cases}1-p(1-\beta)^{k-1}, & \text{for } m=0,\\ p(1-\beta)^{k-1}[1-(1-\beta)](1-\beta)^{m-1}, & \text{for } m=1,2,\cdots.\end{cases}$$

If $\alpha > 0$, the induced Markov chain $\boldsymbol{Y}$ is ergodic and all states are recurrent. Therefore ϕ_{0k}, ϕ_{1k}, and ϕ_{kk} ($\phi_{0k'}$, $\phi_{1k'}$ and $\phi_{k'k}$, in case of type I runs) equal 1 in all the four cases and $P(N_i=\infty)=1$ for $i =$ I, II, III and IV.

For a sequence of Bernoulli random variables $X_1, X_2, \cdots, X_n$, if the number of cumulative counts of "1" is to converge in any of the four types mentioned in the Introduction Chapter or any other ways of counting, a dominating condition is that the number of transitions from state 0 to state 1 must be very moderate, relative to n (as $n \to \infty$). That is to say that $n\alpha$, the mean number of transition from state 0 to state 1, must approach a finite constant as $n \to \infty$. It was proved in Wang and Ji (1993) that in fact it is the only condition needed to assure the existence of the limiting distribution of the number of cumulative counts of "1". The other two parameters p and β can have their own limit behaviors, such as $p \to \pi$ and $\beta \to \rho$, as $n \to \infty$ for $\pi, \rho \in (0,1)$, but will not affect the existence of the limiting distribution of the number of cumulative counts.

CHAPTER THREE

THE ASYMPTOTIC DISTRIBUTION OF THE COUNTING VARIABLES

In this chapter, we shall derive the limiting distributions of the four counting variables under the structure that the original sequence $\boldsymbol{X}$ is a sequence of i.i.d. Bernoulli random variables. For $\boldsymbol{X}$ having a Markov Bernoulli model (1), the results are recently obtained in Wang and Ji (1993) using a different approach.

We shall create another induced Markov chain $\boldsymbol{Z}= \{Z_i : i = 1, 2, \cdots, \}$ from $\boldsymbol{Y}$ by

$$Z_i = I(Y_i = k), \tag{18}$$

so that Z_i is the indicator function of the event that a run of length k occurs time at i. (For type I runs $Z_i = I(Y_i = k')$). With Z_i as defined in (18) it follows that the limiting distributions of N_i are the limiting distributions of $Z_1 + Z_2 + \cdots + Z_n$.

The new process $\boldsymbol{Z}$ is a Markov Bernoulli chain with initial probabilities $P(Z_1 = 1) = p = 1 - P(Z_1 = 0)$ and transition probabilities

$$\begin{aligned} P(Z_i = 1|Z_{i-1} = 1) &= P(Y_i = k|Y_{i-1} = k), \\ &= 1 - P(Z_i = 0|Z_{i-1} = 1). \\ P(Z_i = 1|Z_{i-1} = 0) &= P(Y_i = k|Y_{i-1} \neq k), \\ &= 1 - P(Z_i = 0|Z_{i-1} = 0). \end{aligned} \tag{19}$$

Obviously we have $P(Z_i = 1|Z_{i-1} = 1) = 0$ for types I, II, and III runs and $P(Z_i = 1|Z_{i-1} = 1) = p$ for type IV runs. To compute $P(Z_i = 1|Z_{i-1} = 0)$, we use

$$P(Z_i = 1|Z_{i-1} = 0) = \frac{P(Y_i = k) - P(Y_i = Y_{i-1} = k)}{1 - P(Y_{i-1} = k)}.$$

For $i \geq k + 1$, we have $P(Y_i = k) = p^k(1 - p)$ for type I runs and $P(Y_i = k) = p^k$ for the other three types. $P(Y_i = Y_{i-1} = k) = 0$ for types I and III runs and $P(Y_i = Y_{i-1} = k) = p^{k+1}$ for the other two types. The transition matrices of $\boldsymbol{Z}$ for the four types of runs are summarized as follows:

For type I runs,

$$A_I = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 1-\alpha_1 & \alpha_1 \\ 1 & 0 \end{bmatrix} \end{array}, \qquad 0 < \alpha_1 < 1,$$

where $\alpha_1 = p^k(1-p)/[1-p^k(1-p)]$.

For type II runs,

$$A_{II} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 1-\alpha_2 & \alpha_2 \\ 1 & 0 \end{bmatrix} \end{array}, \qquad 0 < \alpha_2 < 1,$$

where $\alpha_2 = p^k(1-p)/(1-p^k)$.

For type III runs,

$$A_{III} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 1-\alpha_3 & \alpha_3 \\ 1 & 0 \end{bmatrix} \end{array}, \qquad 0 < \alpha_3 < 1,$$

where $\alpha_3 = p^k/(1-p^k)$.

For type IV runs,

$$A_{IV} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1-\alpha_4 & \alpha_4 \\ 1-p & p \end{bmatrix} \end{matrix}, \qquad 0 < \alpha_4 < 1,$$

where $\alpha_4 = p^k(1-p)/(1-p^k)$.

Strictly speaking the above four transition matrices hold only if $i \geq k+1$. Since in this chapter we are interested in the limiting distributions of the four counting variables, without loss of generality, we shall assume all the four transition matrices hold for all $i \geq 1$.

We present the next lemma without proof. It can be derived from the results in Wang (1981) and Gani (1982).

Lemma 1. Let $\{X_i : i = 1, 2, \cdots\}$ be a sequence of completely stationary Markov Bernoulli random variables. If $nP(X_i = 1|X_{i-1} = 0) \to \lambda$, $P(X_i = 1|X_{i-1} = 1) \to 0$, and $P(X_1 = 1) \to 0$, as

$n \to \infty$, then the distribution of the sum $S_n = X_1 + X_2 + \cdots + X_n$ converges to the Poisson distribution with parameter λ.

The next theorem follows from Lemma 1.

Theorem 3. Let $\{X_1, X_2, \cdots\}$ be a sequence of i.i.d. Bernoulli random variables. If $np^k \to \lambda$, as $n \to \infty$, then the limiting distributions of all the four counting variables N_I, N_{II}, N_{III}, and N_{IV} are Poisson with parameter λ.

Even though the limit conditions are the same in all the four cases, for approximation purpose one should use $\hat{\lambda}_i = n\alpha_i$ for $i = 1, 2, 3$, and 4, where each α_i is as specified above.

In the i.i.d. case, the exact distributions of N_{III} and N_{IV} have recently been derived by Hirano (1984) and Ling (1988), and are called the type I binomial distribution of order k and the type II binomial distribution of order k, respectively. Their probability mass functions are:

$$P(N_{III} = x) = \sum_{i=0}^{k-1} \sum_{x_1, \cdots, x_k}{}' \binom{x_1 + \cdots + x_k + x}{x_1, \cdots, x_k, x} p^n \left(\frac{q}{p}\right)^{x_1 + \cdots + x_k},$$

where the summation $\sum'$ is over all nonnegative integers $x_1, \cdots, x_k$ with $x_1 + 2x_2 + \cdots + kx_k = n - i - kx$, and $q = 1 - p$. (See Hirano (1984).)

$$P(N_{IV} = x) = \sum_{i=0}^{n} \sum_{x_1, \cdots, x_n}{}'' \binom{x_1 + \cdots + x_n}{x_1, \cdots, x_n} p^n \left(\frac{q}{p}\right)^{x_1 + \cdots + x_n},$$

where the summation $\sum''$ is over all nonnegative integers $x_1, \cdots, x_k$ with $x_1 + 2x_2 + \cdots + nx_n = n - i$, and $max\{0, i - k + 1\} + \sum_{j=k+1}^{n} (j - k)x_j = x$. (See Ling (1988).)

We have used Theorem 1 and the two expressions of Hirano and Ling above to compute the exact probabilities of N_{III} and N_{IV} for $n = 20$, $k = 2, 3, 4, 5$, and $p = 0.05, 0.1$. As expected the probabilities computed by using Theorem 1 and the two expression above are identical. For comparison purpose, we also use Theorem 1 and Theorem 3 to compute the Poisson approximation of types III and IV runs for $n = 50, 70, 100, 500$, $k = 2, 3, 4, 5$, and $p = 0.05, 0.1$. Their result are compiled in the following tables.

BIBLIOGRAPHY

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

Aki, S., Kuboki, H. and Hirano, K. (1984). On discrete distribution of order k, *the Annals of the Institute of Statistical Mathematics,* **36,** Part A, 431–440

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

Brainerd, B. and Chang, S.M. (1982). Number of occurrences in two-state Markov chains, with an application in linguistics, *Canadian Journals of Statistics,* **10,** 225–231.

Chao, M.T. and Fu, J.C. (1989). A limit theorem of certain repairable systems, *the Annals of the Institute of Statistical Mathematics,* **41,** 809–818.

Chao, M.T. and Fu, J.C. (1991). The reliability of large series system under Markov structure, *Advance in Applied Probability,* 894–908.

Chiang, D.T. and Niu, S.C. (1981). Reliability of consecutive k-out-of-n:F system, *IEEE Transactions on Reliability,* **R-30,** 87–89.

Chung, Kai Lai (1967). *Markov Chains, With Stationary Transition Probabilities,* Second Edition. Springer-Verlag, New York.

Cinlar, E. (1975). *Introduction to Stochastic Processes,* Prentice-Hall, Englewood Cliffs, N.J.

Derman, C., Lieberman, G. J. and Ross, S. M. (1982). On the consecutive-k-out-of-n:F system. *IEEE transactions on Reliability,* **R-31**, 57–63.

Edwards, A. W. F. (1960). The meaning of binomial distribution, *Nature,* London, **186**, 1074.

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

Fu, J.C. (1993). Poisson convergence in reliability and coin tossing, Technical Report, University of Manitoba.

Gani, J. (1982). On the probability generating function of the sum of Markov Bernoulli random variables, *Journal Applied Probability,* **19A**, 321–326.

Godbole, A.P. (1990). Degenerate and Poisson convergence criteria for success runs, *statistics and Probability Letters,* **10**, 247–255.

Godbole, A.P. (1991). Poisson approximations for runs and patterns of rare events, *Advances in Applied Probability,* **23**, 851–865.

Goldstein, L. (1990). Poisson approximation and DNA sequence matching, *Communications in Statistics - Theory and Methodology,* **19**, 4167–4179

Griffith, W.S. (1986). On consecutive k-out-of-n failure systems and their generalizations, In *Reliability and Quality Control,* ed. A.P.Basu, Elsevier, Amrsterdam.

Hirano, K. (1984). Some properties of the distributions of order k, in *Proceedings of First International Conference on Fibonacci Numbers and Their Application,* ed. A. N. Philippou and A. F. Horadam, Gutenbery, Athens.

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

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

Kemeny, J. G. and Snell J. L. (1970) *Finite Markov Chains,* Princeton, N. J. Van Nostrand.

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

Ling, K.D. (1989). A new class of negative binomial distributions of order k, *Statistics Probability Letter,* **7**, 371–376.

Mood, A. (1940). The distribution theory of runs. *The Annals of Mathematical Statistics,* **11**, 367–392.

Panaretos, J. and Xekalaki, E. (1986). On some distributions arising certain generalized sampling schemes. *Communication in Statistics - Theory and Methodology,* **15**, 873–891.

Papastavridis, S. (1991). m-Consecutive-k-out-of-n:F system, *IEEE Transactions on Reliability,* **39**, 386–388.

Pedler, J. (1978). *The occupation time, number of transitions, and waiting time for two-state Markov Chains,* Thesis, The Flinders University of South Australia.

Philippou, A.N. (1986). Distributions and Fibonacci polynomials of order k, longest runs, and reliability of consecutive-k-out-of-n:F systems. *Fibonacci Numbers and Their Applications,* edited by Philippou, A. N. and Horadam, A. F. 203–227, D.Reidel.

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

Philippou, A.N. and Makri, F.S. (1986). Successes, runs and longest runs, *Statistics Probability Letter,* **4**, 101–105.

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

von Mises, R. (1921). Das Problem der Iterationen, *Z. Angew.*

Math. Mech., **1**, 298–307.

Wang, Y.H. (1981). On the limit of the Markov binomial distribution, *Journal of Applied Probability*, **18**, 937–942.

Wang, Y.H. (1989). From Poisson to compound Poisson approximations, *Mathematical Scientist*, **14**, 38–49.

Wang, Y.H. (1992). Approximating the k-th order two-state Markov Bernoulli chains, *Journal of Applied Probability*, **29**, 861–868.

Wang, Y.H. and Ji, Shuixin (1993). Derivations of the compound Poisson distributions and processes, *Statistics Probability Letter*, (To appear.)

Yule, G.U. (1944). *The Statistical Study of Literary Style*. Cambridge University Press.

APPENDIX

Table A. Comparison the Exact Probability of Counting Variable for Type III and Type IV Two S Runs with Poisson Approximation

$k = 2$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
50						
	0.05					
		0	0.88942	0.88803	0.88942	0.88777
		1	0.10458	0.10545	0.09969	0.10569
		2	0.00579	0.00626	0.00994	0.00629
		3	0.00020	0.00025	0.00088	0.00025
		4	0.00000	0.00000	0.00007	0.00000
	0.10					
		0	0.63633	0.63766	0.63633	0.63474
		1	0.29138	0.28692	0.26516	0.28852
		2	0.06291	0.06455	0.07616	0.06557
		3	0.00851	0.00968	0.01787	0.00994
		4	0.00081	0.00109	0.00366	0.00113
		5	0.00006	0.00010	0.00068	0.00010
		6	0.00000	0.00000	0.00012	0.00000
		7	0.00000	0.00000	0.00002	0.00000
70						
	0.05					
		0	0.84792	0.84684	0.84792	0.84648
		1	0.14036	0.14079	0.13375	0.14108
		2	0.01113	0.01170	0.01642	0.01176
		3	0.00056	0.00065	0.00173	0.00065
		4	0.00002	0.00003	0.00016	0.00003
		5	0.00000	0.00000	0.00001	0.00000

k = 2						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
	0.10					
		0	0.52928	0.53263	0.52928	0.52921
		1	0.34113	0.33552	0.31027	0.33677
		2	0.10546	0.10568	0.11555	0.10715
		3	0.02082	0.02219	0.03395	0.02273
		4	0.00295	0.00349	0.00854	0.00362
		5	0.00032	0.00044	0.00192	0.00046
		6	0.00003	0.00005	0.00039	0.00005
		7	0.00000	0.00000	0.00008	0.00000
		8	0.00000	0.00000	0.00001	0.00000
100						
	0.05					
		0	0.78927	0.78860	0.78927	0.78813
		1	0.18742	0.18729	0.17856	0.18765
		2	0.02160	0.02224	0.02807	0.02234
		3	0.00161	0.00176	0.00363	0.00177
		4	0.00009	0.00010	0.00041	0.00011
		5	0.00000	0.00000	0.00004	0.00000
	0.10					
		0	0.40151	0.40661	0.40151	0.40657
		1	0.37117	0.36591	0.33746	0.36591
		2	0.16669	0.16464	0.16869	0.16466
		3	0.04845	0.04939	0.06425	0.04940
		4	0.01025	0.01111	0.02048	0.01111
		5	0.00168	0.00200	0.00573	0.00200
		6	0.00022	0.00030	0.00145	0.00030
		7	0.00002	0.00004	0.00034	0.00004
		8	0.00000	0.00000	0.00007	0.00000
		9	0.00000	0.00000	0.00002	0.00000

$k = 2$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
500						
	0.05					
		0	0.30351	0.30498	0.30351	0.30408
		1	0.36315	0.36217	0.34585	0.36200
		2	0.21598	0.21504	0.21241	0.21547
		3	0.08514	0.08512	0.09300	0.08551
		4	0.02502	0.02527	0.03245	0.02545
		5	0.00585	0.00600	0.00958	0.00606
		6	0.00113	0.00119	0.00248	0.00120
		7	0.00019	0.00020	0.00058	0.00020
		8	0.00003	0.00003	0.00012	0.00003
		9	0.00000	0.00000	0.00002	0.00000
	0.10					
		0	0.01009	0.01111	0.01009	0.01062
		1	0.04698	0.05001	0.04268	0.04825
		2	0.10876	0.11251	0.09371	0.10966
		3	0.16691	0.16874	0.14206	0.16616
		4	0.19101	0.18982	0.16696	0.18881
		5	0.17387	0.17082	0.16200	0.17165
		6	0.13112	0.12810	0.13497	0.13004
		7	0.08427	0.08234	0.09918	0.08444
		8	0.04711	0.04631	0.06553	0.04798
		9	0.02327	0.02315	0.03951	0.02423
		10	0.01029	0.01042	0.02199	0.01101
		11	0.00411	0.00426	0.01139	0.00455
		12	0.00150	0.00160	0.00554	0.00172
		13	0.00050	0.00055	0.00254	0.00060
		14	0.00015	0.00018	0.00111	0.00020
		15	0.00004	0.00005	0.00046	0.00006

$k = 2$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
		16	0.00001	0.00001	0.00018	0.00002
		17	0.00000	0.00000	0.00007	0.00000
		18	0.00000	0.00000	0.00003	0.00000

Table B.

Comparison the Exact Probability of Counting Variable
for Type III and Type IV Three S Runs
with Poisson Approximation

$k = 3$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
50						
	0.05					
		0	0.99431	0.99408	0.99772	0.99438
		1	0.00568	0.00590	0.00216	0.00561
		2	0.00001	0.00002	0.00011	0.00002
	0.10					
		0	0.95750	0.95600	0.95750	0.96031
		1	0.04167	0.04302	0.03763	0.03889
		2	0.00082	0.00097	0.00433	0.00079
		3	0.00000	0.00001	0.00049	0.00001
		4	0.00000	0.00000	0.00005	0.00000
70						
	0.05					
		0	0.99195	0.99172	0.99759	0.99213
		1	0.00802	0.00824	0.00229	0.00783
		2	0.00003	0.00003	0.00012	0.00003
	0.10					
		0	0.94036	0.93894	0.94036	0.93888
		1	0.05795	0.05915	0.05228	0.05921
		2	0.00166	0.00186	0.00647	0.00187
		3	0.00003	0.00004	0.00078	0.00004
		4	0.00000	0.00000	0.00009	0.00000
		5	0.00000	0.00000	0.00001	0.00000

$k = 3$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
100						
	0.05					
		0	0.98842	0.98820	^.99753	0.98878
		1	0.01152	0.01173	0.00234	0.01115
		2	0.00006	0.00007	0.00012	0.00006
	0.10					
		0	0.91523	0.91393	0.91523	0.91393
		1	0.08124	0.08225	0.07326	0.08225
		2	0.00343	0.00370	0.00999	0.00370
		3	0.00009	0.00011	0.00132	0.00011
		4	0.00000	0.00000	0.00017	0.00000
		5	0.00000	0.00000	0.00002	0.00000
500						
	0.05					
		0	0.94255	0.94235	0.94255	0.94235
		1	0.05579	0.05595	0.05301	0.05595
		2	0.00163	0.00166	0.00412	0.00166
		3	0.00003	0.00003	0.00031	0.00003
	0.10					
		0	0.63781	0.63763	0.63781	0.63734
		1	0.28745	0.28693	0.25900	0.28709
		2	0.06415	0.06456	0.07774	0.06466
		3	0.00945	0.00968	0.01976	0.00971
		4	0.00103	0.00109	0.00451	0.00109
		5	0.00009	0.00010	0.00095	0.00010
		6	0.00000	0.00000	0.00019	0.00000
		7	0.00000	0.00000	0.00004	0.00000

Table C.

Comparison the Exact Probability of Counting Variable
for Type III and Type IV Four S Runs
with Poisson Approximat: .n

$k = 4$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
50						
	0.05					
		0	0.99972	0.99970	0.99972	0.99970
		1	0.00028	0.00030	0.00027	0.00030
		2	0.00000	0.00000	0.00001	0.00000
	0.10					
		0	0.99577	0.99551	0.99577	0.99551
		1	0.00422	0.00448	0.00381	0.00448
		2	0.00000	0.00001	0.00038	0.00001
		3	0.00000	0.00000	0.00004	0.00000
70						
	0.05					
		0	0.99960	0.99958	0.99960	0.99958
		1	0.00040	0.00042	0.00038	0.00042
		2	0.00000	0.00000	0.00002	0.00000
	0.10					
		0	0.99398	0.99372	0.99398	0.99372
		1	0.00601	0.00626	0.00542	0.00626
		2	0.00002	0.00002	0.00055	0.00002
		3	0.00000	0.00000	0.00006	0.00000
100						
	0.05					
		0	0.99942	0.99941	0.99942	0.99941

$k = 4$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
		1	0.00058	0.00059	0.00055	0.00059
		2	0.00000	0.00000	0.00003	0.00000
	0.10					
		0	0.99129	0.99104	0.99129	0.99104
		1	0.00867	0.00892	0.00781	0.00892
		2	0.00004	0.00004	0.00080	0.00004
		3	0.00000	0.00000	0.00008	0.00000
500						
	0.05					
		0	0.99705	0.99704	0.99705	0.99704
		1	0.00294	0.00296	0.00280	0.00296
		2	0.00000	0.00000	0.00014	0.00000
	0.10					
		0	0.95623	0.95600	0.95623	0.95599
		1	0.04281	0.04302	0.03854	0.04302
		2	0.00095	0.00097	0.00461	0.00097
		3	0.00001	0.00001	0.00055	0.00001
		4	0.00000	0.00000	0.00006	0.00000

Table D.

Comparison the Exact Probability of Counting Variable
for Type III and Type IV Five S Runs
with Poisson Approximation

$k = 5$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
50						
	0.05					
		0	0.99999	0.99999	0.99999	0.99999
		1	0.00001	0.00001	0.00001	0.00001
	0.10					
		0	0.99959	0.99955	0.99959	0.99955
		1	0.00041	0.00045	0.00037	0.00045
		2	0.00000	0.00000	0.00004	0.00000
70						
	0.05					
		0	0.99998	0.99998	0.99998	0.99998
		1	0.00002	0.00002	0.00002	0.00002
	0.10					
		0	0.99941	0.99937	0.99941	0.99937
		1	0.00059	0.00063	0.00054	0.00063
		2	0.00000	0.00000	0.00005	0.00000
100						
	0.05					
		0	0.99997	0.99997	0.99997	0.99997
		1	0.00003	0.00003	0.00003	0.00003
	0.10					
		0	0.99914	0.99910	0.99914	0.99910
		1	0.00086	0.00090	0.00078	0.00086
		2	0.00000	0.00000	0.00008	0.00000

$k = 5$						
			III		IV	
n	p	N	Exact Distribution	Poisson Approximation	Exact Distribution	Poisson Approximation
450						
	0.05					
		0	0.99987	0.99987	0.99987	0.99987
		1	0.00013	0.00013	0.00013	0.00013