Limit Theorems for the Number of Occurrences of Consecutive k Successes in n Markov Bernoulli Trials

Shuixin Ji

A Thesis
in
The Department
of
Mathematics and Statistics

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

June 1994

National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-612-01338-3

Canada

ABSTRACT

Limit Theorems for the Number of Occurrences of
Consecutive k Successes in n Markov Bernoulli Trials

Shuixin Ji

In this thesis we present a method of deriving the limiting distributions of the number of occurences of success (S) runs of length k for all types of runs under the Markovian structure with stationary transition probabilities. In particular, we consider the following four best-known types of runs:

1. A string of S of exact length k proceded and followed by an F, except the first run which is not proceded by an F, or the last run which may not be followed by an F;

2. A string of S of length k or more;

3. A string of S of exact length k, where recounting starts immediately after a run occurs;

4. A string of S of exact length k, allowing overlapping runs.

It is shown that the limiting distributions are convolutions of two or more distributions with one of them being either Poisson or compound Poisson, depending on the type of runs in question. The completely stationary Markov case and the independent and identically distributed case are also treated.

key words and phrases: Runs, consecutive k successes, Markov chains, stationary transition probabilities, convergence,Poisson, compound Poisson, reliability, consecutive-k out-of n: F system.

Acknowledgements

First I would like to thank Professor 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 encouraging me to do graduate studies, for his precious help to get financial support and for his patience.

Chapter 1

Introduction

Problems on and around RUNS have a long history, beginning, to say the least, with de Moivre' s *The Doctrine of Chances.* (See the English edition De Moivre (1967). The original Latin edition was published in 1717.) In the last decade, the problems propagate into many other fields, like reliability, linguistics, test of randomness in statistics, DNA analysis in biology, etc. Most of the work done have centered around the independent and identically distributed sequence. In this thesis, we shall extend it to the Markov dependent case.

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 in counting the number of occurrences of the pattern

$$S,S,\ldots,S$$

a run of k S in a total of n trials.

There are many way of counting the number of run of S of length k. Four best-known types are:

I: A run of of length k means a string of S of exact length k preceded and followed by an F, except the first run which is not proceded by an F or the last run which may not be followed by an F.

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

III: A run of S of length k means a string of S of exact length k with recounting starts immediately after a run occurs.

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

Runs of type I are the most restricted ones. 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 pattern (a cluster of letters or words) in a randomly selected text of an author is equivalent to counting the number of the "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 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 be 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 is the type II runs 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 equivalent 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 "failure" 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 failure 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 uninterruped 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 a new A binary sequence is obtained by recording S or F for each random drawing according to whether it is a red or a white. In this example, the occurrence of consecutive k successes means that the number of the red balls outside the urn reaching k.

Example B. (See Aki (1985)) An electric bulb is lit. 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 lit for k consecutive days, it is replaced with a new one even if it has not failed. Define a binary sequence by recording S or 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 being replaced with a new one.

Example C. (Counters of Type I.) A sequence of Bernoulli trails 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 $n, n+1, \ldots, n+k-1$ 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 occurence of consecutive k successes means that the counter is locked for a period of k trials, including the initial trail 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 times 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 trail is registered which locks the counter to the $(n+k-1)$–th trail, and another success at the $(n+1)$–th trail is again registered, then the locking period of the counter is prolonged to the $(n+k)$-th trail. 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).)

Let N_I be the number of occurrence of consecutive k successes of type I. the other three counting variables N_{II}, N_{III}, N_{IV} are defined accordingly. (For brevity we suppress the dependence of all four counting variables on k and n.) For example, consider the following realization, with n = 16.

SSSSSSFFFSSSFFSS

If we take k = 3, then $N_I = 1; N_{II} = 2; N_{III} = 3$; and $N_{IV} = 5$. Evidently we have the stochastic ordering of $N_I \leq N_{II} \leq N_{III} \leq N_{IV}$ and if k = 1, then N_{II} is the number of transitions from F to S and $N_{III} = N_{IV}$ is the occupancy time of S. In the recent article of reference, an algorithm for computing the exact probability distributions under the Markovian structure with stationary transition probabilities of the four variables is proposed.

The purpose of this thesis is to give a unified approach to the derivation of the limiting distributions of all the four counting variables for all $k \geq 1$ as n tends to infinity and some minor conditions. We shall show that under the Markovian structure with stationary transition probabilities the limits are convolutions of two or more distributions with one of them Poisson for N_I and N_{II} and compound Poisson for N_{III} and N_{IV}. Similar phenomena occur under the completely stationary Markovian structure, But under the independent and identically (i.i.d.) structure, all the limits are Poisson.

Thoughout this thesis, we adopt the usual convention of denoting "S" by 1 and "F" by 0. Let $X_1, \ldots, X_n$ be a sequence of Markov Bernoulli random variables with the following stationary transition probabilities

$$\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \qquad 0 < \alpha, \beta < 1 \tag{1.1}$$

and the initial probability

$$P(X_1 = 1) = p = 1 - P(X_1 = 0)$$

$0 \leq p \leq 1$.

The model contains the following two special cases:

1). If $p = \alpha/(\alpha + \beta)$, then $\{X_i\}$ is completely stationary in the sense that $P(X_i = 1) = p$ for all $i = 1, \ldots, n$; and with transition probabilities

$$\begin{pmatrix} 1-(1-\pi)p & (1-\pi)p \\ (1-\pi)(1-p) & (1-\pi)p+\pi \end{pmatrix} \tag{1.2}$$

where π is the correlation coefficient of X_i and X_{i+1}. In this model, if $\pi \to 0$ as $n \to \infty$, we call it *asymptotically i.i.d. Markov Bernoulli model.* (See Edwards (1960).)

2). If $\alpha = p$ and $\beta = 1 - p$, then $\{X_i\}$ is the ordinary sequence of i.i.d. Bernoulli random variables. (The parameters α, β, p and π depend upon the index n. For brevity, we shall suppress this index throughout this thesis.)

Historically, Koopman (1950) was the first person to take on the problem of finding limiting distributions of the number of runs under a Markovian structure. He obtained the limiting distribution of N_{III} with k =1 for the asymptotically i.i.d. Markovian Bernoulli model. Later, Dobrusin (1953) obtained many interesting limit results for the same model. Today this model is still being pursued by many authors. See Godbole (1991) and the references cited there. Pedler (1978) was possibly the first person to consider the nonasymptotically independent Markov Bernoulli model (1.1) in his studies of the limiting distributions of N_{III} (and others) also for k =1. The recent papers Wang (1991), Gani (1982), Buhler (1989) and Wang and Buhler (1991) were concerned with the limiting distribution of N_{III}, also for k = 1, of the completely stationary model (1.2).

For $k \geq 2$, some results on the limiting distributions of N_{II}, N_{III} and N_{IV} in the i.i.d. case can be found in Papastavridis (1987), Goldstein (1990), Godble (1991), Barbour et al (1992) and Fu (1993). The recent article Koutras and Papastavridis (1993) deals with related problems in the i.i.d. case.

To derive our results, we shall take an approach different from all the approaches used by everybody else on this topic. Our idea can be traced back to Doeblin (1938). Instead of directly working with the original Markov Bernoulli sequence, we construct an equivalent parallel sequence for which the limiting distributions of all the four types of runs are obtained. The problem is thus to make sure that both sequences are asymptotically equivalent.

Chapter 2

The limiting distributions

Define two sequence of random variables $\{U_i'\}$ and $\{V_i'\}$ by

U_i' = the length of i–th sojourn of $\{X_i\}$ in state 1,

V_i' = the length of i-th sojourn of $\{X_i\}$ in state 0.

The Markovian property and the stationarity of the transition probabilities of $\{X_i\}$ assure that all the components of $\{U_i'\}$ and $\{V_i'\}$ are mutually independent with marginal geometric distribution:

$$\begin{cases} P(U_i' = k) = (1-\beta)^{k-1}\beta, & k = 1,2,\ldots \\ P(V_i' = k) = (1-\alpha)^{k-1}\alpha, & k = 1,2,\ldots \end{cases} \tag{2.1}$$

Evidently observing the sequence $\{X_i\}$ is the same as observing U_1', V_1', U_2', V_2', $\ldots$, if $X_1 = 1$ and V_1', U_1', V_2', U_2', $\ldots$, if $X_1 = 0$. Therefore to study the number of occurrences of success runs of the sequence $\{X_i\}$, it is sufficient to study this alternating process. As in Wang (1992), we shall modify the alternating process as follows:

Let $\{U_i\}$ and $\{V_i\}$ be two independent sequences of random variables with marginal distributions (2.1) and $\{U_i\}$ and $\{V_i\}$ are also independent of each other. If $X_1 = 0$, we let $V_1 = V_1'$. When V_1 terminates, we start two runs U_1 and V_2 immediately and simultaneously. We let U_1 run its course, but when V_2 terminates we start another two runs U_2 and V_3. The process is repeated indefinitely, such that U_i runs its course, but when V_i terminates, two runs U_i and V_{i+1} are started immediately for all $i \geq 1$.

If $X_1 = 1$, let $U_1 = U_1'$, and initiate another run V_1 at the same time i=1. As before, we let U_1 run its course, but when V_1 terminates, two runs U_2 and V_2 are activated immediately and simultaneously, and the process is repeated indefinitely as described above.

Define

$$
\begin{aligned}
W_i^{(I)} &= I(U_i = k) \qquad \text{(I(A) denotes the indicator function of A.)} \\
W_i^{(II)} &= I(U_i \geq k) \\
W_i^{(III)} &= [U_i/k] \qquad \text{([x] denotes the integral part of x.)}
\end{aligned}
$$

$$W_i^{(IV)} = \begin{cases} U_i - k + 1 & \text{if } U_i \geq k \\ 0 & \text{otherwise} \end{cases}$$

Since $\{V_i'\}$, having geometric distribution (2.1), is *memoryless*, i.e.

$$P(V_i' = j | V_i' > l) = P(V_i' = j - l)$$

for all $i = 1, 2, \ldots,$ and $j > l \geq 1$, the original alternating process in which V' and U' alternate is *asymptotically* equivalent to the modified process, provided that

$$P(V_i' > U_i' \quad \text{for all i } | i \leq n) \to 1 \qquad \text{as} \quad n \to \infty.$$

(The proof of the above statement not be given here and shall be found in the proof of Lemma 1.)

For fixed $n \geq 2$, define another stopping time by

$$M_n = \begin{cases} 0 & \text{if } V_1 \geq n, \\ \max\{k : V_1 + \cdots + V_k \leq n - 1\} & \text{otherwise.} \end{cases} \tag{2.2}$$

Then M_n is binomial with parameters $(n-1, \alpha)$ and independent of $W_j^{(i)}$ for all $i = I, II, III$ and IV and $j \geq 1$. Denote

$$S_n^{(i)} = W_1^{(i)} + \cdots + W_{M_n}^{(i)}. \tag{2.3}$$

Because in the modified process U_i may overlap, we have

$$P(S_n^{(i)} \geq N_i) = 1$$

and

$$P(S_n^{(i)} > N_i) > 0$$

for all finite $n \geq 2$.

Lemma 1 *For all $i = I, II, III$ and IV, we have*

$$\lim_{n \to \infty} P(S_n^{(i)} > N_i) = 0$$

if $n\alpha \to \lambda \geq 0$ as $n \to \infty$. (Note that in model (1.1) we assumed $0 \leq \alpha,\ \beta \leq$ 1.)

Proof: First we have

$$\begin{aligned}
P(W_1^{(i)} > V_1) &\leq P(U_1 > V_1) \\
&= \sum_{r=2}^{\infty} \sum_{s=1}^{r-1} P(U_1 = r, V_1 = s) \\
&= \sum_{r > s \geq 1} P(U_1 = r) P(V_1 = s) \\
&= \sum_{r > s \geq 1} (1-\beta)^{r-1} \beta (1-\alpha)^{s-1} \alpha \\
&= \alpha\beta \sum_{r=2}^{\infty} (1-\beta)^{r-1} \sum_{s=1}^{r-1} (1-\alpha)^{s-1}
\end{aligned}$$

$$= \beta \sum_{r=2}^{\infty} (1-\beta)^{r-1}(1-(1-\alpha)^{r-1})$$

$$= 1 - \beta(1-(1-\beta)(1-\alpha))^{-1}$$

$$= [\alpha - \alpha\beta]/[1-(1-\alpha)(1-\beta)]$$

$$\leq \alpha/[1-(1-\alpha)(1-\beta),$$

so that

$$P(S_n^{(i)} > N_i)$$

$$\leq \sum_{j=1}^{n-1} P(W_h^{(i)} > V_h \quad \text{for some h=1, 2, } \ldots \text{, j} | M_n = j) P(M_n = j)$$

$$\leq \sum_{j=1}^{n-1} P(U_1 > V_1) j P(M_n = j)$$

$$\leq P(U_1 > V_1) n\alpha$$

$$\leq n\alpha^2/[1-(1-\alpha)(1-\beta)].$$

The last term tends to 0 if $n\alpha \rightarrow \lambda \geq 0$ as $n \rightarrow \infty$.

According to lemma 1, to find the limiting distributions of N_i it is sufficient to find those of $S_n^{(i)}$.

Denote the probability mass functions (p.m.f.) and probability generating functions (p.g.f.) of $W_j^{(i)}$ to be f_i and g_i, respectively, for $i = I, II, III$, and IV, and all $j = 1, 2, \ldots$. It follows from the definitions of W's that:

$$\begin{cases} f_I(x) = \theta_1^x(1-\theta_1)^{1-x} & \text{for x= 0,1} \\ g_I(t) = 1+\theta_1(t-1) & \text{for all } t \in R \end{cases} \tag{2.4}$$

where $\theta_1 = (1-\beta)^{k-1}\beta$.

$$\begin{cases} f_{II}(x) = \theta_2^x(1-\theta_2)^{1-x} & \text{for x=0,1} \\ g_{II}(t) = 1+\theta_2(t-1) & \text{for all } t \in R \end{cases} \tag{2.5}$$

where $\theta_2 = (1-\beta)^{k-1}$.

$$\begin{cases} f_{III}(x) = \begin{cases} 1-(1-\beta)^{k-1} & \text{if x =0} \\ (1-\beta)^{k-1}(1-\beta)^{k(x-1)}(1-(1-\beta)^k) & \text{if x = 1,2, } \ldots. \end{cases} \\ g_{III}(t) = 1-(1-\beta)^{k-1} + \frac{(1-\beta)^{k-1}(1-(1-\beta)^k)s}{1-s(1-\beta)^k} \text{for} \quad s < (1-\beta)^{-1}. \end{cases} \tag{2.6}$$

and

$$\begin{cases} f_{IV}(x) = \begin{cases} 1-(1-\beta)^{k-1} & \text{if x=0} \\ (1-\beta)^{k-1}\beta(1-\beta)^{x-1} & \text{if x=1,2, } \ldots \end{cases} \\ g_{IV}(t) = 1-(1-\beta)^{k-1} + \frac{(1-\beta)^{k-1}\beta s}{1-s(1-\beta)} \qquad \text{for } s < (1-\beta)^{-1}. \end{cases} \tag{2.7}$$

Let G_i be the p.g.f. of $S_n^{(i)}$. Then by (2.2), (2.3), and conditioning on X_1 we have

$$G_i(t) = [pg_i(t) + (1-p)][1 + \alpha(g_i(t) - 1)]^{n-1} \tag{2.8}$$

It is interesting to note that both p.m.f. in (2.6) and (2.7) are mixtures of two p.m.f. with one of them Bernoulli. In Wang (1989, Theorem 6), it was proved that if a sequence of i.i.d. discrete random variables $\{X_i\}$ whose common distribution is mixture of two p.m.f. with one of them Bernoulli, then its partial sum $S_n = X_1 + \cdots + X_n$ has a compound Poisson limiting distribution.

If the limits of the distributions of N_i are to exist, 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 , $n\alpha$, the mean number of transitions from state 0 to state 1, must approach a constant as $n \to \infty$. We discovered that this is the ***only*** condition needed to assure the existence of the limiting distributions of the number of occurrences of consecutive k (≥ 1) successes in the Markov case (1.1) and in the completely stationary case (1.2). the other parameters p and β in the Markov case and π in the completely stationary case play no roles here. They can stay constant or have their own limits, such as $p \to p_0$, $\beta \to \beta_0$ and $\pi \to \pi_0$, for p_0, β_0, and $\pi_0 \in (0,1)$. Therefore, without loss of generality, we shall assume that the parameters p, β and π stay constant in the next four theorems and corollary

1. (The condition "$n\alpha \to \lambda$ as $n \to \infty$" or its equivalent "$\sum_{i=1}^{n} P(X_{i+1} = 1 | X_i = 0) \to \lambda$ as $n \to \infty$" in the non-stationary case was first used by Koopman (1950) and Dobrusin (1953) and has since become a key condition in Poisson approximation problems for Markovian dependent sequence.)

In the followings we state and prove the four main theorems of this thesis.

Theorem 1 *If $n\alpha \to \lambda > 0$, as $n \to \infty$, then the p.g.f. of N_I converges to*

$$\phi_I(t) = \{1 + p(1-\beta)^{k-1}\beta(t-1)\}\exp\{\lambda(1-\beta)^{k-1}\beta(t-1)\}.$$

Thus the limiting distribution of N_I is the convolution of a Bernoulli with parameter $p(1-\beta)^{k-1}\beta$ and a Poisson with parameter $\lambda(1-\beta)^{k-1}\beta$.

Proof: The p.g.f. of $S_n^{(I)}$ is

$$\begin{aligned} G_I(t) &= [pg_I(t) + (1-p)][1 + \alpha(g_I(t) - 1)]^{n-1} \\ &= [1 + p(1-\beta)^{k-1}\beta(t-1)][1 - \alpha(1 - g_I(t))]^{n-1} \\ &\longrightarrow [1 + p(1-\beta)^{k-1}\beta(t-1)]\exp[-\lambda(1 - g_I(t))] \end{aligned}$$

for $n\alpha \to \lambda$ as $n \to \infty$.

Now, substituting the $g_I(t)$ into the above formula, we get that the limit $\Phi_I(t)$ of $G_I(t)$ is

$$\Phi_I(t) = [1 + p(1-\beta)^{k-1}\beta(t-1)]\exp[\lambda(1-\beta)^{k-1}\beta(t-1)].$$

This completes the proof.

Theorem 2 *If* $n\alpha \to \lambda > 0$, *as* $n \to \infty$, *then the p.g.f. of* N_{II} *converges to*

$$\phi_{II}(t) = [1 + p(1-\beta)^{k-1}(t-1)]\exp[\lambda(1-\beta)^{k-1}(t-1)].$$

Thus the limiting distribution of N_{II} *is the convolution of a Bernoulli with parameter* $p(1-\beta)^{k-1}$ *and a Poisson with parameter* $\lambda(1-\beta)^{k-1}$.

Proof: The p.g.f. of $S_n^{(II)}$ is

$$\begin{aligned} G_{II}(t) &= [pg_{II}(t) + (1-p)][1 + \alpha(g_{II}(t) - 1)]^{n-1} \\ &= [1 + p(1-\beta)^{k-1}(t-1)][1 - \alpha(1 - g_{II}(t))]^{n-1} \\ &\longrightarrow [1 + p(1-\beta)^{k-1}(t-1)]\exp[-\lambda(1 - g_{II}(t))] \end{aligned}$$

as $n \to \infty$.

Now, substituting the $g_{II}(t)$ into the above formula, we get that the limit $\phi_{II}(t)$ of $G_{II}(t)$ is

$$\phi_{II}(t) = [1 + p(1-\beta)^{k-1}(t-1)]\exp[\lambda(1-\beta)^{k-1}(t-1)].$$

This completes the proof.

Theorem 3 *If* $n\alpha \to \lambda > 0$, *as* $n \to \infty$, *then the p.g.f. of* N_{III} *converges to*

$$
\begin{aligned}
\phi_{III}(t) &= [1 - p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}(1-(1-\beta)^k)t}{1-t(1-\beta)^k}] \\
&\times \exp[\lambda(1-\beta)^{k-1}\{\frac{[1-(1-\beta)^k]t}{1-t(1-\beta)^k} - 1\}].
\end{aligned}
$$

Thus the limiting distribution of N_{III} is the convolution of a mixture of Bernoulli distribution with parameter $p(1-\beta)^k$, and a geometric distribution with parameter $(1-\beta)^k$, and a compound Poisson distribution with parameter $\lambda(1-\beta)^{k-1}$ and geometric compounding distribution whose p.g.f. is $[(1-(1-\beta)^k)t]/[1-t(1-\beta)^k]$.

Proof: the p.g.f. of $S_n^{(III)}$ is

$$
\begin{aligned}
G_{III}(t) &= [pg_{III}(t) + (1-p)][1 + \alpha(g_{III}(t) - 1)]^{n-1} \\
&= [1 - p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}(1-(1-\beta)^k)t}{1-t(1-\beta)^k}] \\
&\times [1 - \alpha(1 - g_{III}(t))]^{n-1} \\
&\longrightarrow [1 - p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}(1-(1-\beta)^k)t}{1-t(1-\beta)^k}] \\
&\times \exp[-\lambda(1 - g_{III}(t))]
\end{aligned}
$$

for $n\alpha \to \lambda$ as $n \to \infty$.

Now, substituting the $g_{III}(t)$ into the above formula, we get that the limit $\phi_{III}(t)$ of $G_{III}(t)$ is

$$\begin{aligned}\phi_{III}(t) &= \{1 - p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}(1-(1-\beta)^k)t}{1-t(1-\beta)^k}\} \\ &\times \exp[\lambda(1-\beta)^{k-1}\{\frac{(1-(1-\beta)^k)t}{1-t(1-\beta)^k} - 1\}].\end{aligned}$$

Theorem 4 *If $n\alpha \to \lambda > 0$, as $n \to \infty$, then the p.g.f. of N_{IV} converges to*

$$\phi_{IV}(t) = \{1-p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}\beta t}{1-t(1-\beta)}\}\exp\{\lambda(1-\beta)^{k-1}\{\frac{t\beta}{1-t(1-\beta)} - 1\}\}.$$

Thus the limiting distribution of N_{IV} is the convolution of mixture of Bernoulli distribution with parameter $p(1-\beta)^{k-1}$ and a geometric distribution with parameter $(1-\beta)$, and a compound poisson distribution with parameter $\lambda(1-\beta)^{k-1}$ and a geometric compounding distribution whose p.g.f. is $\frac{\beta t}{1-(1-\beta)t}$.

Proof: the p.g.f. of $S_n^{(IV)}$ is

$$\begin{aligned}G_{IV}(t) &= [pg_{IV}(t) + (1-p)][1+\alpha(g_{IV}(t)-1)]^{n-1} \\ &= [1-p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}\beta t}{1-t(1-\beta)}][1+\alpha(g_{IV}(t)-1)]^{n-1}\end{aligned}$$

$$\longrightarrow \quad [1 - p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}\beta t}{1-t(1-\beta)}]\exp[-\lambda(1 - g_{IV}(t))]$$

for $n\alpha \to \lambda$ as $n \to \infty$.

Now, substituting the $g_{IV}(t)$ into the above formula, we get that the limit $\phi_{IV}(t)$ of $G_{IV}(t)$ is

$$\phi_{IV}(t) = [1 - p(1-\beta)^{k-1} + \frac{p(1-\beta)^{k-1}\beta t}{1-t(1-\beta)}]\exp[\lambda(1-\beta)^{k-1}(\frac{t\beta}{1-t(1-\beta)} - 1)].$$

This completes the proof.

We shall record that the next results as Corollary 1 without proof. We call it a Corollary because the completely stationary model (1.2) is a special case of the Markovian model (1.1). But it is easy to see that it can not be obtained by substituting the parameters p and π in the four Theorems. In the completely stationary case the three parameters p, α and β interweave, and the values of p and β changes as nα tends to λ, while in the other Markov model all the three parameters act independently. Lemma 1 is still applicable to Corollary 1.

Corollary 1 *Let $\{X_i\}$ be a completely stationary Markov Bernoulli sequence with transition probabilities defined by (1.2), if $np \to \lambda > 0$, as $n \to \infty$, then the limiting distribution of:*

1. *N_I is a Poisson with parameter $\lambda(1-\pi)^2\pi^{k-1}$;*

2. N_{II} *is a Poisson with parameter* $\lambda(1-\pi)\pi^{k-1}$*;*

3. N_{III} *is a compound Poisson with parameter* $\lambda(1-\pi)\pi^{k-1}$ *and geometric compounding distribution* $h(x) = (1-\pi^k)\pi^{(x-1)k}$, $x = 1, 2, \ldots$*;*

4. N_{IV} *is a compound Poisson with parameter* $\lambda(1-\pi)\pi^{k-1}$ *and geometric compounding distribution* $h(x) = (1-\pi)\pi^{x-1}$, $x = 1, 2, \ldots$.

As noted earlier, if k =1, then $S_n = X_1 + \cdots + X_n = N_{III} = N_{IV}$. In such a case the limiting distributions in corollary 1) - 3) and 4) are identical and coincide with the one obtained in Wang(1981), Gani(1982), Buhler(1989) and Wang and Buhler(1991). Recently Godbole (1991, Theorem3) obtained limiting distributions of N_{III} and N_{IV} under the condition "$n\alpha(1-\beta)^{k-1} \to \lambda$ as $n \to \infty$". (α and β are according to our notations.) His limit is a Poisson. To be so, one must have "$(1-\beta) \to 0$ as $n \to \infty$", which is not necessarily implied by the condition "$n\alpha(1-\beta)^{k-1} \to \lambda$ as $n \to \infty$". Therefore the proper way to state the limit condition in Godbole's theorem 3 is "$n\alpha(1-\beta)^{k-1} \to \lambda$, $\alpha \to 0$, and $\beta \to 1$, as $n \to \infty$". Thus Godbole is dealing with an asymptotically independent, completely stationary Markov Bernoulli chain, and it has the same limit behavior as in the next i.i.d. Bernoulli case. (See Corollary 2.) In probability theorey, most of the results concerning asymptotically independent cases are identical to those concerning i.i.d. cases. Godbole's result is of no exception. Furthermore, because of lemma 1, it goes without saying that throughout this thesis all of our results are under condition "$\alpha \to 0$ as $n \to \infty$". The same condition evidently also applies to Godbole's result, otherwise counter examples can be easily constructed.

In the i.i.d. case, one might think that the limit results would follow from the corresponding results in the completely stationary case simply by letting the correlation coefficient of X_i and X_{i+1} $\pi = 0$. Surprisingly, this is not so. For example, for all $k > 1$, if $np \to \lambda$ as $n \to \infty$, then the *limiting distributions of all the four counting variable N are degenerate at 0* in the i.i.d. case. Therefore different limit conditions are required to obtain non-degenerate limits. We shall present the i.i.d. case as Corollary 2 below. Since Lemma 1 is no longer applicable in this case, the next lemma is needed.

lemma 2 *If the sequence $\{X_i\}$ of Bernoulli random variables are i.i.d., then for all $i = I, II, III$ and IV,*

$$\lim_{n \to \infty} P(S_n^{(i)} > N_i) = 0$$

if $np^k \to \lambda \geq 0$ as $n \to \infty$.

Proof: We note here that

$$P(W_1^{(i)} > V_1) = \begin{cases} 0 & i = I, II \\ p^{2k}/(1-(1-p)p^k) & i = III \\ p^{k+1}/(1-p(1-p)) & i = IV \end{cases}$$

and

$$\begin{aligned} & P(S_n^{(i)} > N_i) \\ \leq & \sum_{j=1}^{n-1} P(W_h^{(i)} > V_h \quad \text{for some h=1, \ldots, j} \mid M_n = j) P(M_n = j) \\ \leq & \sum_{j=1}^{n-1} P(W_1^{(i)} > V_1) j P(M_n = j) \\ = & P(W_1^{(i)} > V_1)(n-1)\alpha. \end{aligned}$$

Thus,

$$P(S_n^{(i)} > N_i) \quad = \quad \begin{cases} 0 & \text{i = I, II} \\ p^{2k}(n-1)\alpha/(1-(1-p)p^k) & \text{i = III} \\ p^{k+1}(n-1)\alpha/(1-p(1-p)) & \text{i = IV} \end{cases}$$

$$\longrightarrow \; 0 \qquad \text{as } n \rightarrow \infty.$$

This completes the proof.

Corollary 2 *Let $\{X_i\}$ be a sequence of i.i.d. Bernoulli random variables, with $P(X_i = 1) = p = 1 - P(X_i = 0)$, if $np^k \to \lambda > 0$, as $n \to \infty$, the distributions of $N_1, \ldots, N_{IV}$ all converge to the same Poisson with parameter λ.*

It follows from Corollary 2 that for a sequence of i.i.d. or asymptotically independent Bernoulli random variables, the limiting distributions of the number of occurences of consecutive k successes are all Poisson. Furthermore, for all moderate values of p, say $\leq 1/2$ and of k, say ≥ 5, the Poisson distribution provides very good approximation for all the distributions of the number of occurrences of consecutive k successes.

Godbole (1991, Theorems 1 and 2) obtained the limiting distributions for N_{III} and N_{IV} identical to corollary 2). As we noted earlier, the study of a consecutive k-out-of-n:F system in reliability is identical to the study of the type II way of counting the number of "failure" runs of length k. Papastavridis (1987), and Chao and Fu (1989) showed that the reliability of a consecutive k-out-of-n:F system converges to $e^{-\lambda}$ if $nq^k \to \lambda > 0$ as $n \to \infty$. Their results are special case of Corollary 2), interchanging q = 1-p with p. A more general result along the lines was obtained recently by Fu (1993, theorem 2.1). Fu's result is the same as Corollary 2) for type II way of counting.

An extension of the consecutive k–out–of–n: F system is "m–consecutive–k–out–of–n:F system". It is a consecutive k–out–of–n: F system with $m-1$ identical back-up systems so that the whole system fails if and only if m or more of the systems fail. Fu (1993, Theorem 2.2) proved that the reliability of such a system converges to $\sum_{x\leq m-1} e^{-\lambda}\lambda^{x}/x!$ if $n(1-p)^{k} \rightarrow \lambda > 0$ as $n \rightarrow \infty$. Again, this result is an immediate consequence of corollary 2)

To test the hypothesis whether a binary sequence of length n is randomly arranged, one often uses the statistic R, the total number of runs. (See Mood (1940).) For k=1, $N_{II} = R/2$, if R is even, and $N_{II} = (R+1)/2$, if R is odd and $X_1 = 1$, and $N_{II} = (R-1)/2$, if R is odd and $X_1 = 0$. Thus if n_1, the number of 1 in the sequence ($X_1, X_2, \ldots, X_n$) is small relative to n, and the asymptotical distribution H of R is

$$H(x) = \begin{cases} \exp(-\lambda)\lambda^{k}/k! & \text{for } x = 2k \\ p\lambda^{k+1}e^{-\lambda}/(k+1)! + (1-p)\lambda^{k-1}e^{-\lambda}/(k-1)! & \text{for } x = 2k+1 \end{cases}$$

where $\lambda = n_1$ and $p = n_1/n$.

Chapter 3

Discussions

The results obtained in this thesis can be easily extended to many other problems. As an illustration consider two examples as follows:

A. A combination padlock with a circular dial can be unlocked only if the dial is turned, from the center, say, to the left k_1 times and then to the right k_2 times. What is the probability of unlocking the padlock if the dial is turned left and right in the random manner?

B. In a simplified genetics set up, consider a process of random mating in which there exists a certain arrangemant of genes of two types, A and B in the chromosomes of each cell of an offspring. A certain arrangement of genotypes, say, k_1 of A–type followed by k_2 of B–type would reflect a certain characteristic of the offspring. What would be the probability of observing such a characteristic in a large population?

These two examples lead Huang and Tsai (1989) to define a "binomial distribution of order $(k_1,\ k_2)$". Their distribution is an extension of Hirano's binomial distribution of order k, because by letting one of k_1 or k_2 equal 0, theirs is reduced to Hirano's. (See Hirano 1984.) In the combination padlock example and take $k_1 = 3$ and $k_2 = 4$. As anyone who has ever tried a combination padlock knows, the padlock can be unlocked only if is turned exactly three times to the left and then exactly four times to the right. If it is turned three times to the left and the five times to the right, it will not do. On the other hand, if one wants to reach a telephone number such as 848-3233 but mistakenly dials two extra digits such as 848-323333, it still works.

Therefore "a run of k_1 F and then k_2 S" can mean two things: One is that "a string of consecutive k_1 F, proceded by an S, is followed by another string of consecutive k_2 S, which procedes an F" and the other is that "a string of consecutive k_1 F, proceded by an S, is followed by k_2 or more S".

Evidently the first kind of runs corresponds to the type I runs while the second kind corresponds to the type II runs. We shall denote the number of occurrences of the first kind by M_1 and the second by M_2, and call the distribution of M_1 *the type I bionomial distribution of order* $(k_1,\ k_2)$, and the distribution of M_2 *the type II binomial distribution of order* $(k_1,\ k_2)$. Huang and Tsai's distribution is of the type II.

To find the limit of the type I binomial distribution of order $(k_1,\ k_2)$ we follow the approach taken above: define

$$W_i^{(I)} = I(V_i = k_1, U_i = k_2), \tag{3.1}$$

so that

$$P(W_i^{(I)} = 1) = \{(1-\alpha)^{k_1-1}\alpha\}\{(1-\beta)^{k_2-1}\beta\}$$

for all $i \geq 1$. Using (3.1), it follows that:

A). Under the Markov Bernoulli structure (1.1), if $n\alpha^2(1-\alpha)^{k_1-1} \to \lambda$, as $n \to \infty$, the p.g.f. of the type I binomial distribution of order $(k_1,\ k_2)$ converges to $\Phi_I(t) = e^{-\lambda\theta(t-1)}$, where $\theta = (1-\beta)^{k_2-1}\beta$, which is the p.g.f. of the Poisson distribution with parameter $\lambda\theta$.

B). Under the completely stationary structure, the limit of the type I binomial distribution of order $(k_1,\ k_2)$ is Poisson with parameter $\lambda(1-\pi)^3\pi^{k_2-1}$, if $np^2\{1-(1-\pi)p\}^{k_1-1} \to \lambda$, as $n \to \infty$.

C). Under the i.i.d. structure, the limit of the type I Binomial distribution of order $(k_1,\ k_2)$ is Poisson with parameter λ, if $np^{k_2+1}(1-p)^{k_1} \to \lambda$ as $n \to \infty$.

For finding the limiting distribution of the Type II binomial distribution of order $(k_1,\ k_2)$, we define

$$W_i^{(II)} = I(V_i = k_1, U_i \geq k_2), \tag{3.2}$$

so that

$$P(W_i^{(II)} = 1) = (1-\alpha)^{k_1-1}\alpha(1-\beta)^{k_2-1}$$

for all $i \geq 1$

Using (3.2), we can verify that:

A'). Under the Markov Bernoulli structure (1.1), if $n\alpha^2(1-\alpha)^{k_1-1} \to \lambda$, as $n \to \infty$, the p.g.f. of the type II binomial distribution of order (k_1, k_2) converges to $\Phi_I(t) = e^{-\lambda\theta(t-1)}$, where $\theta = (1-\beta)^{k_2-1}$, which is the p.g.f. of the Poisson distribution with parameter $\lambda\theta$.

B'). Under the completely stationary structure, the limit of the type II binomial distribution of order (k_1, k_2) is Poisson with parameter $\lambda(1-\pi)^2\pi^{k_2-1}$, if $np^2\{1-(1-\pi)p\}^{k_1-1} \to \lambda$, as $n \to \infty$.

C'). Under the i.i.d. structure , the limit of the type II binomial distribution of order $(k_1,\ k_2)$ is Poisson with parameter λ, if $np^{k_2}(1-p)^{k_1} \to \lambda$ as $n \to \infty$.

The result in C') coincides with Corollary 1 in Huang and Tsai (1991). The two lemmas in section 2 and 3 apply also to the derivation of the limiting distributions of M_1 and M_2 . Therefore in A) , B) and C) the limit conditions imply "$\alpha \to 0$ and $p \to 0$, as $n \to \infty$".

References.

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

Aki, S. Kuboki, H. and Hirano, K. (1984). On discrete distributions of order k, *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. *The Canadian Joural of Statistics,* 10, 225–231.

Buhler, W.J. (1989). probabilistic derivation for the limit of the Markov binomial distribution. *Math. Scientist.* 14, 117–119.

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

Chao, M.T. and Fu, J.C. (1991). The reliability of large series system under Markov structure. *Advances in Applied Probability*, 23, 894-908.

Chiang, D.T. and Niu, S.C. (1981). Reliability of consecutive k out of -n: F system. *IEEE Transactions on Reliability*, R-33, 87-9.

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.

De Moivre, A. (1967). *The Doctrine of Chances: or A Method of Calculating the Probabilities of Events in Play.* Chelesea Publishing Co., New York.

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.

Dobrusin, R.L. (1953). Limit theorems for a Markov chain of two states. *Izvestia Akad. Nauk. S.S.S. R. Ser. Mat.*, 17, 291–330.

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 of 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 – Theorey 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 applications,* 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, Dordreecht.

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

Huang, W.T. and Tsai, C.S. (1991). On a modified binomial distribution of order k. *Statist. Probab. Lett.* 11, 125–131.

Koopman, B.O. (1950) A generalization of Poisson's distribution for Markov chains. *proc. Natl. Academy of Sc.* 36, 202-207.

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

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

Ling, K.D. (1989). A new class of negative binomial distributions of order k, *Statistics and Probability letters,* 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. (1987). A limit theorem for reliability of a consecutive-k-out-of-n:F system. *Adv. Appl. Probab.* 19, 746-748.

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 Philliou, 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 and Probability letters,* 4, 101–105.

Rajarshi, M.B. (1974). Success runs in a two-state Markov chain, *Journal of Applied Probability,* 11, 119–192.

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. *Math. Scientist,* 14, 38–49.

Wang, Y.H. (1992). Approximating the k-th order two-state Markov Bernoulli chains, *J. Appl. Prob.* 29, 861-868.

Wang, Y.H. and W.J. Buhler. (1991) Renewal process proof for the limit of the Markov binomial distribution. *Math. Scientist.* 16, 66-68.

Wang, Y.H. and Liang, Z.Y. (1993). The probability of occurrences of runs of length k in n Markov Bernoulli trials. *Math. Scientist.* 18, 105–112

Yule, G.U. (1944). *The statistical study of literary style.* Cambridge University. Press.