The Annals of Applied Probability
1997, Vol. 7, No. 3, 666–678

EXPLICIT DISTRIBUTIONAL RESULTS IN PATTERN FORMATION

BY V. T. STEFANOV[1] AND A. G. PAKES

University of Western Australia

A new and unified approach is presented for constructing joint generating functions for quantities of interest associated with pattern formation in binary sequences. The method is based on results on exponential families of Markov chains. The tools we use are not only new for this area; they seem to be the *right* approach for deriving general explicit distributional results.

1. Introduction. The concept of runs, and more generally patterns, have been used in various areas: hypothesis testing, reliability theory, DNA sequencing [see Bishop, Williamson and Skolnick (1983) and the references therein], computer science and others. In the present paper we derive explicitly the distributions (joint, in general) of various quantities associated with the time of first reaching an arbitrary and fixed pattern of 0's and 1's, although the methodology we use allows much wider applicability. The observed sequences are derived from either independent Bernoulli trials or some dependent trials, including Markov dependent ones. The quantities of interest are the time of first seeing the pattern itself and the counts of appearances of all of its subpatterns until seeing the pattern for the first time. Throughout the paper we understand a subpattern to be any initial substring of its elements (e.g., $x_1x_2x_3$ in the pattern $x_1x_2\cdots x_5$). We will derive the joint Laplace transform of these random quantities. Therefore, we will have the Laplace transform of any linear transformation of these; note that linear combinations of these random quantities produce, for example, the counts of one symbol runs of an arbitrary length and of a length not smaller than a preassigned number and other quantities that might be of interest.

Relevant literature is widely scattered, with overlaps and rediscoveries not at all uncommon. We intend publishing a survey elsewhere, and here we mention just a few sources on occurrence of patterns. Pioneers such as De Moivre and Laplace solved problems about success runs, and Feller (1950) showed how recurrent event theory could be deployed in a systematic manner. The key to handling complex patterns was provided by Conway's *leading numbers*. See Gardner (1988) for a popular account. Formalization of this notion arose almost simultaneously (in various guises) at the hands of Li (1980) (using optional stopping), Gerber and Li (1981) (using Markov chains), Guibas and

Received February 1996; revised January 1997.

[1]Research partially supported by a grant from the Faculty of Engineering and Mathematical Sciences of UWA.

AMS 1991 *subject classifications*. Primary 60E10, 60J20; secondary 60G40, 62H10, 68R99.

Key words and phrases. Run, Markov chain, stopping time, exponential family.

Odlyzko (1981) (using elementary arguments) and Blom and Thorburn (1982). The last makes connections with Markov renewal theory, and this is systematically exploited to obtain very general results for Markov dependent trials by Biggins and Cannings (1987). Chryssaphinou and Papastavridis (1990) independently considered Markov dependent trials by suitably extending the methodology of Guibas and Odlyzko (1981).

Gerber and Li (1981) [cf. also Li (1980)] have found an expression for the Laplace transform of the time to first occurrence of a given pattern in repetitive drawings from a finite set of symbols (finite alphabet). Other related references on the time of first seeing a pattern are Guibas and Odlyzko (1980), Breen, Waterman and Zhang (1985) and Biggins and Cannings (1987). Limiting distribution results for counts of appearances of a given pattern in alphabetical sequences can be found in Chryssaphinou and Papastavridis (1988), Godbole and Schaeffner (1993) and Schbath (1995). Recently Aki and Hirano (1994, 1995) derived joint distributional results for the time of first seeing a run of successes and all its subpatterns (being also runs of successes) in binary sequences.

In the present paper we illustrate a new methodology by extending the results of Aki and Hirano (1994, 1995) to more general patterns. Our methodology is based on first imbedding the problem into a more general one for an appropriate finite-state Markov chain with one absorbing state, and second, treating that chain by the tools of exponential families. The first step of imbedding the problem into a similar one for Markov chains is natural and used in earlier as well as many recent treatments, for example, Li (1980), Gerber and Li (1981), Fu and Koutras (1994), Banjevic (1994). More specifically, the process of reaching a pattern can be modelled by a Markov chain with the states recording the progress towards achieving it, with actually reaching it being an absorbing state. Clearly this approach is applicable to reaching patterns formed from finite alphabets when the trials are independent or Markov-dependent. However our second step based on exponential families is new for this area. It allows a routine derivation of the Laplace transform of a whole collection of variables of interest. A brief description of the exponential family technology follows. In a Markov chain stopped at a finite random time τ let $N_{i,j}(\tau)$ be the number of the one-step transitions from i to j up to time τ. Clearly the sequential likelihood function is (with suitable conventions to cover $p_{i,j} = 0$)

$$\prod_{i,j} p_{i,j}^{N_{i,j}(\tau)} = \exp\left(\sum_{i,j} N_{i,j}(\tau) \ln p_{i,j} \right)$$

and we have a general exponential family. If the chain has k states, there are k linear constraints on the $p_{i,j}$, namely $\sum_j p_{i,j} = 1$, for each i. So the parameter space is $(k^2 - k)$-dimensional in general. For a finite stopping time τ there might be linear constraints on the $N_{i,j}$ as well. For most stopping times, there are either no linear constraints or fewer than k. Then the random variables in the exponent above number more than the free parameters.

Such exponential families are called curved exponential families. However, for some Markov chains there are a few finite stopping times for which there are exactly k linear constraints on the $N_{i,j}$. This issue has been discussed in Stefanov (1991) for ergodic Markov chains. Also the time to absorption in a chain with one absorbing state is a stopping time that possesses this nice property (see Remark 1 in Section 2). So the number of linearly independent $N_{i,j}$ is equal to the number of free parameters. Such an exponential family is called a noncurved exponential family. For rigorous definitions of a curved and a noncurved exponential family and their basic properties one may refer to Barndorff-Nielsen (1978, 1980) and Brown (1986). For noncurved cases, standard theory identifies the Laplace transform of the linearly independent $N_{i,j}$. More specifically, after a suitable reparametrization and denoting the linearly independent $N_{i,j}$ by $X_1, X_2, \ldots$, the noncurved exponential family takes the form

$$\exp\left(\sum_i \theta_i X_i + \varphi(\theta)\right)$$

and the Laplace transform of the X's is $\exp(\varphi(\theta) - \varphi(\theta + s))$. So this provides us with an implicit formula for the Laplace transform of the variables of interest. If the absorbing Markov chain we use to model our problem has a *relatively simple* transition probability matrix, we have the possibility of an explicit formula. In particular, this is the case for various patterns formed by binary sequences. The aim of this article is to promote this technology through applying it to a relatively simple example. On the other hand, the explicit expressions we derive here are general enough to cover all existing results in this direction for binary sequences as well as providing closed explicit expressions for some patterns which were not available earlier. Also this technology is applicable in other areas where similarly patterned Markov chains are used as underlying models. Moreover, the revealed noncurved exponential structure of the random quantities of interest leads to explicit limit results for the counts of appearances of the pattern and all of its subpatterns via an appropriate extension of the results of Stefanov (1995).

The paper is organized as follows. In Section 2 we introduce a special finite-state Markov chain with one absorbing state. Explicit distributional results are found when absorption occurs. These embrace explicit distributional results in pattern formation when independent or some dependent (including Markov dependent) Bernoulli trials are performed. In Section 3 we discuss briefly the case of first seeing a fixed number of consecutive successes; this is the case that has been extensively studied in recent literature.

2. General theory. Let $\{Z(t)\}_{t\geq 0}$, $t \in \{0, 1, 2, \ldots\}$ be a homogeneous $(n+2)$-state Markov chain defined as follows. The set of states is $I = \{0, 1, \ldots, n+1\}$. Assume that

$$n + 1 = k_1 + k_2$$

for some preassigned positive integers k_1, k_2. Let

$$0 < p_i,\ q_i < 1 \quad \forall\, i = 0, 1, \ldots, n,$$
$$p_i + q_i = 1,$$

and let

$$I_0 =_{\text{def}} \{i \in I \colon i = 0, 1, \ldots, k_1 - 1\}$$
$$I_1 =_{\text{def}} I \setminus \big(I_0 \cup \{k_1, n+1\}\big) = \{k_1 + 1, \ldots, n\}.$$

The one-step transition probabilities p_{ij} are defined as follows:

$$
\begin{aligned}
p_{i,\,i+1} &= p_i, && i = 0, 1, \ldots, n,\\
p_{i,\,0} &= q_i, && i \in I_0,\\
p_{k_1,\,k_1} &= q_{k_1}, &&\\
p_{i,\,1} &= q_i, && i \in I_1,\\
p_{n+1,\,n+1} &= 1, &&\\
p_{i,\,j} &= 0 \quad \text{otherwise.} &&
\end{aligned}
\tag{1}
$$

Therefore, the one-step transition probability matrix has the following form:

$$
\begin{pmatrix}
q_0 & p_0 & 0 & \cdots & \ldots & \ldots & \ldots & 0\\
q_1 & 0 & p_1 & 0 & \ldots & \ldots & \ldots & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
q_{k_1-1} & 0 & \ldots & 0 & p_{k_1-1} & 0 & \ldots & 0\\
0 & 0 & 0 & \ldots & q_{k_1} & p_{k_1} & 0 & 0\\
0 & q_{k_1+1} & 0 & \ldots & \ldots & 0 & p_{k_1+1} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
0 & q_{k_1+k_2-1} & 0 & \ldots & \ldots & \ldots & 0 & p_{k_1+k_2-1}\\
0 & 0 & \ldots & \ldots & \ldots & \ldots & 0 & 1
\end{pmatrix}
$$

Assume also that

$$P(Z(0) = 0) = 1.$$

Denote by $N_{i,\,j}(t)$ the number of the one-step transitions from state i to state j up to time t. Let

$$\tau =_{\text{def}} \inf\{t \colon Z(t) = n+1\}$$
$$= \inf\{t \colon N_{n,\,n+1}(t) = 1\}.$$

That is, τ is the time to absorption in state $n+1$. The motivation for discussing the above Markov chain comes from the following particular case.

EXAMPLE 1. Let

$$
\begin{aligned}
p_i &= p, && i \in I_0,\\
p_i &= q, && i \in I_1,\ i = k_1,\\
q_i &= q, && i \in I_0,\\
q_i &= p, && i \in I_1,\ i = k_1,
\end{aligned}
\tag{2}
$$

where p_i, q_i were given in (1) and $p + q = 1$. In this case the stopping time τ marks the moment of first seeing the following pattern of 0's and 1's, by independent Bernoulli trials with probability of success p:

$$\frac{1\cdots 1}{k_1}\quad\frac{0\cdots 0}{k_2},$$

where k_i denotes the number of digits in the ith block. In fact it is easy to see that an entry to the state i, $0 \le i \le k_1 + k_2$, records an attainment of the subpattern consisting of the first i digits of the above pattern; consequently the entry to the absorbing state means reaching the pattern.

EXAMPLE 2 (The Markov-dependent extension of Example 1). Let

$$
\begin{aligned}
p_0 &= p^{(0)}, & q_0 &= q^{(0)},\\
p_1 &= p, & q_1 &= q,\\
&\vdots & &\vdots\\
p_{k_1-1} &= p, & q_{k_1-1} &= q,\\
p_{k_1} &= q, & q_{k_1} &= p,\\
p_{k_1+1} &= q^{(0)}, & q_{k_1+1} &= p^{(0)},\\
&\vdots & &\vdots\\
p_{k_1+k_2-1} &= q^{(0)}, & q_{k_1+k_2-1} &= p^{(0)},
\end{aligned}
$$

where $q^{(0)} = 1 - p^{(0)}$, $q = 1 - p$ and p is the probability of getting a success given a success in the previous trial, while $p^{(0)}$ is the probability of getting a success given a failure in the previous trial. The initial state is 0.

Generally, to each pattern of 0's and 1's,

$$\frac{1\cdots 1}{k_1}\quad\frac{0\cdots 0}{k_2}\quad\frac{1\cdots 1}{}\cdots\frac{0\cdots 0}{k_m},$$

we can introduce a Markov chain with $n + 2$ states ($n + 1 = k_1 + \cdots + k_m$) with the last state being an absorbing state, such that its hitting time is the moment of first seeing this pattern by independent and some dependent (including Markov dependent) Bernoulli trials. Each row (except for the last one) of the transition probability matrices of these Markov chains has exactly two nonzero entries; one of them is always $p_{i,\,i+1},\ i = 1, \ldots, n$ and the other one, say $p_{i,\,j}$, depends on the particular pattern. Note that in the latter case $j \le i$ and j is the length of the longest achieved subpattern when the chain leaves state i.

For the sake of clarity, we shall demonstrate our methodology on the Markov chain introduced at the beginning of this section; this covers the case of a pattern consisting of two blocks only. The methodology allows the derivation of explicit solutions for any fixed pattern of 0's and 1's.

Consider the chain $\{Z(t)\}_{t\ge 0}$ introduced above. For the sake of brevity we shall denote $N_{i,\,j}(\tau)$ by $N_{i,\,j}$ and even more briefly $N_{j-1,\,j}(\tau)$ by N_j. Then the Radon–Nikodym derivative of the measure generated by the chain on the time interval $[0, \tau]$, with respect to some σ-finite measure is [Stefanov (1991)]:

$$\exp\left\{\sum_{i=0}^{n} N_{i+1}\ln p_i + \sum_{i=0}^{k_1-1} N_{i,\,0}\ln q_i + N_{k_1,\,k_1}\ln q_{k_1} + \sum_{i=k_1+1}^{n} N_{i,\,1}\ln q_i\right\}. \tag{3}$$

In view of Stefanov's (1991) results (Proposition 1; cf. also Remark 1 below), the family given by (3) is a noncurved exponential family of order $(n+1)$.

REMARK 1. Actually, Stefanov's (1991) paper treats ergodic Markov chains only. However, it is straightforward to extend his results to absorbing Markov chains with one absorbing state. Now the stopping time representing the time of absorption is the one for which a noncurved exponential family is obtained. Observe that in order to obtain a noncurved exponential family it is essential that the evolution of the chain begin at a fixed state and terminate at a fixed state.

Furthermore, we shall find a minimal canonical representation of the family given by (3). Assume first that $k_2 > 1$. In view of Stefanov (1991) [see (3) on page 355] we have the following linear relationships between $N_{i+1},\ i = 0, \ldots, n,\ N_{i,\,0},\ i \in I_0$ and $N_{i,\,1},\ i \in I_1$:

$$\begin{aligned}
N_{i,\,0} = N_{i-1,\,i} - N_{i,\,i+1} &= N_i - N_{i+1} \quad \text{if } i = 2, 3, \ldots, k_1 - 1,\\
N_{k_1} - N_{k_1+1} &= 0,\\
N_{i,\,1} = N_{i-1,\,i} - N_{i,\,i+1} &= N_i - N_{i+1} \quad \text{if } i = k_1 + 1, \ldots, n,\\
N_{n,\,n+1} &= N_{n+1} = 1,\\
N_{0,\,1} &= N_1 = 1 + \sum_{i=1}^{k_1-1} N_{i,\,0},
\end{aligned} \tag{4}$$

with the convention that $\sum_{i=1}^{j}(\cdot) = 0$ if $j = 0$. This convention is applicable throughout the paper unless otherwise explicitly stated.

Specifically, these are derived by counting the number of entries to and exits from a state, up to time τ. The last equality can also be expressed as

$$
\begin{aligned}
N_1 &= 1 + N_{1,0} + \sum_{i=2}^{k_1-1} (N_i - N_{i+1}) \\
&= 1 + N_{1,0} + N_2 - N_{k_1}.
\end{aligned} \tag{5}
$$

Hence

$$N_{1,0} = N_1 - N_2 + N_{k_1} - 1. \tag{6}$$

Now, replacing N_{k_1} by N_{k_1+1}, $N_{1,0}$, $N_{i,0}$ $(i = 2, \ldots, k_1 - 1)$ and $N_{i,1}$ $(i = k_1 + 1, \ldots, n)$ in (3) by the expressions in (4) and (6), we get the following representation of this exponential family, where ρ_i stands for $p_i q_{i+1}/q_i$:

$$
\begin{aligned}
&\exp\bigg\{ N_{00} \ln q_0 + N_1 \ln(p_0 q_1) + \sum_{i=1}^{k_1-2} N_{i+1} \ln \rho_i \\
&\qquad + N_{k_1,k_1} \ln q_{k_1} + N_{k_1+1} \ln\left(\frac{p_{k_1-1} p_{k_1} q_{k_1+1} q_1}{q_{k_1-1}} \right) \\
&\qquad + \sum_{i=k_1+1}^{n-1} N_{i+1} \ln \rho_i + N_{n+1} \ln \frac{p_n}{q_n} - \ln q_1 \bigg\}.
\end{aligned} \tag{7}
$$

Since $N_{n+1} = 1$, one of the minimal canonical representations of the non-curved exponential family of size $n + 1$ given by (7) is

$$\exp\left\{ \sum_{i=0}^{n} \theta_i \ X_i + \varphi(\theta) \right\}, \qquad \theta = (\theta_0, \ldots, \theta_n) \in \Theta \subset R^{n+1}, \tag{8}$$

where $X_0 = N_{00}$, $X_i = N_i$, $i \neq k_1$, $X_{k_1} = N_{k_1,k_1}$, Θ is an open set and

$$
\begin{aligned}
\theta_0 &= \ln q_0, \\
\theta_1 &= \ln(p_0,\ q_1), \\
\theta_{k_1} &= \ln q_{k_1}, \\
\theta_{k_1+1} &= \ln(p_{k_1-1} p_{k_1} q_{k_1+1} q_1 / q_{k_1-1}), \\
\theta_i &= \ln \rho_{i-1} \quad \text{otherwise}
\end{aligned} \tag{9}
$$

and $\varphi(\theta) = \ln[p_n/(q_n q_1)]$, whose explicit form will be found below. Let us introduce the following notation:

$$
\begin{aligned}
\sigma_1 &:=_{\text{def}} 1 - \exp(\theta_0) - \exp(\theta_1) - \exp(\theta_1 + \theta_2) - \cdots \\
&\qquad - \exp(\theta_1 + \cdots + \theta_{k_1-1}), \\
\sigma_{2,1} &:=_{\text{def}} -\exp\left(\sum_{i=1}^{k_1-1} \theta_i \right) \left[\sum_{l=1}^{k_2-1} \exp\left(\sum_{j=1}^{l} \theta_{k_1+j} \right) \right],
\end{aligned} \tag{10}
$$

$$\sigma_{2,2} :=_{\text{def}} -\exp\left(\sum_{i=1}^{k_1-1} \theta_i\right)\left[\sum_{l=1}^{k_2-2} \exp\left(\sum_{j=1}^{l} \theta_{k_1+j}\right)\right],$$

$$\nu_i :=_{\text{def}} \exp\left(\sum_{j=1,\ j\neq k_1}^{i} \theta_j\right), \qquad i = 1, 2, \dots .$$

From (9) we find that

$$q_0 = e^{\theta_0}, \qquad p_0 = 1 - e^{\theta_0},$$
$$q_1 = \frac{e^{\theta_1}}{1 - e^{\theta_0}}, \qquad p_1 = \frac{1 - e^{\theta_0} - e^{\theta_1}}{1 - e^{\theta_0}} \tag{11}$$

and after further calculation we find that

$$\begin{aligned} q_{k_1-1} &= \frac{\nu_{k_1-1}}{\sigma_1}, \\ q_{k_1} &= \exp(\theta_{k_1}), \\ q_{k_1+1} &= \frac{\nu_{k_1+1}}{p_{k_1} q_1 \sigma_1}, \\ q_n &= \frac{\nu_n}{p_{k_1} q_1 \sigma_1 + \sigma_{2,2}}. \end{aligned} \tag{12}$$

From (12) we find that

$$p_n = \frac{p_{k_1} q_1 \sigma_1 + \sigma_{2,1}}{p_{k_1} q_1 \sigma_1 + \sigma_{2,2}}$$

because

$$\sigma_{2,1} = \sigma_{2,2} - \nu_n,$$

whence

$$\exp(-\varphi(\theta)) = \frac{q_1 \nu_n}{p_{k_1} q_1 \sigma_1 + \sigma_{2,1}}. \tag{13}$$

Here p_{k_1} and q_1 are functions of θ given explicitly above and the other quantities in (13) are also functions of θ, which were introduced in (10). In the case when $k_2 = 1$ the expression given in (7) takes the following form:

$$\exp\left\{N_{00} \ln q_0 + N_1 \ln(p_0 q_1) + \sum_{i=1}^{k_1-2} N_{i+1} \ln \rho_i \right.$$
$$\left. + N_{k_1,k_1} \ln q_{k_1} + \ln \frac{p_{k_1-1} p_{k_1}}{q_{k_1-1}}\right\}.$$

Then $\varphi(\theta) = \ln(p_{k_1-1} p_{k_1} / q_{k_1-1})$ and along the same lines we find that

$$\exp(-\varphi(\theta)) = \frac{\sum_{i=1}^{k_1-1} \exp(\theta_i)}{p_{k_1} \sigma_1},$$

with the convention that $\sum_{i=1}^{j} (e^{(\cdot)}) = 1$ if $j = 0$.

From well-known properties of noncurved exponential families [cf. Brown (1986) or Barndorff-Nielsen (1978), page 114], the moment generating function of the random vector $X = (X_0, X_1, X_2, \dots, X_n)$ is

$$MGF_X(t) = \exp[\varphi(\theta) - \varphi(\theta + t)], \tag{14}$$

where $t =_{\text{def}} (t_0, t_1, \dots, t_n)$, $\theta = (\theta_0, \theta_1, \dots, \theta_n)$ and $\theta + t \in \Theta$. From (13) we find an explicit expression for $\exp(\varphi(\theta + t))$. In the latter we replace the θ_i's by their expressions in terms of p_i's and q_i's [cf. (9) above] to get ($t(i)$ stands for $t_1 + t_2 + \cdots + t_i$):

$$\exp(\varphi(\theta + t)) = \frac{PQA_1 + A_2}{Q \exp(\sum_{i=1,\, i\neq k_1}^{n} t_i) p_0 p_1 \dots p_{n-1} q_n q_1}, \tag{15}$$

if $k_2 > 1$, and

$$\exp(\varphi(\theta + t)) = \frac{PA_1}{\exp(\sum_{i=1}^{k_1-1} t_i) p_0 p_1 \dots p_{k_1-1} q_{k_1}}, \tag{16}$$

if $k_2 = 1$, where

$$A_1 = 1 - q_0 e^{t_0} - p_0 q_1 e^{t_1} - p_0 p_1 q_2 e^{t(2)} - \cdots - p_0 p_1 \dots p_{k_1-2} q_{k_1-1} \exp(t(k_1 - 1)),$$

$$\begin{aligned} A_2 = -q_1 \exp(-t_{k_1})\big(&p_0 p_1 \dots p_{k_1} q_{k_1+1} \exp(t(k_1 + 1)) + \cdots \\ &+ p_0 p_1 \dots p_{k_1+k_2-2} q_{k_1+k_2-1} \exp(t(k_1 + k_2 - 1))\big), \end{aligned}$$

$$P = 1 - \exp(\theta_{k_1} + t_{k_1}) = 1 - q_{k_1} \exp(t_{k_1})$$

and

$$Q = \frac{\exp(\theta_1 + t_1)}{1 - \exp(\theta_0 + t_0)} = \frac{p_0 q_1 \exp(t_1)}{1 - q_0 \exp(t_0)}.$$

Then from (14), (15) and (16) we get the following.

THEOREM 1. *The moment generating function of X is*

$$MGF_X(t) = \frac{p_0 p_1 \dots p_n \exp(\sum_{i=1,\, i\neq k_1}^{n} t_i)}{(1 - q_{k_1} \exp(t_{k_1}))A_1 + A_2/Q}, \tag{17}$$

where Q, A_1 and A_2 are given above and $A_2 = 0$, if $k_2 = 1$.

The interpretation of the X's in terms of quantities related to first seeing a pattern is that X_i, $(1 \le i < k_1$ and $k_1 < i \le n)$ is the count of appearances of the subpattern consisting of the first i digits of the pattern. Also

$$\begin{aligned}\tau &= X_0 + 2X_1 + X_2 + \cdots + X_{k_1} + 3X_{k_1+1} + X_{k_1+2} + \cdots + X_n - 1\\ &= \sum_{i=0}^{n} X_i + X_1 + 2X_{k_1+1} - 1\end{aligned}$$

because

$$\tau = \sum_{i,\, j} N_{i,\, j}(\tau)$$

and

$$\sum_{i,\, i\neq 0} N_{i,\,0}(\tau) = N_1 - 1, \qquad \sum_{i,\, i>1} N_{i,\,1}(\tau) = N_{k_1+1} - 1,$$
$$N_{k_1} = N_{k_1+1} \quad \text{and} \quad N_{n,\, n+1}(\tau) = 1.$$

EXAMPLE 1 (Continued). Consider the case of independent Bernoulli trials introduced in Example 1 above. In this case

$$A_2 = -p^{k_1+1} q \exp(-t_{k_1}) \left(\sum_{i=1}^{k_2-1} q^i \exp(t(k_1 + i)) \right),$$

if $k_1 > 1$, and

$$A_2 = -p^3 \exp(-t_1) \left(\sum_{i=1}^{k_2-1} q^i \exp(t(1+i)) \right),$$

if $k_1 = 1$; note that in the latter case $q_1 = p$. However, for both cases the expressions for A_2/Q have the same form; that is,

$$A_2/Q = -p^{k_1} \exp(-t_1 - t_{k_1})(1 - q\exp(t_0)) \left(\sum_{i=1}^{k_2-1} q^i \exp(t(k_1+i)) \right),$$

if $k_1 > 1$, and

$$A_2/Q = -p \exp(-2t_1)(1 - q\exp(t_0)) \left(\sum_{i=1}^{k_2-1} q^i \exp(t(1+i)) \right),$$

if $k_1 = 1$. Therefore, from (17) we have

$$\begin{aligned} &MGF_X(t)\\ &\quad = \left[p^{k_1} q^{k_2} \exp\left(\sum_{i=1,\, i\neq k_1}^{n} t_i \right) \right]\\ &\qquad \times \Bigg\{ (1 - p\exp(t_{k_1})) \left[1 - q\exp(t_0) - q\left(\sum_{i=1}^{k_1-1} p^i \exp(t(i)) \right) \right]\\ &\qquad\quad - p^{k_1} \exp(-t_1 - t_{k_1})(1 - q\exp(t_0)) \left(\sum_{i=1}^{k_2-1} q^i \exp(t(k_1+i)) \right) \Bigg\}^{-1}. \end{aligned}$$

EXAMPLE 2 (Continued). From (17) we have

$$MGF_X(t) = \left[p^{(0)} q p^{k_1-1} (q^{(0)})^{k_2-1} \exp\left(\sum_{i=1,\, i\neq k_1}^{n} t_i\right)\right] \times \left\{(1 - p\exp(t_{k_1}))\left[1 - q^{(0)}\exp(t_0) - qp^{(0)}\left(\sum_{i=1}^{k_1-1} p^{i-1}\exp(t(i))\right)\right] - qp^{(0)} p^{k_1-1} \exp(-t_1 - t_{k_1})(1 - q^{(0)}\exp(t_0)) \times \left(\sum_{i=1}^{k_2-1} (q^{(0)})^{i-1} \exp(t(k_1+i))\right)\right\}^{-1}.$$

3. The case $k_2 = 0$. The case $k_2 = 0$, which reduces the pattern to a run of k_1 1's, can be treated along the same lines with simplified algebra. Because of its central importance in recent literature [see Aki and Hirano (1995) and Mohanty (1994) and the references therein] we give a brief account of this case as well.

The one-step transition probability matrix of the corresponding Markov chain has the success run form

$$\begin{pmatrix} q_0 & p_0 & 0 & \dots & \dots & 0 \\ q_1 & 0 & p_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{k-1} & 0 & \dots & \dots & 0 & p_{k-1} \\ 0 & 0 & \dots & \dots & 0 & 1 \end{pmatrix}.$$

When $p_i = p$ $(i = 0, \dots, k-1)$, the stopping time τ represents the moment of first seeing a run of k successes in independent Bernoulli trials with probability of success p. Suitable selections of the p_i's cover the case of Markov dependent Bernoulli trials as well as the so-called binary sequences of order k [for the definition of the latter see Aki and Hirano (1994)].

Equations (3), (4) and (9) take the forms

$$\exp\left\{\sum_{i=0}^{k-1} N_{i+1}\ln p_i + \sum_{i=0}^{k-1} N_{i,0}\ln q_i\right\}, \tag{18}$$

where

$$\begin{aligned} N_{i,0} &= N_{i-1,i} - N_{i,i+1} = N_i - N_{i+1} \quad \text{if } i = 1, 2, \dots, k-1, \\ N_{k-1,k} &= N_k = 1, \\ N_{0,1} &= N_1 = 1 + \sum_{i=1}^{k-1} N_{i,0}, \end{aligned} \tag{19}$$

$$\begin{aligned} \theta_0 &= \ln q_0, \\ \theta_1 &= \ln(p_0 q_1), \\ \theta_i &= \ln \rho_i, \qquad i = 2, \dots, k-1, \end{aligned} \tag{20}$$

and $\varphi(\theta) = \ln[p_{k-1}/q_{k-1}]$. Along the same lines as above, with substantially simplified algebra, we derive the moment generating function of $N = (N_{00}, N_1, \dots, N_{k-1})$. The result is stated in the following.

THEOREM 2. *The moment generating function of N is*

$$MGF_N(t) = \frac{p_0 p_1 \dots p_{k-1} \exp(\sum_{i=1}^{k-1} t_i)}{A_1}$$

where A_1 is as given in the previous section with k_1 replaced by k.

There is a useful interpretation of the components of the vector N in terms of runs and their length. Since there are several different ways of counting success runs [see Fu and Koutras (1994)], we shall first recall these via an example. Let 0110011110 be a sequence of 1's (success) and 0's (failure). Then the run 11 appears once if Mood's (1940) counting is used, three times in Feller's (1950) counting, and four times in Ling's (1988) overlapping counting. Now it is straightforward to see that according to Mood's counting, N_i represents the number of success runs of length at least i in a sequence which is stopped after reaching k consecutive successes. It is easy to see that

$$\tau = N_{00} + 2N_1 + N_2 + N_3 + N_4 + \dots + N_{k-1}.$$

Also $N_i - N_{i+1}$ represents the number of runs of length exactly i. Of course, from the moment generating function of N, we find immediately the moment generating function of any linear transformation of N. In particular, we can find the moment generating function of the vector $(\tau, N_1 - N_2, \dots, N_{k-1} - N_k)$, that is, the vector consisting of the moment of first seeing consecutive k successes and numbers of success runs of length exactly i, $i = 1, \dots, k-1$, in Mood's counting, for each of the three cases (i) independent Bernoulli trials, (ii) Markov dependent Bernoulli trials and (iii) binary sequences of order k. By the same token, we can derive immediately the moment generating function of the same vector when any of the other above countings of runs is used; note that the number of runs of a fixed length, in either Feller's (1950) or Ling's (1988) counting, is a simple linear function of $N_1, \dots, N_{k-1}$. Therefore, all results presented by Aki and Hirano (1995) can be derived from Theorem 2 above.

Acknowledgment. We are grateful to both referees for their careful reading of the paper and comments which improved the presentation.

REFERENCES

AKI, S. and HIRANO, K. (1994). Distributions of numbers of failures and successes until the first consecutive k successes. *Ann. Inst. Statist. Math.* **46** 193–202.

AKI, S. and HIRANO, K. (1995). Joint distributions of numbers of success-runs and failures until the first consecutive k successes. *Ann. Inst. Statist. Math.* **47** 225–235.

BANJEVIC, D. (1994). Pattern recognition in Markov chains. In *Runs and Patterns in Probability* (A. P. Godbole and S. G. Papastavridis, eds.) 213–230. Kluwer, Dordrecht.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families*. Wiley, Chichester.

BARNDORFF-NIELSEN, O. (1980). Conditionality resolutions. *Biometrika* **67** 293–310.

BIGGINS, J. D. and CANNINGS, C. (1987). Markov renewal processes, counters and repeated sequences in Markov chains. *Adv. in Appl. Probab.* **19** 521–545.

BISHOP, D. T., WILLIAMSON, J. A. and SKOLNICK, M. H. (1983). A model for restriction fragment length distributions. *American Journal of Human Genetics* **35** 795–815.

BLOM, G. and THORBURN, D. (1982). How many random digits are required until given sequences are obtained? *J. Appl. Probab.* **19** 518–531.

BREEN, S., WATERMAN, M. S. and ZHANG, N. (1985). Renewal theory for several patterns. *J. Appl. Probab.* **22** 228–234.

BROWN, L. (1986). *Fundamentals of Statistical Exponential Families*. IMS, Hayward, CA.

CHRYSSAPHINOU, O. and PAPASTAVRIDIS, S. (1988). A limit theorem on the number of overlapping appearances of a pattern in a sequence of independent trials. *Probab. Theory Related Fields* **79** 129–143.

CHRYSSAPHINOU, O. and PAPASTAVRIDIS, S. (1990). The occurrence of a sequence of patterns in repeated dependent experiments. *Theory Probab. Appl.* **35** 167–173.

FELLER, W. (1950). *An Introduction of Probability Theory and Its Applications* **1**. Wiley, New York.

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

GARDNER, M. (1988). *Time Travel and Other Mathematical Bewilderments*. W. H. Freeman, New York.

GERBER, H. and LI, S-Y. R. (1981). The occurrence of sequence patterns in repeated experiments and hitting times in a Markov chain. *Stochastic Process. Appl.* **11** 101–108.

GODBOLE, A. P. and SCHAEFFNER, A. A. (1993). Improved Poisson approximation for word patterns. *Adv. in Appl. Probab.* **25** 334–347.

GUIBAS, L. J. and ODLYZKO, A. M. (1980). Long repetitive patterns in random sequences. *Z. Wahrsch. Verw. Gebiete* **53** 241–262.

GUIBAS, L. J. and ODLYZKO, A. M. (1981). String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory A* **30** 183–208.

LI, S-Y. R. (1980). A martingale approach to the study of occurrence of sequence patterns in repeated experiments. *Ann. Probab.* **8** 1171–1176.

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

MOHANTY, S. G. (1994). Success runs of length k in Markov dependent trials. *Ann. Inst. Statist. Math.* **46** 777–796.

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

SCHBATH, S. (1995). Compound Poisson approximation of word counts in DNA sequences. *ESAIM: Probability and Statistics* **1** 1–16.

STEFANOV, V. T. (1991). Noncurved exponential families associated with observations over finite state Markov chains. *Scand. J. Statist.* **18** 353–356.

STEFANOV, V. T. (1995). Explicit limit results for minimal sufficient statistics and maximum likelihood estimators in some Markov processes: exponential families approach. *Ann. Statist.* **23** 1073–1101.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WESTERN AUSTRALIA
NEDLANDS 6907, W.A.
AUSTRALIA
E-MAIL: stefanov@maths.uwa.edu.au
pakes@maths.uwa.edu.au