J. Japan Statist. Soc.
Vol. 37 No. 2 2007 285–298

JOINT DISTRIBUTIONS ASSOCIATED WITH COMPOUND PATTERNS IN A SEQUENCE OF MARKOV DEPENDENT MULTISTATE TRIALS AND ESTIMATION PROBLEMS

Kiyoshi Inoue* and Sigeo Aki**

Let Λ_i, $1 \leq i \leq \ell$ be simple patterns, i.e., finite sequences of outcomes from a set $\Gamma = \{b_1, b_2, \ldots, b_m\}$ and let Λ be a compound pattern (a set of ℓ distinct simple patterns). In this paper, we study joint distributions of the waiting time until the r-th occurrence of the compound pattern Λ, and the numbers of each simple pattern observed at that time in the multistate Markov dependent trials. We provide methods for deriving the probability generating functions of the joint distributions under two types of counting schemes (non-overlap counting and overlap counting) for the compound pattern Λ. Besides, the present work is useful in elucidating the primary difference between non-overlap counting and overlap counting. As applications, when Λ is a set of runs, the corresponding joint distributions are investigated and a practical example is mentioned. Also, the Chen-Stein approximation is derived for the waiting time distribution, and its asymptotic behaviour is discussed. Finally, we address the parameter estimation in the waiting time distributions of the compound pattern along with problems of identifiability.

Key words and phrases: Chen-Stein approximations, Markov chain, maximum likelihood estimate, multistate trials, non-overlap counting, overlap counting, parameter estimation, patterns, probability generating function, runs.

1. Introduction

The distribution theory of patterns has recently received attention in various areas of statistics and applied probability, for example, the reliability of engineering systems, hypothesis testing, continuity measurement in health care and quality control (see Antzoulakos (2001), Fu (1996), Inoue (2004), Fu and Lou (2003), Inoue and Aki (2002), Fu and Chang (2002) and Han and Hirano (2003)). Especially the waiting time distribution of patterns has been broadly used in a wide range of areas such as moving window detection, machine maintenance, start-up demonstration tests and matching in DNA sequence (see Chao *et al.* (1995), Shmueli and Cohen (2000), Ewens and Grant (2001) and Robin and Daudin (1999, 2001)).

Traditionally, the distributions of patterns were studied via combinatorial analysis. For example, Mood (1940) wrote: “The distribution problem is, of

Received June 30, 2006. Revised September 16, 2006. Accepted September 21, 2006.

*Faculty of Economics, Seikei University, 3-3-1 Kichijoji-Kitamachi, Musasino-shi, Tokyo 180-8633, Japan.

**Department of Mathematics, Faculty of Engineering, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan.

course, a combinatorial one, and the whole development depends on some identities in combinatory analysis". However, it is very difficult to find the appropriate combinatorial identities to derive the probability distributions. This perhaps is the main reason why many exact distributions of patterns remain unknown even for the simple case where the underlying sequence is identically and independently distributed (i.i.d.).

With the exception of Aki and Hirano's (1989) study, very little work has appeared in the statistical literature on the parameter estimation for the distributions of patterns. In parameter estimation, there are not any papers which treat the problems of identifiability. Since the estimation problems have so far been restricted to simple cases where the problems of identifiability do not arise, it has not been neccessary to consider whether the parameters are identifiable or not.

In this paper, departing from the traditional combinatorial approach, we present methods for deriving the distributions of patterns based on the method of conditional probability generating functions. The problems of identifiability are also discussed under additional constraints.

Let $Z_1, Z_2, \ldots$ be a time homogeneous Markov chain defined on the state space $\Gamma = \{b_1, b_2, \ldots, b_m\}$, transition probabilities $p_{b_i b_j} = P(Z_{t+1} = b_j \mid Z_t = b_i)$ for $t \geq 1$, $i, j = 1, 2, \ldots, m$ and initial probabilities $p_{b_i} = P(Z_1 = b_i)$ for $i = 1, 2, \ldots, m$.

According to Fu and Lou (2003) (see Fu and Chang (2002) and Fu (2001)), we will define a simple pattern and a compound pattern, respectively.

Definition 1.1. *We say that Λ is a simple pattern if Λ is composed of a specified sequence of length k; i.e. $\Lambda = (a_1, a_2, \ldots, a_k)$, $a_i \in \Gamma$, $1 \leq i \leq k$ (k, the length of the pattern is fixed, and the elements in the pattern are allowed to be repeated). Let Λ_1 and Λ_2 be two simple patterns with length k_1 and k_2 respectively. We say that Λ_1 and Λ_2 are distinct if neither is a subsequence (segment) of the other.*

Definition 1.2. *We say that Λ is a compound pattern if it is a set of $\ell (\geq 2)$ distinct simple patterns; i.e. $\Lambda = \{\Lambda_i : 1 \leq i \leq \ell\}$, where $\Lambda_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,k_i})$. When $\ell = 1$, we identify the compound pattern Λ with the simple pattern Λ_1.*

In the sequel, we assume that the length of the simple pattern is greater than 1. This is the most common situation in practice.

We define

$$T_r = \inf\{n : \text{number of trials required} \\ \text{to have } r \text{ simple patterns in total among } \Lambda\}$$

as the waiting time for the r-th ($r \geq 1$) occurrence of a compound pattern Λ. For Λ_i ($i = 1, 2, \ldots, \ell$), let $N(T_r : \Lambda_i)$ be the number of occurrences of Λ_i in $Z_1, Z_2, \ldots, Z_{T_r}$.

In Section 2, we investigate the joint distribution of $(T_r, N(T_r : \Lambda_1), \ldots, N(T_r : \Lambda_\ell))$ under two types of counting schemes (non-overlap counting and overlap counting) for the compound pattern Λ, and propose methods for deriving the joint probability generating function (p.g.f.) of $(T_r, N(T_r : \Lambda_1), \ldots, N(T_r : \Lambda_\ell))$. The approach departs from the traditional combinatorial approach and provides a very efficient computational tool, which is also useful in elucidating the primary difference between non-overlap counting and overlap counting. In Section 3, we study the special case where Λ is a set of runs of certain lengths. Some explicit closed-form expressions for joint p.g.f.'s are given, which are, to the best of our knowledge, new. Also, the Chen-Stein approximation for the waiting time distribution is provided, which is a useful tool for investigating its asymptotic behaviour. In Section 4, we address parameter estimation in the waiting time distributions of the compound pattern and discuss problems of identifiability.

2. Main results

In this section, we are going to study the joint distribution of $(T_r, N(T_r : \Lambda_1), \ldots, N(T_r : \Lambda_\ell))$ under non-overlap counting and overlap counting. Each one of the two counting schemes is treated separately.

For a simple pattern Λ_i $(i = 1, \ldots, \ell)$, let $\Lambda_{i,j} = (a_{i,1}, \ldots, a_{i,j})$, $1 \leq j \leq k_i - 1$. Then, for a compound pattern Λ, we define a set by

$$\Omega = \{\emptyset\} \bigcup \{\Lambda_{i,j} : 1 \leq i \leq \ell, 1 \leq j \leq k_i - 1\} \bigcup \Lambda.$$

Here, by relabelling the states in the set Ω, we rewrite Ω as

$$\Omega = \{\alpha_0, \alpha_1, \ldots, \alpha_s, \Lambda_1, \ldots, \Lambda_\ell\},$$

(convention: $\alpha_0 = \emptyset$, which is regarded as an empty subpattern of length 0)

where $\{\alpha_1, \ldots, \alpha_s\} = \{\Lambda_{i,j} : 1 \leq i \leq \ell, 1 \leq j \leq k_i - 1\}$ and α_ω is a subpattern of length x_{i_ω} i.e., $\alpha_\omega = (\alpha_{i_\omega,1}, \ldots, \alpha_{i_\omega,x_{i_\omega}})$ for $\omega = 1, 2, \ldots, s$.

2.1. Non-overlap counting

The joint p.g.f. of $(T_r, N(T_r : \Lambda_1), \ldots, N(T_r : \Lambda_\ell))$ will be denoted $\phi_r(t, \boldsymbol{u})$ by

$$\phi_r(t, \boldsymbol{u}) = E[t^{T_r} u_1^{N(T_r:\Lambda_1)} \cdots u_\ell^{N(T_r:\Lambda_\ell)}],$$

where $\boldsymbol{u} = (u_1, \ldots, u_\ell)$. Let t_0 be any positive integer. Suppose that we observe currently α_ω $(\omega = 0, 1, \ldots, s)$ at the $(t_0 - 1)$-th trial and we have $Z_{t_0-1} = b_j$ $(j = 1, \ldots, m)$. Then, we denote by $\phi_r^{(b_j)}(\alpha_\omega; t, \boldsymbol{u})$ the p.g.f. of the conditional distribution of $(T_r, N(T_r : \Lambda_1), \ldots, N(T_r : \Lambda_\ell))$ from the $(t_0 - 1)$-th trial given the above conditions. Let P_{ω,b_v} be the longest pattern among $\{(\alpha_{i_\omega,1}, \ldots, \alpha_{i_\omega,x_{i_\omega}}, b_v), (\alpha_{i_\omega,2}, \ldots, \alpha_{i_\omega,x_{i_\omega}}, b_v), \ldots, (b_v)\} \bigcap \Omega$ and let P_{0,b_v} be the longest pattern among $\{(b_v)\} \bigcap \Omega$. We define a mapping $f : (\Omega \setminus \Lambda) \times \Gamma \to \Omega$ by $f(\alpha_\omega, b_v) = P_{\omega,b_v}$.

From the definitions of $\phi_r(t, \boldsymbol{u})$ and $\phi_r^{(b_j)}(\alpha_\omega; t, \boldsymbol{u})$, we obtain the next theorem.

THEOREM 2.1. *Under non-overlap counting, the p.g.f. and the conditional p.g.f.'s of* $(T_r, N(T_r : \Lambda_1), \ldots, N(T_r : \Lambda_\ell))$ *satisfy the following recursive relations*:

$$\phi_r(t, \boldsymbol{u}) = \sum_{v=1}^{m} p_{b_v} t \phi_r^{(b_v)}(f(\alpha_0, b_v); t, \boldsymbol{u}) \qquad r \geq 1, \tag{2.1}$$

$$\phi_r^{(b_j)}(\alpha_\omega; t, \boldsymbol{u}) = \sum_{v=1}^{m} p_{b_j b_v} t \phi_r^{(b_v)}(f(\alpha_\omega, b_v); t, \boldsymbol{u}) \qquad r \geq 1,\ 0 \leq \omega \leq s,\ 1 \leq j \leq m, \tag{2.2}$$

$$\phi_r^{(a_{i,k_i})}(\Lambda_i; t, \boldsymbol{u}) = u_i \phi_{r-1}^{(a_{i,k_i})}(\alpha_0; t, \boldsymbol{u}) \qquad r \geq 2, \quad 1 \leq i \leq \ell, \tag{2.3}$$

$$\phi_1^{(a_{i,k_i})}(\Lambda_i; t, \boldsymbol{u}) = u_i \qquad 1 \leq i \leq \ell. \tag{2.4}$$

PROOF. The proof of (2.1) is immediately completed by observing that

$$E[t^{T_r} u_1^{N(T_r:\Lambda_1)} \cdots u_\ell^{N(T_r:\Lambda_\ell)}] = \sum_{v=1}^{m} p_{b_v} E[t^{T_r} u_1^{N(T_r:\Lambda_1)} \cdots u_\ell^{N(T_r:\Lambda_\ell)} \mid Z_1 = b_v].$$

Suppose that we observe currently α_ω ($\omega = 0, 1, \ldots, s$) at the $(t_0 - 1)$-th trials and we have $Z_{t_0-1} = b_j$ ($j = 1, \ldots, m$). Given the condition, we observe the t_0-th trial. For every b_j ($j = 1, \ldots, m$), the conditional probability that we observe $Z_{t_0} = b_v$ is $p_{b_j b_v}$. If we have $Z_{t_0} = b_v$, then the p.g.f. of the conditional distribution of $(T_r, N(T_r : \Lambda_1), \ldots, N(T_r : \Lambda_\ell))$ from the t_0-th trial becomes $\phi_r^{(b_v)}(f(\alpha_\omega, b_v); t, \boldsymbol{u})$. Hence we have (2.2). From the definitions of $\phi_r^{(b_j)}(\alpha_\omega; t, \boldsymbol{u})$, it is easy to check the equations (2.3) and (2.4). The proof is completed. $\square$

2.2. Overlap counting

Let Λ be a simple pattern of length k. Suppose that the simple pattern Λ is formed. If overlap counting is employed, we should count the next pattern Λ which is overlapping with the previous pattern Λ, by at most length $(k - 1)$, which is a primary difference between non-overlap counting and overlap counting (see Fu and Lou (2003)). For example, when $\Gamma = \{0, 1\}$ and $\Lambda = (1, 1, 0, 1, 1)$, in the sequence

0 1 1 0 1 1 0 1 1 1 0 1 1

the pattern $(1, 1, 0, 1, 1)$ is counted as occurring in positions 2–6, 5–9 and 9–13 for overlap counting, then we have $(T_1, T_2, T_3) = (6, 9, 13)$.

For $i = 1, 2, \ldots, \ell$, let Λ_i^* be the longest pattern among $\{(a_{i,2}, \ldots, a_{i,k_i}), (a_{i,3}, \ldots, a_{i,k_i}), \ldots, (a_{i,k_i})\} \bigcap (\Omega \setminus \Lambda)$. Suppose we have a simple pattern Λ_i ($i = 1, 2, \ldots, \ell$). Taking account of the overlapping structure of the simple pattern Λ_i ($i = 1, 2, \ldots, \ell$), easily we see that

$$\phi_r^{(a_{i,k_i})}(\Lambda_i; t, \boldsymbol{u}) = u_i \phi_{r-1}^{(a_{i,k_i})}(\Lambda_i^*; t, \boldsymbol{u}) \qquad r \geq 2, \quad 1 \leq i \leq \ell. \tag{2.5}$$

We have the following theorem by replacing equation (2.3) by equation (2.5) under overlap counting.

THEOREM 2.2. *Under overlap counting, the p.g.f. and the conditional p.g.f.'s of* $(T_r, N(T_r : \Lambda_1), \dots, N(T_r : \Lambda_\ell))$ *satisfy the recursive relations*: (2.1), (2.2), (2.4), (2.5).

Remark that Λ_i^* $(i = 1, 2, \dots, \ell)$ is equal to $\emptyset$ when the simple pattern Λ_i $(i = 1, 2, \dots, \ell)$ has no overlapping structure. Apparently, the results presented in Theorems 2.1 and 2.2 are the same since equation (2.5) reduces to equation (2.3).

3. Distributions of runs as special cases

In this section we assume that $\Gamma = \{F_1, \dots, F_\lambda, S_1, \dots, S_\nu\}$ and $\Lambda = \{\Lambda_i : 1 \le i \le \nu\}$ with $\Lambda_i = (\underbrace{S_i, S_i, \dots, S_i}_{k_i})$, $i = 1, 2, \dots, \nu$. Then we see that

$$\Omega = \left\{ \emptyset, (S_i), (S_i, S_i), \dots, (\underbrace{S_i, S_i, \dots, S_i}_{k_i}) : i = 1, 2, \dots, \nu \right\}.$$

Let $Z_1, Z_2, \dots$ be a time homogeneous Markov chain defined on the state space Γ with the transition probabilities

$$p_{b_i b_j} = P(Z_{t+1} = b_j \mid Z_t = b_i), \qquad \text{for} \quad t \ge 1,\ \ b_i, b_j \in \Gamma,\ \ i, j = 1, 2, \dots, \lambda + \nu$$

and the initial probabilities

$$p_{b_i} = P(Z_1 = b_i), \qquad \text{for} \quad b_i \in \Gamma,\ \ i = 1, 2, \dots, \lambda + \nu.$$

In this section, when no confusion is likely to arise, we will use $\phi_r^{(F_i)}(t, \boldsymbol{u})$ and $\phi_r^{(S_i)}(t, \boldsymbol{u})$ instead of $\phi_r^{(F_i)}(\emptyset; t, \boldsymbol{u})$ and $\phi_r^{(S_i)}((S_i); t, \boldsymbol{u})$, respectively.

3.1. Non-overlap counting

Using Theorem 2.1, we can obtain the following corollary.

COROLLARY 3.1. *The p.g.f.* $\phi_r(t, \boldsymbol{u})$ *and the conditional p.g.f.'s* $\phi_r^{(F_i)}(t, \boldsymbol{u})$, $i = 1, 2, \dots, \lambda$, $\phi_r^{(S_i)}(t, \boldsymbol{u})$, $i = 1, 2, \dots, \nu$ *satisfy the following recursive relations*:

$$\text{(3.1)} \quad \phi_r(t, \boldsymbol{u}) = \sum_{i=1}^{\lambda} p_{F_i} t \phi_r^{(F_i)}(t, \boldsymbol{u}) + \sum_{i=1}^{\nu} p_{S_i} t \phi_r^{(S_i)}(t, \boldsymbol{u}),$$

$$\text{(3.2)} \quad \phi_r^{(F_i)}(t, \boldsymbol{u}) = \sum_{j=1}^{\lambda} p_{F_i F_j} t \phi_r^{(F_j)}(t, \boldsymbol{u}) + \sum_{j=1}^{\nu} p_{F_i S_j} t \phi_r^{(S_j)}(t, \boldsymbol{u}), \qquad i = 1, 2, \dots, \lambda,$$

$$\begin{aligned}(3.3)\quad \phi_r^{(S_i)}(t,\boldsymbol{u}) = {} & R_{k_i-2}(p_{S_iS_i}t)\left(\sum_{j=1}^{\lambda} p_{S_iF_j}t\phi_r^{(F_j)}(t,\boldsymbol{u}) + \sum_{j\neq i} p_{S_iS_j}t\phi_r^{(S_j)}(t,\boldsymbol{u})\right) \\ & + R_{k_i-1}(p_{S_iS_i}t)\left(\sum_{j=1}^{\lambda}\sum_{\ell=1}^{r-1}(p_{S_iS_i}t)^{k_i\ell-1}p_{S_iF_j}tu_i^{\ell}\phi_{r-\ell}^{(F_j)}(t,\boldsymbol{u})\right. \\ & \left. + \sum_{j\neq i}\sum_{\ell=1}^{r-1}(p_{S_iS_i}t)^{k_i\ell-1}p_{S_iS_j}tu_i^{\ell}\phi_{r-\ell}^{(S_j)}(t,\boldsymbol{u})\right) \\ & + (p_{S_iS_i}t)^{k_ir-1}u_i^r, \qquad i=1,2,\dots,\nu,\end{aligned}$$

where

$$(3.4)\qquad R_x(t) = \begin{cases} 1+t+t^2+\cdots+t^x, & x=0,1,2,\dots \\ 0, & \textit{otherwise.}\end{cases}$$

We will define the double generating functions $\Phi(t,\boldsymbol{u},z)$ and $\Phi^{(b_i)}(t,\boldsymbol{u},z)$ ($b_i \in \Gamma$, $i=1,2,\dots,\lambda+\nu$) as

$$\Phi(t,\boldsymbol{u},z) = \sum_{r=1}^{\infty}\phi_r(t,\boldsymbol{u})z^r$$

and

$$\Phi^{(b_i)}(t,\boldsymbol{u},z) = \sum_{r=1}^{\infty}\phi_r^{(b_i)}(t,\boldsymbol{u})z^r.$$

PROPOSITION 3.1. *The double generating functions* $\Phi(t,\boldsymbol{u},z)$ *and* $\Phi^{(F_i)}(t,\boldsymbol{u},z)$, $i=1,2,\dots,\lambda$ $\Phi^{(S_i)}(t,\boldsymbol{u},z)$, $i=1,2,\dots,\nu$ *satisfy the following system of equations*:

$$(3.5)\quad \Phi(t,\boldsymbol{u},z) = \sum_{j=1}^{\lambda}p_{F_j}t\Phi^{(F_j)}(t,\boldsymbol{u},z) + \sum_{j=1}^{\nu}p_{S_j}t\Phi^{(S_j)}(t,\boldsymbol{u},z),$$

$$\begin{aligned}(3.6)\quad \Phi^{(F_i)}(t,\boldsymbol{u},z) = {} & \sum_{j=1}^{\lambda}p_{F_iF_j}t\Phi^{(F_j)}(t,\boldsymbol{u},z) \\ & + \sum_{j=1}^{\nu}p_{F_iS_j}t\Phi^{(S_j)}(t,\boldsymbol{u},z), \qquad i=1,2,\dots,\lambda,\end{aligned}$$

$$\begin{aligned}(3.7)\quad \Phi^{(S_i)}(t,\boldsymbol{u},z) = {} & \frac{R_{k_i-2}(p_{S_iS_i}t)+(p_{S_iS_i}t)^{k_i-1}u_iz}{1-(p_{S_iS_i}t)^{k_i}u_iz} \\ & \times\left(\sum_{j=1}^{\lambda}p_{S_iF_j}t\Phi^{(F_j)}(t,\boldsymbol{u},z) + \sum_{j\neq i}p_{S_iS_j}t\Phi^{(S_j)}(t,\boldsymbol{u},z)\right) \\ & + \frac{(p_{S_iS_i}t)^{k_i-1}u_iz}{1-(p_{S_iS_i}t)^{k_i}u_iz}, \qquad i=1,2,\dots,\nu.\end{aligned}$$

For the special case of $\lambda = 1$ and $\nu = 1$, Koutras (1997) studied the marginal distribution of T_r, which is called *Type I Markov negative binomial distribution of order k* (see Inoue and Aki (2003)).

In the case of i.i.d. trials,

$$p_{b_i b_j} = P(Z_{t+1} = b_j \mid Z_t = b_i) = p_{b_j}, \quad \text{for} \quad t \geq 1, \ b_i, b_j \in \Gamma, \ i, j = 1, 2, \ldots, \lambda + \nu, \tag{3.8}$$

so the double generating function $\Phi(t, \boldsymbol{u}, z)$ may be expressed in an appealing form.

PROPOSITION 3.2. *Under the condition* (3.8), *the double generating function $\Phi(t, \boldsymbol{u}, z)$ is given by*

$$\Phi(t, \boldsymbol{u}, z) = \frac{\sum_{i=1}^{\nu} \dfrac{(p_{S_i} t)^{k_i} u_i z}{R_{k_i - 1}(p_{S_i} t)}}{1 - \sum_{i=1}^{\lambda} p_{F_i} t - \sum_{i=1}^{\nu} \dfrac{p_{S_i} t R_{k_i - 2}(p_{S_i} t) + (p_{S_i} t)^{k_i} u_i z}{R_{k_i - 1}(p_{S_i} t)}},$$

or equivalently

$$\Phi(t, \boldsymbol{u}, z) = \frac{\sum_{i=1}^{\nu} \dfrac{(p_{S_i} t)^{k_i} u_i z}{R_{k_i - 1}(p_{S_i} t)}}{1 - t + \sum_{i=1}^{\nu} \dfrac{(1 - u_i z)(p_{S_i} t)^{k_i}}{R_{k_i - 1}(p_{S_i} t)}}.$$

Expanding the double generating function $\Phi(t, \boldsymbol{u}, z)$ in a Taylor series around $z = 0$ and picking out the coefficient of z^r, we get the explicit expression for the joint probability generating function $\phi_r(t, \boldsymbol{u})$. More specifically, we have the following result.

PROPOSITION 3.3. *Under the condition* (3.8), *the joint probability generating function $\phi_r(t, \boldsymbol{u})$ can be expressed as*

$$\phi_r(t, \boldsymbol{u}) = \left[\frac{\sum_{i=1}^{\nu} \dfrac{(p_{S_i} t)^{k_i} u_i}{R_{k_i - 1}(p_{S_i} t)}}{1 - \sum_{i=1}^{\lambda} p_{F_i} t - \sum_{i=1}^{\nu} \dfrac{p_{S_i} t R_{k_i - 2}(p_{S_i} t)}{R_{k_i - 1}(p_{S_i} t)}} \right]^r,$$

or equivalently

$$\phi_r(t, \boldsymbol{u}) = \left[\frac{\sum_{i=1}^{\nu} \dfrac{(p_{S_i} t)^{k_i} u_i}{R_{k_i - 1}(p_{S_i} t)}}{1 - t + \sum_{i=1}^{\nu} \dfrac{(p_{S_i} t)^{k_i}}{R_{k_i - 1}(p_{S_i} t)}} \right]^r.$$

For the special case of $r = 1$, the marginal distribution of T_1 is called *the sooner geometric distribution of order* $(k_1, \ldots, k_\nu)$ (see Balakrishnan and Koutras (2002)).

3.2. Overlap counting

Using Theorem 2.2, we have the following equations

$$\begin{aligned}
(3.9)\quad \phi_r^{(S_i)}(t,\boldsymbol{u}) &= R_{k_i-2}(p_{S_iS_i}t)\left(\sum_{j=1}^{\lambda} p_{S_iF_j}t\phi_r^{(F_j)}(t,\boldsymbol{u}) + \sum_{j\neq i} p_{S_iS_j}t\phi_r^{(S_j)}(t,\boldsymbol{u})\right)\\
&\quad + \sum_{\ell=1}^{r-1}(p_{S_iS_i}t)^{k_i+\ell-2}u_i^{\ell}\\
&\quad \times\left(\sum_{j=1}^{\lambda} p_{S_iF_j}t\phi_{r-\ell}^{(F_j)}(t,\boldsymbol{u}) + \sum_{j\neq i} p_{S_iS_j}t\phi_{r-\ell}^{(S_j)}(t,\boldsymbol{u})\right)\\
&\quad + (p_{S_iS_i}t)^{k_i+r-2}u_i^r, \qquad i=1,2,\ldots,\nu,\\
(3.10)\quad \Phi^{(S_i)}(t,\boldsymbol{u},z) &= \frac{R_{k_i-2}(p_{S_iS_i}t) - p_{S_iS_i}tu_izR_{k_i-3}(p_{S_iS_i}t)}{1-p_{S_iS_i}tu_iz}\\
&\quad \times\left(\sum_{j=1}^{\lambda} p_{S_iF_j}t\Phi^{(F_j)}(t,\boldsymbol{u},z) + \sum_{j\neq i} p_{S_iS_j}t\Phi^{(S_j)}(t,\boldsymbol{u},z)\right)\\
&\quad + \frac{(p_{S_iS_i}t)^{k_i-1}u_iz}{1-p_{S_iS_i}tu_iz}, \qquad i=1,2,\ldots,\nu.
\end{aligned}$$

Replacing (3.3) by (3.9) and (3.7) by (3.10), respectively, in a similar fashion as in the conclusion of Subsection 3.1, we could treat the case where overlap counting is employed. Easily we can state the following two results.

COROLLARY 3.2. *The p.g.f.* $\phi_r(t,\boldsymbol{u})$ *and the conditional p.g.f.'s* $\phi_r^{(F_i)}(t,\boldsymbol{u})$, $i=1,2,\ldots,\lambda$, $\phi_r^{(S_i)}(t,\boldsymbol{u})$, $i=1,2,\ldots,\nu$ *satisfy the recursive relations*: (3.1), (3.2) *and* (3.9).

PROPOSITION 3.4. *The double generating functions* $\Phi(t,\boldsymbol{u},z)$ *and* $\Phi^{(F_i)}(t,\boldsymbol{u},z)$, $i=1,2,\ldots,\lambda$ $\Phi^{(S_i)}(t,\boldsymbol{u},z)$, $i=1,2,\ldots,\nu$ *satisfy the system of equations*: (3.5), (3.6) *and* (3.10).

In the special case of $\lambda=1$ and $\nu=1$, Koutras (1997) derived the double generating function of the marginal distribution of T_r, which is called *Type III Markov negative binomial distribution of order k* (see Inoue and Aki (2003)). Under the condition (3.8), we can establish a compact formula for the double generating function $\Phi(t,\boldsymbol{u},z)$.

PROPOSITION 3.5. *Under the condition* (3.8), *the double generating function* $\Phi(t,\boldsymbol{u},z)$ *is given by*

$$\Phi(t,\boldsymbol{u},z) = \frac{\sum_{i=1}^{\nu}\dfrac{(p_{S_i}t)^{k_i}u_iz}{R_{k_i-1}(p_{S_i}t) - p_{S_i}tu_izR_{k_i-2}(p_{S_i}t)}}{1-\sum_{i=1}^{\lambda}p_{F_i}t - \sum_{i=1}^{\nu}\dfrac{p_{S_i}tR_{k_i-2}(p_{S_i}t) - (p_{S_i}t)^2u_izR_{k_i-3}(p_{S_i}t)}{R_{k_i-1}(p_{S_i}t) - p_{S_i}tu_izR_{k_i-2}(p_{S_i}t)}},$$

or equivalently

$$\Phi(t,\boldsymbol{u},z)=\frac{\sum_{i=1}^{\nu}\dfrac{(p_{S_i}t)^{k_i}u_iz}{R_{k_i-1}(p_{S_i}t)-p_{S_i}tu_izR_{k_i-2}(p_{S_i}t)}}{1-t+\sum_{i=1}^{\nu}\dfrac{(1-u_iz)(p_{S_i}t)^{k_i}}{R_{k_i-1}(p_{S_i}t)-p_{S_i}tu_izR_{k_i-2}(p_{S_i}t)}}.$$

Expanding the double generating function $\Phi(t,\boldsymbol{u},z)$ in a Taylor series around $z=0$ and picking out the coefficient of z^r, we may obtain the explicit form of the joint probability generating function $\phi_r(t,\boldsymbol{u})$. The next proposition provides the details.

PROPOSITION 3.6. *Under the condition* (3.8), *the joint probability generating function $\phi_r(t,\boldsymbol{u})$ can be expressed as*

$$\begin{aligned}\phi_r(t,\boldsymbol{u}) &= \sum_{n=0}^{r-1}\sum_{x_1+2x_2+\cdots+rx_r=n}\binom{x_1+x_2+\cdots+x_r}{x_1,x_2,\ldots,x_r}\\ &\quad\times\frac{P_{r-n}Q_1^{x_1}\cdots Q_r^{x_r}}{\left(1-\sum_{i=1}^{\lambda}p_{F_i}t-\sum_{i=1}^{\nu}\dfrac{p_{S_i}tR_{k_i-2}(p_{S_i}t)}{R_{k_i-1}(p_{S_i}t)}\right)^{1+x_1+\cdots+x_r}},\end{aligned}$$

or equivalently

$$\begin{aligned}\phi_r(t,\boldsymbol{u}) &= \sum_{n=0}^{r-1}\sum_{x_1+2x_2+\cdots+rx_r=n}\binom{x_1+x_2+\cdots+x_r}{x_1,x_2,\ldots,x_r}\\ &\quad\times\frac{P_{r-n}Q_1^{x_1}\cdots Q_r^{x_r}}{\left(1-t+\sum_{i=1}^{\nu}\dfrac{(p_{S_i}t)^{k_i}}{R_{k_i-1}(p_{S_i}t)}\right)^{1+x_1+\cdots+x_r}},\end{aligned}$$

where

$$\begin{aligned}P_n&=\sum_{i=1}^{\nu}\frac{(p_{S_i}t)^{k_i+n-1}[R_{k_i-2}(p_{S_i}t)]^{n-1}u_i^n}{[R_{k_i-1}(p_{S_i}t)]^n}, \qquad n\ge 1,\\ Q_n&=\sum_{i=1}^{\nu}\frac{(p_{S_i}t)^{k_i+n-1}[R_{k_i-2}(p_{S_i}t)]^{n-1}u_i^n}{[R_{k_i-1}(p_{S_i}t)]^{n+1}}, \qquad n\ge 1,\end{aligned}$$

and $R_x(\cdot)$ is as given in (3.4).

As indicated by Koutras and Alexandrou (1997), the waiting time distributions of runs play an important role in a wide range of areas. Especially sooner waiting time distributions of runs were applied to a variety of different areas (see Balakrishnan and Koutras (2002), Shmueli and Cohen (2000) and Balakrishnan *et al.* (1997)).

We would like to mention a class of multiple failure mode (MFM) systems (see Chao *et al.* (1995)). According to Boutsikas and Koutras (2002), the consecutive

$k_1, k_2, \ldots, k_\nu$-out-of-n:MFM system consists of n linearly arranged components, and enter failure mode s whenever at least k_s consecutive components are failed in mode s, $(s = 1, 2, \ldots, \nu)$. Clearly, this reliability system is associated with sooner waiting time problems in a sequence of multistate trials. The exact reliability of the consecutive $k_1, k_2, \ldots, k_\nu$-out-of-n:MFM system with Markov dependent components can be evaluated through the results in case of $\lambda = 1$ in this section. Here, we regard F_1 as a working state and $S_1, \ldots, S_\nu$ as failure modes.

In closing, we establish the approximation formula for the tail probability $P(T_1 > n)$ of the sooner geometric distribution of order $(k_1, \ldots, k_\nu)$ by applying the Chen-Stein method. For fixed n, let W_x, $x \in I = \{1, 2, \ldots, n - k_{\min} + 1\}$ be indicator variables taking on the value 1 if and only if $\prod_{t=x}^{x+k_1-1} Z_t = S_1^{k_1}$ or $\prod_{t=x}^{x+k_2-1} Z_t = S_2^{k_2}$ or, $\prod_{t=x}^{x+k_\nu-1} Z_t = S_\nu^{k_\nu}$ (if $n - k_{\max} + 1 < x < n - k_{\min} + 1$, the conditions involving indices exceeding n are disregarded), where $k_{\min} = \min_{1 \le i \le \nu} k_i$ and $k_{\max} = \max_{1 \le i \le \nu} k_i$. It is evident that $P(T_1 > n) = P(\sum_{x \in I} W_x = 0)$. Introducing the neighborhood of the dependence $B_x = \{y \in I : |y - x| < k_{\max}\}$, taking into account the relations,

$$E[W_x] \le \sum_{i=1}^{\nu} p_{S_i}^{k_i}, \qquad E[W_x W_y] \le (p_{S_1} + \cdots + p_{S_\nu}) E[W_x], \qquad |B_x| \le 2k_{\max} - 1,$$

and applying Theorem 1 of Arratia *et al.* (1989), we deduce that

$$\begin{aligned} &|P(T_1 > n) - e^{-\mu}| \\ &\quad \le \frac{(1 - e^{-\mu})}{\mu} \sum_{x=1}^{n-k_{\min}+1} \left\{ \sum_{y \in B_x} E(W_x) E(W_y) + \sum_{y \in B_x \setminus \{x\}} E(W_x W_y) \right\}, \end{aligned}$$

where $\mu = \sum_{i=1}^{\nu} (n - k_i + 1) p_{S_i}^{k_i}$. Therefore, we have the following approximation formula for the tail probability $P(T_1 > n)$ of the sooner geometric distribution of order $(k_1, \ldots, k_\nu)$:

$$\begin{aligned} &|P(T_1 > n) - e^{-\mu}| \\ &\quad \le (1 - e^{-\mu}) \left\{ (2k_{\max} - 1) \sum_{i=1}^{\nu} p_{S_i}^{k_i} + (2k_{\max} - 2) \sum_{i=1}^{\nu} p_{S_i} \right\}. \end{aligned} \tag{3.11}$$

The formula (3.11) offers a useful tool for the investigation of the asymptotic behaviour of $P(T_1 > n)$. For example, assume that $0 < p_{S_i} < 1$, $(i = 1, 2, \ldots, \nu)$ are fixed and $k_i = k_{i,n}$, $(i = 1, 2, \ldots, \nu)$ are functions of n such that $\lim_{n \to \infty} k_{i,n} = \infty$, $(i = 1, 2, \ldots, \nu)$. Under the conditions $\lim_{n \to \infty} (n - k_{i,n} + 1) p_{S_i}^{k_{i,n}} = \mu_i$, $(i = 1, 2, \ldots, \nu)$, we readily obtain

$$\lim_{n \to \infty} P(T_1 > n) = \exp\left(-\sum_{i=1}^{\nu} \mu_i \right).$$

For several choices n, k_1, k_2, p_1 and p_2, the performance of the aforementioned bounds (3.11) is illustrated in Table 1.

Table 1. Exact and approximate values for the tail probability.

n	(k_1, k_2)	(p_1, p_2)	Lower Bound	$P(T_1 > n)$	Upper Bound
100	$(10, 9)$	$(0.20, 0.25)$	0.9967	0.9997	1.0026
100	$(13, 9)$	$(0.40, 0.25)$	0.9844	0.9994	1.0137
200	$(12, 10)$	$(0.35, 0.20)$	0.9914	0.9996	1.0073
200	$(13, 10)$	$(0.40, 0.20)$	0.9803	0.9992	1.0172
300	$(15, 10)$	$(0.35, 0.25)$	0.9943	0.9998	1.0050
300	$(16, 14)$	$(0.42, 0.20)$	0.9948	0.9998	1.0047

4. Estimation problems and numerical examples

In this section we assume that $\Gamma = \{0, 1, 2, \ldots, \nu\}$ and the sequence $Z_1, Z_2, \ldots$ are i.i.d. trials with probabilities $p_i = P(Z_t = i)$ for $i = 0, 1, \ldots, \nu$ and $t \geq 1$. We denote the p.g.f. of the marginal distribution of T_r by $\varphi_r(t)$. It is easy to see that $\varphi_r(t) = \phi_r(t, \boldsymbol{u}) \mid_{u_1 = \cdots = u_\ell = 1}$. Let $T_r^{(1)}, T_r^{(2)}, \ldots, T_r^{(N)}$ be the random sample of size N from the waiting time distribution. Then we will calculate the maximum likelihood estimates (MLE's) of the parameters p_i, $(i = 1, 2, \ldots, \nu)$ in the waiting time distributions of compound patterns based on the independent observations. However, the problems of identifiability arise.

4.1. The problems of identifiability

To begin with, we introduce the notion of identifiability: the parameter θ is said to be identifiable, if for all pairs of distinct parameter values, say θ and $\theta^{'}$, the sample distributions, say $\mathcal{P}_\theta$ and $\mathcal{P}_{\theta'}$, are also distinct, i.e., $\mathcal{P}_\theta = \mathcal{P}_{\theta'}$ implies $\theta = \theta^{'}$. The problems of identifiability often arise when we treat the estimation of parameters in the waiting time distributions of the patterns.

Example 4.1. Let $\Gamma = \{0, 1, 2\}$ and let $\Lambda = \{\Lambda_1, \Lambda_2\}$ be a compound pattern with $\Lambda_1 = (1, 2, 0)$ and $\Lambda_2 = (2, 1, 0)$. Using Theorem 2.1, we can obtain the p.g.f. of sooner waiting time distribution as

$$\varphi_1(t) = \frac{2p_0tp_1tp_2t}{1 - t + 2p_0tp_1tp_2t}. \tag{4.1}$$

The p.g.f. is invariant under the symmetry $p_1 \leftrightarrow p_2$. This implies that $\mathcal{P}_{p_1} = \mathcal{P}_{p_2}$. In the sequel, for known parameters $a_i (> 0)$, $(i = 2, 3, \ldots, \nu)$, we assume the additional constraints

$$p_i = a_i p_1, \qquad i = 2, 3, \ldots, \nu, \tag{4.2}$$

which restore the parameter identifiability.

4.2. The likelihood function and maximum likelihood estimate

From Theorems 2.1 and 2.2, we can obtain the p.g.f. $\varphi_r(t)$ of the waiting time distribution of the compound pattern. Let $C(t, i)$ denote the coefficient of

t^i in the Taylor expansion of t for $\varphi_r(t)$ around $t = 0$. Under the constraint (4.2), the likelihood function of p_1 based on $T_r^{(1)}, T_r^{(2)}, \ldots, T_r^{(N)}$ can be written as

$$L(p_1) = \prod_{i=1}^{N} C(t, T_r^{(i)}). \tag{4.3}$$

It is well known that the MLE is obtained by maximizing the likelihood function. The following is an example illustrating how to obtaining the MLE.

Example 4.2 (Continuation of Example 4.1). We consider the estimation of parameter p_1 under the constraint $p_2 = p_1$. Table 2 is a simulated data set of sooner waiting time with $p_1 = p_2 = 0.33$ and $N = 20$.

Expanding the p.g.f. (4.1) in a Taylor series around $t = 0$, picking out the coefficient of t^i based on the data in Table 2 and using the formula (4.3), we can obtain the likelihood function $L(p_1)$. However, the likelihood function $L(p_1)$ is omitted here since it is not represented in a simple form. In Fig. 1, we give the graph of the likelihood function based on the data in Table 2.

By maximizing the likelihood function numerically, we have the MLE $\hat{p}_1 = 0.3333$.

Table 2. A simulated data set of sooner waiting time with $p_1 = p_2 = 0.33$.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$T_1^{(i)}$	16	5	6	9	11	5	4	3	14	4	14	3	10	15	54	10	11	3	21	21

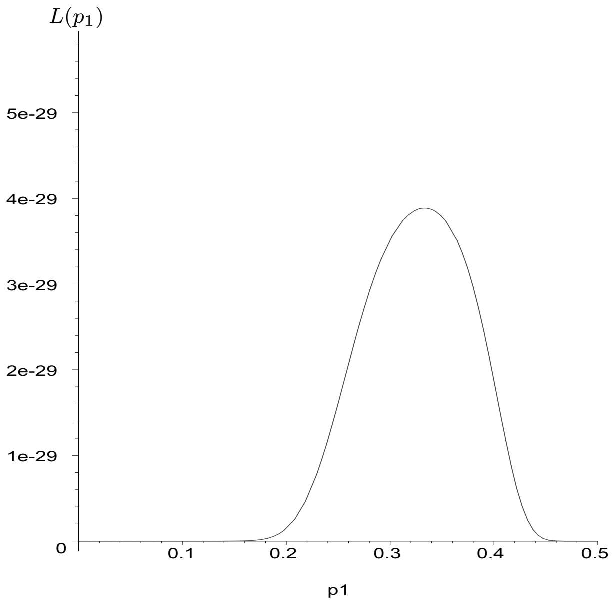

Figure 1. The likelihood function $L(p_1)$ based on the data in Table 2.

Acknowledgements

We wish to thank the editor and the referees for careful reading of our paper and helpful suggestions which led to improvements to the original version of the manuscript.

References

Aki, S. and Hirano, K. (1989). Estimation of parameters in the discrete distributions of order k, *Annals of the Institute of Statistical Mathematics*, **41**, 47–61.

Antzoulakos, D. L. (2001). Waiting times for patterns in a sequence of multistate trials, *Journal of Applied Probability*, **38**, 508–518.

Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method, *Annals of Probability*, **17**, 9–17.

Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*, John Wiley, New York.

Balakrishnan, N., Mohanty, S. G. and Aki, S. (1997). Start-up demonstration tests under Markov dependence model with corrective actions, *Annals of the Institute of Statistical Mathematics*, **49**, 155–169.

Boutsikas, M. V. and Koutras, M. V. (2002). On a class of multiple failure mode systems, *Naval Research Logistics*, **49**, 167–185.

Chao, M. T., Fu, J. C. and Koutras, M. V. (1995). Survey of reliability studies of consecutive-k-out-of-n: F & related systems, *IEEE Transactions on Reliability*, **40**, 120–127.

Ewens, W. J. and Grant, G. R. (2001). *Statistical Methods in Bioinformatics*: *An Introduction*, Springer.

Fu, J. C. (1996). Distribution theory of runs and patterns associated with a sequence of multi-state trials, *Statistica Sinica*, **6**, 957–974.

Fu, J. C. (2001). Distribution of the scan statistic for a sequence of bistate trials, *Journal of Applied Probability*, **38**, 908–916.

Fu, J. C. and Chang, Y. M. (2002). On probability generating functions for waiting time distributions of compound patterns in a sequence of multistate trials, *Journal of Applied Probability*, **39**, 70–80.

Fu, J. C. and Lou, W. Y. W. (2003). *Distribution Theory of Runs and Patterns and Its Applications*: *A Finite Markov Chain Imbedding Approach*, World Scientific, Singapore.

Han, Q. and Hirano, K. (2003). Sooner and later waiting time problems for patterns in Markov dependent trials, *Journal of Applied Probability*, **40**, 73–86.

Inoue, K. (2004). Joint distributions associated with patterns, successes and failures in a sequence of multi-state trials, *Annals of the Institute of Statistical Mathematics*, **56**, 143–168.

Inoue, K. and Aki, S. (2002). Generalized waiting time problems associated with pattern in Polya's urn scheme, *Annals of the Institute of Statistical Mathematics*, **54**, 681–688.

Inoue, K. and Aki, S. (2003). Generalized binomial and negative binomial distributions of order k by the ℓ-overlapping enumeration scheme, *Annals of the Institute of Statistical Mathematics*, **55**, 153–167.

Koutras, M. V. (1997). Waiting time distributions associated with runs of fixed length in two-state Markov chain, *Annals of the Institute of Statistical Mathematics*, **49**, 123–139.

Koutras, M. V. and Alexandrou, V. A. (1997). Sooner waiting time problems in a sequence of trinary trials, *Journal of Applied Probability*, **34**, 593–609.

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

Robin, S. and Daudin, J. J. (1999). Exact distribution of word occurrences in a random sequences of letters, *Journal of Applied Probability*, **36**, 179–193.

Robin, S. and Daudin, J. J. (2001). Exact distribution of the distances between any occurrences of a set of words, *Annals of the Institute of Statistical Mathematics*, **53**, 895–905.

Shmueli, G. and Cohen, A. (2000). Run-related probability functions applied to sampling inspection, *Technometrics*, **42**, 188–202.