J. Appl. Prob. **44,** 71–81 (2007)
Printed in England
© *Applied Probability Trust* 2007

ON GENERATING FUNCTIONS OF WAITING TIMES AND NUMBERS OF OCCURRENCES OF COMPOUND PATTERNS IN A SEQUENCE OF MULTISTATE TRIALS

KIYOSHI INOUE,* *Seikei University*
SIGEO AKI,** *Kansai University*

Abstract

In this paper we study two distributions, namely the distribution of the waiting times until given numbers of occurrences of compound patterns and the distribution of the numbers of occurrences of compound patterns in a fixed number of trials. We elucidate the interrelation between these two distributions in terms of the generating functions. We provide perspectives on the problems related to compound patterns in statistics and probability. As an application, the waiting time problem of counting runs of specified lengths is considered in order to illustrate how the distributions of waiting times can be derived from our theoretical results.

Keywords: Compound pattern; multistate trial; sooner waiting time; later waiting time; run; enumeration scheme; probability function; probability generating function; double generating function

2000 Mathematics Subject Classification: Primary 60E05
Secondary 60J10

1. Introduction

Recently, the distribution theory associated with patterns has received a great deal of attention (see Fu and Chang (2003), Fu and Lou (2003), Inoue and Aki (2002), and Inoue (2004)). Two important distributions are related to patterns and applied in a wide range of areas (for example quality control, reliability theory, psychology, genome sequence analysis, etc.). The first is the distribution of the waiting times until given numbers of occurrences of compound patterns. An especially interesting class of waiting time distributions are referred to as sooner and later waiting time distributions (see Fu and Chang (2002), (2003)). The second is the distribution of the numbers of occurrences of patterns in a fixed number of trials.

In this paper we study these distributions and elucidate the relationship between the two. Koutras (1997) has investigated the relationship between the distributions of the waiting times and the number of occurrences of a simple pattern. We generalize the results of Koutras (1997). Our generalizations will not only be of theoretical interest but will also have some important applications. Although the later waiting time problems for compound patterns are closely related to many important applications and there is a need to study the later waiting time distributions of compound patterns, currently the development of the relevant distribution theory is very slow and there are not enough results to tackle practical problems in probability and statistics. We

Received 19 August 2005; revision received 13 December 2006.

* Postal address: Faculty of Economics, Seikei University, 3-3-1 Kichijoji-Kitamachi, Musasino-shi, Tokyo, 180-8633, Japan. Email address: kinoue@econ.seikei.ac.jp

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

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press

provide perspectives on the later waiting time problems for compound patterns and offer a very efficient computational tool.

Let $\{Z_n,\ n \geq 1\}$ be a sequence of multistate trials defined on the state space $\Gamma = \{0, \dots, m\}$. Following Fu and Lou (2003) (see also Fu and Chang (2002) and Fu (1996)), we define a simple pattern and a compound pattern.

Definition 1.1. We say that ε is a simple pattern if ε is composed of a specified sequence of k states, i.e. $\varepsilon = (a_1, \dots, a_k)$, $a_i \in \Gamma$, $1 \leq i \leq k$ (k, the length of the pattern, is fixed, and the states in the pattern are allowed to repeat).

Let ε_1 and ε_2 be two simple patterns of lengths k_1 and k_2, respectively. We say that ε_1 and ε_2 are distinct if neither is a subsequence (segment) of the other. We define the union $\{\varepsilon_1, \varepsilon_2\}$ to be the occurrence of either the pattern ε_1 or the pattern ε_2.

Definition 1.2. We say that ε is a compound pattern if it is a union of $c \geq 2$ distinct simple patterns (a set of c distinct simple patterns). For $c = 1$, we identify the compound pattern with the simple pattern.

Let $\varepsilon_i = \{\varepsilon_{i,j},\ j = 1, \dots, c_i\}$, $i = 1, \dots, \nu$, be compound patterns. We assume that the simple patterns $\varepsilon_{i,j}$, $i = 1, \dots, \nu$, $j = 1, \dots, c_i$, are distinct from each other. For $i = 1, \dots, \nu$, let $X_n^{\varepsilon_i}(\alpha_i)$ be the total number of occurrences of compound pattern ε_i in the trials $Z_1, \dots, Z_n$ under $\alpha_i \in \{N, O\}$ counting, where α_i represents the type of counting scheme employed: $\alpha_i = N$ will indicate nonoverlapping counting and $\alpha_i = O$ will indicate overlapping counting. Note that the compound patterns are observed independently and that we allow overlapping counting of compound patterns. For $i = 1, \dots, \nu$, we denote by $E_{r_i}^{\varepsilon_i}(\alpha_i)$ the event that r_i occurrences of the compound pattern ε_i are observed in the sequence of multistate trials under type-α_i counting. Let $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha})$ be the waiting time until the occurrence of the xth event among $E_{r_i}^{\varepsilon_i}(\alpha_i)$, $i = 1, \dots, \nu$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_\nu)$, $\boldsymbol{r} = (r_1, \dots, r_\nu)$, and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_\nu)$. Note that each compound pattern ε_i is observed only r_i times; that is, after its r_ith occurrence we are no longer interested in ε_i and are instead interested in when the remaining events occur. It is clear that

$$T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1; \boldsymbol{\alpha}) \leq T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(2; \boldsymbol{\alpha}) \leq \dots \leq T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha}).$$

In the special cases in which $x = 1$ and, respectively, $x = \nu$, the distributions of $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1; \boldsymbol{\alpha})$ and $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha})$ are respectively called the *sooner waiting time distribution* and the *later waiting time distribution*.

In Section 2 we introduce necessary definitions and notation. In Section 3 we study the distribution of the waiting times $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha})$, $x = 1, \dots, \nu$, using the distribution of $(X_n^{\varepsilon_1}(\alpha_1), \dots, X_n^{\varepsilon_\nu}(\alpha_\nu))$. We also elucidate the relationship between the distributions of $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha})$ and $(X_n^{\varepsilon_1}(\alpha_1), \dots, X_n^{\varepsilon_\nu}(\alpha_\nu))$. Section 4 serves as an illustration of how the general theory presented in Section 3 can be used to derive probability functions and probability generating functions. We apply the general results to the special case in which the compound patterns ε_i, $i = 1, \dots, \nu$, are sets of runs of certain lengths.

2. Definitions and notation

As above, let $\{Z_n,\ n \geq 1\}$ be a sequence of multistate trials defined on the state space $\Gamma = \{0, \dots, m\}$, let $\varepsilon_i = \{\varepsilon_{i,j},\ j = 1, \dots, c_i\}$, $i = 1, \dots, \nu$, be compound patterns (under the assumption that the simple patterns $\varepsilon_{i,j}$, $i = 1, \dots, \nu$, $j = 1, \dots, c_i$, are distinct from each other), and let $X_n^{\varepsilon_i}(\alpha_i)$, $i = 1, \dots, \nu$, be the numbers of occurrences of compound pattern

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press

ε_i in $Z_1, \dots, Z_n$ under $\alpha_i \in \{N, O\}$ counting. We define the probability generating function and the double generating function of $(X_n^{\varepsilon_1}(\alpha_1), \dots, X_n^{\varepsilon_\nu}(\alpha_\nu))$ by

$$\begin{aligned}\phi_n^{\boldsymbol{\varepsilon}}(\boldsymbol{z}; \boldsymbol{\alpha}) &= \mathrm{E}[z_1^{X_n^{\varepsilon_1}(\alpha_1)} \cdots z_\nu^{X_n^{\varepsilon_\nu}(\alpha_\nu)}] \\ &= \sum_{x_1,\dots,x_\nu \ge 0} \mathrm{P}(X_n^{\varepsilon_1}(\alpha_1) = x_1, \ \dots, \ X_n^{\varepsilon_\nu}(\alpha_\nu) = x_\nu) z_1^{x_1} \cdots z_\nu^{x_\nu}, \\ \Phi^{\boldsymbol{\varepsilon}}(\boldsymbol{z}, t; \boldsymbol{\alpha}) &= \sum_{n=0}^{\infty} \phi_n^{\boldsymbol{\varepsilon}}(\boldsymbol{z}; \boldsymbol{\alpha}) t^n \\ &= \sum_{n=0}^{\infty} \sum_{x_1,\dots,x_\nu \ge 0} \mathrm{P}(X_n^{\varepsilon_1}(\alpha_1) = x_1, \ \dots, \ X_n^{\varepsilon_\nu}(\alpha_\nu) = x_\nu) z_1^{x_1} \cdots z_\nu^{x_\nu} t^n,\end{aligned}$$

respectively, where $\boldsymbol{z} = (z_1, \dots, z_\nu)$. Clearly, the probability generating function and double generating function of $(X_n^{\varepsilon_{i_1}}(\alpha_{i_1}), \dots, X_n^{\varepsilon_{i_j}}(\alpha_{i_j}))$ can be expressed, for $j = 1, \dots, \nu$, as

$$\begin{aligned}\phi_n^{\varepsilon_{i_1},\dots,\varepsilon_{i_j}}(z_{i_1}, \dots, z_{i_j}; \alpha_{i_1}, \dots, \alpha_{i_j}) &= \phi_n^{\boldsymbol{\varepsilon}}(\boldsymbol{z}; \boldsymbol{\alpha})|_{z_{i_u}=1,\, u \ne 1,\dots,j}, \\ \Phi^{\varepsilon_{i_1},\dots,\varepsilon_{i_j}}(z_{i_1}, \dots, z_{i_j}, t; \alpha_{i_1}, \dots, \alpha_{i_j}) &= \Phi^{\boldsymbol{\varepsilon}}(\boldsymbol{z}, t; \boldsymbol{\alpha})|_{z_{i_u}=1,\, u \ne 1,\dots,j}.\end{aligned}$$

Again, let $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha})$ be the waiting time until the occurrence of the xth event among $E_{r_i}^{\varepsilon_i}(\alpha_i)$, $i = 1, \dots, \nu$. The probability generating function and the double generating function of $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha})$, $r_i \ge 0$, $i = 1, \dots, \nu$, are defined respectively by

$$\begin{aligned}H_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(t, x; \boldsymbol{\alpha}) &= \mathrm{E}[t^{T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x;\boldsymbol{\alpha})}] = \sum_{n=0}^{\infty} \mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha}) = n) t^n, \\ H^{\boldsymbol{\varepsilon}}(t, \boldsymbol{z}, x; \boldsymbol{\alpha}) &= \sum_{r_1,\dots,r_\nu \ge 0} H_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(t, x; \boldsymbol{\alpha}) z_1^{r_1} \cdots z_\nu^{r_\nu} \\ &= \sum_{r_1,\dots,r_\nu \ge 0} \sum_{n=0}^{\infty} \mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha}) = n) t^n z_1^{r_1} \cdots z_\nu^{r_\nu}.\end{aligned}$$

3. Main results

In this section we study the distribution of $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha})$ using the distribution of $(X_n^{\varepsilon_1}(\alpha_1), \dots, X_n^{\varepsilon_\nu}(\alpha_\nu))$ and elucidate the relationship between the two distributions. We consider the three cases in which $x = 1$, $2 \le x \le \nu - 1$, and, respectively, $x = \nu$, and treat them separately.

3.1. Sooner waiting time distribution

Here we study the distribution of the sooner waiting time, $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1; \boldsymbol{\alpha})$. Note that the dual relationship between the random variables $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1; \boldsymbol{\alpha})$ and $(X_n^{\varepsilon_1}(\alpha_1), \dots, X_n^{\varepsilon_\nu}(\alpha_\nu))$, namely

$$\{T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1; \boldsymbol{\alpha}) > n\} \quad \Longleftrightarrow \quad \{X_n^{\varepsilon_1}(\alpha_1) < r_1, \ \dots, \ X_n^{\varepsilon_\nu}(\alpha_\nu) < r_\nu\},$$

gives the probability identity

$$\begin{aligned}\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1; \boldsymbol{\alpha}) = n) = {}& \mathrm{P}(X_{n-1}^{\varepsilon_1}(\alpha_1) < r_1, \ \dots, \ X_{n-1}^{\varepsilon_\nu}(\alpha_\nu) < r_\nu) \\ & - \mathrm{P}(X_n^{\varepsilon_1}(\alpha_1) < r_1, \ \dots, \ X_n^{\varepsilon_\nu}(\alpha_\nu) < r_\nu), \qquad n, r_1, \dots, r_\nu \ge 1. \quad (3.1)\end{aligned}$$

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press

We set

$$\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1;\boldsymbol{\alpha})=0)=\begin{cases}1 & \text{if } r_i=0 \text{ for some } i=1,\ldots,\nu,\\ 0 & \text{otherwise.}\end{cases} \tag{3.2}$$

Theorem 3.1. *The double generating function* $H^{\boldsymbol{\varepsilon}}(t,z,1;\boldsymbol{\alpha})$ *can be expressed in terms of the double generating function* $\Phi^{\boldsymbol{\varepsilon}}(z,t;\boldsymbol{\alpha})$ *as follows:*

$$H^{\boldsymbol{\varepsilon}}(t,z,1;\boldsymbol{\alpha})=\frac{1}{\prod_{i=1}^{\nu}(1-z_i)}\biggl(1-\prod_{i=1}^{\nu}z_i(1-t)\Phi^{\boldsymbol{\varepsilon}}(z,t;\boldsymbol{\alpha})\biggr). \tag{3.3}$$

Proof. By virtue of (3.1) and (3.2), we have

$$\begin{aligned}
&H^{\boldsymbol{\varepsilon}}(t,z,1;\boldsymbol{\alpha})\\
&\quad=\sum_{r_1,\ldots,r_\nu\ge 0}\sum_{n=0}^{\infty}\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1;\boldsymbol{\alpha})=n)t^n z_1^{r_1}\cdots z_\nu^{r_\nu}\\
&\quad=\sum_{r_1,\ldots,r_\nu\ge 0}\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1;\boldsymbol{\alpha})=0)z_1^{r_1}\cdots z_\nu^{r_\nu}\\
&\qquad+\sum_{r_1,\ldots,r_\nu\ge 1}\sum_{n=1}^{\infty}\sum_{\substack{0\le i_j\le r_j-1\\ j=1,\ldots,\nu}}\mathrm{P}(X_{n-1}^{\varepsilon_1}(\alpha_1)=i_1,\ \ldots,\ X_{n-1}^{\varepsilon_\nu}(\alpha_\nu)=i_\nu)t^n z_1^{r_1}\cdots z_\nu^{r_\nu}\\
&\qquad-\sum_{r_1,\ldots,r_\nu\ge 1}\sum_{n=1}^{\infty}\sum_{\substack{0\le i_j\le r_j-1\\ j=1,\ldots,\nu}}\mathrm{P}(X_{n}^{\varepsilon_1}(\alpha_1)=i_1,\ \ldots,\ X_{n}^{\varepsilon_\nu}(\alpha_\nu)=i_\nu)t^n z_1^{r_1}\cdots z_\nu^{r_\nu}.
\end{aligned}$$

Using (3.2), for the three terms on the right-hand side of this expression we obtain

$$\sum_{r_1,\ldots,r_\nu\ge 0}\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1;\boldsymbol{\alpha})=0)z_1^{r_1}\cdots z_\nu^{r_\nu}=\prod_{i=1}^{\nu}\frac{1}{1-z_i}-\prod_{i=1}^{\nu}\frac{z_i}{1-z_i},$$

$$\begin{aligned}
&\sum_{r_1,\ldots,r_\nu\ge 1}\sum_{n=1}^{\infty}\sum_{\substack{0\le i_j\le r_j-1\\ j=1,\ldots,\nu}}\mathrm{P}(X_{n-1}^{\varepsilon_1}(\alpha_1)=i_1,\ \ldots,\ X_{n-1}^{\varepsilon_\nu}(\alpha_\nu)=i_\nu)t^n z_1^{r_1}\cdots z_\nu^{r_\nu}\\
&\quad=\prod_{i=1}^{\nu}\frac{z_i}{1-z_i}\sum_{n=1}^{\infty}\sum_{i_1,\ldots,i_\nu\ge 0}\mathrm{P}(X_{n-1}^{\varepsilon_1}(\alpha_1)=i_1,\ \ldots,\ X_{n-1}^{\varepsilon_\nu}(\alpha_\nu)=i_\nu)t^n z_1^{i_1}\cdots z_\nu^{i_\nu}\\
&\quad=\prod_{i=1}^{\nu}\frac{z_i}{1-z_i}\sum_{n=1}^{\infty}\phi_{n-1}^{\boldsymbol{\varepsilon}}(z;\boldsymbol{\alpha})t^n,
\end{aligned}$$

$$\begin{aligned}
&\sum_{r_1,\ldots,r_\nu\ge 1}\sum_{n=1}^{\infty}\sum_{\substack{0\le i_j\le r_j-1\\ j=1,\ldots,\nu}}\mathrm{P}(X_{n}^{\varepsilon_1}(\alpha_1)=i_1,\ \ldots,\ X_{n}^{\varepsilon_\nu}(\alpha_\nu)=i_\nu)t^n z_1^{r_1}\cdots z_\nu^{r_\nu}\\
&\quad=\prod_{i=1}^{\nu}\frac{z_i}{1-z_i}\sum_{n=1}^{\infty}\phi_{n}^{\boldsymbol{\varepsilon}}(z;\boldsymbol{\alpha})t^n,
\end{aligned}$$

where we have interchanged the order of certain summations. The proof is thus complete.

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press

It should be noted that the inversion of (3.3) produces the following expression for $\Phi^{\boldsymbol{\varepsilon}}(\boldsymbol{z}, t; \boldsymbol{\alpha})$ in terms of $H^{\boldsymbol{\varepsilon}}(t, \boldsymbol{z}, 1; \boldsymbol{\alpha})$:

$$\Phi^{\boldsymbol{\varepsilon}}(\boldsymbol{z}, t; \boldsymbol{\alpha}) = \frac{1}{\prod_{i=1}^{\nu} z_i(1-t)}\biggl(1 - \prod_{i=1}^{\nu}(1-z_i)H^{\boldsymbol{\varepsilon}}(t, \boldsymbol{z}, 1; \boldsymbol{\alpha})\biggr).$$

3.2. Later waiting time distribution

Now we consider the distribution of the later waiting time, $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha})$. Note that the dual relationship between the random variables $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha})$ and $(X_n^{\varepsilon_1}(\alpha_1), \dots, X_n^{\varepsilon_\nu}(\alpha_\nu))$, namely

$$\{T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha}) \le n\} \quad \Longleftrightarrow \quad \{X_n^{\varepsilon_1}(\alpha_1) \ge r_1, \ \dots, \ X_n^{\varepsilon_\nu}(\alpha_\nu) \ge r_\nu\},$$

gives the probability identity

$$\begin{aligned} \mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha}) = n) &= \mathrm{P}(X_n^{\varepsilon_1}(\alpha_1) \ge r_1, \ \dots, \ X_n^{\varepsilon_\nu}(\alpha_\nu) \ge r_\nu) \\ &\quad - \mathrm{P}(X_{n-1}^{\varepsilon_1}(\alpha_1) \ge r_1, \ \dots, \ X_{n-1}^{\varepsilon_\nu}(\alpha_\nu) \ge r_\nu), \qquad n \ge 1, \ r_1, \dots, r_\nu \ge 0. \end{aligned} \tag{3.4}$$

We set

$$\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha}) = 0) = \begin{cases} 1 & \text{if } r_i = 0 \text{ for all } i = 1, \dots, \nu, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Using (3.4) and (3.5) and working in the same fashion as we did in the proof of Theorem 3.1, we arrive at the following theorem.

Theorem 3.2. *The double generating function $H^{\boldsymbol{\varepsilon}}(t, \boldsymbol{z}, \nu; \boldsymbol{\alpha})$ can be expressed in terms of the double generating functions $\Phi^{\varepsilon_{i_1}, \dots, \varepsilon_{i_j}}(z_{i_1}, \dots, z_{i_j}, t; \alpha_{i_1}, \dots, \alpha_{i_j})$, $j = 1, \dots, \nu$, as follows:*

$$\begin{aligned} H^{\boldsymbol{\varepsilon}}(t, \boldsymbol{z}, \nu; \boldsymbol{\alpha}) = \frac{1}{\prod_{i=1}^{\nu}(1-z_i)}\biggl(1 + \sum_{j=1}^{\nu}(-1)^j \sum_{1 \le i_1 < \cdots < i_j \le \nu} \prod_{u=1}^{j} z_{i_u}(1-t) \\ \times \Phi^{\varepsilon_{i_1}, \dots, \varepsilon_{i_j}}(z_{i_1}, \dots, z_{i_j}, t; \alpha_{i_1}, \dots, \alpha_{i_j})\biggr). \end{aligned} \tag{3.6}$$

It is noteworthy that the inversion of (3.6) produces the following expression for $\Phi^{\boldsymbol{\varepsilon}}(\boldsymbol{z}, t; \boldsymbol{\alpha})$ in terms of $H^{\varepsilon_{i_1}, \dots, \varepsilon_{i_j}}(t, z_{i_1}, \dots, z_{i_j}, j; \alpha_{i_1}, \dots, \alpha_{i_j})$, $j = 1, \dots, \nu$, the double generating functions of the later waiting times $T_{r_{i_1}, \dots, r_{i_j}}^{\varepsilon_{i_1}, \dots, \varepsilon_{i_j}}(j; \alpha_{i_1}, \dots, \alpha_{i_j})$:

$$\begin{aligned} \Phi^{\boldsymbol{\varepsilon}}(\boldsymbol{z}, t; \boldsymbol{\alpha}) = \frac{1}{\prod_{i=1}^{\nu} z_i(1-t)}\biggl(1 + \sum_{j=1}^{\nu}(-1)^j \sum_{1 \le i_1 < \cdots < i_j \le \nu} \prod_{k=1}^{j}(1 - z_{i_k}) \\ \times H^{\varepsilon_{i_1}, \dots, \varepsilon_{i_j}}(t, z_{i_1}, \dots, z_{i_j}, j; \alpha_{i_1}, \dots, \alpha_{i_j})\biggr). \end{aligned}$$

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press

3.3. The waiting time for the occurrence of the xth event

Finally, in this section we examine the distribution of the waiting time $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha})$ for $1 \le x \le \nu$. Observe that

$$
\begin{aligned}
&\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha}) \le n) \\
&\quad = \sum_{j=x}^{\nu} \sum_{\substack{1 \le i_1 < \cdots < i_j \le \nu \\ \{i_{j+1}, \dots, i_\nu\} \subset \{1,\dots,\nu\} \setminus \{i_1,\dots,i_j\}}} \mathrm{P}(X_n^{\varepsilon_{i_1}}(\alpha_{i_1}) \ge r_{i_1}, \; \dots, \; X_n^{\varepsilon_{i_j}}(\alpha_{i_j}) \ge r_{i_j}, \\
&\qquad\qquad\qquad\qquad X_n^{\varepsilon_{i_{j+1}}}(\alpha_{i_{j+1}}) < r_{i_{j+1}}, \; \dots, \; X_n^{\varepsilon_{i_\nu}}(\alpha_{i_\nu}) < r_{i_\nu}) \\
&\quad = \sum_{j=x}^{\nu} \sum_{w=j}^{\nu} \sum_{1 \le i_1 < \cdots < i_w \le \nu} (-1)^{w-j} \binom{w}{j} \mathrm{P}(X_n^{\varepsilon_{i_1}}(\alpha_{i_1}) \ge r_{i_1}, \; \dots, \; X_n^{\varepsilon_{i_w}}(\alpha_{i_w}) \ge r_{i_w}),
\end{aligned}
$$

yielding

$$
\begin{aligned}
\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha}) = n) = \sum_{j=x}^{\nu} \sum_{w=j}^{\nu} & \sum_{1 \le i_1 < \cdots < i_w \le \nu} (-1)^{w-j} \binom{w}{j} \\
& \times (\mathrm{P}(X_n^{\varepsilon_{i_1}}(\alpha_{i_1}) \ge r_{i_1}, \; \dots, \; X_n^{\varepsilon_{i_w}}(\alpha_{i_w}) \ge r_{i_w}) \\
& \quad - \mathrm{P}(X_{n-1}^{\varepsilon_{i_1}}(\alpha_{i_1}) \ge r_{i_1}, \; \dots, \; X_{n-1}^{\varepsilon_{i_w}}(\alpha_{i_w}) \ge r_{i_w})), \\
& \qquad n \ge 1, \; r_1, \dots, r_\nu \ge 0. \qquad (3.7)
\end{aligned}
$$

We set

$$
\begin{aligned}
&\mathrm{P}(T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(x; \boldsymbol{\alpha}) = 0) \\
&\quad = \begin{cases} 1 & \text{if } r_{i_1} = \cdots = r_{i_j} = 0 \text{ for } 1 \le i_1 < \cdots < i_j \le \nu \text{ and } j = x, x+1, \dots, \nu, \\ 0 & \text{otherwise.} \end{cases} \qquad (3.8)
\end{aligned}
$$

Using (3.7) and (3.8), we can express the double generating function $H^{\boldsymbol{\varepsilon}}(t, z, x; \boldsymbol{\alpha})$ in terms of the double generating functions $\Phi^{\varepsilon_{i_1}, \dots, \varepsilon_{i_j}}(z_{i_1}, \dots, z_{i_j}, t; \alpha_{i_1}, \dots, \alpha_{i_j})$, $j = 1, \dots, \nu$. Working in a similar fashion to in Sections 3.1 and 3.2, we can establish a formula for the double generating function. The details can be worked out easily and are thus omitted here.

Theorem 3.3. *The double generating function $H^{\boldsymbol{\varepsilon}}(t, z, x; \boldsymbol{\alpha})$ is given, for $x = 1, \dots, \nu$, by*

$$
\begin{aligned}
H^{\boldsymbol{\varepsilon}}(t, z, x; \boldsymbol{\alpha}) = 1 &+ \sum_{j=1}^{\nu - x} \sum_{1 \le i_1 < \cdots < i_j \le \nu} \prod_{u=1}^{j} \frac{z_{i_u}}{1 - z_{i_u}} \\
&+ \frac{1}{\prod_{i=1}^{\nu}(1 - z_i)} \sum_{j=x}^{\nu} \sum_{w=j}^{\nu} \sum_{v=1}^{w} \sum_{1 \le i_1 < \cdots < i_v \le \nu} (-1)^{w-j+v} \binom{\nu - v}{\nu - w} \binom{w}{j} \\
&\qquad \times \prod_{u=1}^{v} z_{i_u} ((1 - t)\, \Phi^{\varepsilon_{i_1}, \dots, \varepsilon_{i_v}}(z_{i_1}, \dots, z_{i_v}, t; \alpha_{i_1}, \dots, \alpha_{i_v}) - 1). \\
&\qquad\qquad (3.9)
\end{aligned}
$$

Needless to say, for $x = 1$ and $x = \nu$, (3.9) corresponds to (3.3) and (3.6), respectively, giving alternative formulae for the double generating functions $H^{\boldsymbol{\varepsilon}}(t, z, 1; \boldsymbol{\alpha})$ and $H^{\boldsymbol{\varepsilon}}(t, z, \nu; \boldsymbol{\alpha})$.

4. Applications

For a sequence of Bernoulli trials (with two possible outcomes, '1' or '0'), in the literature there are various ways of counting the number of '1'-runs of length k (see Fu and Koutras (1994) and Balakrishnan and Koutras (2002)). The important and frequently used ways of counting the number of '1'-runs of length k are as follows:

(i) the type-I enumeration scheme, namely the way of counting the number of nonoverlapping and recurrent '1'-runs of length k, in the sense of Feller's (1968) counting;

(ii) the type-II enumeration scheme, namely the way of counting the number of '1'-runs of length at least k, in the sense of Goldstein's (1990) counting;

(iii) the type-III enumeration scheme, namely the way of counting the number of overlapping '1'-runs of length k, in the sense of Ling's (1988) counting,

For $i = 1, 2, \dots, \nu$, let $\varepsilon_i = \{(i, i, \dots, i)\}$ be the 'i'-run of length k_i. As stated previously, α_i represents the scheme employed in counting the 'i'-run of length k_i and here takes the value I, II, or III as appropriate.

We will propose extensions to the sooner and later waiting time problems. In this section we assume that $Z_1, Z_2, \dots$ are independent and identically distributed random variables taking values in $\Gamma = \{0, 1, \dots, m\}$ with probabilities $p_i = \Pr(Z_t = i)$, $1 \le t$, $i = 0, 1, \dots, m$.

4.1. Sooner waiting time distributions for runs

For the 'i'-run of length k_i, $i = 1, \dots, \nu$, we will study the distribution of the sooner waiting time, $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(1; \boldsymbol{\alpha})$, using the three different counting schemes (types I, II, and III). For $\nu = 2$, the corresponding waiting time distribution is known as the *type-*(α_1, α_2) *sooner negative binomial distribution of order* (k_1, k_2) and has been studied by several authors (see Ebneshahrashoob and Sobel (1990), Aki *et al.* (1996), Aki and Hirano (1993), Uchida and Aki (1995), Han and Aki (2000), and Balakrishnan and Koutras (2002)).

Inoue and Aki (2005) derived the double generating function of $(X_n^{\varepsilon_1}(\alpha_1), \dots, X_n^{\varepsilon_\nu}(\alpha_\nu))$ and found it to be

$$\Phi^{\boldsymbol{\varepsilon}}(\boldsymbol{z}, t; \boldsymbol{\alpha}) = \frac{1}{1 - p_0 t - \sum_{i=1}^{\nu} Q(z_i, p_i t, \alpha_i)}, \tag{4.1}$$

where

$$Q(z_i, p_i t, \alpha_i) = \begin{cases} \dfrac{p_i t - (p_i t)^{k_i} + (p_i t)^{k_i} z_i (1 - p_i t)}{1 - (p_i t)^{k_i}}, & \alpha_i = \mathrm{I}, \\[2ex] \dfrac{p_i t - (p_i t)^{k_i}(1 - z_i)}{1 - (p_i t)^{k_i}(1 - z_i)}, & \alpha_i = \mathrm{II}, \\[2ex] \dfrac{p_i t - (p_i t)^{k_i}(1 - z_i) - (p_i t)^2 z_i}{1 - p_i t z_i - (p_i t)^{k_i}(1 - z_i)}, & \alpha_i = \mathrm{III}, \end{cases} \tag{4.2}$$

for $i = 1, 2, \dots, \nu$. From this, we obtain the following proposition.

Proposition 4.1. *The double generating function* $H^{\boldsymbol{\varepsilon}}(t, \boldsymbol{z}, 1; \boldsymbol{\alpha})$ *is given by*

$$H^{\boldsymbol{\varepsilon}}(t, \boldsymbol{z}, 1; \boldsymbol{\alpha}) = \frac{1}{\prod_i^{\nu}(1 - z_i)} \left(1 - \frac{\prod_{i=1}^{\nu} z_i (1 - t)}{1 - p_0 t - \sum_{i=1}^{\nu} Q(z_i, p_i t, \alpha_i)}\right), \tag{4.3}$$

where $Q(z_i, p_i t, \alpha_i)$, $\alpha_i = \mathrm{I}, \mathrm{II}, \mathrm{III}$, $i = 1, \dots, \nu$, *is as defined in (4.2).*

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press

By expanding the double generating function (4.3) in a Taylor series around $\boldsymbol{z} = \boldsymbol{0}$ and picking out the coefficient of $z_1 \cdots z_\nu$, we obtain the following explicit expression for the probability generating function $H^{\boldsymbol{\varepsilon}}_{1,\dots,1}(t, 1; \boldsymbol{\alpha})$, known as the *sooner geometric distribution of order* $(k_1, k_2, \dots, k_\nu)$:

$$H^{\boldsymbol{\varepsilon}}_{1,\dots,1}(t, 1; \boldsymbol{\alpha}) = \frac{\sum_{i=1}^{\nu} (p_i t)^{k_i}(1 - p_i t)/(1 - (p_i t)^{k_i})}{1 - t + \sum_{i=1}^{\nu}(p_i t)^{k_i}(1 - p_i t)/(1 - (p_i t)^{k_i})}. \tag{4.4}$$

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

Remark 4.1. Aki and Hirano (2000) introduced a generalized enumeration scheme known as ℓ-overlapping counting (see also Inoue and Aki (2003)). By setting

$$Q(z_i, p_i t, \alpha_i) = \frac{p_i t - (p_i t)^{k_i} + (p_i t)^{k_i} z_i - (p_i t)^{k_i - \ell_i - 1} z_i}{1 - (p_i t)^{k_i} + (p_i t)^{k_i} z_i - (p_i t)^{k_i - \ell_i} z_i}, \qquad 0 \le \ell_i \le k_i - 1,$$

in (4.1), the results presented in this section can easily be extended to cover this case.

Example 4.1. (*Birthday problem.*) Assume that $p_0 = 0$ and $\varepsilon_i = \{(i)\}$, $i = 1, \dots, \nu$. By expanding the double generating function (4.3) in a Taylor series around $\boldsymbol{z} = \boldsymbol{0}$ and picking out the coefficient of $z_1^{r_1} \cdots z_\nu^{r_\nu}$, we obtain an explicit expression for the probability generating function $H^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(t, 1; \boldsymbol{\alpha})$. Furthermore, we obtain an explicit expression for the expected value of the waiting time $T^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(1; \boldsymbol{\alpha})$ by differentiating $H^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(t, 1; \boldsymbol{\alpha})$ with respect to t. We find that

$$H^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(t, 1; \boldsymbol{\alpha}) = 1 - \sum_{\substack{0 \le i_j \le r_j - 1 \\ j = 1, \dots, \nu}} \binom{i_1 + i_2 + \cdots + i_\nu}{i_1, i_2, \dots, i_\nu} p_1^{i_1} p_2^{i_2} \cdots p_\nu^{i_\nu} (1 - t) t^{i_1 + i_2 + \cdots + i_\nu}, \tag{4.5}$$

$$\mathrm{E}[T^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(1; \boldsymbol{\alpha})] = \sum_{\substack{0 \le i_j \le r_j - 1 \\ j = 1, \dots, \nu}} \binom{i_1 + i_2 + \cdots + i_\nu}{i_1, i_2, \dots, i_\nu} p_1^{i_1} p_2^{i_2} \cdots p_\nu^{i_\nu}. \tag{4.6}$$

For $\nu = 365$, $p_1 = \cdots = p_{365} = \frac{1}{365}$, and $r_1 = \cdots = r_{365} = r \ge 2$, the waiting time problem is known as the *birthday problem*. Suppose that we interview people at random, one by one, until we find r people with a common birthday. How many people will we have to interview? The case in which $r = 2$ has been investigated by many authors (see, for example, Johnson and Kotz (1977) and references therein). However, there are relatively few papers dealing with the general case ($r > 2$) and general arbitrary probabilities $p_1, \dots, p_\nu$. Equations (4.5) and (4.6) provide useful clues to the general birthday problems. In Table 1 we present illustrative numerical results for the expected value $\mathrm{E}[T^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(1; \boldsymbol{\alpha})]$.

Klamkin and Newman (1967) gave the following asymptotic expression for $\mathrm{E}[T^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(1; \boldsymbol{\alpha})]$:

$$\mathrm{E}[T^{\boldsymbol{\varepsilon}}_{\boldsymbol{r}}(1; \boldsymbol{\alpha})] \sim (r!)^{1/r} \Gamma\biggl(1 + \frac{1}{r}\biggr) 365^{1 - 1/r}.$$

For $r = 3, 4, 5$, their expression gives the respective values 82.87, 167.53, and 268.28. Comparing these results with those in Table 1 suggests that the asymptotic expression is not particularly accurate.

TABLE 1: The expected value of $T_r^{\varepsilon}(1; \boldsymbol{\alpha})$ in the birthday problem.

r	$\mathrm{E}[T_r^{\varepsilon}(1; \boldsymbol{\alpha})]$
2	24.62
3	88.74
4	187.05
5	311.45
6	456.02
7	616.62
8	790.30
9	974.89

4.2. Later waiting time distributions for runs

In this section we consider the distribution of the later waiting time, $T_r^{\varepsilon}(\nu; \boldsymbol{\alpha})$, under the three different counting schemes (types I, II, and III). For $\nu = 2$, the corresponding waiting time distribution is known as the *type-(α_1, α_2) later negative binomial distribution of order* (k_1, k_2).

Proposition 4.2. *The double generating function $H^{\varepsilon}(t, \boldsymbol{z}, \nu; \boldsymbol{\alpha})$ is given by*

$$H^{\varepsilon}(t, \boldsymbol{z}, \nu; \boldsymbol{\alpha}) = \frac{1}{\prod_{i=1}^{\nu}(1 - z_i)} \Bigg(1 + \sum_{j=1}^{\nu}(-1)^j \sum_{1 \le i_1 < \cdots < i_j \le \nu} \prod_{u=1}^{j} z_{i_u}(1 - t) \times \frac{1}{1 - (1 - \sum_{u=1}^{j} p_{i_u})t - \sum_{u=1}^{j} Q(z_{i_u}, p_{i_u}t, \alpha_{i_u})}\Bigg), \tag{4.7}$$

where $Q(z_i, p_i t, \alpha_i)$, $\alpha_i = \mathrm{I, II, III}$, $i = 1, \ldots, \nu$, is as defined in (4.2).

By expanding the double generating function (4.7) in a Taylor series around $\boldsymbol{z} = \mathbf{0}$ and picking out the coefficient of $z_1 \cdots z_\nu$, we obtain the following explicit expressions for the probability generating function $H_{1,\ldots,1}^{\varepsilon}(t, \nu; \boldsymbol{\alpha})$, known as the *later geometric distribution of order* $(k_1, k_2, \ldots, k_\nu)$:

$$\begin{aligned} &H_{1,\ldots,1}^{\varepsilon}(t, \nu; \boldsymbol{\alpha}) \\ &\quad = 1 + \sum_{j=1}^{\nu}(-1)^j \sum_{1 \le i_1 < \cdots < i_j \le \nu} \frac{1 - t}{1 - t + \sum_{u=1}^{j}(p_{i_u}t)^{k_{i_u}}(1 - p_{i_u}t)/(1 - (p_{i_u}t)^{k_{i_u}})}, \end{aligned} \tag{4.8}$$

or, equivalently,

$$\begin{aligned} &H_{1,\ldots,1}^{\varepsilon}(t, \nu; \boldsymbol{\alpha}) \\ &\quad = \sum_{j=1}^{\nu}(-1)^{j-1} \sum_{1 \le i_1 < \cdots < i_j \le \nu} \frac{\sum_{u=1}^{j}(p_{i_u}t)^{k_{i_u}}(1 - p_{i_u}t)/(1 - (p_{i_u}t)^{k_{i_u}})}{1 - t + \sum_{u=1}^{j}(p_{i_u}t)^{k_{i_u}}(1 - p_{i_u}t)/(1 - (p_{i_u}t)^{k_{i_u}})}. \end{aligned} \tag{4.9}$$

Balakrishnan and Koutras (2002) have also given formulae for the probability generating functions (4.4), (4.8), and (4.9). Their derivation was based on a completely different technique.

Example 4.2. (*Coupon collector problems.*) Assume that $p_0 = 0$ and $\varepsilon_i = \{(i)\}$, $i = 1, \dots, \nu$. By expanding the double generating function (4.7) in a Taylor series around $z = \mathbf{0}$ and picking out the coefficient of $z_1^{r_1} \cdots z_\nu^{r_\nu}$, we obtain an explicit expression for the probability generating function $H_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(t, \nu; \boldsymbol{\alpha})$. Furthermore, we obtain an explicit expression for the expected value of the waiting time $T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha})$ by differentiating $H_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(t, \nu; \boldsymbol{\alpha})$ with respect to t. We find that

$$\begin{aligned} H_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(t, z, \nu; \boldsymbol{\alpha}) = 1 + \sum_{j=1}^{\nu} (-1)^j \sum_{1 \le i_1 < \cdots < i_j \le \nu} \sum_{\substack{0 \le y_{i_u} \le r_{i_u} - 1 \\ u = 1, \dots, j}} \binom{y_{i_1} + \cdots + y_{i_j}}{y_{i_1}, \dots, y_{i_j}} & \\ \times \frac{p_{i_1}^{y_{i_1}} \cdots p_{i_j}^{y_{i_j}} (1-t) t^{y_{i_1} + \cdots + y_{i_j}}}{(1 - (1 - \sum_{u=1}^{j} p_{i_u}) t)^{y_{i_1} + \cdots + y_{i_j} + 1}}, & \end{aligned} \tag{4.10}$$

$$\begin{aligned} \mathrm{E}[T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha})] = \sum_{j=1}^{\nu} (-1)^{j+1} \sum_{1 \le i_1 < \cdots < i_j \le \nu} \sum_{\substack{0 \le y_{i_u} \le r_{i_u} - 1 \\ u = 1, \dots, j}} \binom{y_{i_1} + \cdots + y_{i_j}}{y_{i_1}, \dots, y_{i_j}} & \\ \times \frac{p_{i_1}^{y_{i_1}} \cdots p_{i_j}^{y_{i_j}}}{(p_{i_1} + \cdots + p_{i_j})^{y_{i_1} + \cdots + y_{i_j} + 1}}. & \end{aligned} \tag{4.11}$$

Assume that $r_1 = \cdots = r_\nu = 1$. The waiting time problem in this case is well known as the *coupon collector problem*: suppose that there are ν distinct types of coupons bearing the numbers $1, \dots, \nu$ and that the coupon of type i is collected with probability p_i, $i = 1, \dots, \nu$. We are interested in the number of coupons we need to collect in order to have at least one of each type. From (4.10) and (4.11), we have

$$H_{1,\dots,1}^{\boldsymbol{\varepsilon}}(t, \nu; \boldsymbol{\alpha}) = 1 + (-1)^\nu (1-t) + \sum_{j=1}^{\nu} \sum_{1 \le i_1 < i_2 < \cdots < i_j \le \nu} \frac{(-1)^{\nu - j}(1-t)}{1 - (p_{i_1} + \cdots + p_{i_j}) t},$$

$$\mathrm{E}[T_{\boldsymbol{r}}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha})] = \sum_{j=1}^{\nu} \sum_{1 \le i_1 < i_2 < \cdots < i_j \le \nu} \frac{(-1)^{j+1}}{p_{i_1} + p_{i_2} + \cdots + p_{i_j}}.$$

In closing, we would like to mention a generalization of the coupon collector problem which is called the *coupon subset collection problem* (see Adler and Ross (2001)). Let $S = \{1, \dots, m\}$, and let $S_i = \{s_{i,1}, \dots, s_{i,c_i}\}$, $i = 1, \dots, \nu$, be subsets of S such that $\bigcup_{i=1}^{\nu} S_i = S$. Suppose that we may choose subset S_i with probability q_i, $i = 1, \dots, \nu$, with $\sum_{i=1}^{\nu} q_i = 1$. We are interested in the number of subsets that must be chosen before the coupons $1, \dots, m$ are all contained in at least one of these subsets. Clearly, when all subsets are of size 1 the waiting time problem corresponds to the coupon collector problem.

Since we readily find the double generating function to be

$$\Phi^{\boldsymbol{\varepsilon}}(z, t; \boldsymbol{\alpha}) = \frac{1}{1 - \left(\sum_{i=1}^{m} q_i \prod_{j=1}^{m} z_j^{\mathbf{1}(s_{i,1} = j) + \cdots + \mathbf{1}(s_{i,c_i} = j)} \right) t},$$

where $\mathbf{1}(\cdots)$ denotes the indicator function, the distribution of the waiting time $T_{1,\dots,1}^{\boldsymbol{\varepsilon}}(\nu; \boldsymbol{\alpha})$ can be easily evaluated by direct application of Theorem 3.2.

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press

References

Adler, I. and Ross, S. (2001). The coupon subset collection problem. *J. Appl. Prob.* **38,** 737–746.

Aki, S. and Hirano, K. (1993). Discrete distributions related to succession events in a two-state Markov chain. In *Statistical Sciences and Data Analysis* (Proc. Third Pacific Area Statist. Conf.), eds K. Matusita, M. L. Puri and T. Hayakawa, VSP International Science Publishers, Zeist, pp. 467–474.

Aki, S. and Hirano, K. (2000). Numbers of success-runs of specified length until certain stopping time rules and generalized binomial distributions of order k. *Ann. Inst. Statist. Math.* **52,** 767–777.

Aki, S., Balakrishnan, N. and Mohanty, S. G. (1996). Sooner and later waiting time problems and failure runs in higher order Markov dependent trials. *Ann. Inst. Statist. Math.* **48,** 773–787.

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. *Ann. Inst. Statist. Math.* **49,** 155–169.

Ebneshahrashoob, M. and Sobel, M. (1990). Sooner and later waiting time problems for Bernoulli trials: frequency and run quotas. *Statist. Prob. Lett.* **9,** 5–11.

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

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

Fu, J. C. and Chang, Y. M. (2003). On ordered series and later waiting time distributions in a sequence of Markov dependent multistate trials. *J. Appl. Prob.* **40,** 623–642.

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

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.

Goldstein, L. (1990). Poisson approximation and DNA sequence matching. *Commun. Statist. Theory Meth.* **19,** 4167–4179.

Han, Q. and Aki, S. (2000). Waiting time problems in a two-state Markov chain. *Ann. Inst. Statist. Math.* **52,** 778–789.

Inoue, K. (2004). Joint distributions associated with patterns, successes and failures in a sequence of multi-state trials. *Ann. Inst. Statist. Math.* **56,** 143–168.

Inoue, K. and Aki, S. (2002). Generalized waiting time problems associated with pattern in Pólya's urn scheme. *Ann. Inst. Statist. Math.* **54,** 681–688.

Inoue, K. and Aki, S. (2003). Generalized binomial and negative binomial distributions of order k by the ℓ-overlapping enumeration scheme. *Ann. Inst. Statist. Math.* **55,** 153–167.

Inoue, K. and Aki, S. (2005). A generalized Pólya urn model and related multivariate distributions. *Ann. Inst. Statist. Math.* **57,** 49–59.

Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Applications*. John Wiley, New York.

Klamkin, M. S. and Newman, D. J. (1967). Extensions of the birthday surprise. *J. Combinatorial Theory* **3,** 279–282.

Koutras, M. V. (1997). Waiting times and number of appearances of events in a sequence of discrete random variables. In *Advances in Combinatorial Methods and Applications to Probability and Statistics*, ed. N. Balakrishnan, Birkhäuser, Boston, MA, pp. 363–384.

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

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

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

Uchida, M. and Aki, S. (1995). Sooner and later waiting time problems in a two-state Markov chain. *Ann. Inst. Statist. Math.* **47,** 415–433.

https://doi.org/10.1239/jap/1175267164 Published online by Cambridge University Press