Josai Mathematical Monographs
vol. 8 (2015), pp. 3-16

Distributions of numbers of runs and scans on higher order Markov directed acyclic graphs with generation

Kiyoshi INOUE

Abstract. In this paper, we introduce a class of a directed acyclic graph on the assumption that the collection of random variables indexed by the vertices has a higher order Markov dependence. A method for the study of the exact distributions of the numbers of runs and scans on the higher order Markov directed acyclic graph is presented by the method of conditional probability generating functions. We show that our theoretical results can easily be carried out through some computer algebra systems and give some numerical results for run and scan statistics in order to demonstrate the feasibility of our theoretical results. As applications, two special reliability systems are considered, which are closely related to our general results. Finally, we address the parameter estimation problems in the distributions of runs and scans through the maximum likelihood estimation.

Key words and phrases: Run, scan, overlapping enumeration scheme, graphical model, directed acyclic graph, Markov property, reliability, parameter estimation, probability generating function.

1. Introduction

The distribution theory of runs and scans has been successfully developed by many authors. Several generalizations of the run and scan statistics can be found in the statistical literature and have been useful in many applications, such as reliability theory, statistical quality control, molecular biology and epidemiology (see Feller [**8**], Ebneshahrashoob and Sobel [**7**], Fu and Koutras [**9**], Fu nad Lou [**10**], Koutras and Alexandrou [**17**], Balakrishnan and Koutras [**4**], Glaz and Balakrishnan [**11**], Glaz *et al.* [**12**], [**13**]. In recent years, considerable attention is paid to the theory of runs and scans in the graphical models. Especially, the theory and applications in directed acyclic graphs (DAG's) have been of great interest to researchers in a wide range of subjects (see Aki [**1**], [**2**], Inoue and Aki [**14**], [**16**] and Jonczy and Haenni [**18**]).

In this paper, we introduce a higher order Markov *directed acyclic graph with generation* (DAG with generation) and present a method for deriving the exact distributions of the numbers of runs and scans on the DAG with generation by extensive use of the method of conditional probability generating functions (p.g.f.'s).

In Section 2, we introduce the concept of the m-th order Markov DAG with generation. We also introduce all necessary definitions and notations that will be used remaining sections. In Section 3, we give a general result for deriving the exact distributions of runs and scans on the DAG with generation by extending the method of conditional p.g.f.'s. Our method presented in the paper will provide a feasible algorithm for the distributions of runs and scans. Our model is closely related to some interesting reliability systems called consecutive-k-out-of-n:F and k-within-consecutive-w-out-of-n:F (see Aki and Hirano [**3**], Inoue and Aki [**15**], [**16**], Chao *et al.* [**5**] and Chang *et al.* [**6**]). In Section 4, some applications to practical problems in the reliability theory are given in order to show our theoretical results, which illustrate the potential use of run and scan statistics. In Section 5, we address the parameter estimation in the distributions of runs and scans.

2. Definitions and notations

According to Lauritzen [**19**] and Ripley [**21**], we shall use some basic notations in graphical models. Let $G = (V, E)$ be a directed acyclic graph, which is often abbreviated as DAG, where V is a finite set of vertices and the set of edges E is a subset of the set $V \times V$ of ordered pairs of distinct vertices. The edges are all directed. Here, an edge $(u, v) \in E$ is directed if $(v, u) \notin E$. The condition of acyclic means that there does not exist a sequence of vertices $v_1, ..., v_n$ such that (v_1, v_2), (v_2, v_3),...,(v_n, v_1) are edges in E.

We will define the directed acyclic graph with generation (see Inoue and Aki [**16**]).

DEFINITION 1. (Directed acyclic graph with generation)

The directed acyclic graph will be called a directed acyclic graph with generation if

(i) *for given positive integers* h_1,...,h_n, *there exists a sequence of disjoint sets* $V(1), V(2), ..., V(g)$ *of the vertices such as* $V(i) = \{v_1(i), ..., v_{h_i}(i)\}$, $i = 1, 2, ..., g$ *and* $V = \cup_{i=1}^{g} V(i)$, *(convention:* $V(i) = \varnothing$ *and* $h_i = 0$ *for* $i \leq 0$*),*

(ii) $(u, v) \notin E$, *for every* $u \in V(i)$, $v \in V(j)$, $i \geq j$, $i, j = 1, 2, ..., g$,

(iii) $(v, u) \notin E$, *for every* $u \in V(i)$, $v \in V(j)$, $i - j \geq 2$, $i, j = 1, 2, ..., g$.

For $i = 1, 2, ..., g$, *each set* $V(i)$ *is called* i*-th generation.*

Suppose that we have a collection of $\{0, 1\}$-valued random variables $\{X_v, v \in V\}$ (we say success and failure for the outcomes "1" and "0", respectively). Let $\boldsymbol{X}(n) = (X_{v_1}(n), ..., X_{v_{h_n}}(n))$ for $n = 1, 2, ..., g$ and let $E(d) = \{0, 1\}^d$ for $d = 1, 2, ...,$.

We will define a higher order Markov directed acyclic graph with generation.

DEFINITION 2. (the m-th order Markov directed acyclic graph with generation)

The collection of random variables $\{\boldsymbol{X}(n), n = 1, 2, ..., g\}$ *will be called a homogeneous* m*-th order Markov directed acyclic graph with generation if*

(i) *there exists the initial distribution at first generation* $P(\boldsymbol{X}(1) = e(1)) = p(e(1))$, $e(1) \in E(h_1)$
(ii) *for every sequence* $e(i) = (e_1(i), ..., e_{h_i}(i)) \in E(h_i)$, $i = 2, ..., g$, *there exists the conditional probability which satisfies the following condition:*
for every $i = 2, 3, ..., g$, *if* $i \leq m$, *then*

$$P\left(\boldsymbol{X}(i) = e(i) \mid \boldsymbol{X}(1) = e(1), ..., \boldsymbol{X}(i-1) = e(i-1)\right) = p\left(e(i) \mid e(1), ..., e(i-1)\right),$$

for every $i = 2, 3, ..., g$, *if* $i > m$, *then*

$$\begin{aligned} &P\left(\boldsymbol{X}(i) = e(i) \mid \boldsymbol{X}(1) = e(1), ..., \boldsymbol{X}(i-1) = e(i-1)\right) \\ &= P\left(\boldsymbol{X}(i) = e(i) \mid \boldsymbol{X}(i-m) = e(i-m), ..., \boldsymbol{X}(i-1) = e(i-1)\right) \\ &= p\left(e(i) \mid e(i-m), ..., e(i-1)\right). \end{aligned}$$

The distributions of the numbers of runs and scans are investigated based on the overlapping enumeration scheme (Type III enumeration scheme). We enumerate the number of overlapping success runs of length k in the sense of Ling's [**20**] counting and enumerate the number of overlapping windows of length w containing at least k successes each, observed on the DAG with generation (see Inoue and Aki [**16**]). It is evident that the scan statistics reduce to the run statistics in the special case when $w = k$.

To make the definition of overlapping enumeration scheme transparent to the reader, we illustrate how to enumerate the number of overlapping success runs and the number of overlapping scans on the DAG with generation by using Figure 1.

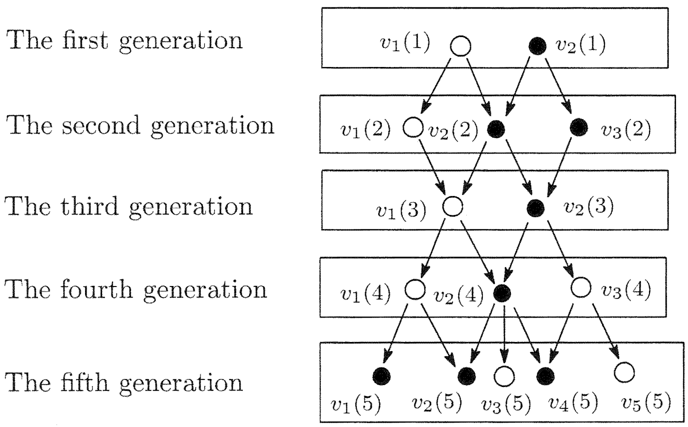

Fig. 1. An example to illustrate overlapping enumeration scheme on a DAG with generation.

Then we can find out the following 6 overlapping "•"-runs of length 3 on DAG with generation in Figure 1: $R_1 = (v_2(1), v_2(2), v_2(3))$, $R_2 = (v_2(1), v_3(2), v_2(3))$, $R_3 = (v_2(2), v_2(3), v_2(4))$, $R_4 = (v_3(2), v_2(3), v_2(4))$, $R_5 = (v_2(3), v_2(4), v_2(5))$, and

$R_6 = (v_2(3), v_2(4), v_4(5))$. For example, the case where $(w, k) = (4, 3)$ is considered. Then we can find out the following 16 overlapping scans on DAG with generation in Figure 1: $S_1 = (v_1(1), v_2(2), v_2(3), v_2(4))$, $S_2 = (v_2(1), v_2(2), v_1(3), v_2(4))$, $S_3 = (v_2(1), v_2(2), v_2(3), v_2(4))$, $S_4 = (v_2(1), v_2(2), v_2(3), v_3(4))$, $S_5 = (v_2(1), v_3(2), v_2(3), v_2(4))$, $S_6 = (v_2(1), v_3(2), v_2(3), v_3(4))$, $S_7 = (v_2(2), v_1(3), v_2(4), v_2(5))$, $S_8 = (v_2(2), v_1(3), v_2(4), v_4(5))$, $S_9 = (v_2(2), v_2(3), v_2(4), v_2(5))$, $S_{10} = (v_2(2), v_2(3), v_2(4), v_3(5))$, $S_{11} = (v_2(2), v_2(3), v_2(4), v_4(5))$, $S_{12} = (v_2(2), v_2(3), v_3(4), v_4(5))$, $S_{13} = (v_3(2), v_2(3), v_2(4), v_2(5))$, $S_{14} = (v_3(2), v_2(3), v_2(4), v_3(5))$, $S_{15} = (v_3(2), v_2(3), v_2(4), v_4(5))$ and $S_{16} = (v_3(2), v_2(3), v_3(4), v_4(5))$.

3. Distributions of runs and scans

To begin with, we study the distribution of the number of the overlapping scans on the m-th order Markov DAG with generation. Let $\phi^{(w,k)}(t)$ be the probability generating function (p.g.f.) of the distribution of the number of overlapping scans. Let $w^* = \max\{w, m+1\}$ and for $n = 1, ..., g$, let $\phi_n^{(w,k)}([\boldsymbol{e}]_{n-w^*+2}^n; t)$ be the p.g.f. of the conditional distribution of the number of occurrences of the overlapping scans from the $(n+1)$-th generation to the g-th generation along the direction in $\{X_v, v \in V\}$ given that at the generation $V(n-w^*+2), ..., V(n)$ the outcomes $e(n-w^*+2) \in E(h_{n-w^*+2}), ..., e(n) \in E(h_n)$ are observed, where

$$[\boldsymbol{e}]_i^j = \begin{cases} (e(i), ..., e(j)) & \text{if } i \le j, \\ e(j) & \text{if } i > j \\ (e(1), ..., e(j)) & \text{if } i \le 0 < j. \end{cases}$$

Given that at the generation $V(n-w^*+2), ..., V(n+1)$ the outcomes $e(n-w^*+2) \in E(h_{n-w^*+2}), ..., e(n+1) \in E(h_{n+1})$ are observed, we will denote the number of occurrences of overlapping scans at the vertex v by $N(v, [\boldsymbol{e}]_{n-w+2}^{n+1})$.

THEOREM 3. *The p.g.f.* $\phi^{(w,k)}(t)$ *and the conditional p.g.f.'s* $\phi_n^{(w,k)}([\boldsymbol{e}]_{n-w^*+2}^n; t)$, $n = 1, ..., g$, *satisfy the following recursive relation:*

$$\phi^{(w,k)}(t) = \sum_{e(1) \in E(h_1)} p(e(1))\, t^{\sum_{v \in V(1)} N(v, e(1))}\, \phi_1^{(w,k)}(e(1); t), \tag{1}$$

$$\phi_n^{(w,k)}([\boldsymbol{e}]_{n-w^*+2}^n; t) = \sum_{e(n+1) \in E(h_{n+1})} p(e(n+1) \mid [\boldsymbol{e}]_{n-m+1}^n) t^{\sum_{v \in V(n+1)} N(v, [\boldsymbol{e}]_{n-w+2}^{n+1})} \times \phi_{n+1}^{(w,k)}([\boldsymbol{e}]_{n-w^*+3}^{n+1}; t), \quad 1 \le n \le g-1, \tag{2}$$

$$\phi_n^{(w,k)}([\boldsymbol{e}]_{n-w^*+2}^n; t) = 1, \qquad n = g. \tag{3}$$

Proof By considering the outcomes at the first generation, we can obtain the first equation (1) immediately. Let $Y_n^{(w,k)}$ be the number of occurrences of the overlapping scans from the $(n+1)$-th generation to the g-th generation along the direction in $\{X_v, v \in V\}$. Then the proof of the equation (2) is completed by observing that

$$
\begin{aligned}
&\phi_n^{(w,k)}([\boldsymbol{e}]_{n-w^*+2}^n; t) = E[t^{Y_n^{(w,k)}} | \boldsymbol{X}(n) = e(n), ..., \boldsymbol{X}(n-w^*+2) = e(n-w^*+2)] \\
&= P(\boldsymbol{X}(n+1) = e(n+1) | \boldsymbol{X}(n) = e(n), ..., \boldsymbol{X}(n-w^*+2) = e(n-w^*+2)) \\
&\quad \times E[t^{Y_n^{(w,k)}} | \boldsymbol{X}(n+1) = e(n+1), ..., \boldsymbol{X}(n-w^*+2) = e(n-w^*+2)] \\
&= \sum_{e(n+1)\in E(h_{n+1})} p(e(n+1) \mid [\boldsymbol{e}]_{n-m+1}^n) \prod_{v \in V(n+1)} \left(t^{N(v,[\boldsymbol{e}]_{n-w+2}^{n+1})}\right) \\
&\quad \times E\left[t^{Y_n^{(w,k)} - \sum_{v\in V(n+1)} N(v,[\boldsymbol{e}]_{n-w+2}^{n+1})} \middle| \begin{array}{l} \boldsymbol{X}(n+1) = e(n+1), ... \\ ..., \boldsymbol{X}(n-w^*+2) = e(n-w^*+2) \end{array} \right] \\
&= \sum_{e(n+1)\in E(h_{n+1})} p(e(n+1) \mid [\boldsymbol{e}]_{n-m+1}^n) \, t^{\sum_{v\in V(n+1)} N(v,[\boldsymbol{e}]_{n-w+2}^{n+1})} \\
&\quad \times E[t^{Y_{n+1}^{(w,k)}} | \boldsymbol{X}(n+1) = e(n+1), ..., \boldsymbol{X}(n-w^*+3) = e(n-w^*+3)].
\end{aligned}
$$

It is easy to see the last equation (3). The proof is completed. □

REMARK 3.1. *As already mentioned, it is clear that in the special case where $w=k$, the p.g.f. $\phi^{(w,k)}$ reduces to the p.g.f. of the number of overlapping success runs of length k. The evaluation of the p.g.f. can be readily performed through the recurrence relations (1), (2) and (3) by setting $w=k$.*

We illustrate how to derive the p.g.f. through the following example.

Example 3.1 : The number of overlapping runs of length 2

Consider the case where $w=k$. It is clear that the p.g.f. $\phi^{(w,k)}$ reduces to the p.g.f. of the number of overlapping success runs of length k. The evaluation of the p.g.f. can be readily performed through the recurrence relations (1), (2) and (3) by setting $w=k$.

In the case where $(w,k)=(2,2)$, we derive the p.g.f. of the distribution of the number of overlapping runs of length 2 on DAG with generation in Fig 2. Since the first generation has the vertex $\{v_1(1)\}$, we have from the equation (1)

$$\phi^{(2,2)}(t) = p((1))\phi_1^{(2,2)}((1);t) + p((0))\phi_1^{(2,2)}((0);t).$$

Next, noting that the second generation has the vertices $\{v_1(2), v_2(2)\}$ and the third generation has the vertices $\{v_1(3)\}$, we have

$$
\begin{aligned}
&\phi_1^{(2,2)}((1);t) = p((1,1)|(1))t^2\phi_2^{(2,2)}((1),(1,1);t) + p((1,0)|(1))t\phi_2^{(2,2)}((1),(1,0);t) \\
&+p((0,1)|(1))t\phi_2^{(2,2)}((1),(0,1);t) + p((0,0)|(1))\phi_2^{(2,2)}((1),(0,0);t), \\
&\phi_1^{(2,2)}((0);t) = p((1,1)|(0))\phi_2^{(2,2)}((0),(1,1);t) + p((1,0)|(0))\phi_2^{(2,2)}((0),(1,0);t)
\end{aligned}
$$

$$+p((0,1)|(0))\phi_2^{(2,2)}((0),(0,1);t)+p((0,0)|(0))\phi_2^{(2,2)}((0),(0,0);t).$$

From the equation (2), we have

$$\begin{aligned}
&\phi_2^{(2,2)}((1),(1,1);t)\\
&\quad=p((1)|(1),(1,1))t^2\phi_3^{(2,2)}((1,1),(1);t)+p((0)|(1),(1,1))\phi_3^{(2,2)}((1,1),(0);t),\\
&\phi_2^{(2,2)}((1),(1,0);t)\\
&\quad=p((1)|(1),(1,0))t\phi_3^{(2,2)}((1,0),(1);t)+p((0)|(1),(1,0))\phi_3^{(2,2)}((1,0),(0);t),\\
&\phi_2^{(2,2)}((1),(0,1);t)\\
&\quad=p((1)|(1),(0,1))t\phi_3^{(2,2)}((0,1),(1);t)+p((0)|(1),(0,1))\phi_3^{(2,2)}((0,1),(0);t),\\
&\phi_2^{(2,2)}((1),(0,0);t)\\
&\quad=p((1)|(1),(0,0))\phi_3^{(2,2)}((0,0),(1);t)+p((0)|(1),(0,0))\phi_3^{(2,2)}((0,0),(0);t),\\
&\phi_2^{(2,2)}((0),(1,1);t)\\
&\quad=p((1)|(0),(1,1))t^2\phi_3^{(2,2)}((1,1),(1);t)+p((0)|(0),(1,1))\phi_3^{(2,2)}((1,1),(0);t),\\
&\phi_2^{(2,2)}((0),(1,0);t)\\
&\quad=p((1)|(0),(1,0))t\phi_3^{(2,2)}((1,0),(1);t)+p((0)|(0),(1,0))\phi_3^{(2,2)}((1,0),(0);t),\\
&\phi_2^{(2,2)}((0),(0,1);t)\\
&\quad=p((1)|(0),(0,1))t\phi_3^{(2,2)}((0,1),(1);t)+p((0)|(0),(0,1))\phi_3^{(2,2)}((0,1),(0);t),\\
&\phi_2^{(2,2)}((0),(0,0);t)\\
&\quad=p((1)|(0),(0,0))\phi_3^{(2,2)}((0,0),(1);t)+p((0)|(0),(0,0))\phi_3^{(2,2)}((0,0),(0);t).
\end{aligned}$$

From the equation (3), we have

$$\begin{aligned}
&\phi_3^{(2,2)}((1,1),(1);t)=\phi_3^{(2,2)}((1,0),(1);t)=\phi_3^{(2,2)}((0,1),(1);t)\\
&=\phi_3^{(2,2)}((0,0),(1);t)=\phi_3^{(2,2)}((1,1),(0);t)=\phi_3^{(2,2)}((1,0),(0);t)\\
&=\phi_3^{(2,2)}((0,1),(0);t)=\phi_3^{(2,2)}((0,0),(0);t)=1.
\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}
&\phi_2^{(2,2)}((1),(1,1);t)=p((1)|(1),(1,1))t^2+p((0)|(1),(1,1)),\\
&\phi_2^{(2,2)}((1),(1,0);t)=p((1)|(1),(1,0))t+p((0)|(1),(1,0)),\\
&\phi_2^{(2,2)}((1),(0,1);t)=p((1)|(1),(0,1))t+p((0)|(1),(0,1)),\\
&\phi_2^{(2,2)}((1),(0,0);t)=\phi_2^{(2,2)}((0),(0,0);t)=1,\\
&\phi_2^{(2,2)}((0),(1,1);t)=p((1)|(0),(1,1))t^2+p((0)|(0),(1,1)),\\
&\phi_2^{(2,2)}((0),(1,0);t)=p((1)|(0),(1,0))t+p((0)|(0),(1,0)),
\end{aligned}$$

$$\phi_2^{(2,2)}((0),(0,1);t) = p((1)|(0),(0,1))t + p((0)|(0),(0,1)).$$

Then we have

$$\begin{aligned}
\phi_1^{(2,2)}((1);t) &= p((1,1)|(1))t^2\left[p((1)|(1),(1,1))t^2 + p((0)|(1),(1,1))\right] \\
&\quad + p((1,0)|(1))t\left[p((1)|(1),(1,0))t + p((0)|(1),(1,0))\right] \\
&\quad + p((0,1)|(1))t\left[p((1)|(1),(0,1))t + p((0)|(1),(0,1))\right] + p((0,0)|(1)), \\
\phi_1^{(2,2)}((0);t) &= p((1,1)|(0))\left[p((1)|(0),(1,1))t^2 + p((0)|(0),(1,1))\right] \\
&\quad + p((1,0)|(0))\left[p((1)|(0),(1,0))t + p((0)|(0),(1,0))\right] \\
&\quad + p((0,1)|(0))\left[p((1)|(0),(0,1))t + p((0)|(0),(0,1))\right] + p((0,0)|(0)).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\phi^{(2,2)}(t) = p((1))\,&\Big[p((1,1)|(1))t^2\left[p((1)|(1),(1,1))t^2 + p((0)|(1),(1,1))\right] \\
&+p((1,0)|(1))t\left[p((1)|(1),(1,0))t + p((0)|(1),(1,0))\right] + p((0,1)|(1))t \\
&\times\left[p((1)|(1),(0,1))t + p((0)|(1),(0,1))\right] + p((0,0)|(1))\Big] \\
+p((0))\,&\Big[p((1,1)|(0))\left[p((1)|(0),(1,1))t^2 + p((0)|(0),(1,1))\right] \\
&+p((1,0)|(0))\left[p((1)|(0),(1,0))t + p((0)|(0),(1,0))\right] + p((0,1)|(0)) \\
&\times\left[p((1)|(0),(0,1))t + p((0)|(0),(0,1))\right] + p((0,0)|(0))\Big].
\end{aligned}$$

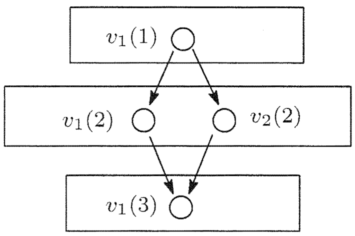

Fig. 2. Overlapping runs on the Markov DAG with generation.

REMARK 3.2. *Letting $w = k = 1$ in Theorem 3, we can obtain the distribution of the number of successes in the DAG with generation. In the special case when $k = 1$ and $P(X_{v_j(i)} = 1) = p(= 1-q)$, $j = 1,2,...,h_i$, $i = 1,2,...,g$, it is easy to see that $\phi(t) = (pt+q)^N$, where N is the cardinality of V. This is the usual binomial distribution.*

REMARK 3.3. *When $h_i = 1$ for $i = 1,2,...,g$, the DAG with generation reduces to a time homogeneous two-state Markov chain. Then the distributions of the number of overlapping success runs and scans are called Type III Markov binomial distribution of order k and Type III Markov binomial distribution of order k/w, respectively (see Balakrishnan and Koutras* [**4**]*).*

Example 3.2 : The number of overlapping success runs and scans

We consider the distributions of numbers of overlapping success runs and scans on the second order Markov DAG with generation given in Figure 3, respectively. For $i = 1, 2, ..., 7$, let $S(i)$ be the number of successes in the i-th generation, that is, $S(i) = X_1(i) + X_2(i) + \cdots + X_{h_i}(i)$ and $F(i)$ be the number of failures in the i-th generation, that is, $F(i) = h_i - S(i)$. For the second order Markov DAG with generation given in Figure 3, we assume that the conditional independence

$$P\left(\begin{matrix} X_{v_1(i+1)} = e_1(i+1), ... \\ ..., X_{v_{h_{i+1}}(i+1)} = e_{h_{i+1}}(i+1) \end{matrix} \middle| \boldsymbol{X}(i) = \boldsymbol{e}(i), \boldsymbol{X}(i-1) = \boldsymbol{e}(i-1)\right)$$
$$= \prod_{j=1}^{h_{i+1}} P\left(X_{v_j(i+1)} = e_j(i+1) \mid \boldsymbol{X}(i) = \boldsymbol{e}(i), \boldsymbol{X}(i-1) = \boldsymbol{e}(i-1)\right),$$

the initial probabilities

$$P(X_{v_1(1)} = 0) = q, \tag{4}$$

and the transition probabilities

$$P(X_{v_j(2)} = 0 | X_{v_1(1)} = 0) = q^2, \qquad j = 1, 2, \tag{5}$$
$$P(X_{v_j(2)} = 0 | X_{v_1(1)} = 1) = q^{\frac{1}{2}}, \qquad j = 1, 2, \tag{6}$$

$$P\left(X_{v_j(i+1)} = 0 \mid \boldsymbol{X}(i) = \boldsymbol{e}(i),\ \boldsymbol{X}(i-1) = \boldsymbol{e}(i-1)\right) \tag{7}$$
$$= \begin{cases} q^2,\ S(i) \geq F(i),\ S(i-1) \geq F(i-1), \\ q^{\frac{1}{2}},\ S(i) < F(i),\ S(i-1) \geq F(i-1), \\ q^4,\ S(i) \geq F(i),\ S(i-1) < F(i-1), \\ q^{\frac{1}{4}}\ S(i) < F(i),\ S(i-1) < F(i-1), \end{cases}$$

for $j = 1, 2, ..., h_{i+1}$, $i = 1, 2, ..., 6$.

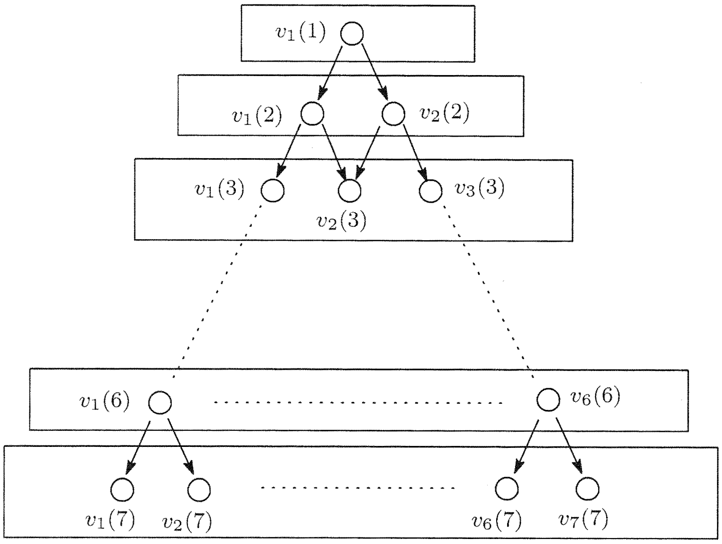

Fig. 3. An example of the DAG with generation.

Then we can derive the p.g.f.'s of the numbers of overlapping runs and scans by using the algorithm presented in Theorem 3. In Figures 4 and 5, We have presented the graphs of the distributions of numbers of overlapping success runs of length 3 and scans in the case where $(w, k) = (3, 2)$, respectively.

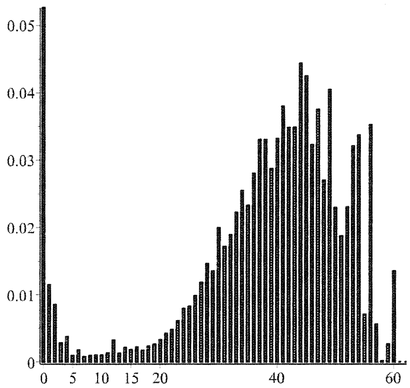

Fig. 4. The number of overlapping runs ($q = 0.3$)

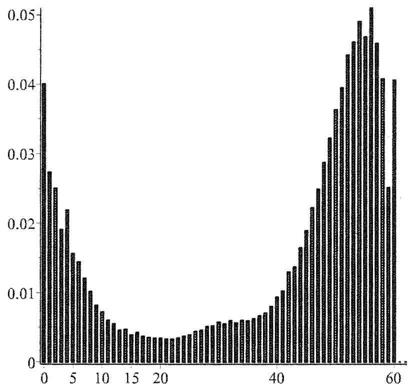

Fig. 5. The number of overlapping scans ($q = 0.5$)

4. Application to reliability systems

Let N be the cardinality of V. Suppose that a reliability system with N components are allocated at the vertices V one by one. Then we will consider two special reliability systems, which are closely related to the distributions of runs and scans.

One is called the *consecutive-k-out-of-N:F system* where the system fails if and only if k consecutive components fail. The other is called the *k-within-consecutive-w-out-of-N:F system* where the system fails if and only if there is k failed components within the moving window of length w in the system (see Chao *et al.* [**5**], Aki and Hirano [**3**], Chang *et al.* [**6**] and Fu and Lou [**10**]). Here, each component can only be in one of two states, either operating or failed, and so also the entire system. The reliability evaluation of the consecutive systems has become an important and integral part of the planning, design and operation of engineering systems such as radar stations, fluid transportation networks and atomic power plants. Originally, the system consisting of N components placed in a line (labeled as first, second, and so on, up to N-th) is considered and the N components in the system are assumed to work independently of each other. Several extensions to the directed trees of the systems were subsequently studied by Aki [**1**], [**2**], Inoue and Aki [**14**], [**15**]. Furthermore, the systems on a Markov DAG with generation were discussed by Inoue and Aki [**16**].

In this section, we will study more complex consecutive systems on a m-th order Markov DAG with generation. However, the reliability of the system can be easily calculated by making use of Theorem 3. Then we can obtain useful information for the more efficient study of these systems, which provides clues to the maintenance. Assume that each vertex is a component of a system and can be in one of two states, either operating or failed (we define a "success" as a failed component) and so also the entire system on the DAG with generation.

For simplicity, the reliability of the system is investigated under the assumption that all the components work according to a second order Markov chain, that is, the collection of random variables $\{X_v, v \in V\}$ is a second order Markov DAG with generation. We consider the reliability of the systems on the second order Markov DAG with generation given in Figure 3 under the assumptions (4), (5), (6) and (7). The probability q means the reliability value of each component.

4.1. The reliabilities of consecutive systems

The reliability of the whole system is given by $\phi^{(w,k)}(0)$, which is a function of q. We define the reliability $R(q)$ of the system by

$$R(q) = \sum_{i=0}^{\ell^{(w,k)}} a_i q^i, \tag{8}$$

where $\ell^{(w,k)}$ is the degree of the polynomial $\phi^{(w,k)}(0)$ in q.

We consider the reliability of consecutive-k-out-of-N:F system on the DAG with generation, which can be effectively evaluated by the aid of Theorem 3.

Example 4.1 : The consecutive-k-out-of-N:F system

In the case where $k = 3$, the consecutive system on the DAG with generation in Figure 3 is considered. We can obtain the reliability $R(q)$ of the whole system. However, the expression of $R(q)$ is omitted here since it is very lengthy. In Figure 6, we give the graph of the reliability of the consecutive system.

We treat the reliability of k-within-consecutive-w-out-of-N:F system on the DAG with generation, which can be easily obtained by the aid of Theorem 3.

Example 4.2 : The k-within-consecutive-w-out-of-N:F system
In the case where $(w, k) = (3, 2)$, the consecutive system on the DAG with generation given in Figure 3 is considered. The reliability $R(q)$ of the whole system can be easily obtained by making use of Theorem 3. However, the expression of $R(q)$ is omitted here since it is not represented in a simple form.

Figure 6 is the graph of the reliability of the consecutive system.

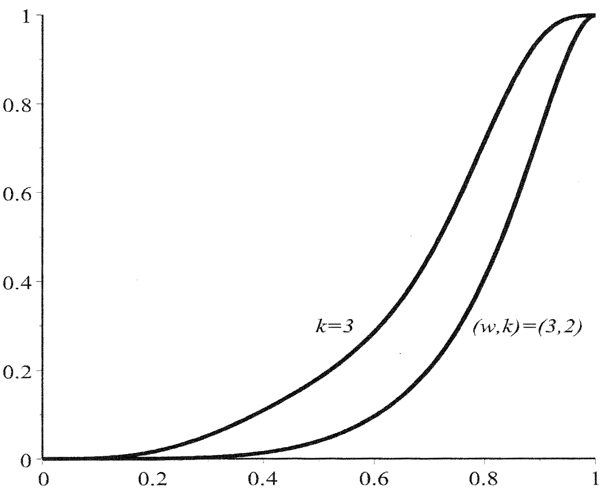

Fig. 6. Reliabilities of the two consecutive systems in Examples 4.1 and 4.2.

5. Estimation problems

We discuss the maximum likelihood estimation for the distribution of the number of runs and scans. For simplicity, the parameter estimations are investigated on the second order Markov DAG with generation given in Figure 3 under the assumptions (4), (5), (6) and (7). We address the parameter estimation of q based on the number of run and scans. Let $n_1, n_2, ..., n_l$ be independent observations of the number of runs and scans. We consider the statistical estimation of the parameter q based on $n_1, n_2, ..., n_l$. We write $\phi^{(w,k)}(t) = \sum_{i \geq 0} c_i(q)t^i$. Here, we use the notation $c_i = [t^i]\phi^{(w,k)}(t)$ to extract the coefficient of t^i in the probability generating function.

Then the likelihood function $L(q)$ of q based on $n_1, n_2, ..., n_l$ can be written as

$$L(q) = \prod_{i=1}^{l} [t^{n_i}]\phi^{(w,k)}(t). \tag{9}$$

The following are examples illustrating how to obtain the MLE.

Example 5.1 : Estimation based on runs

We investigate the estimation problem arising from the distribution of the number of success runs of length 3. Table 1 is a simulated data set of the number of success runs of length 3 on the DAG with generation given in Figure 3 in the case where $q = 0.4$ and $l = 10$.

Table 1: A simulated data set in the case where $q = 0.4$.

i	1	2	3	4	5	6	7	8	9	10
n_i	48	22	37	46	21	38	27	27	35	40

From the equation (9), we can obtain the likelihood function $L(q)$ easily. However, the likelihood function $L(q)$ is omitted here since it is not represented in a simple form. In Figure 7, we give the graph of the likelihood function based on the data in Table 1.

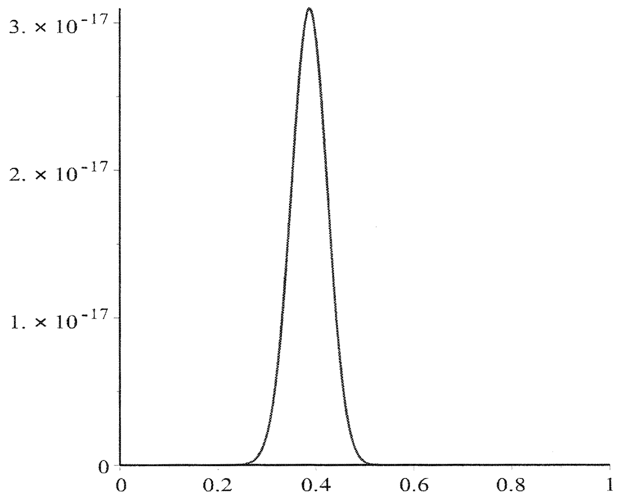

Fig. 7. The likelihood function $L(q)$ based on the data in Table 1.

By maximizing the likelihood function numerically, we have the MLE $\hat{q} = 0.3869$.

Example 5.2 : Estimation based on scans

We address the estimation problem arising from the distribution of the number of scans in the case where $(w, k) = (3, 2)$. Table 2 is a simulated data set of the number of scans on the DAG with generation given in Figure 3 in the case where $(w, k) = (3, 2)$, $q = 0.7$ and $l = 10$.

Table 2: A simulated data set in the case where $q = 0.7$.

i	1	2	3	4	5	6	7	8	9	10
n_i	6	1	23	34	6	17	2	32	1	4

Working in the same fashion as we did in Example 5.1, from the equation (9), we can obtain the likelihood function $L(q)$ immediately. However, the likelihood function $L(q)$ is omitted here for the same reason as in Example 5.1. In Figure 8, we give the graph of the likelihood function based on the data in Table 2.

As already mentioned before, the MLE is obtained by maximizing the likelihood function.

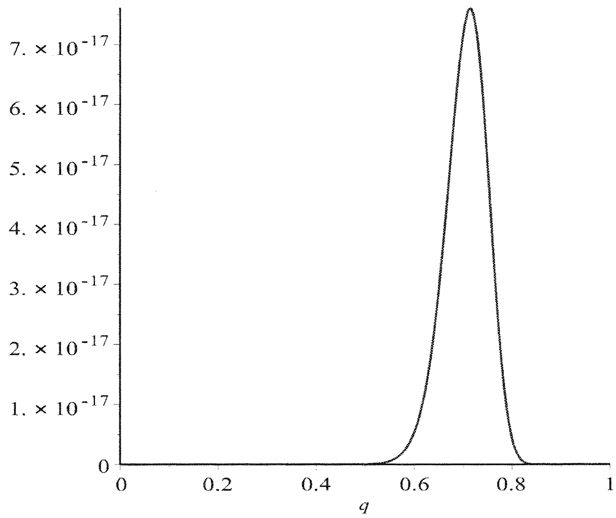

Fig. 8. The likelihood function $L(q)$ based on the data in Table 2.

By maximizing the likelihood function numerically, we have the MLE $\hat{q} = 0.7145$.

Acknowledgements

The author will wish to thank the editor and the referees for careful reading of our paper and helpful suggestions which led to improved results. This study was supported in part by Grant-in-Aid for Scientific Research 24540141 from Japan Society for the Promotion of Science.

References

[1] Aki, S. (1999). Distributions of runs and consecutive systems on directed trees, *Annals of the Institute of Statistical Mathematics*, **51**, 1–15.

[2] Aki, S. (2001). Exact reliability and lifetime of consecutive systems, *Handbook of Statistics*, (N. Balakrishnan and C. R. Rao Eds.), **20**, 281–300, Elsevier.

[3] Aki, S. and Hirano, K. (1996). Lifetime distribution and estimation problems of consecutive-k-out-of-n:F systems, *Annals of the Institute of Statistical Mathematics*, **48**, 185–199.

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

[5] 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.

[6] Chang, G. J., Cui, L. and Hwang, F. K. (2000). *Reliabilities of Consecutive-k Systems*, Kluwer Academic Publishers.

[7] Ebneshahrashoob, M. and Sobel, M. (1990). Sooner and later waiting time problems for Bernoulli trials: frequency and run quotas, *Statst. Probab. Lett.*,**9**, 5–11.

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

[9] Fu, J. C. and Koutras, M. V. (1994). Distribution theory of runs : a Markov chain approach, *Journal of the American Statistical Association*, **89**, 1050–1058.

[10] 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.

[11] Glaz, J. and Balakrishnan, N. (1999). *Scan Statistics and Applications*, Birkhäuser, Boston.

[12] Glaz, J., Naus, J. and Wallenstein, S. (2001). *Scan Statistics*, Springer-Verlag New York.

[13] Glaz, J., Pozdnyakov, V. and Wallenstein, S. (2009). *Scan Statistics: Methods and Applications*, Birkhäuser, Boston.

[14] Inoue, K. and Aki, S. (2009). Distributions of runs and scans on higher order Markov trees, *Communications in Statistics–Theory and Methods*, **38(5)**, 621–641.

[15] Inoue, K. and Aki, S. (2010). On the conditional and unconditional distributions of the number of success runs on a circle with applications, *Statistics & Probability Letters*, **80**, 874–885.

[16] Inoue, K. and Aki, S. (2013). Distributions of numbers of runs and scans on directed acyclic graphs with generation, *Computational Statistics*, **28**, 1133-1150.

[17] Koutras, M. V. and Alexandrou, V. A. (1995). Runs, scans and urn model distributions: A unified Markov chain approach, *Annals of the Institute of Statistical Mathematics,* **47**, 743–766.

[18] Jonczy, J. and Haenni, H. (2008). Network Reliability Evaluation with Propositional Directed Acyclic Graphs, *Advances in Mathematical Modeling for Reliability (eds. Bedford, T. et al.)*, 25–31, IOS Press.

[19] Lauritzen, S. L. (1996). *Graphical Models* , Clarendon Press, Oxford.

[20] Ling, K. D. (1988). On binomial distributions of order k, *Statistics & Probability Letters,* **6**, 247–250.

[21] Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge.

Kiyoshi Inoue

Faculty of Economics, Seikei University,

3-3-1 Kichijoji-Kitamachi, Musasino-shi,

Tokyo, 180-8633, Japan

kinoue@econ.seikei.ac.jp