

A Markov chain approach to quality control

F. Bartolucci

Istituto di Scienze Economiche

Università di Urbino

www.stat.unipg.it/~bart

Preliminaries

- We use the *finite Markov chain imbedding* approach of Alexandrou, Fu and Koutras to deal with the distribution of a statistic (called *window statistic*) defined as the number of the times that an m -length subsequence of an arbitrary type appears in a sequence of R independent and discrete random variables.
- On the basis of this approach, one may compute:
 - ★ the distribution of a window statistic even in presence of rather large values of m and R ;
 - ★ just the probability of never observing a subsequence of a certain kind even for rather large values of m and any R ;
 - ★ the expected waiting time till a subsequence of a certain kind appears (ARL).

Notation

- We assume that there is a sequence

$$X_1, \dots, X_R$$

of independent and discrete random variables each of them can assume a value in the set $\{0, \dots, C - 1\}$

- Let also λ_{ic} is the probability of the event $X_i = c$,

$$\lambda_{ic} = P(X_i = c)$$

and $\boldsymbol{\lambda}_i = (\lambda_{i0} \ \cdots \ \lambda_{i,C-1})'$.

- A subsequence of length m out from the overall sequence is denoted by $\boldsymbol{x} = (x_1 \ \cdots \ x_m)'$.

- The set of all the possible values of \boldsymbol{x} , is denoted by \mathcal{X} .

- The cardinality of \mathcal{X} is C^m and its elements are ordered lexicographically with x_m running faster; so, the element \boldsymbol{x} of \mathcal{X} has index

$$I(\boldsymbol{x}) = 1 + \sum_{j=1}^m x_j C^{m-j}.$$

Window statistic

- Given the sequence of random variables

$$X_1, \dots, X_R \tag{1}$$

and any $m \leq R$ and $\mathcal{A} \subseteq \mathcal{X}$ we define $N_{R,m}(\mathcal{A})$ (*window statistic*) as the number of the times that a subsequence of length m out from the (1) belongs to \mathcal{A} .

- The kind of subsequences we are interested in depends on the structure of \mathcal{A} .

Example

- If subsequences of three Bernoulli trials are considered ($m = 3, C = 2$) we have that

$$\mathcal{X} = \{(0 \ 0 \ 0)', (0 \ 0 \ 1)', (0 \ 1 \ 0)', \\ (0 \ 1 \ 1)', (1 \ 0 \ 0)', (1 \ 0 \ 1)', \\ (1 \ 1 \ 0)', (1 \ 1 \ 1)'\}.$$

- The cardinality of the set is $C^m = 2^3 = 8$.
- if we are interested in runs with all success, it is enough to set $\mathcal{A} = \{(1 \ 1 \ 1)'\}$.
- if we are interested in the subsequences containing at least 2 successes we have to set

$$\mathcal{A} = \{(0 \ 1 \ 1)', (1 \ 0 \ 1)', (1 \ 1 \ 0)', (1 \ 1 \ 1)'\}.$$

Distribution of a window statistic

- Our aim is that of computing the distribution of $N_{R,m}(\mathcal{A})$, that is the vector $\boldsymbol{\pi} = (\pi_0 \ \cdots \ \pi_{T+1})'$, where

$$\pi_n = P(N_{R,m}(\mathcal{A}) = n).$$

- To this aim, a method based on the finite imbeddable Markov chain approach is used.
- **Definition:** The random variable Z , with support $\{0, \dots, Z^*\}$, is a *Markov chain imbeddable variable* (MVB) if (Koutras and Alexandrou, 1995):

1. There exists a Markov chain $\{\mathbf{Y}_t, t = 1, \dots, T\}$ defined on a state space \mathcal{Y} , such that

$$P(Z = z) = P(\mathbf{Y}_T \in \mathcal{Y}_z), \quad z = 0, \dots, Z^*,$$

where $\{\mathcal{Y}_0, \dots, \mathcal{Y}_{Z^*}\}$ is a suitable partition of \mathcal{Y} .

2. For each $t = 1, \dots, T$ and each $z' \neq z, z + 1$,

$$P(\mathbf{Y}_t \in \mathcal{Y}_{z'} | \mathbf{Y}_{t-1} \in \mathcal{Y}_z) = 0.$$

- Let s the maximum number of elements in each subset \mathcal{Y}_z of \mathcal{Y} and consider the vector

$$\mathbf{f}_t(z) = \{P(\mathbf{Y}_t = \mathbf{y}_{zi}), i = 1, \dots, s\}$$

where \mathbf{y}_{zi} is the i -th element of \mathcal{Y}_z .

- Consider also the matrix of the *within state* one-step transition probabilities

$$\mathbf{A}_t(z) = \{P(\mathbf{Y}_t = \mathbf{y}_{zj} | \mathbf{Y}_{t-1} = \mathbf{y}_{zi}), i, j = 1, \dots, s\}$$

and that of the *between state* transition probabilities

$$\mathbf{B}_t(z) = \{P(\mathbf{Y}_t = \mathbf{y}_{z+1,j} | \mathbf{Y}_{t-1} = \mathbf{y}_{zi}), i, j = 1, \dots, s\}.$$

both of dimension $s \times s$.

- **Theorem 1** (Koutras and Alexandrou, 1995): If Z is a MVB, the following relations hold:

$$\mathbf{f}_t(0) = \mathbf{A}'_t(0)\mathbf{f}_{t-1}(0),$$

$$\mathbf{f}_t(z) = \mathbf{A}'_t(z)\mathbf{f}_{t-1}(z) + \mathbf{B}'_t(z-1)\mathbf{f}_{t-1}(z-1)$$

We also have

$$P(Z = z) = \mathbf{f}'_T(z)\mathbf{1}, \quad z = 0, \dots, Z^*.$$

- **Theorem 2** (Bartolucci, 2000). $N_{R,m}(\mathcal{A})$ is a MVB for any $\mathcal{A} \subseteq \mathcal{X}$ and $1 \leq m \leq R$.

- The underlying Markov chain has index set $\{0, \dots, T\}$, $T = R - m$, and state space

$$\mathcal{Y} = \{(n \quad \mathbf{x}')'\}, \quad n = 0, \dots, T + 1, \quad \mathbf{x} \in \mathcal{X}.$$

- Let \mathbf{X}_t be the t -th subsequence we can observe

$$\mathbf{X}_t = (X_t \quad \cdots \quad X_{t+m-1})'$$

- Let u_i denote the probability that the first subsequence of length m is equal to the i -th element of \mathcal{X} , \mathbf{x}_i ,

$$u_i = P(\mathbf{X}_1 = \mathbf{x}_i) = P(X_1 = x_{i1}) \cdots P(X_m = x_{im})$$

and note that the vector $\mathbf{u} = \{u_i\}$ is equal to $\otimes_{i=1}^m \boldsymbol{\lambda}_i$.

- Let v_{tij} denote the transition probability from \mathbf{x}_i to \mathbf{x}_j when we observe the t -th subsequence,

$$v_{tij} = P(\mathbf{X}_t = \mathbf{x}_j | \mathbf{X}_{t-1} = \mathbf{x}_i) = \{P(X_{t+m} = x_{jm}) \text{ or } 0\}$$

and note that $\mathbf{V}_t = \{v_{tij}\}$ is equal to $\mathbf{1}_C \otimes \mathbf{I}_{C^{m-1}} \otimes \boldsymbol{\lambda}'_{t+m}$.

Structure of $\mathbf{f}_0(n)$, $\mathbf{A}_t(n)$ and $\mathbf{B}_t(n)$

- Let $\mathcal{I}(\mathcal{A})$ be the set with elements $I(\mathbf{x})$ for each $\mathbf{x} \in \mathcal{A}$.
- The i -th entry of $\mathbf{f}_0(0)$ is equal to u_i if $i \notin \mathcal{I}(\mathcal{A})$ and to 0 otherwise.
- The i -th entry of $\mathbf{f}_0(1)$ is equal to u_i if $i \in \mathcal{I}(\mathcal{A})$ and to 0 otherwise and $\mathbf{f}_0(n) = \mathbf{0}$ for any $n > 1$.
- For each n , the i -th column of $\mathbf{A}_t(n)$ is equal to \mathbf{v}_{ti} if $i \notin \mathcal{I}(\mathcal{A})$ and to $\mathbf{0}$ otherwise, where \mathbf{v}_{ti} is the i -th column of \mathbf{V}_t .
- The i -th column of $\mathbf{B}_t(n)$ is equal to \mathbf{v}_{ti} if $i \in \mathcal{I}(\mathcal{A})$ and to $\mathbf{0}$ otherwise.
- Note that matrices $\mathbf{A}_t(n)$ and $\mathbf{B}_t(n)$ do not depend on n ; therefore it may be omitted.

Example

- If we are interested in subsequences of 2 random variables with support $\{0, 1, 2\}$, so that $m = 2$ and $C = 3$, we have

$$\mathcal{Y}_n = \left\{ (n \ 0 \ 0)', (n \ 0 \ 1)', (n \ 0 \ 2)', \right. \\ \left. (n \ 1 \ 0)', (n \ 1 \ 1)', (n \ 1 \ 2)', \right. \\ \left. (n \ 2 \ 0)', (n \ 2 \ 1)', (n \ 2 \ 2)' \right\}.$$

- If we are interested in the subsequences

$$\mathcal{A} = \left\{ (1 \ 2)', (2 \ 1)', (2 \ 2)' \right\},$$

then $\mathcal{I}(\mathcal{A}) = \{6, 8, 9\}$.

- Then we have

$$\mathbf{f}_0(0) = \begin{pmatrix} \lambda_{10}\lambda_{20} \\ \lambda_{10}\lambda_{21} \\ \lambda_{10}\lambda_{22} \\ \lambda_{11}\lambda_{20} \\ \lambda_{11}\lambda_{21} \\ 0 \\ \lambda_{12}\lambda_{20} \\ 0 \\ 0 \end{pmatrix}' \quad \mathbf{f}_0(1) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \lambda_{11}\lambda_{22} \\ 0 \\ \lambda_{12}\lambda_{21} \\ \lambda_{12}\lambda_{22} \end{pmatrix}', \quad \mathbf{f}_0(n) = \mathbf{0}, n > 1$$

- The transition matrix \mathbf{A}_t is equal to

$$\begin{pmatrix} \lambda_{t+m,0} & \lambda_{t+m,1} & \lambda_{t+m,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{t+m,0} & \lambda_{t+m,1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,0} & 0 & 0 \\ \lambda_{t+m,0} & \lambda_{t+m,1} & \lambda_{t+m,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{t+m,0} & \lambda_{t+m,1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,0} & 0 & 0 \\ \lambda_{t+m,0} & \lambda_{t+m,1} & \lambda_{t+m,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{t+m,0} & \lambda_{t+m,1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,0} & 0 & 0 \end{pmatrix}$$

- The transition matrix \mathbf{B}_t is equal to

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,1} & \lambda_{t+m,2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,1} & \lambda_{t+m,2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{t+m,1} & \lambda_{t+m,2} \end{pmatrix}.$$

- Computing the distribution of $N_{R,m}(\mathcal{A})$ using the recursions of Theorem 1 implies the need of the sequences of matrices $\{\mathbf{A}_t\}$ and $\{\mathbf{B}_t\}$.
- However each of these matrices has dimension $C^m \times C^m$ which may become huge even for small values of C and m ; this may give rise to computational problems.
- We propose an algorithm which avoids the use of such sequences of matrices and which can be used saving much memory and computational time.
- The approach allows also to compute the expected waiting time till a subsequence equal to an element of \mathcal{A} appears in the hypothetical infinite sequence

$$X_1, X_2, \dots$$

- We denote this quantity by $E(W_m(\mathcal{A}))$ where $W_m(\mathcal{A})$ is the random variable such that $W_m(\mathcal{A}) = w$ means that the first subsequence belonging to \mathcal{A} is

$$X_{w-m+1}, X_{w-m+2}, \dots, X_w.$$

The algorithm for computing $\boldsymbol{\pi}$

- Let \mathbf{F}_t be the matrix $C^m \times (T + 2)$ obtained by stacking the vectors $\mathbf{f}_t(n)$ for $n = 0, \dots, T + 1$,

$$\mathbf{F}_t = (\mathbf{f}_t(0) \quad \cdots \quad \mathbf{f}_t(T + 1)).$$

- After having prepared the initial matrix \mathbf{F}_0 perform the following operations for $t = 1, \dots, T$:

1. Compute $\mathbf{L}_t = (\mathbf{G}_{t-1}\mathbf{H}) \otimes \boldsymbol{\lambda}_{t+m}$ where $\mathbf{H} = \mathbf{I}_{T+2} \otimes \mathbf{1}_C$, \mathbf{I}_{T+2} is the identity matrix of order $T + 2$, $\mathbf{1}_C$ is a column vector of C ones and \mathbf{G}_{t-1} is the matrix $C^{m-1} \times [C(T + 2)]$ obtained by reshaping \mathbf{F}_{t-1} .
2. Obtain the matrices $\tilde{\mathbf{A}}_t$ and $\tilde{\mathbf{B}}_t$ by substituting the i -th row of \mathbf{L}_t with $\mathbf{0}'$ if, respectively, $i \in \mathcal{I}(\mathcal{A})$ and $i \notin \mathcal{I}(\mathcal{A})$.
3. Compute \mathbf{F}_t as $(\tilde{\mathbf{A}}_t + \tilde{\mathbf{C}}_t)$ where $\tilde{\mathbf{C}}_t = (\mathbf{0} \quad \tilde{\mathbf{B}}_t^*)$ and $\tilde{\mathbf{B}}_t^*$ is obtained by removing the last column from $\tilde{\mathbf{B}}_t$.

- Finally the vector $\boldsymbol{\pi}$ is equal to $\mathbf{F}'_T \mathbf{1}$.

- In presence of Bernoulli trials we can deal with subse-

- If one is interested in computing just some elements of the vector $\boldsymbol{\pi}$, further memory may be saved. It may be interesting to compute just

$$\pi_0 = P(N_{R,m}(\mathcal{A}) = 0);$$

note that $1 - \pi_0$ is the probability of observing at least once a subsequence belonging to \mathcal{A} .

- In this case it is not necessary to use the entire matrix \mathbf{F}_t and we have the following algorithm.
- After having computed the initial vector $\mathbf{f}_0(0)$ perform the following operations for $t = 1, \dots, T$:
 1. Compute $\mathbf{l}_t = (\mathbf{G}_{t-1}\mathbf{1}) \otimes \boldsymbol{\lambda}_{t+m}$, where \mathbf{G}_{t-1} is obtained by reshaping $\mathbf{f}_{t-1}(0)$ as a $C^{m-1} \times C$ matrix.
 2. Obtain $\mathbf{f}_t(0)$ by substituting the i -th element of \mathbf{l}_t by 0 for each $i \in \mathcal{I}(\mathcal{A})$.
- Finally $\pi_0 = P(N_{R,m}(\mathcal{A}) = 0)$ is equal to $\mathbf{f}'_T(0)\mathbf{1}$.
- This algorithm permits to deal with subsequences of Bernoulli trials with length up to 22 regardless of R .

Expected waiting time

- About the expected waiting time till the first subsequence of length m belonging to \mathcal{A} appears, the following rule holds

$$\begin{aligned}
 E(W_m(\mathcal{A})) &= \sum_{w=0}^{\infty} P(W_m(\mathcal{A}) > w) = \\
 &= \sum_{w=0}^{\infty} P(N_{w,m}(\mathcal{A}) = 0) = \\
 &= m + \sum_{t=0}^{\infty} P(N_{t+m,m}(\mathcal{A}) = 0)
 \end{aligned}$$

- To compute it the following algorithm below may be used.
- After having computed the initial value $\mathbf{f}_0(0)$ and w as $m + \mathbf{1}'\mathbf{f}_0(0)$ perform the following operations until the convergence on w :
 1. Compute $\mathbf{l}_t = (\mathbf{G}_{t-1}\mathbf{1}) \otimes \boldsymbol{\lambda}_{t+m}$;
 2. Obtain $\mathbf{f}_t(0)$ by substituting the i -th element of \mathbf{l}_t by 0 for each $i \in \mathcal{I}(\mathcal{A})$.
 3. Update w by adding $\mathbf{f}'_t(0)\mathbf{1}$.
- The last value of w is equal to $E(W_m(\mathcal{A}))$.

Obviously this algorithm may require many iterations to reach the convergence on w . However, in practice, this does not give rise to problems as will be clarified in the next section.

1 Software implementation of the algorithms

All the algorithms previously shown have been implemented in MATLAB functions which are available from the author on request.

The first of these functions provides the index set $\mathcal{I}(\mathcal{A})$ given the conditions satisfied by the vectors belonging to \mathcal{A} . This function generates sequentially, by an appropriate cycle of instructions, all the entries of \mathcal{X} and, from time to time, puts in $\mathcal{I}(\mathcal{A})$ the index of that which belongs to \mathcal{A} . However, if the conditions satisfied by each vector in \mathcal{A} depend on the sum of its elements, such a function creates much faster $\mathcal{I}(\mathcal{A})$ since just a vector whose i -th entry is the sum of the entries of the i -th vector contained in the set \mathcal{X} is needed. So, much computing time is saved since this vector, which will shall denoted by \mathbf{z}_m , may be quickly generated exploiting an iterative rule which, starting from $\mathbf{z}_1 = (0 \ \cdots \ C - 1)'$, consists of computing

$$\mathbf{z}_l = \mathbf{z}_1 \otimes \mathbf{1}_{C^{l-1}} + \mathbf{1}_C \otimes \mathbf{z}_{l-1} \quad (2)$$

for $l = 2, \dots, m$. A justification of this rule may be found in appendix.

The other three MATLAB functions provide, respectively, the vector $\boldsymbol{\pi}$ (definition 3), just π_0 (definition 4) and the expected waiting time $E(W_m(\mathcal{A}))$ (definition 5) given the set $\mathcal{I}(\mathcal{A})$ and the sequence of vectors $\{\boldsymbol{\lambda}_i\}$. The implementation and the use of such functions do not give rise to relevant problems. This is true also for the third algorithm which may require many iterations since a small tolerance (10^{-8}) is used to establish if the convergence on w is reached. In fact, within this algorithm, the only object which depends on the number of iterations is the sequence $\{\boldsymbol{\lambda}_i\}$ but, in practice, $\boldsymbol{\lambda}_i$ is independent of i , namely $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}, \forall i$ or $\boldsymbol{\lambda}_i$ may be expressed in function of i and so this sequence is not necessary. Moreover, even if it was necessary, we can easily deal with huge values of the length of this sequence by using nowadays PC. A similar consideration holds for the algorithm in definition 4; then it may be used for any R .

In the following section some examples of application of the algorithms are provided. Also the computing times necessary to obtain each numerical result are shown. They concerns the use of the MATLAB functions indicated above on a PC with a PENTIUM II 400 Mhz and 64 Mb of RAM.

2 Some applications

The first example is within quality control (see also Mosteller, 1941 and Anscombe *et al.*, 1947). Suppose that the products of an assembly line are classified in the following way: 0 for no defects, 1 for defect of type A and 2 for defect of type B (the two defects cannot appear together). It may

appears. This situation can be easily handled using the algorithms shown in section 3 also if, as usual, the probability of the defects decreases in time since the assembly line improves.

In this situation, if $m = 6$, then $s = 3^6 = 729$ and if we are interested in subsequences with at least two defects of type A and one of type B, \mathcal{A} has cardinality equal to 473. The distribution of $N_{R,m}(\mathcal{A})$ is shown in the following table for $R = 10$, $C = 2$, $\lambda_{i1} = 0.1(0.05)$, $\lambda_{i2} = 1/(i/2 + 3)$ and $\lambda_{i0} = 1 - (\lambda_{i1} + \lambda_{i2})$.

Table 1: Distribution of $N_{10,6}(\mathcal{A})$ with $C = 2$, $\lambda_{i1} = 0.1(0.05)$, $\lambda_{i2} = 1/(i/2 + 3)$, $\lambda_{i0} = 1 - (\lambda_{i1} + \lambda_{i2})$

$N_{10,6}(\mathcal{A})$	Probability ($\lambda_{i1} = 0.1$)	Probability ($\lambda_{i1} = 0.05$)
0	0.7791	0.9304
1	0.0617	0.0226
2	0.0540	0.0185
3	0.0449	0.0142
4	0.0342	0.0095
5	0.0260	0.0047
$E(N_{10,6}(\mathcal{A}))$	0.5713	0.1639
$E(W_6(\mathcal{A}))$	34.5912	108.4562

In the same table the expected value of $N_{10,6}(\mathcal{A})$ is shown together with the expected waiting time till the first subsequence belonging to \mathcal{A} appears. The computing time for the distribution of $N_{R,m}(\mathcal{A})$ was equal to 0.22 seconds in both cases ($\lambda_{i1} = 0.1$ and $\lambda_{i1} = 0.05$) whereas to compute $E(W_6(\mathcal{A}))$ were necessary 583 iterations and 1.81 seconds in the first case and 1,940 iterations and 5.88 seconds in the latter one.

As a second example let's consider the statistic $S_t = \sum_{i=1}^m X_{t+i}$. This statistic has been studied, among others, by Naus (1982) and Glaz and Naus (1991). Moreover Karwe and Naus (1997) proposed and used excellent tight bounds and approximations for $G_{k,m}(R) = P(\max_{t \in \mathcal{T}} S_t < k)$ which are useful when m and R are large. However, note that the same probability can be computed also by using the method proposed here for values of m not very large and for any R . In fact, if \mathcal{A} contains all the subsequences \mathbf{x} such that $\sum_{i=1}^m x_i \geq k$, $G_{k,m}(R) = P(N_{R,m}(\mathcal{A}) = 0)$. The value of $G_{k,m}(R)$ is indicated for some values of m , k and R in the following table together with the corresponding computing time (see also table 3 in Karwe and Naus, 1997).

Table 2: Some values of $G_{k,m}(R)$ with $C = 2$ and $\boldsymbol{\lambda} = (0.6 \ 0.3 \ 0.1)'$

m	k	$G_{m,k}(30)$	Comp. time (s.)	$G_{m,k}(100)$	Comp. time (s.)
10	10	0.8431	6.75	0.5080	26.58
10	12	0.9762	6.26	0.9058	25.26
10	15	0.9996	6.20	0.9984	24.27

classification of amino acids based on the letters: acidic, neutral and basic. Namely they deal with three possible charges which are usually indicated with X_i^* equal, respectively, to -1 , 0 and 1 while $S_t^* = \sum_{i=1}^m X_{t+i}^*$ is the combined net charge in m trials starting from $t+1$. Note that also the distribution of the maximum of S_t^* , which may be denoted as $G_{k,m}^*(R)$, can be simply computed with the algorithms proposed here since $G_{k,m}^*(R) = G_{k+m,m}(R)$.

Finally consider that other useful applications of these algorithms can be found within the hypothesis testing when we have to verify if there is a changed segment in a sequence of Bernoulli trials (see Fu and Curnow, 1990 and Wallenstein *et al.*, 1994). For other applications see also Naus (1982), Glaz and Naus (1991) and the introduction of Koutras and Alexandrou (1995).

3 Conclusions

The general approach proposed here shown itself to be very flexible since it may be used to deal with many well-known run and scan statistics as well as with analogous statistics not yet discussed in the literature but which may be useful at a practical level. Moreover, using the algorithms provided here, it is possible to apply the approach in many real situations also if the limits concerning the length of the subsequences (m) and of the overall sequence (R) have to be taken into account. In particular, the limit about m seems more serious since the memory requirement increases exponentially with this quantity. In contrast, the memory requirement increases with R just for the algorithm in definition 3 and in a linear way while the other algorithm (definition 4) can be used regardless of R . This is due to the use of a state space of the underlying Markov chain which contains all the possible outcomes of the subsequences of a certain length. Such a state space is necessary to guarantee generality to the method. However, in some standard situations, smaller state spaces (Koutras and Alexandrou, 1995) may be used so that larger values of m and R may be treated. This is true, for instance, for the second example shown in the previous section where, with the proper imbedding, also subsequences of length $m = 100$ may be examined. In less standard situations, a reduction of the state space seems difficult. In other words, there is a trade-off between flexibility and computational efficiency. So a further improvement of the algorithms proposed here is an interesting challenge for future research.

Appendix

Proof of Theorem 2

Consider the stochastic process $\{\mathbf{Y}_t, t \in \mathcal{T}\}$ where $\mathcal{T} = \{0, \dots, T\}$ and $\mathbf{Y}_t = (n \ \mathbf{x}')' = (n \ x_1 \ x_2 \ \dots \ x_m)'$ means that at time $t+m$ there have been already n subsequences in \mathcal{A} and $X_{t+1} = x_1, X_{t+2} = x_2, \dots, X_{t+m} = x_m$. This obviously is a Markov chain since, for any

To prove the structure of $\mathbf{f}_0(n)$ it is enough to consider that the i -th entry of such a vector, namely $P(\mathbf{Y}_0 = \mathbf{y}_{ni})$, is equal to

$$P(\mathbf{Y}_0 = (n \quad \mathbf{x}'_i)') = \begin{cases} u_i & \text{if } (i \notin \mathcal{I}(\mathcal{A}) \text{ and } n = 0) \text{ or } (i \in \mathcal{I}(\mathcal{A}) \text{ and } n = 1) \\ 0 & \text{otherwise} \end{cases}$$

where

$$u_i = P(X_1 = x_{i1}, X_2 = x_{i2}, \dots, X_m = x_{im}) = \prod_{j=1}^m \lambda_{jx_{ij}}$$

and \mathbf{x}_i and \mathbf{y}_{ni} are, respectively, the i -th vector in \mathcal{X} and that in \mathcal{Y}_n whereas x_{ij} is the j -th entry of \mathbf{x}_i . Note also that, with $s = C^m$,

$$\mathbf{u} = (u_1 \quad \dots \quad u_s)' = \bigotimes_{i=1}^m \boldsymbol{\lambda}_i. \quad (3)$$

This can be proven by induction. Denote, for instance, by $\mathbf{u}^{(l)}$ the vector \mathbf{u} computed considering subsequences of length l . So $\mathbf{u}^{(1)}$ is obviously equal to $\boldsymbol{\lambda}_1$ whereas, supposing that the (3) holds for $l - 1$ and because of the ordering of the elements of \mathcal{X} , we have that $\mathbf{u}^{(l)} = \mathbf{u}^{(l-1)} \otimes \boldsymbol{\lambda}_l = \bigotimes_{i=1}^l \boldsymbol{\lambda}_i$.

In similar way it can be proven the structure of $\mathbf{A}_t(n)$ and $\mathbf{B}_t(n)$. In fact the i, j -th entry of the first matrix, namely $P(\mathbf{Y}_t = \mathbf{y}_{nj} | \mathbf{Y}_{t-1} = \mathbf{y}_{ni})$, is equal to

$$\begin{cases} \lambda_{t+m, x_{jm}} & \text{if } x_{i2} = x_{j1}, x_{i3} = x_{j2}, \dots, x_{im} = x_{j, m-1} \text{ and } j \notin \mathcal{I}(\mathcal{A}) \\ 0 & \text{otherwise} \end{cases}$$

whereas the i, j -th entry of $\mathbf{B}_t(n)$, namely $P(\mathbf{Y}_t = \mathbf{y}_{n+1, j} | \mathbf{Y}_{t-1} = \mathbf{y}_{ni})$, is

$$\begin{cases} \lambda_{t+m, x_{jm}} & \text{if } x_{i2} = x_{j1}, x_{i3} = x_{j2}, \dots, x_{im} = x_{j, m-1} \text{ and } j \in \mathcal{I}(\mathcal{A}) \\ 0 & \text{otherwise} \end{cases}$$

So $\mathbf{A}_t(n)$ and $\mathbf{B}_t(n)$ are independent of n and may be denoted simply by \mathbf{A}_t and \mathbf{B}_t . Consider then the matrix $\mathbf{V}_t = (\mathbf{v}_{t1} \quad \dots \quad \mathbf{v}_{ts})$ whose i, j -th entry is

$$\begin{cases} \lambda_{t+m, x_{jm}} & \text{if } x_{i2} = x_{j1}, x_{i3} = x_{j2}, \dots, x_{im} = x_{j, m-1} \\ 0 & \text{otherwise} \end{cases}$$

Then the i -th column of \mathbf{A}_t is equal to \mathbf{v}_{ti} if $i \notin \mathcal{I}(\mathcal{A})$ and to $\mathbf{0}$ otherwise whereas that of \mathbf{B}_t is equal to \mathbf{v}_{ti} if $i \in \mathcal{I}(\mathcal{A})$ and to $\mathbf{0}$ otherwise. It remains to prove that \mathbf{V}_t can be directly obtained, using Kronecker products, as $\mathbf{V}_t = \mathbf{1}_C \otimes \mathbf{I}_{C^{m-1}} \otimes \boldsymbol{\lambda}'_{t+m}$. With this aim consider the i -th row of \mathbf{V}_t with $1 \leq i \leq C^{m-1}$. Recalling the (??), when $c = j - C(i - 1) - 1$ is between 0 and $C - 1$, the j -th entry of such a row is equal to $\lambda_{t+m, c}$ whereas in all the other cases this entry is equal to 0. So this row has the form $(\mathbf{0}'_1 \quad \boldsymbol{\lambda}'_t \quad \mathbf{0}'_2)$ where the $\mathbf{0}_1$ is a column vector of $C(i - 1)$ zeros and the block of the first C^{m-1} rows of \mathbf{V}_t is equal to $\mathbf{I}_{C^{m-1}} \otimes \boldsymbol{\lambda}'_{t+m}$. Then, considering that the row i and i' of \mathbf{V}_t are equal if the last $m - 1$ elements of \mathbf{x}_i are equal to those of $\mathbf{x}_{i'}$, it results that

Correctness of the algorithms in section 3

About the algorithm in definition 3, consider that the independence of transition matrices of n and Theorem 1 imply that

$$\mathbf{F}_t = \tilde{\mathbf{A}}_t + \tilde{\mathbf{C}}_t$$

where $\tilde{\mathbf{A}}_t = \mathbf{A}'_t \mathbf{F}_{t-1}$, $\tilde{\mathbf{C}}_t = (\mathbf{0} \quad \tilde{\mathbf{B}}_t^*)$ and $\tilde{\mathbf{B}}_t^*$ is the matrix obtained from $\tilde{\mathbf{B}}_t = \mathbf{B}'_t \mathbf{F}_{t-1}$ by removing the last column. Then, because of the structure of \mathbf{A}_t and \mathbf{B}_t , $\tilde{\mathbf{A}}_t$ and $\tilde{\mathbf{B}}_t$ can also be obtained from $\mathbf{L}_t = \mathbf{V}'_t \mathbf{F}_{t-1}$ by letting the i -th row equal to $\mathbf{0}'$ if, respectively, $i \in \mathcal{I}(\mathcal{A})$ and $i \notin \mathcal{I}(\mathcal{A})$. Moreover, recalling that $\mathbf{V}_t = \mathbf{1}_C \otimes \mathbf{I}_{C^{m-1}} \otimes \boldsymbol{\lambda}'_{t+m}$, we have that

$$\mathbf{L}_t = \mathbf{V}'_t \mathbf{F}_{t-1} = [(\mathbf{1}'_C \otimes \mathbf{I}_{C^{m-1}}) \mathbf{F}_{t-1}] \otimes \boldsymbol{\lambda}_{t+m} = (\mathbf{G}_{t-1} \mathbf{H}) \otimes \boldsymbol{\lambda}_{t+m}$$

and this completes the proof.

To prove the correctness of the algorithms in definition 4 and 5 it is enough to consider that $\mathbf{f}_t(0)$ is the first column of the matrix \mathbf{F}_t . In particular the convergence of the second one is ensured because π_0 decreases as T increases.

Correctness of the iterative relation (2)

Denote for the moment by \mathcal{X}_l the set of all the vectors \mathbf{x} with dimensions $l \times 1$ ordered in the usual way. Then the correctness of the (2) may be proven by considering that, in general, \mathbf{z}_l is a vector whose i -th entry is the sum of the elements of the i -th vector of \mathcal{X}_l . This is may be realised recalling that the elements of \mathcal{X}_l are ordered lexicographically with the last entry of \mathbf{x} running faster. So \mathbf{z}_1 is obviously equal to $(0 \quad \dots \quad C-1)'$; moreover, for $l > 1$, consider that $x_{i1} = 0$ for $i = 1, \dots, C^{l-1}$, $x_{i1} = 1$ for $i = C^{l-1} + 1, \dots, 2C^{l-1}$ and so on where x_{i1} is the first element of the i -th vector of \mathcal{X}_l . Then the block of the first C^{l-1} elements of \mathbf{z}_l is equal to \mathbf{z}_{l-1} , that of the second C^{l-1} elements is equal to $\mathbf{1} + \mathbf{z}_{l-1}$ and so on and the (2) holds.

References

- Anscombe, F. J., Godwin, H. J. and Plackett, R. L., 1947. Methods of deferred sentencing in Testing, *Journal of the Royal Statistical Society*, **7B**, pp. 198-217.
- Brendel, V. and Karlin, S., 1989. Association of charge clusters with functional domains of cellular transcription factors, *Proceedings of the National Academy of Science U.S.A*, **86**, pp. 5698-5702.
- Chang, G. J., Cui, L. and Hwang, F. K., 1999. Reliabilities for (n, f, k) systems, *Statistics and Probability Letters*, **43**, pp. 237-242.
- Fu, Y. and Curnow, R. N., 1990. Locating a changed segment in a sequence of Bernoulli variables, *Biometrika*, **77**, pp. 295-304.

- Glaz, J. and Naus J. I., 1991. Tight bounds and approximations for scan statistic probabilities for discrete data, *The Annals of Applied Probability*, **1**, pp. 306-318.
- Karlin, S., Blaisdell, B., Mocarski, E. and Brendel, V., 1989. A method to identify distinctive charge configurations in protein sequences, with applications to human Herpesvirus polypeptides, *Journal of Molecular Biology*, **205**, pp. 165-177.
- Karwe, V. V. and Naus, J. I., 1997. New recursive methods for scan statistic probabilities, *Computational Statistics & Data Analysis*, **23**, pp. 389-402.
- 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**, pp. 743-766.
- Ling, K. D., 1988. On binomial distribution of order k , *Statistics and Probability Letters*, **6**, pp. 247-250.
- Mood, A. M., 1940. The distribution theory of runs, *Annals of Mathematical Statistics*, **11**, pp. 367-392.
- Mosteller, F., 1941. Note on an application of runs to quality control charts, *Annals of Mathematical Statistics*, **12**, pp. 228-232.
- Naus, J., 1982. Approximations for distribution of scan statistics, *Journal of American Statistical Association*, **77**, pp. 177-183.
- Wallenstien, S., Naus, J. and Glaz, J., 1994. Power of the scan statistics in detecting a changed segment in a Bernoulli sequence, *Biometrika*, **81**, pp. 595-601.
- Wolfowitz, J., 1943. On the theory of runs with some applications to quality control, *Annals of Mathematical Statistics*, **14**, pp. 280-288.