Ann Inst Stat Math (2011) 63:559–584
DOI 10.1007/s10463-009-0240-6

# Binary consecutive covering arrays

**A. P. Godbole · M. V. Koutras · F. S. Milienos**

Received: 25 June 2008 / Revised: 19 January 2009 / Published online: 9 June 2009


**Abstract** A $k \times n$ array with entries from a $q$-letter alphabet is called a $t$-covering array if each $t \times n$ submatrix contains amongst its columns each one of the $q^t$ different words of length $t$ that can be produced by the $q$ letters. In the present article we use a probabilistic approach based on an appropriate Markov chain embedding technique, to study a $t$-covering problem where, instead of looking at all possible $t \times n$ submatrices, we consider only submatrices of dimension $t \times n$ with its rows being consecutive rows of the original $k \times n$ array. Moreover, an exact formula is established for the probability distribution function of the random variable, which enumerates the number of deficient submatrices (i.e., submatrices with at least one missing word, amongst their columns), in the case of a $k \times n$ binary matrix ($q = 2$) obtained by realizing $kn$ Bernoulli variables.

**Keywords** $t$-Covering arrays · Orthogonal arrays · Consecutive covering arrays · Markov chains · Random matrices · Complete factorial designs

A. P. Godbole
Department of Mathematics, East Tennessee State University,
Box 70663, Johnson City, TN 37614, USA
e-mail: godbolea@mail.etsu.edu

M. V. Koutras (✉) · F. S. Milienos
Department of Statistics and Actuarial Science, University of Piraeus,
80 Karaoli and Dimitriou Str., 18534 Piraeus, Greece
e-mail: mkoutras@unipi.gr

F. S. Milienos
e-mail: milienos@yahoo.com

## 1 Introduction

Imagine that a complex circuit needs to be tested for resilience under a wide variety of conditions. In the simplest model considered by Dalal and Mallows (1998) and Hartman (2006), the circuit is switch based. There are $k$ switches, each of which may be kept on or off, and there is a critical number $t$ $(2 \leq t \leq k)$ of switches that measures the maximal comparison we may wish to make between switches. Usually $t$ is held fixed while $n, k$ may grow arbitrarily large. Now, when we run a series of tests on the circuit, we want to ensure that no matter which $t$ switches we want to compare, there is a test in which every possible on–off combination for the chosen switches is represented. So, if we represent the $k$ switches by the rows, and the $n$ tests by the columns, of a binary $k \times n$ matrix, we want to make sure that for each choice of $t$ rows, each of the $2^t$ binary $t$-ples appear at least once among the columns of the selected rows. Of course, cost is a major concern so we wish to use the smallest number of tests.

A second application of the same nature arises if we imagine that a computer scientist has developed software that she/he believes will enjoy almost universal adaptability. Of course, the software needs to be subject to extensive and expensive testing on a variety of computers before it is marketed. There are $k$ factors that may be relevant, such as the computer's cost, operating system, size of hard drive, RAM, country of origin, temperature at which it will operate, age of machine, etc. Also, for cost reasons, we wish to test the software on every combination of $t$ of these factors. Finally, let us suppose that the number of levels of each factor is $q$. More specifically, if the computer's cost is less than 500, between 500 and 1,000, between 1,000 and 1,500, between 1,500 and 2,000, or over 2,000; the operating system is either Windows XP, Windows 2000, Windows Vista, UNIX, or Macintosh; etc., then $q = 5$. Let us imagine that $t = 3$, i.e., we can afford to compare at most three factors at a time. We want to test the software on the smallest number of computers $n$ so that for each choice of $t = 3$ factors (say operating system, country of origin and age of machine) there is a computer that represents each combination of levels of these factors, (e.g. UNIX machine that is made in China and is over 5 years old), that the software is tested on.

We next formalize the above two examples into a key definition. Consider a two-dimensional array of $k$ rows and $n$ columns whose entries are chosen from an alphabet with $q$ letters. If we choose $t$ rows from the $k \times n$ array ($t$ is a fixed integer, $1 < t \leq k$) then across the columns of the selected $t$ rows, $t$-letter words are formed. The original array is called *$t$-covering array*, when every set of $t$ rows contains amongst its columns, all the $q^t$ possible words of length $t$.

The problem of enumerating the $t$-covering arrays has attracted considerable research interest (see, e.g. Renyi 1971; Katona 1973; Kleitman and Spencer 1973; Sloane 1993; Godbole et al. 1996; Colbourn 2004). The determination of the minimum number of columns or the maximum number of rows, so that a $t$-covering array exists has special appeal, since it is closely related to circuit, software or hardware testing problems such as the ones described above. In addition, the rich history of $t$-covering arrays is also justified by their close relationship with the classical combinatorial theory (see, e.g. Kleitman and Spencer 1973; Carey and Godbole 2008).

If every $t \times n$ submatrix contains, amongst its columns, all the $q^t$ possible words of length $t$ exactly $\lambda$ times (where, $\lambda \in \{1, 2, \ldots\}$ and $n = \lambda q^t$), then the original $(k \times n)$ array is called *orthogonal array*. Arrays of this type have also attracted considerable attention from the research community (see, e.g. Hedayat et al. 1999) due to their interesting applications in the area of design of experiments.

A closely related problem to the $t$-covering framework ensues if we restrict the selection of the $t$ rows, by considering only consecutive rows. More specifically, in this model we are looking at all row selections $\{i, i+1, \ldots, i+t-1\}, i = 1, 2, \ldots, k-t+1$ and ask all the $q^t$ possible words of length $t$ to be present in each of the resulted submatrices. A $k \times n$ array possessing this property will be called *$t$-consecutive covering array* ($t$-CCA). The shift from the classical covering array set up to consecutive covering arrays is analogous to the one established when one-dimensional binary sequences are considered and we are looking at run quotas instead of frequency quotas (see, e.g. the monograph by Balakrishnan and Koutras (2002) and the articles Aki (1992), Aki and Hirano (1989), Aki and Hirano (1995) or Aki and Hirano (2004), for the two dimensional case). The analogy to the relationship between covering arrays and consecutive covering arrays is almost obvious. The above CCA model, introduced for the first time in this paper, can be justified at several levels. Most importantly, it often makes little sense to consider non-contiguous rows. The rows may represent consecutive days of the year, or they may represent consecutive switches in a series circuit. In such cases, comparing rows that are far apart is not necessary; the covering array model introduces redundancies and/or irrelevant data.

In the present article we consider a random $k \times n$ array $\mathbf{X} = (X_{ij})$ obtained by realizing $kn$ independent and identically distributed (i.i.d.) Bernoulli trials, with $P(X_{ij} = 1) = p$, and $P(X_{ij} = 0) = 1 - p$. Moreover, we are exploring the problem of determining the probability distribution function of the random variable $W_{k,n,t}$, which enumerates the number of $t \times n$ submatrices (obtained by selecting $t$ consecutive rows) that are deficient (i.e., at least one word of length $t$, is not included amongst its columns). The main tool of our approach is an appropriate Markov chain embedding technique (note that, $W_{k,n,t} \in \{0, 1, \ldots, k-t+1\}$), which could be effortlessly extended to cope up with the non-i.i.d. case as well.

It is worth mentioning that in case of independent and symmetric Bernoulli variables ($p = 1/2$), it holds

$$P(W_{k,n,t} = 0) = \frac{C_{k,n,t}}{2^{kn}}$$

where $C_{k,n,t}$ is the cardinality of the family of $t$-CCA's and should we be able to compute $P(W_{k,n,t} = 0)$, the number of different $t$-CCA's can be easily evaluated as

$$C_{k,n,t} = 2^{kn} P(W_{k,n,t} = 0).$$

Besides, it is easily ascertained that for any values of the parameters $k, n, t$ (with $1 < t \leq k$ and $n \geq 2^t$), there is at least one $t$-CCA.

A potential application of the $t$-CCA's arises from the theory of factorial designs as follows. Assume that we are interested in the influence of the $t = 3$ two-level factors

**Table 1** Complete factorial design

| | | Factor levels for run $i$ | | | | | Response for run $i$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Day | Factor | 1 | 2 | 3 | ... | $n$ | 1 | 2 | 3 | ... | $n$ |
| 1 | $A$ | 0 | 1 | 0 | ... | 0 | | | | | |
| 2 | $B$ | 1 | 1 | 0 | ... | 0 | | | | | |
| 3 | $C$ | 1 | 0 | 1 | ... | 1 | $z_{11}$ | $z_{12}$ | $z_{13}$ | ... | $z_{1n}$ |
| 4 | $A$ | 0 | 1 | 1 | ... | 1 | $z_{21}$ | $z_{22}$ | $z_{23}$ | ... | $z_{2n}$ |
| 5 | $B$ | 1 | 0 | 0 | ... | 0 | $z_{31}$ | $z_{32}$ | $z_{33}$ | ... | $z_{3n}$ |
| 6 | $C$ | 0 | 0 | 0 | ... | 1 | $z_{41}$ | $z_{42}$ | $z_{43}$ | ... | $z_{4n}$ |
| 7 | $A$ | 1 | 1 | 0 | ... | 1 | $z_{51}$ | $z_{52}$ | $z_{53}$ | ... | $z_{5n}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | ... | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | ... | $\vdots$ |
| $k$ | $D$ | 1 | 0 | 1 | ... | 0 | $z_{k-2,1}$ | $z_{k-2,2}$ | $z_{k-2,3}$ | ... | $z_{k-2,n}$ |

$A$, $B$, $C$ on a continuous response variable $Z$. Suppose that we use a random design with $n$ runs, i.e., in each run, we simply assign the levels for every factor by a random mechanism (c.f. Dalal and Mallows 1998), and denote by $p$, $1-p$ the probability of setting a factor at level 1, 0, respectively. Moreover, let us consider the case where the random assignment of the factor levels is repeated every $t = 3$ time periods (e.g. days) in a cyclic fashion, that is, after establishing, on days 1, 2 and 3 an initial random assignment to the $t \times n$ binary level values on all factors $A$, $B$, $C$ on day 4 we repeat the random assignment of the factor levels on factor $A$ only, then (on day 5) we repeat the random assignment on factor $B$ only, then (on day 6) we reassign levels on factor $C$ and so on.

If we denote by $Z_{i-2,j}$, $j = 1, 2, \ldots, n$ the response associated with the $j$th run on the $i$th day, $i = 3, 4, \ldots, k$, our experimental environment may be described by Table 1 ($D \in A, B, C$). Should we wish to carry out, for each day, a complete factorial study (i.e., to investigate all the main effects, all the two-factor interactions and the three-factor interactions) it is clear that the resulting $k \times n$ design matrix must form a $t$-CCA with $t = 3$.

As a consequence, $P(W_{k,n,t} = 0)$ provides the probability to have our task fulfilled in a $k$ days period and the following questions are of prominent importance for the practitioner:

a. for a given time period $k$, what is the minimum number $n$ of runs so that the probability of carrying out a complete factorial study exceeds a prespecified level?
b. for a given number of runs $n$, what is the maximum number of time periods that could be exercised so that the probability of carrying out a complete factorial study exceeds a prespecified level?
c. what are the respective values of $n$ (for question (a)) or $k$ (for question (b)) if we allow for a few incomplete factorial designs? For example we might ask $P(W_{k,n,t} \leq 1)$ to be greater than a prespecified level.

It goes without saying that in the case where $t$ factors (instead of 3) were available, the general setup of the $t$-CCA problem will be needed to treat analogous questions.

In Sect. 2, we illustrate in detail how the probability of the event $W_{k,n,t} = 0$ can be calculated by the aid of an appropriate Markov chain embedding technique, in the special case of $kn$ i.i.d. Bernoulli($p$) trials ($p \in (0, 1)$). In Sect. 3, observing that the random variable $W_{k,n,t}$, belongs to the class of Markov chain embeddable variables of Binomial type (see Koutras and Alexandrou 1995), we establish a recursive formula for computing the whole probability distribution function of $W_{k,n,t}$, by considering an appropriate extension of the embedding exploited in Sect. 2. Section 4 presents briefly how the case of Markov dependent trials can be tackled. Finally, in Sect. 5 we present some numerical results (for specific values of the parameters $k, n, t$) that can be exploited to provide answers to the three questions raised above.

## 2 Consecutive $t$-covering arrays for binary codes

Let $\mathbf{X} = (X_{ij})_{k\times n}$ be a random $k \times n$ matrix with $X_{ij}$ being i.i.d. binary outcomes, with

$$P(X_{ij} = 1) = p, \quad P(X_{ij} = 0) = 1 - p$$

for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n$. In this section, we shall focus on evaluating the probability that all the $t \times n$ submatrices deduced by selecting $t$ consecutive rows, contain (columnwise) at least one each of the $2^t$ possible words. On introducing the enumerating random variable

$$W_{k,n,t} = \sum_{i=1}^{k-t+1} I_i, \tag{1}$$

where

$$I_i = \begin{cases} 1, & \text{if at least one word is missing from the } i\text{th submatrix consisting of} \\ & \text{rows } i, i+1, \ldots, i+t-1 \\ 0, & \text{otherwise} \end{cases}$$

with $i = 1, 2, \ldots, k-t+1$, we may write the probability of interest as $P(W_{k,n,t} = 0)$.

Viewing the formation of the random $k \times n$ matrix as a sequential procedure starting off with a random binary $(t-1) \times n$ matrix and subsequently appending the rest of the rows one-by-one so that the desired requirement is met in each individual step, we may readily give birth to a Markov chain, provided that we keep track of the composition of the last $t - 1$ rows in each step. Then the computation of $P(W_{k,n,t} = 0)$ can be easily carried out by looking at the $(k - t + 1)$-step probabilities of the chain. The interested reader on Markov chain approaches for studying waiting time distributions or the number of occurrences of runs, scans, patterns etc. is referred to Fu and Koutras (1994), Koutras (2003), Koutras and Alexandrou (1995) or the monographs by Fu and Lou (2003) and Balakrishnan and Koutras (2002).

Before stating the results relating to the general case ($t \geq 2$), it would prove to be very useful to illustrate the application of the Markov chain embedding technique, for the special case $t = 2$. Let us consider a Markov chain $\{Y_r, r = 0, 1, \ldots\}$ with state space $\Omega = \Omega_1 \cup \{x_{\text{abs}}\}$ where

$$\Omega_1 = \left\{ (x_1, x_2, \ldots, x_n) : x_i \in \{0, 1\} \quad \text{for } i = 1, 2, \ldots, n \text{ and } 2 \leq \sum_{i=1}^{n} x_i \leq n - 2 \right\}$$

defined as follows: $Y_r = (x_1, x_2, \ldots, x_n)$, $r \geq 1$, if and only if the $(r + 1)$th row of the random matrix $\mathbf{X}$ equals $(x_1, x_2, \ldots, x_n)$ and the submatrix of $\mathbf{X}$ consisting of rows $1, 2, \ldots, r + 1$ meets the consecutive two-covering array criterion. All the other configurations $(x_1, x_2, \ldots, x_n)$ not included in $\Omega_1$ have been accumulated in the absorbing state $x_{\text{abs}}$. Apparently, the number of states in $\Omega_1$ equals

$$s = 2^n - 2(n + 1)$$

while the cardinality of the Markov chain state space $\Omega$ is $|\Omega| = s + 1$. The non-zero transition probabilities among the states of $\Omega_1$ are described as follows:

$$P(Y_r = (x_1, x_2, \ldots, x_n) | \, Y_{r-1} = (x_1', x_2', \ldots, x_n')) = p^{\sum_{j=1}^{n} x_j} (1 - p)^{n - \sum_{j=1}^{n} x_j} \quad (2)$$

if and only if the $2 \times n$ array

$$\begin{pmatrix} x_1' \; x_2' \ldots x_n' \\ x_1 \; x_2 \ldots x_n \end{pmatrix}$$

contains (columnwise) all $2^2 = 4$ possible binary words. Then the probability of the event $P(W_{k,n,2} = 0)$ will be given by the expression

$$P(W_{k,n,2} = 0) = P(Y_{k-1} \neq x_{\text{abs}}) = \boldsymbol{\pi}_1 \Lambda^{k-1} \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}$$

where

$$\Lambda = \begin{pmatrix} P & \mathbf{h}' \\ \mathbf{0} & 1 \end{pmatrix}$$

while the $s \times s$ matrix $P$ refers to the states of $\Omega_1$ [and therefore contains the transition probabilities described in (2)], $\mathbf{0} = (0, 0, \ldots, 0)$ is a $1 \times s$ vector, $\mathbf{h}$ is given by

$$\mathbf{h}' = \mathbf{1}' - P\mathbf{1}' = (I - P)\mathbf{1}', \quad \mathbf{1} = (1, 1, \ldots, 1),$$

and $\boldsymbol{\pi}_1$ is a $1 \times (s + 1)$ vector containing the initial probabilities of the Markov chain. It is worth noting that the above partition of the transition probability matrix $\Lambda$ leads to the next expression

$$\Lambda^{k-1} = \begin{pmatrix} P^{k-1} & (I + P + \cdots + P^{k-2})\mathbf{h}' \\ \mathbf{0} & 1 \end{pmatrix}$$

and $P(W_{k,n,2} = 0)$ admits the more attractive form

$$P(W_{k,n,2}=0)=\boldsymbol{\pi}_0 P^{k-1}\mathbf{1}', \tag{3}$$

where $\boldsymbol{\pi}_0$ is the $1\times s$ vector of initial probabilities

$$\boldsymbol{\pi}_0=(p_1,p_2,\ldots,p_s).$$

In the last vector, $p_i$, $i=1,2,\ldots,s$, stands for the probability of the state $i$, e.g., the probability of the state $(1,1,0,\ldots,0)$ is $p^2(1-p)^{n-2}$. It should be stressed that $\boldsymbol{\pi}_0$ is not a proper distribution vector since the entry corresponding to the absorbing state has been removed from $\boldsymbol{\pi}_1$.

The main disadvantage of the approach used before is that the state space cardinality $s+1=2^n-2(n+1)+1$ increases rapidly with $n$, due to the exponential term $2^n$. Fortunately, an appropriate grouping of the states may result in a much smaller state space. The key point for achieving that is to take advantage of the fact that the transition probabilities depend only on the number of ones (equivalently, zeros) in the state space vector $(x_1,x_2,\ldots,x_n)$ and not on the actual position of them. With this in mind we may merge all the states $(x_1,x_2,\ldots,x_n)$ with $\sum_{i=1}^n x_i=x$ in a single state that will be labeled hereafter as state $x$. Recalling that the configurations with $\sum_{i=1}^n x_i<2$ or $\sum_{i=1}^n x_i>n-2$ have been incorporated to the absorbing state $x_{\text{abs}}$, the state space reduces to $\Omega=\Omega_2\cup x_{\text{abs}}$ where

$$\Omega_2=\{2,3,\ldots,n-2\}$$

and its dimensionality becomes now $|\Omega|=s+1$, with $s=n-3$. Exploiting direct combinatorial arguments, it is not difficult to verify that the non-zero transition probabilities between the states of $\Omega_2$, for the respective Markov chain $\{Y_r, r=0,1,\ldots\}$, are given by

$$P(Y_r=x\,|\,Y_{r-1}=x')=p^x(1-p)^{n-x}\sum_{j=\max\{1,x+x'+1-n\}}^{\min\{x-1,x'-1\}}\binom{x'}{j}\binom{n-x'}{x-j},$$

where $2\le x,x'\le n-2$. The entries of the initial probability vector $\boldsymbol{\pi}_1$ will now be given by

$$p_x=\binom{n}{x+1}p^{x+1}(1-p)^{n-x-1},\quad x=1,2,\ldots,n-3.$$

*Example 1* Let us treat the special case $n=6$ and $p=1/2$ (recall that we are still dealing with the case $t=2$). The reduced state space $\Omega$ becomes

$$\Omega=\Omega_2\cup\{x_{\text{abs}}\}=\{2,3,4,x_{\text{abs}}\}$$

while the transition probabilities between the states of $\Omega_2$ will be given by

$$\begin{aligned}
&P(Y_r = 2 | \, Y_{r-1} = 2) = 8p^6, \quad P(Y_r = 3 | \, Y_{r-1} = 2) = 12p^6, \\
&P(Y_r = 4 | \, Y_{r-1} = 2) = 8p^6, \quad P(Y_r = 2 | \, Y_{r-1} = 3) = 9p^6, \\
&P(Y_r = 3 | \, Y_{r-1} = 3) = 18p^6, \quad P(Y_r = 4 | \, Y_{r-1} = 3) = 9p^6, \\
&P(Y_r = 2 | \, Y_{r-1} = 4) = 8p^6, \quad P(Y_r = 3 | \, Y_{r-1} = 4) = 12p^6, \\
&P(Y_r = 3 | \, Y_{r-1} = 4) = 8p^6.
\end{aligned}$$

The transition probability matrix takes on the form

$$P = \begin{pmatrix} 8p^6 & 12p^6 & 8p^6 \\ 9p^6 & 18p^6 & 9p^6 \\ 8p^6 & 12p^6 & 8p^6 \end{pmatrix}$$

while the vector $\boldsymbol{\pi}_0$ reads

$$\boldsymbol{\pi}_0 = \begin{pmatrix} 15p^6 \; 20p^6 \; 15p^6 \end{pmatrix}.$$

Applying formula (3) for $k = 2, 3, \ldots$ we may easily compute the probabilities associated with the $t$-CCA problem; e.g. for $k = 5$ and $k = 10$ we have

$$P(W_{5,6,2} = 0) = \boldsymbol{\pi}_0 P^4 \mathbf{1}' = 0.0464, \quad P(W_{10,6,2} = 0) = \boldsymbol{\pi}_0 P^9 \mathbf{1}' = 0.0014$$

and the respective numbers $C_{k,6,2}$ of consecutive two-covering arrays for binary codes will be given by

$$\begin{aligned}
C_{5,6,2} &= 2^{5 \cdot 6} P(W_{5,6,2} = 0) = 49{,}774{,}080, \\
C_{10,6,2} &= 2^{10 \cdot 6} P(W_{10,6,2} = 0) = 160{,}109 \times 10^{10}.
\end{aligned}$$

We shall now proceed with the investigation of the general case $t \geq 2$. Let $a = 2^{t-1}$ and denote by $w_1, w_2, \ldots, w_a$ the $a$ different binary words of length $t - 1$ arranged them in the lexicographic order. Arguing in exactly the same way as in the case $t = 2$, we may easily verify that, in order to have appropriate control when shifting from time (row) $r - 1$ to time $r$, the only information needed is the number $x_i$ of appearances of word $w_i$, $i = 1, 2, \ldots, a$, in the last $t - 1$ rows. Therefore, a convenient state space for studying the general case may be introduced as $\Omega = \Omega_2 \cup \{x_{\text{abs}}\}$ where $\Omega_2$ denotes the set of integer solutions of the linear equation

$$x_1 + x_2 + \ldots + x_a = n, \text{ with } x_i \geq 2, \text{ for } i = 1, 2, \ldots, a. \tag{4}$$

The restriction $x_i \geq 2$, $i = 1, 2, \ldots, a$ was imposed because every $(t - 1) \times n$ submatrix of a $t \times n$ matrix that fulfills the $t$-CCA criterion, contains amongst its columns the words $w_1, w_2, \ldots, w_a$ at least two times; as a consequence, should this restriction not be satisfied, the configuration we are looking at is incorporated at the absorbing state.

Taking into account that the number of integer solutions of the aforementioned linear equation equals (see, e.g. Charalambides 2002)

$$s = |\Omega_2| = \binom{a + (n-2a) - 1}{n-2a} = \binom{n-a-1}{a-1},$$

we have $|\Omega| = |\Omega_2| + 1 = s + 1$.

We shall now proceed with the derivation of the transition probabilities $P(Y_r = (x_1, x_2, \ldots, x_a)|Y_{r-1} = (x_1', x_2', \ldots, x_a'))$ for $(x_1, x_2, \ldots, x_a)$, $(x_1', x_2', \ldots, x_a') \in \Omega_2$.

The next lemma provides a condition so that the transition probabilities do not vanish while the theorem following it offers an exact formula for evaluating these probabilities.

**Lemma 1** *A necessary and sufficient condition so that*

$$P(Y_r = (x_1, x_2, \ldots, x_a)|Y_{r-1} = (x_1', x_2', \ldots, x_a')) \neq 0$$

*is given by*

$$x_{2i-1} + x_{2i} = x_i' + x_{i+a/2}', \textit{ for all } i = 1, 2, \ldots, a/2. \tag{5}$$

*Proof* Having arranged the $a = 2^{t-1}$ district words of length $t-1$

$$w_1, w_2, \ldots, w_a$$

in the lexicographic order it can be easily verified that for each $i = 1, 2, \ldots, a/2$ the words $w_{2i-1}$ and $w_{2i}$ have exactly the same first block of length $t-2$; the meaning of this statement is that, if we denote by $w_i(j)$, $j = 1, 2, \ldots, t-1$ the coordinates of the word $w_i$, the following conditions hold true

$$\begin{aligned} &w_{2i-1}(j) = w_{2i}(j), \quad j = 1, 2, \ldots, t-2 \\ &w_{2i-1}(t-1) = 0, \qquad w_{2i}(t-1) = 1. \end{aligned}$$

Likewise, for $i = 1, 2, \ldots, a/2$, the ending blocks of length $t-2$ of the words $w_i$ and $w_{i+a/2}$ coincide, i.e

$$\begin{aligned} &w_i(j) = w_{i+a/2}(j), \quad j = 2, 3, \ldots, t-1 \\ &w_i(1) = 0, \qquad\qquad w_{i+a/2}(1) = 1. \end{aligned}$$

It is also immediate that the starting blocks of the words $w_{2i-1}$, $w_{2i}$ and the ending blocks of the words $w_i$, $w_{i+a/2}$ are exactly the same. Next note that the transition probability $P(Y_r = (x_1, x_2, \ldots, x_a)|Y_{r-1} = (x_1', x_2', \ldots, x_a'))$ describes a situation where from a $(t-1) \times n$ array containing $x_i' + x_{i+a/2}'$ words of the type $w_i$, $w_{i+a/2}$ we obtain (by removing its first row and adding a new row after the last one) a $(t-1) \times n$

array containing $x_{2i-1}+x_{2i}$ words of type $w_{2i-1}, w_{2i}$. Manifestly, such a transition will be feasible if and only if the quantities $x'_i + x'_{i+a/2}$ and $x_{2i-1}+x_{2i}$ coincide. □

We are now ready to prove the main result of this section that offers an exact formula for the evaluation of the non-vanishing transition probabilities for the $t$-CCA problem.

**Theorem 1** *If* $(x_1, x_2, \ldots, x_a) \in \Omega_2$, $(x'_1, x'_2, \ldots, x'_a) \in \Omega_2$ *and*

$$x_{2i-1}+x_{2i} = x'_i + x'_{i+a/2}, \quad \textit{for all } i = 1, 2, \ldots, a/2.$$

*then*

$$\begin{aligned} &P(Y_r = (x_1, x_2, \ldots, x_a) | Y_{r-1} = (x'_1, x'_2, \ldots, x'_a)) \\ &\quad = (p^{\sum_{i=1}^{a/2} x_{2i}} (1-p)^{n-\sum_{i=1}^{a/2} x_{2i}}) \prod_{i=1}^{a/2} c_i \end{aligned} \tag{6}$$

*where*

$$c_i = \binom{x'_{i+a/2}+x'_i}{x_{2i}} - \binom{x'_{i+a/2}}{m_i}\binom{x'_i}{x_{2i}-m_i} - \binom{x'_{i+a/2}}{m'_i}\binom{x'_i}{x_{2i}-m'_i}, \tag{7}$$

$$m_i = \max\{0, x_{2i}-x'_i\}, \quad m'_i = \min\{x_{2i}, x'_{i+a/2}\},$$

*for* $i = 1, 2, \ldots, a/2$.

*Proof* First note that the new $(t-1)\times n$ array whose specification is described by $(x_1, x_2, \ldots, x_a)$ will contain a new row that is placed after the last row of the $(t-1)\times n$ array obtained by removing the first row of the array described by the event $Y_{r-1} = (x'_1, x'_2, \ldots, x'_a)$. The new row includes exactly $b = \sum_{i=1}^{a/2} x_{2i}$ ones and $\sum_{i=1}^{a/2} x_{2i-1} = n - \sum_{i=1}^{a/2} x_{2i} = n-b$ zeros, a fact that justifies the use of the term $p^b(1-p)^{n-b}$ in the expression (6).

Let us next look at the columns of the new $(t-1)\times n$ array that have exactly the same first block of length $t-2$. For a fixed $i \in \{1, 2, \ldots, a/2\}$, the $x_{2i}+x_{2i-1}$ columns containing the words $w_{2i-1}, w_{2i}$ will result from the $x'_i + x'_{i+a/2}$ columns of the old $(t-1)\times n$ array that use the words $w_i, w_{i+a/2}$. Therefore, what we need to do is to distribute the $x_{2i-1}$ zeros and $x_{2i}$ ones in a new row that will assign at least one 0 and at least one 1 in the columns associated with each of the words $w'_i$ and $w'_{i+a/2}$.

Manifestly, the number of different ways to accomplish that equals

$$c_i = \sum_{j=1}^{x_{2i}-1} \binom{x'_{i+a/2}}{j}\binom{x'_i}{x_{2i}-j},$$

where the convention

$$\binom{u}{v} = 0, \quad u \le v,$$

has been used. Without that convention, the lower and upper limits in the summation should be replaced by $\max\{1, x_{2i} - x'_i + 1\}$, and $\min\{x_{2i} - 1, x'_{i+a/2} - 1\}$, respectively. Applying the well known combinatorial identity (Cauchy's formula)

$$\binom{u_1 + u_2}{v} = \sum_{j=\max\{0, v-u_2\}}^{\min\{u_1, v\}} \binom{u_1}{j}\binom{u_2}{v-j},$$

for

$$u_1 = x'_{i+a/2}, u_2 = x'_i, v = x_{2i}$$

we may easily verify that $c_i$ can be expressed equivalently by formula (7).

Since, $c_i$ enumerates the number of distinct ways that the transition from the $x'_i + x'_{i+a/2}$ words $w_i, w_{i+a/2}$ (at time $r-1$) to the $x_{2i-1} + x_{2i}$ words $w_{2i-1}, w_{2i}$ (at time $r$) can be realized (under the conditions described earlier), the transition probability from configuration $(x'_1, x'_2, \ldots, x'_a)$ to $(x_1, x_2, \ldots, x_a)$, can be achieved by $\prod_{i=1}^{a/2} c_i$ different ways. This completes the proof of the theorem. □

In order to obtain the probability of observing a $k \times n$ $t$-CCA, we can apply formula

$$P(W_{k,n,t} = 0) = \boldsymbol{\pi}_0 P^{k-t+1}\mathbf{1}', \quad k \geq t \tag{8}$$

where $\boldsymbol{\pi}_0 = (p_1, p_2, \ldots, p_s)$ is a $1 \times s$ initial vector accounting for the $(t-1) \times n$ binary matrix we are starting with. The entries $p_1, p_2, \ldots, p_s$ correspond to the $s$ integer solutions of the linear equation (4) and provide the probabilities of observing a $(t-1) \times n$ binary matrix with composition $(x_1, x_2, \ldots, x_a)$. The last probabilities equal

$$\frac{n!}{x_1!x_2!\cdots x_a!} p^{\sum_{i=1}^{a} x_i|w_i|}(1-p)^{n(t-1)-\sum_{i=1}^{a} x_i|w_i|} \tag{9}$$

where $|w_i|$, $i = 1, 2, \ldots, a$ denote the number of 1's contained in word $i$; a practically minded reader may alternatively view $\sum_{i=1}^{a} x_i|w_i|$ as the total number of 1's in the $(t-1) \times n$ binary matrix containing $x_i$ times the word $w_i$, for $i = 1, 2, \ldots, a$.

We shall now illustrate the previous approach by considering a special case and presenting the details for obtaining the probabilities of the respective $t$-CCA problem.

*Example 2* Let us assume that we are looking at $k \times n$ arrays with $n = 10$ and we are interested on $t$-CCA probabilities with $t = 3$. Then $a = 2^{t-1} = 4$ and the lexicographically ordered binary words of length $t - 1 = 2$ read

$$w_1 = (0, 0)', \quad w_2 = (0, 1)', \quad w_3 = (1, 0)', \quad w_4 = (1, 1)'.$$

The state space $\Omega$ is now of the form $\Omega = \Omega_2 \cup \{x_{\text{abs}}\}$ where $\Omega_2$ contains the integer solutions of the linear equation

$$x_1 + x_2 + x_3 + x_4 = 10,$$

subject to the conditions $x_i \geq 2$, for $i = 1, 2, 3, 4$. The cardinality of $\Omega_2$ equals

$$s = |\Omega_2| = \binom{n-a-1}{a-1} = \binom{10-4-1}{4-1} = 10$$

and let us denote by $\omega_i$, $i = 1, 2, \ldots, 10$ its element namely

$$\begin{array}{llll} \omega_1 = (4,2,2,2), & \omega_2 = (2,2,4,2), & \omega_3 = (2,4,2,2), & \omega_4 = (2,2,2,4), \\ \omega_5 = (3,2,3,2), & \omega_6 = (3,3,2,2), & \omega_7 = (3,2,2,3), & \omega_8 = (2,2,3,3), \\ \omega_9 = (2,3,2,3), & \omega_{10} = (2,3,3,2). & & \end{array}$$

A direct application of Lemma 1 reveals that the transition probabilities $P(Y_r = (x_1, x_2, \ldots, x_a) | Y_{r-1} = (x'_1, x'_2, \ldots, x'_a))$ are not vanishing if $(x_1, x_2, \ldots, x_a) \in C$ and $(x'_1, x'_2, \ldots, x'_a) \in D$ where

a. $C = \{\omega_1, \omega_3, \omega_6\}$ and $D = \{\omega_1, \omega_2, \omega_5\}$
b. $C = \{\omega_2, \omega_4, \omega_8\}$ and $D = \{\omega_3, \omega_4, \omega_9\}$
c. $C = \{\omega_5, \omega_7, \omega_9, \omega_{10}\}$ and $D = \{\omega_6, \omega_7, \omega_8, \omega_{10}\}$.

For example, in the case $(x_1, x_2, x_3, x_4) = \omega_1 = (4, 2, 2, 2)$ and $(x'_1, x'_2, x'_3, x'_4) = \omega_5 = (3, 2, 3, 2)$, we have

$$x_{2i-1} + x_{2i} = \begin{cases} x_1 + x_2 = 4 + 2 = 6, & i = 1 \\ x_3 + x_4 = 2 + 2 = 4, & i = 2 \end{cases}$$

and

$$x'_i + x'_{i+a/2} = \begin{cases} x'_1 + x'_3 = 3 + 3 = 6, & i = 1 \\ x'_2 + x'_4 = 2 + 2 = 4, & i = 2 \end{cases}$$

which ascertains that condition (5) of Lemma 1 holds true. In order to compute the value of the respective transition probability, it suffices to calculate the quantities (7); since

$$m_i = \begin{cases} \max\{0, 2-3\} = 0, & i = 1 \\ \max\{0, 2-2\} = 0, & i = 2, \end{cases} \qquad m'_i = \begin{cases} \min\{2, 3\} = 2, & i = 1 \\ \min\{2, 2\} = 2, & i = 2, \end{cases}$$

we deduce

$$c_1 = \binom{3+3}{2} - \binom{3}{0}\binom{3}{2-0} - \binom{3}{2}\binom{3}{2-2} = 9$$
$$c_2 = \binom{2+2}{2} - \binom{2}{0}\binom{2}{2-0} - \binom{2}{2}\binom{2}{2-2} = 4$$

and therefore, formula (6) yields

$$P(Y_r = (4,2,2,2)|Y_{r-1} = (3,3,2,2)) = 36p^4(1-p)^6.$$

Working in a similar fashion, we get the next transition probabilities

$$P = \begin{pmatrix} 32b_1 & 0 & 32b_3 & 0 & 0 & 48b_2 & 0 & 0 & 0 & 0 \\ 32b_1 & 0 & 32b_3 & 0 & 0 & 48b_2 & 0 & 0 & 0 & 0 \\ 0 & 32b_1 & 0 & 32b_3 & 0 & 0 & 0 & 48b_2 & 0 & 0 \\ 0 & 32b_1 & 0 & 32b_3 & 0 & 0 & 0 & 48b_2 & 0 & 0 \\ 36b_1 & 0 & 36b_3 & 0 & 0 & 72b_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36b_1 & 0 & 36b_2 & 0 & 36b_3 & 36b_2 \\ 0 & 0 & 0 & 0 & 36b_1 & 0 & 36b_2 & 0 & 36b_3 & 36b_2 \\ 0 & 0 & 0 & 0 & 36b_1 & 0 & 36b_2 & 0 & 36b_3 & 36b_2 \\ 0 & 36b_1 & 0 & 36b_3 & 0 & 0 & 0 & 72b_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36b_1 & 0 & 36b_2 & 0 & 36b_3 & 36b_2 \end{pmatrix}$$

where $b_i = p^{3+i}(1-p)^{7-i}, i = 1,2,3$. The initial vector $\boldsymbol{\pi}_0$ which appears in the formula (8) can be easily evaluated by a direct application of formula (9) as $\boldsymbol{\pi}_0 = (p_1, p_2, \ldots, p_{10})$ where

$$p_1 = \alpha p^8(1-p)^{12}, \quad p_2 = p_3 = \alpha p^{10}(1-p)^{10}, \quad p_4 = \alpha p^{12}(1-p)^8,$$
$$p_5 = p_6 = \beta p^9(1-p)^{11}, \quad p_7 = p_{10} = \beta p^{10}(1-p)^1, \quad p_8 = p_9 = \beta p^{11}(1-p)^9$$

with

$$\alpha = \frac{10!}{4!(2!)^3} = 18{,}900, \quad \beta = \frac{10!}{(2!)^2(3!)^2} = 25{,}200.$$

Finally, the calculation of the $t$-CCA probabilities for the special case $n = 10$, $t = 3$ can be carried out by virtue of the expression

$$P(W_{k,10,3} = 0) = \boldsymbol{\pi}_0 P^{k-2}\mathbf{1}', \quad k \geq 3.$$

It is of interest to note that the last formula may also be exploited to establish a recursive scheme for the sequence of probabilities $a_k = P(W_{k,10,3} = 0), k = 3, 4, \ldots$. To achieve that, it suffices to calculate the generating function

$$\sum_{k=3}^{\infty} a_k z^k = \sum_{k=3}^{\infty} \boldsymbol{\pi}_0 z^2 (zP)^{k-2} \mathbf{1}' = z^3 \boldsymbol{\pi}_0 (I - zP)^{-1} P \mathbf{1}' = z^2 (\boldsymbol{\pi}_0 (I - zP)^{-1} \mathbf{1}' - \boldsymbol{\pi}_0 \mathbf{1}')$$

as a rational function of $z$, i.e., $\sum_{k=3}^{\infty} a_k z^k = P_1(z)/P_2(z)$, write the last identity in the form $P_2(z) \left(\sum_{k=3}^{\infty} a_k z^k\right) = P_1(z)$ and then compare the coefficients of the resulting power series in the right and left hand side.

In the special case of a symmetric $t$-CCA ($p = 1/2$), the transition probability matrix $P$ reduces to

$$P = \frac{1}{2^{10}} \begin{pmatrix} 32 & 0 & 32 & 0 & 0 & 48 & 0 & 0 & 0 & 0 \\ 32 & 0 & 32 & 0 & 0 & 48 & 0 & 0 & 0 & 0 \\ 0 & 32 & 0 & 32 & 0 & 0 & 0 & 48 & 0 & 0 \\ 0 & 32 & 0 & 32 & 0 & 0 & 0 & 48 & 0 & 0 \\ 36 & 0 & 36 & 0 & 0 & 72 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36 & 0 & 36 & 0 & 36 & 36 \\ 0 & 0 & 0 & 0 & 36 & 0 & 36 & 0 & 36 & 36 \\ 0 & 0 & 0 & 0 & 36 & 0 & 36 & 0 & 36 & 36 \\ 0 & 36 & 0 & 36 & 0 & 0 & 0 & 72 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36 & 0 & 36 & 0 & 36 & 36 \end{pmatrix} \tag{10}$$

while the initial vector $\boldsymbol{\pi}_0$ becomes

$$\boldsymbol{\pi}_0 = \frac{1}{2^{20}} (\alpha, \alpha, \alpha, \alpha, \beta, \beta, \beta, \beta, \beta, \beta). \tag{11}$$

Then, the $t$-CCA probabilities $P(W_{k,10,3} = 0)$ for $k = 3, 4, 5$ are readily computed to be

$$P(W_{3,10,3} = 0) = \frac{3.024 \times 10^7}{2^{30}} = 0.02816,$$
$$P(W_{4,10,3} = 0) = \frac{4.08361 \times 10^9}{2^{40}} = 0.00371,$$
$$P(W_{5,10,3} = 0) = \frac{5.53977 \times 10^{11}}{2^{50}} = 0.00049.$$

Moreover, from the expression

$$\sum_{k=3}^{\infty} a_k z^k = \frac{P_1(z)}{P_2(z)} = \frac{-4,725(-409,600z^3 + 384z^4 + 243z^5)}{65,536(1,048,576 - 13,9264z - 576z^2 + 81z^3)}$$

the following simple recurrence relation can be established

$$a_k = \frac{1}{1,048,576}(139,264 a_{k-1} + 576 a_{k-2} - 81 a_{k-3}), \quad k \geq 6$$

which may effortlessly produce all the values of $a_k = P(W_{k,10,3} = 0)$, for $k = 6, 7, \ldots$

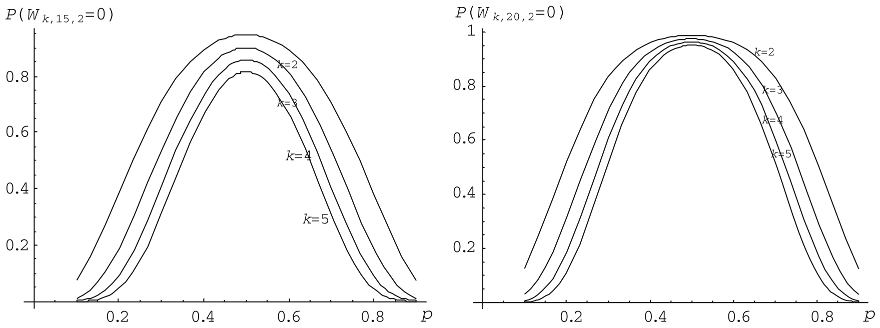


**Fig. 1** The probability $P(W_{k,n,2} = 0)$ as a function of $p$

Two obvious conclusions to be drawn from the numerical values obtained for $P(W_{k,10,3} = 0)$, are the following:

a. $P(W_{k,10,3} = 0)$ is monotically decreasing as $k$ increases (it is not difficult to prove that $\lim_{k\to\infty} P(W_{k,n,t} = 0)$=0, for any $n$, $t$; the details are left to the reader).
b. The numerators in the expressions of $P(W_{k,10,3} = 0)$ as ratios provide the number $C_{k,10,3}$ of distinct binary $t$-CCA's that could be constructed for $n = 10$, $t = 3$ and $k = 3, 4, 5$.

It is straightforward that every word of length $t$ has the same probability of appearance if and only if $p = 0.5$. Therefore, the probability $P(W_{k,n,t} = 0)$ would be expected to attain its maximum value when $p = 0.5$ (for given $k$, $n$, $t$). This is clearly illustrated in Fig. 1, for $t = 2$ and several choices of the parameters $n$, $k$.

As already mentioned in Sect. 1, classical orthogonal array is a $k \times n$ matrix containing, amongst its columns, exactly $\lambda$ times ($\lambda \in \{1, 2, \ldots\}$), each of the $q^t$ possible words of length $t$, in every $t \times n$ submatrix. Manifestly, we could extend the notion of orthogonal arrays by demanding every $t \times n$ submatrix consisting of $t$ consecutive rows of the original $k \times n$ array, to contain amongst its columns all the $q^t$ possible words of length $t$, exactly $\lambda$ times ($n = \lambda q^t$). A random matrix that obeys the aforementioned property, will be called *consecutive orthogonal array*. The probability that a binary random array is a consecutive orthogonal array (event, COA$(k, \lambda, t)$) could be evaluated by the next simple formula which may be effortlessly established either by a Markov chain embedding technique or by the aid of classical combinatorial arguments (recall that, $a = 2^{t-1}$)

$$P(\mathrm{COA}(k, \lambda, t)) = \frac{(2^t\lambda)!}{((2\lambda)!)^a} \cdot \binom{2\lambda}{\lambda}^{a(k-t+1)} ((1-p)\cdot p)^{\lambda a k}.$$

## 3 The distribution of the number of deficient submatrices

In this section, we shall investigate the whole distribution of the enumerating random variable $W_{k,n,t}$ defined in (1). On using the nomenclature "deficient submatrix" to describe a submatrix consisting of rows $i, i+1, \ldots, i+t-1$ $(1 \le i \le k-t+1)$ in which at least one of the $2^t$ possible words is missing, it is apparent that $W_{k,n,t}$ may be interpreted as the number of deficient submatrices in a $k \times n$ binary array.

To elucidate the distribution of $W_{k,n,t}$ let us consider $k = 6, n = 10$ and assume that a realization of the $k \times n$ binary array was

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Then $W_{6,10,2} = 2$ since in the $t \times n = 2 \times 10$ submatrix consisting of rows

a. $\{1, 2\}$ the word $\binom{1}{1}$ is missing
b. $\{5, 6\}$ the words $\binom{0}{1}$, $\binom{1}{0}$ are missing

while the submatrices consisting of rows {2, 3}, {3, 4} contain all the possible binary words of length 2. Likewise, $W_{6,10,3} = 3$ since in the $t \times n = 3 \times 10$ submatrices consisting of rows {1, 2, 3}, {3, 4, 5} and {4, 5, 6} there is at least one missing binary word of length 3.

In order to capture the whole distribution of $W_{k,n,t}$ we shall extend the state space $\Omega_2$ so that we keep track not only of the composition of the $(t-1) \times n$ array which are under inspection at time $r$ but also of the number of deficient submatrices observed up to time $r$ as well. Then, $W_{k,n,t}$ can be viewed as a Markov chain embeddable variable of binomial type (MVB) and the machinery available for them may be exploited to establish the exact distribution of $W_{k,n,t}$.

The formal definition of a MVB is as follows (see Koutras and Alexandrou 1995)

**Definition 1** A family of non-negative (integer valued) random variables $W_v$ defined for $v = 0, 1, \ldots$, will be called a *MVB* if

(a) there exists a Markov chain $\{Y_r, r = 0, 1, \ldots\}$ defined on a discrete state space $\Omega$ which can be partitioned as

$$\Omega = \bigcup_{m \geq 0} C_m, \quad C_m = \{c_{m0}, c_{m1}, \ldots, c_{m,s-1}\}.$$

(without loss of generality we may assume that the state subspaces $C_m$, have the same cardinality $s$)

(b) $P(Y_r \in C_{m_1} | Y_{r-1} \in C_{m_2}) = 0$, for all $m_1 \neq m_2, m_2 + 1$ and $r \geq 1$.

(c) The event $W_v = m$ is equivalent to $Y_v \in C_m$, i.e.,

$$P(W_v = m) = P(Y_v \in C_m), \quad m = 0, 1, \ldots, l_v.$$

where $l_v = \max\{m : P(W_v = m) > 0\}$

From the above definition it is conceivable that the state subclasses $C_m, m \geq 0$ can be ordered in such a way that once the chain enters state $C_m$, the one step transitions may lead exclusively either to the same state subclass or to the next one. These two types of transition give rise to the next $s \times s$ transition probability matrices

$$A_r(m) = (P(Y_r \in C_m | Y_{r-1} \in C_m)), \quad B_r(m) = (P(Y_r \in C_{m+1} | Y_{r-1} \in C_m)),$$

where the first one describes the transitions between states of the same state subclass $C_m$ (one step transition matrix), while the second refers to the transitions between states of two consecutive state subclasses $C_m$ and $C_{m+1}$ (note that, due to condition (b) of the definition, matrix $A_r(m) + B_r(m)$ is a stochastic matrix).

It is sufficient for the needs of the present article to restrict ourselves to the case of homogeneous MVBs where $A_r(m)$ and $B_r(m)$ do not depend on $r$ and $m$, that is

$$A_r(m) = A, \quad B_r(m) = B, \quad \text{for all } r \text{ and } m.$$

In this case, the double sequence of vectors

$$\mathbf{f}_r(m) = \big(P(Y_r = c_{m0}), \ldots, P(Y_r = c_{m,s-1})\big), \quad 0 \le r \le v \text{ and } 0 \le m \le l_v.$$

satisfies the next set of recurrence relations

$$\begin{aligned} \mathbf{f}_r(0) &= \mathbf{f}_{r-1}(0)A, \\ \mathbf{f}_r(m) &= \mathbf{f}_{r-1}(m)A + \mathbf{f}_{r-1}(m-1)B, \quad 1 \le m \le l_v. \end{aligned} \tag{12}$$

for $r = 1, 2, \ldots, v$. A direct application of this scheme, with initial condition

$$\mathbf{f}_0(m) = (P(Y_0 = c_{m0}), \ldots, P(Y_0 = c_{m,s-1}))$$

may then be used to capture the whole distribution of $W_v$ by the aid of the following simple formula (c.f. condition (c) in the definition of the MVB)

$$P(W_v = m) = \mathbf{f}_v(m)\mathbf{1}'$$

($\mathbf{1} = (1, 1, \ldots, 1) \in \Re^s$). To embed the enumerating random variable $W_{k,n,t}$ into a Markov chain let us first use, apart from the $s$ states of $\Omega_2$ introduced in Sect. 2, an additional state $\omega_{s+1}$ indicated that in the $(t-1) \times n$ binary matrix under inspection, there is at least one word (of length $t-1$) which appears less than two times. Incorporating a counter, say $m$, that will keep track of the number of deficient submatrices appeared so far, we arrive at the enhanced state space

$$\Omega^* = (\Omega_2 \cup \{\omega_{s+1}\}) \times \{0, 1, \ldots, k-t+1\}$$

with cardinality

$$|\Omega^*| = (s+1)(k-t+2).$$

Finally, let us define a Markov chain $\{Y_r, r = 0, 1, \ldots\}$ on $\Omega^*$, as follows:

a. $Y_r = (\omega, m)$, with $\omega = (x_1, x_2, \ldots, x_a) \in \Omega_2$ and $0 \le m \le k-t+1$, if and only if the number of appearances of word $w_i$, in the $(t-1) \times n$ submatrix, ending up at row $r+t-1$ equals $x_i$ (with $x_i \ge 2$ for $i = 1, 2, \ldots, a$) and exactly $m$ deficient $t \times n$ submatrices were registered in the binary matrix consisting of rows $1, 2, \ldots, r+t-1$.
b. $Y_r = (\omega, m)$ with $\omega = \omega_{s+1}$ and $0 \le m \le k-t+1$, if and only if at least one of the $a = 2^{t-1}$ words $w_1, w_2, \ldots, w_a$ has appeared less than two times in the $(t-1) \times n$ submatrix, ending up at row $r+t-1$ and exactly $m$ deficient $t \times n$ submatrices were registered in the binary matrix consisting of rows $1, 2, \ldots, r+t-1$.

Considering the natural partition of the state space $\Omega^*$ in the state subspaces

$$\Omega^* = \bigcup_{m \ge 0} C_m, \quad C_m = \{(\omega, m) : \omega \in \Omega_2 \cup \{\omega_{s+1}\}\}, \quad m = 0, 1, \ldots, k-t+1.$$

it is readily ascertained that the $W_{k,n,t}$ becomes a MVB with transition probability matrix $A$ of the form

$$A = \left(\begin{array}{c|c} P & \mathbf{0}' \\ \hline \mathbf{0} & 0 \end{array}\right)_{(s+1)\times(s+1)} \tag{13}$$

where $P$ is the transition probability matrix used in Sect. 2. This follows immediately on observing that the first $s$ rows and $s$ columns of $A$ account for transitions from states of $\Omega_2$ to states of $\Omega_2$ and therefore do not give birth to deficient matrices. On the other hand, if $Y_{r-1} = (\omega_{s+1}, m)$ it is obvious that $Y_r$ cannot enter any of the states of the form $(\omega, m)$ with $\omega \in \Omega_2$, which justifies vanishing probabilities in the last row of $A$.

Before shifting to the investigation of transition probability matrix $B$, we mention that the first $s$ entries of the initial probability vector

$$\mathbf{f}_0(0) = \left(P(Y_0 = (\omega_1, 0)), \ldots, P(Y_0 = (\omega_{s+1}, 0))\right),$$

can be evaluated by direct application of formula (9), while $P(Y_0 = (\omega_{s+1}, 0))$ equals

$$P(Y_0 = (\omega_{s+1}, 0)) = 1 - \sum_{i=1}^{s} P(Y_0 = (\omega_i, 0)). \tag{14}$$

Let us next identify the entries of the transition probability matrix $B$, which may also be partitioned in a similar fashion as follows

$$B = \left(\begin{array}{c|c} Q & \mathbf{c}' \\ \hline \mathbf{b} & \rho \end{array}\right)_{(s+1)\times(s+1).}$$

The transition probabilities appearing in $Q$ are of the form

$$P(Y_r = (\omega, m+1) | Y_{r-1} = (\omega', m)),$$

where

$$\omega = (x_1, x_2, \ldots, x_a) \in \Omega_2 \quad \text{and} \quad \omega' = (x'_1, x'_2, \ldots, x'_a) \in \Omega_2.$$

In view of Lemma 1, if condition (5) is not satisfied, state $\omega$ is unreachable (in one step) from $\omega'$ and therefore $P(Y_r = (\omega, m+1)|Y_{r-1} = (\omega', m)) = 0$. On the other hand, if condition (5) is satisfied, the probability $P(Y_r = (\omega, m)|Y_{r-1} = (\omega', m))$, which coincides with quantity given in (6) has been included in the submatrix $P$ of $A$; consequently the submatrix $Q$ of $B$ will receive the remaining probability which may be easily expressed as

$$P(Y_r = (\omega, m+1)|Y_{r-1} = (\omega', m)) \\ = p^{\sum_{i=1}^{a/2} x_{2i}} (1-p)^{n-\sum_{i=1}^{a/2} x_{2i}} \left(\prod_{i=1}^{a/2} d_i - \prod_{i=1}^{a/2} c_i\right), \tag{15}$$

where

$$d_i = \binom{x_{2i} + x_{2i-1}}{x_{2i}}, \quad i = 1, 2, \ldots, a/2.$$

Having established the entries of submatrix $Q$, we can evaluate the entries of the vector $\mathbf{c}$ by the aid of formula

$$\mathbf{c}' = \mathbf{1}' - Q\mathbf{1}' - P\mathbf{1}' \tag{16}$$

which is readily ascertainable form the fact that $A + B$ is stochastic and the respective entries of $A$ are vanishing.

Let us finally look at the vector $\mathbf{b}$, which consists of transition probabilities of the form

$$P(Y_r = (\omega, m+1)|Y_{r-1} = (\omega_{s+1}, m)) \tag{17}$$

where $\omega = (x_1, x_2, \ldots, x_a) \in \Omega_2$. Denote by $\omega_{s+1,j} = (x'_{1j}, x'_{2j}, \ldots, x'_{aj})$, $j =$ 1,2, $\ldots$, $h$ the solutions of the next integer equation

$$x'_{1j} + x'_{2j} + \cdots + x'_{aj} = n,$$

under the condition $x'_i < 2$, for at least one $i \in \{1, 2, \ldots, a\}$ and $x'_{ij} + x'_{i+a/2,j} = x_{2i-1} + x_{2i}$ for all $i \in \{1, 2, \ldots, a\}$. Then

$$h = \prod_{i=1}^{a/2}(x_{2i} + x_{2i-1} + 1) - \prod_{i=1}^{a/2}(x_{2i} + x_{2i-1} - 3),$$

and (17) can be expressed as

$$P(Y_r = (\omega, m+1)|Y_{r-1} = (\omega_{s+1}, m)) = \frac{p_0}{P(Y_0 = (\omega_{s+1}, 0))} \sum_{j=1}^{h} p_{\omega_{s+1,j}}, \quad (18)$$

where $p_{\omega_{s+1,j}}$ are probabilities evaluated by a formula that parallels formula (9), $P(Y_0 = (\omega_{s+1}, 0))$ is given by formula (14) and

$$p_0 = p^{\sum_{i=1}^{a/2} x_{2i}} (1-p)^{n - \sum_{i=1}^{a/2} x_{2i}} \prod_{i=1}^{a/2} d_i.$$

It is worth mentioning that the sum appearing in (18) may be alternative expressed as

$$\sum_{\substack{0 \le y_i \le x_{2i-1} + x_{2i}, \\ \text{for } i=1,2,\ldots,a/2}} \frac{n!}{\prod_{i=1}^{a/2} y_i!(x_{2i} + x_{2i-1} - y_i)!} \pi_\omega$$
$$- \sum_{\substack{2 \le y_i \le x_{2i-1} + x_{2i} - 2, \\ \text{for } i=1,2,\ldots,a/2}} \frac{n!}{\prod_{i=1}^{a/2} y_i!(x_{2i} + x_{2i-1} - y_i)!} \pi_\omega$$

where

$$\pi_\omega = p^{\sum_{i=1}^{a/2} (x_{2i} + x_{2i-1})|w_i| + n - \sum_{i=1}^{a/2} y_i} (1-p)^{n(t-1) - \sum_{i=1}^{a/2} (x_{2i} + x_{2i-1})|w_i| - n + \sum_{i=1}^{a/2} y_i}.$$

Manifestly, in the case of i.i.d. symmetric Bernoulli trials, $\pi_\omega$ does not depend on $\omega \in \Omega_2$ ($\pi_\omega = 2^{-n(t-1)}$ for all $\omega \in \Omega_2$), while for $t = 2$ the transition probability (18) simplifies to

$$P(Y_r = (\omega, m+1)|Y_{r-1} = (\omega_{s+1}, m)) = \binom{n}{x_2} p^{x_2} (1-p)^{n-x_2}.$$

(since, $P(Y_0 = (\omega_{s+1}, 0)) = \sum_{j=1}^{h} p_{\omega_{s+1,j}}$). The foregone analysis is further elucidated in the next example.

*Example 3* Let us assume that $n = 10$, $t = 3$ and $p = 1/2$ (i.i.d. symmetric Bernoulli trials). Then recalling the notation and analysis carried out earlier in the Example 2, we may write the state space $\Omega^*$ as

$$\Omega^* = \{(\omega, m) : \omega = (x_1, x_2, x_3, x_4) \in \Omega_2 \cup \{\omega_{s+1}\} \text{ and } 0 \le m \le k-2\},$$

where $\Omega_2 = \{\omega_1, \omega_2, \ldots, \omega_{10}\}$. The first 10 coordinates of the initial probability vector $\mathbf{f}_0(0)$ are the entries of $\boldsymbol{\pi}_0$ in formula (1) and therefore $\mathbf{f}_0(0) = (\boldsymbol{\pi}_0, 1 - \boldsymbol{\pi}_0 \mathbf{1}')$ which yields

$$\mathbf{f}_0(0) = \frac{1}{2^{20}} (\alpha, \alpha, \alpha, \alpha, \beta, \beta, \beta, \beta, \beta, \beta, \gamma),$$

with, $\alpha = 18,900$, $\beta = 25,200$ and $\gamma = 2^{20} - 4\alpha - 6\beta$. The transition probability matrix $A$ will have a blocked-matrix representation described in (13) with $P$ provided by (10).

The entries of the submatrix $Q$ of the transition probability $B$ are easily computed by virtue of (15). In the special case we are dealing with, formula (15) reads

$$P(Y_r = (\omega, m+1)|Y_{r-1} = (\omega', m)) = \frac{1}{2^{10}}\binom{x_1 + x_2}{x_2}\binom{x_3 + x_4}{x_4} - p(\omega, \omega')$$

where $p(\omega, \omega')$ is the respective entry in matrix $P$. Thus, for $\omega_1 = (4, 2, 2, 2)$, we get

$$P(Y_r = (\omega_1, m+1)|Y_{r-1} = (\omega_1, m)) = \frac{1}{2^{10}}\left(\binom{6}{2}\binom{4}{2} - 32\right) = \frac{58}{2^{10}}$$

and working in a similar manner for the rest non-vanishing entries of $Q$ we obtain

$$Q = \frac{1}{2^{10}}\begin{pmatrix} 58 & 0 & 58 & 0 & 0 & 72 & 0 & 0 & 0 & 0 \\ 58 & 0 & 58 & 0 & 0 & 72 & 0 & 0 & 0 & 0 \\ 0 & 58 & 0 & 58 & 0 & 0 & 0 & 72 & 0 & 0 \\ 0 & 58 & 0 & 58 & 0 & 0 & 0 & 72 & 0 & 0 \\ 54 & 0 & 54 & 0 & 0 & 48 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 & 64 & 0 & 64 & 64 \\ 0 & 0 & 0 & 0 & 64 & 0 & 64 & 0 & 64 & 64 \\ 0 & 0 & 0 & 0 & 64 & 0 & 64 & 0 & 64 & 64 \\ 0 & 54 & 0 & 54 & 0 & 0 & 0 & 48 & 0 & 0 \\ 0 & 0 & 0 & 0 & 64 & 0 & 64 & 0 & 64 & 64 \end{pmatrix}.$$

The vector $\mathbf{c}$ appearing in the last column of $B$ is easily evaluated by exploiting the formula (16) as

$$\mathbf{c}' = \mathbf{1}' - Q\mathbf{1}' - P\mathbf{1}' = \frac{1}{2^{10}}(\delta, \delta, \delta, \delta, \epsilon, \epsilon, \epsilon, \delta, \epsilon)'$$

where $\delta = 724$, $\epsilon = 624$. Finally, the vector $\mathbf{b}$ results by a direct application of (18), while for $n = 10$, $t = 3$ we may easily identify $\mathbf{b}$ as

$$\mathbf{b} = \frac{1}{2^{10}}(\zeta, \zeta, \zeta, \zeta, \eta, \theta, \eta, \theta, \eta, \eta)$$

where $\zeta = 16.6513$, $\eta = 19.1351$ and $\theta = 22.2017$. Needles to say, the quantity $\rho$ appearing in the last row and column of matrix $B$, equals

$$\rho = 1 - \mathbf{b1}' = 1 - 2^{-10}(4\zeta + 4\eta + 2\theta) = \frac{1}{2^{10}}836,451.$$

Launching the recurrence scheme (12) we deduce immediately the whole distribution of $W_{k,n,t}$ for $n = 10, t = 3, k = 5$ as follows

$$
\begin{aligned}
P(W_{5,10,3} = 0) &= \mathbf{f}_0(0)A^3\mathbf{1}' = 0.00049\\
P(W_{5,10,3} = 1) &= \mathbf{f}_0(0)(BA^2 + ABA + A^2B)\mathbf{1}' = 0.00722\\
P(W_{5,10,3} = 2) &= \mathbf{f}_0(0)(B^2A + BAB + AB^2)\mathbf{1}' = 0.06858\\
P(W_{5,10,3} = 3) &= \mathbf{f}_0(0)B^3\mathbf{1}' = 0.92371.
\end{aligned}
$$

## 4 Consecutive $t$-covering arrays for Markov dependent trials

The results detailed in the previous section can be effortlessly extended for the case of binary arrays where the trials in the same column share a Markov dependence while the columns itself are independent. In the present section we shall briefly illustrate the necessary modifications that should be carried out in order to accommodate in our model a first order Markov dependence, in the special case $t = 2$. The consideration of the special case $t = 2$ was for notation convenience, while the ideas can be manifestly extended to cover the case of any $t \geq 2$ and Markov dependence of order (at most) $t - 1$.

Let $\mathbf{X} = (X_{ij})_{k\times n}$, be a binary random matrix with independent columns and assume that, for a fixed $j \in \{1, 2, \ldots, n\}$, the sequence $\{X_{ij}, i = 1, 2, \ldots, k\}$ is a time homogeneous two-state Markov chain with transition probabilities

$$
\begin{aligned}
P(X_{ij} = 1|X_{i-1,j} = 0) = p_{01}, \quad P(X_{ij} = 0|X_{i-1,j} = 0) = p_{00},\\
P(X_{ij} = 1|X_{i-1,j} = 1) = p_{11}, \quad P(X_{ij} = 0|X_{i-1,j} = 1) = p_{10},
\end{aligned}
$$

for $i = 1, 2, \ldots, k$. Then the number of deficient $2 \times n$ submatrices can be treated as a MVB on the state space

$$\Omega^* = \{(x, m) : x \in \{2, 3, \ldots, n-1\}, \quad m = 0, 1, \ldots, k-t+1\},$$

where $x \in \{2, 3, \ldots, n-2\}$ keeps track of the number of 1's in the row under investigation, while $x = n - 1$ declares that the number of 1's is 0, 1, $n - 1$ or $n$. Moreover, the counter $m$ indicates the number of deficient $2 \times n$ submatrices observed so far.

The transition probabilities associated with submatrix $P$ of transition probability matrix $A$ are now expressed as

$$
\begin{aligned}
&P(Y_r = (x, m)|Y_{r-1} = (x', m))\\
&\quad = \sum_{j=\max\{1,x+x'+1-n\}}^{\min\{x-1,x'-1\}} \binom{x'}{j}\binom{n-x'}{x-j} p_{11}^{j} p_{01}^{x-j} p_{10}^{x'-j} p_{00}^{n-x'-x+j}
\end{aligned}
$$

for $x, x' \in \{2, 3, \ldots, n-2\}$ while the rest entries of $A$ are vanishing. The transition probabilities of submatrix $Q$ of transition probability matrix $B$ take on the form

$$\begin{aligned} P(Y_r &= (x, m+1) | Y_{r-1} = (x', m)) \\ &= \sum_{j=0}^{x} \binom{x'}{j} \binom{n-x'}{x-j} p_{11}^{j} p_{01}^{x-j} p_{10}^{x'-j} p_{00}^{n-x'-x+j} \\ &\quad - \sum_{j=\max\{1, x+x'+1-n\}}^{\min\{x-1, x'-1\}} \binom{x'}{j} \binom{n-x'}{x-j} p_{11}^{j} p_{01}^{x-j} p_{10}^{x'-j} p_{00}^{n-x'-x+j} \end{aligned}$$

for $x, x' \in \{2, 3, \ldots, n-2\}$. The coordinates of the vector $\mathbf{b}$ appearing in the last row of $B$ can be evaluated by the aid of the next formula

$$\begin{aligned} P(Y_r &= (x, m+1) | Y_{r-1} = (n-1, m)) \\ &= \sum_{y \in \{0, 1, n-1, n\}} \sum_{j=0}^{x} \binom{y}{j} \binom{n-y}{x-j} p_{11}^{j} p_{01}^{x-j} p_{10}^{y-j} p_{00}^{n-y-x+j}, \\ &\quad x \in \{2, 3, \ldots, n-2\}. \end{aligned}$$

Finally, if we assume that the binary entries of the first row are i.i.d. random variables with success (failure) probability $p$ $(1-p)$, then the initial row vector $\mathbf{f}_0(0)$ needed to launch the recursive scheme (12) will contain the respective binomial probabilities as entries

$$\binom{n}{i+1} p^{i+1} (1-p)^{n-i-1}, \quad i = 1, 2, \ldots, n-2,$$

(with the last entry being the complementary probability of the sum of them, with respect to 1). Needless to say, should the binary entries be non i.i.d., the entries of $\mathbf{f}_0(0)$ would be probabilities associated with the generalized binomial distribution.

## 5 Numerical calculations

In this section, we shall illustrate how the results developed in the previous sections can be exploited to treat the complete factorial design problem mentioned in Sect. 1.

Let us first state the problem in its general setup. Assume that we are interested in investigating the influence of $t$ two-level (0-1) factors $A_1, A_2, \ldots, A_t$ on a continuous response variable $Z$. A random design with $n$ runs in each time period (e.g. day) is created by changing only the levels of a single factor per day, using a random mechanism that assigns level 1 with probability $p$ and level 0 with probability $1-p$; this assignment is repeated $n$ times for each day, while the factors used in consecutive days are rotating in a cyclic fashion as follows: $A_1, A_2, \ldots, A_t, A_1, A_2, \ldots$ and so on. On day $i$, we collect $n$ measurements of the response variable $Z$, say $z_{ij}$, $j = 1, 2, \ldots, n$ which correspond to the $n$ treatments consisting of the factor levels of the $t$ factors,

**Table 2** Case $t = 2$ and $p = 1/2$ ($P(W_{k,n,2} = 0)$)

| $k$ | $n = 12$ | $n = 13$ | $n = 14$ | $n = 15$ | $n = 16$ | $n = 17$ | $n = 18$ | $n = 19$ | $n = 20$ |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.8748 | 0.9057 | 0.9291 | 0.9467 | 0.9600 | 0.9700 | 0.9775 | 0.9831 | 0.9873 |
| 3 | 0.7782 | 0.8293 | 0.8694 | 0.9004 | 0.9244 | 0.9427 | 0.9566 | 0.9673 | 0.9753 |
| 4 | 0.6923 | 0.7593 | 0.8135 | 0.8564 | 0.8900 | 0.9161 | 0.9363 | 0.9517 | 0.9634 |
| 5 | 0.6158 | 0.6953 | 0.7612 | 0.8145 | 0.8570 | 0.8903 | 0.9163 | 0.9363 | 0.9517 |
| 6 | 0.5478 | 0.6366 | 0.7122 | 0.7747 | 0.8252 | 0.8653 | 0.8968 | 0.9212 | 0.9401 |
| 7 | 0.4873 | 0.5829 | 0.6664 | 0.7368 | 0.7945 | 0.8409 | 0.8777 | 0.9064 | 0.9287 |
| 8 | 0.4335 | 0.5337 | 0.6236 | 0.7007 | 0.7650 | 0.8172 | 0.8589 | 0.8918 | 0.9174 |

**Table 3** The minimum number of $n$, for which $P(W_{k,n,3} = 0) \geq 1 - a$ ($p = 1/2$)

| $k$ | $a = 0.10$ | $a = 0.05$ | $a = 0.025$ | $a = 0.01$ | $k$ | $a = 0.10$ | $a = 0.05$ | $a = 0.025$ | $a = 0.01$ |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 33 | 38 | 44 | 50 | 10 | 48 | 54 | 59 | 66 |
| 4 | 38 | 43 | 48 | 55 | 11 | 49 | 54 | 60 | 66 |
| 5 | 41 | 46 | 52 | 58 | 12 | 50 | 55 | 61 | 67 |
| 6 | 43 | 48 | 54 | 60 | 13 | 51 | 56 | 61 | 68 |
| 7 | 45 | 50 | 55 | 62 | 14 | 51 | 57 | 62 | 69 |
| 8 | 46 | 51 | 57 | 63 | 15 | 52 | 57 | 62 | 69 |
| 9 | 47 | 53 | 58 | 65 | 16 | 52 | 58 | 63 | 70 |

that have been established in the last $t$ days (i.e., days $i - t + 1, i - t + 2, \ldots, i$). Manifestly, no measurements are collected on the first $t - 1$ days, since the assignment process is not complete until the $t$th day is reached.

If we wish to carry out a complete factorial study (i.e., to investigate all the main effects and interactions of all higher orders), for days $t, t + 1, \ldots, k$ it is clear that the resulting $k \times n$ design matrix should form a $t$-CCA. Then for fixed $k$, the minimum number of runs ($n$) needed to achieve our goal with probability at least $1 - a$ ($0 < a < 1$) can be formally expressed as

$$n_{\min} = \min\{n : P(W_{k,n,t} = 0) \geq 1 - a\}.$$

Table 2, which displays the values of the probability $P(W_{k,n,t} = 0)$ for $t = 2$, $p = 1/2$ can be exploited to provide the answer to question (a) that was raised in Sect. 1. For example, if we wish to secure a 90% probability that complete factorial studies will be feasible for $k = 5$, we should have at least $n_{\min} = 18$ runs in each day. Extending the time frame to $k = 8$ will require $n_{\min} = 20$ runs, to achieve the same goal.

For the practical minded reader, a table that will provide the minimum number $n$ of runs so that $P(W_{k,n,t} = 0) \geq 1 - a$ for the typical choices of $a$, will be much more useful. Table 3 provides this information for $t = 3$, $p = 1/2$, $k \in \{3, 4, \ldots, 16\}$ and $a = 0.01, 0.025, 0.05, 0.10$. Needles to say the range of the tabulated parameters can be effortlessly extended by exploiting the results developed in the previous paragraphs.

**Table 4** Case $t = 3$ and $p = 1/2$ (CDF : $P(W_{k,n,3} \leq i)$)

| $k$ | $i$ | $n = 25$ | $n = 26$ | $n = 27$ | $n = 28$ | $n = 29$ | $n = 30$ | $n = 31$ | $n = 32$ | $n = 33$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 0 | 0.5734 | 0.6157 | 0.6549 | 0.6910 | 0.7240 | 0.7541 | 0.7813 | 0.8058 | 0.8279 |
| | 1 | 0.8999 | 0.9184 | 0.9336 | 0.9461 | 0.9562 | 0.9645 | 0.9713 | 0.9767 | 0.9812 |
| | 2 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 5 | 0 | 0.4485 | 0.4962 | 0.5419 | 0.5851 | 0.6255 | 0.6631 | 0.6978 | 0.7296 | 0.7586 |
| | 1 | 0.7995 | 0.8334 | 0.8622 | 0.8865 | 0.9067 | 0.9236 | 0.9376 | 0.9491 | 0.9585 |
| | 2 | 0.9619 | 0.9714 | 0.9787 | 0.9841 | 0.9881 | 0.9912 | 0.9935 | 0.9951 | 0.9963 |
| | 3 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 6 | 0 | 0.3508 | 0.3999 | 0.4483 | 0.4953 | 0.5404 | 0.5831 | 0.6232 | 0.6606 | 0.6952 |
| | 1 | 0.7004 | 0.7470 | 0.7878 | 0.8230 | 0.8531 | 0.8786 | 0.9001 | 0.9180 | 0.9329 |
| | 2 | 0.9095 | 0.9308 | 0.9474 | 0.9603 | 0.9701 | 0.9775 | 0.9832 | 0.9875 | 0.9907 |
| | 3 | 0.9860 | 0.9903 | 0.9934 | 0.9955 | 0.9970 | 0.9979 | 0.9986 | 0.9991 | 0.9994 |
| | 4 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 7 | 0 | 0.2743 | 0.3223 | 0.3709 | 0.4194 | 0.4668 | 0.5127 | 0.5566 | 0.5981 | 0.6370 |
| | 1 | 0.6065 | 0.6628 | 0.7134 | 0.7582 | 0.7973 | 0.8310 | 0.8599 | 0.8843 | 0.9048 |
| | 2 | 0.8468 | 0.8809 | 0.9081 | 0.9296 | 0.9464 | 0.9594 | 0.9695 | 0.9771 | 0.9828 |
| | 3 | 0.9608 | 0.9726 | 0.9809 | 0.9868 | 0.9909 | 0.9938 | 0.9958 | 0.9971 | 0.9980 |
| | 4 | 0.9948 | 0.9968 | 0.9980 | 0.9987 | 0.9992 | 0.9995 | 0.9997 | 0.9998 | 0.9998 |
| | 5 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

The answer to question (b) raised in Sect. 1, which can be formally stated as

$$k_{\max} = \max\{k : P(W_{k,n,t} = 0) \geq 1 - a\},$$

is also extracted from the entries of Table 2 (in the spacial case $t = 2$, $p = 1/2$ ). For example, if only $n = 15$ runs are to be carried out in each day, the longest time period in which complete factorial designs are feasible with probability at least 90%, corresponds to $k = 3$ days. If we could afford for $n = 19$ days runs per day, the time period for achieving the same goal, increases to $k = 7$. Table 3 can also be used for spotting out the maximum value of $k$ for $t = 3$, $p = 1/2$ and several choices of $n$ and $a$.

In closing, let us turn our attention to question (c) raised in Sect. 1. We are now interested in calculating

$$n_{\min}(r) = \min\{n : P(W_{k,n,t} \leq r) \geq 1 - a\},$$

for prespecified values of $r \in \{1, 2, \ldots\}$ and $a \in (0, 1)$. Table 4 offers the cumulative distribution function (CDF) of $W_{k,n,t}$ for $t = 3$, $p = 1/2$ and a variety of choices of $n$ and $k$. According to this table, should we wish to achieve a 90% probability that the number of deficient submatrices in our factorial design does not exceed $r = 1$, we have to carry out at least $n_{\min}(1) = 26$ runs per day if our time frame is described by $k = 4$, $n_{\min}(1) = 29$ runs if $k = 5$ and so on. It is evident that the value of $n$ arising under this set up will always be smaller than the respective value of $n$ for a complete factorial design (provided that the rest parameters are kept fixed).

**Acknowledgments** Research of the last author was supported by the National Scholarship Foundation of Greece.

## References

Aki, S. (1992). Waiting time problems for a sequence of discrete random variables. *Annals of the Institute of Statistical Mathematics, 44*, 363–378.

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

Aki, S., Hirano, K. (1995). Joint distributions of numbers of success-runs and failures until the first consecutive $k$ successes. *Annals of the Institute of Statistical Mathematics, 47*, 225–235.

Aki, S., Hirano, K. (2004). Waiting time problems for a two-dimensional pattern. *Annals of the Institute of Statistical Mathematics, 56*, 169–182.

Balakrishnan, N., Koutras, M. V. (2002). *Runs, scans and applications*. New York: Wiley.

Carey, P. A., Godbole, A. P. (2008). Partial covering arrays and a generalized Erdös-Ko-Rado property (under revision).

Charalambides, Ch. A. (2002). *Enumerative combinatorics*. Boca Raton: Chapman & Hall.

Colbourn, C. J. (2004). Combinatorial aspects of covering arrays. *Le Matematiche (Catania), 58*, 121–167.

Dalal, S. R., Mallows, C. L. (1998). Factor-covering designs for testing software. *Technometrics, 40*, 234–243.

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

Fu, J. C., Lou, W. Y. W. (2003). *Distribution theory of runs and patterns and its applications: A finite markov chain imbedding approach*. Singapore: World Scientific.

Godbole, A. P., Skipper, D. E., Sunley, R. A. (1996). $t$-covering arrays: Upper bounds and Poisson approximations. *Combinatorics, Probability and Computing, 5*, 105–118.

Hartman, A. (2006). Software and hardware testing using combinatorial covering suites. In M.C. Golumbic, A. Hartman (Eds.), *Interdisciplinary applications of graph theory, combinatorics and algorithms* (pp. 237–266). Berlin: Springer.

Hedayat, A. S., Sloane, N. J. A., Stufken, J. (1999). *Orthogonal arrays*. New York: Springer.

Katona, G. (1973). Two applications (for search theory and truth functions) of Sperner type theorems. *Periodica Mathematica Hungarica, 3*, 19–26.

Kleitman, D., Spencer, J. (1973). Families of $k$-independent sets. *Discrete Mathematics, 6*, 255–262.

Koutras, M. V. (2003). Applications of Markov chains to the distribution theory of runs and patterns. In D.N. Shanbhag, C. R. Rao (Eds.), *Handbook of statistics: Vol. 21. Stochastic processes, modeling and simulation*. Amsterdam: Elsevier.

Koutras, M. V., 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.

Renyi, A. (1971). *Foundations of probability*. New York: Wiley.

Sloane, N. J. A. (1993). Covering arrays and intersecting codes. *Journal of Combinatorial Designs, 1*, 51–63.