THE DISTRIBUTION OF FIRST OCCURRENCE OF PATTERNS OF OUTCOMES OF BERNOULLI EXPERIMENTS

By J.P. McDermott and J.N. Sheahan*
Department of Mathematics, National University of Ireland, Galway

[Received 4 September 2001. Read 12 September 2002. Published November 2004.]

Abstract

Let X_1, $X_2, \ldots$ be a sequence of independent Bernoulli random variables. For any fixed integer $K \geq 1$, let $\underline{x}_K = (x_1, x_2, \ldots, x_K)$ be any one of the 2^K vectors each of whose components is 0 or 1. We study the first occurrence of the pattern $\underline{x}_K$ in $X_1, X_2, \ldots$, and characterise the factors that explain why some sequences $\underline{x}_K$ are more likely to occur before other sequences of the same length. Explicit formulae and a simple recursion are derived for the distribution of the first occurrence of the vector $\underline{x}_K$ in the sequence $X_1, X_2, \ldots$.

1. Introduction

This paper is concerned with the distribution of the first occurrence of any given vector in a sequence of outcomes of independent Bernoulli random variables. Let X_1, $X_2, \ldots$ be a sequence of independent random variables with the Bernoulli distribution $P(X = x) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$. In coin tossing, for example, we can set X_i equal to 1 or 0 according to whether a head (H) or a tail (T) results from the ith toss. Let K be any positive integer and let $\underline{x}_K = (x_1, x_2, \ldots, x_K)$ be any one of the 2^K possible values of $\underline{X} = (X_1, X_2, \ldots, X_K)$. For $i = 1,2,\ldots$, we denote by block i the K positions occupied by the variables X_i, X_{i+1}, $\ldots$, X_{i+K-1}, and we let $N = N(\underline{x}_K)$ be the number of the block at which $\underline{x}_K$ occurs for the first time in the sequence X_1, $X_2, \ldots$. In queueing terminology, N may be referred to as the *waiting time* until the sequence first occurs. For example, if the vector of interest is $\underline{x}_3 = (1, 1, 0)$, then the waiting time N would have the value $n = 7$ if we observe a sequence beginning with 1,0,1,0,0,0,1,1,0,…, and N would have the value 2 for any sequence that commences with either 0,1,1,0 or 1,1,1,0. In his applications of renewal theory, Ross [6] refers to $\underline{x}_K$ as a *pattern*, and N (denoted T in [6]) is called the first *renewal time* of the pattern.

For $K = 1$ and $\underline{x}_1 = (1)$, the density of N is, of course, geometric: $P(N = n) = (1-\theta)^{n-1}\theta$, $n = 1,2,\ldots$. The case $K = 2$ is solved in Hombas [3]. Our goal is the derivation of the distribution N for any given vector $\underline{x}_K$ of arbitrary length K.

The motivation for the present work is as follows. In the context of an experiment involving tossing of a fair coin ($\theta = 1/2$), Konold [4] describes an interesting simulation study to compare the probabilities of first occurrence of the sequences HHHHH and HTHHT, i.e. $\underline{x}_5 = (1, 1, 1, 1, 1)$ and $\underline{x}_5 = (1, 0, 1, 1, 0)$, and points out that Engel [1] gives an algorithm to determine that the expected number of

*Corresponding author; e-mail: jerome.sheahan@nuigalway.ie

tosses until $(1,1,1,1,1)$ occurs is 62, while for $(1,0,1,1,0)$ it is only 36. Even in the comparison of first occurrence of sequences of length $K=2$, one's intuition may prove unreliable. Indeed while each one of the sequences HH and HT has equal probability (1/2) of occurring before the other, the same is not the case for the sequences HH and TH. In fact TH has a 75% chance of occurring before HH. To see this, a little reflection shows that if the event HH is to occur before TH, then the first two flips must both result in H (because any occurrence of T in either of the first two flips means that with probability 1, TH will occur before HH.

Classroom experience suggests that this result is counter-intuitive to most students, perhaps because in any two flips, the occurrence of two heads is just as likely as the occurrence of a tail followed by a head. Similarly, the probability with which HHH precedes HHT can be seen to be 1/2, but the probability with which HHH precedes HTH is only 2/5. It was found that repetitions by students of either simulations (as described in [4]) or empirical studies of experiments to verify the above results, gave numbers that prove useful in clarifying the laws of large numbers and in verifying theoretical results relating to the magnitude of variation to be expected in sample results.

The above results may seem surprising since in n repetitions the expected number of occurrences is the same for any two sequences of equal length with the same number of ones. In general, if $\underline{x}_K$ is any sequence consisting of K_1 ones and K_2 zeros $(K_1+K_2=K\leq n)$, then the expected number of occurrences of $\underline{x}_K$ in $(X_1,\ldots,X_n)$ is:

$$(n-K+1)\theta^{K_1}(1-\theta)^{K_2}. \tag{1}$$

This result has the following simple derivation if one employs an indicator (characteristic) variables method and linearity of the expectation operator.

For $i=1,\,2,\ldots,n-K+1$, define

$$L_i=\begin{cases}1, & \text{if } X_i=x_1, X_{i+1}=x_2,\ldots,X_{i+K-1}=x_K\\ 0, & \text{otherwise.}\end{cases}$$

Note that $\underline{x}_K$ occurs at the ith block if and only if $L_i=1$. Then since the number of occurrences of $\underline{x}_K$ in $(X_1,\,\ldots,\,X_n)$ is evidently $\sum_{i=1}^{n-K+1}L_i$, the expected number of occurrences is $E(\sum_{i=1}^{n-K+1}L_i)=(n-K+1)P(X_j=x_1,\ X_{j+1}=x_2,\ldots,\ X_{j+K-1}=x_K)$ for any $j=1,2,\ldots,\ n-K+1$, and hence (1) holds.

If we specialise (1) to our coin tossing example and put $K=3$, we find that the expected number of occurrences of each of the outcomes HHT and HTH in n flips of the coin is $(n-2)\left(\frac{1}{2}\right)^3$. Why then, one may ask, is the event HTH more likely than HHT to occur before HHH? Note that of course the first occurrence of an event is quite a different concept from the number of occurrences of that event in a specified number of trials. This paper explains why certain sequences are more likely to precede other sequences and derives an explicit formula for the distribution of N along with a simple recursion for the distribution.

2. Factors that determine the probability of first occurrence

As in Section 1, let N be the number of the block at which a given sequence $\underline{x}_K = (x_1, x_2, \ldots, x_K)$ occurs for the first time in the sequence $X_1, X_2 \ldots$. For each $n = 1, 2, \ldots$, it is clear that N takes the value n if and only if block n is the given sequence and no previous block constitutes the sequence. Thus if we let A_i be the event that the ith block is the (given) sequence $\underline{x}_K$, we have

$$P(N = n) = P\left(\bigcup_{i=1}^{n} A_i\right) - P\left(\bigcup_{i=1}^{n-1} A_i\right). \tag{2}$$

Now in (2) we first employ the principle of inclusion–exclusion, then the obvious fact that there is a 'one-dependence' between the A_i (e.g. that the conditional probabilities $P(A_{i_1}|A_{i_2} \cap A_{i_3})$ and $P(A_{i_1}|A_{i_2})$ are equal for $i_1 > i_2 > i_3$), and finally the fact that $P(A_i)$ is independent of i (so that we may write $P(A_i) = P(A)$).

We then obtain

$$\begin{aligned}
P(N = n) &= \left\{\sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) + \sum_{i<j<l}^{n} P(A_i \cap A_j \cap A_l)\right. \\
&\quad \left. - \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)\right\} - \left\{\sum_{i=1}^{n-1} P(A_i) - \sum_{i<j}^{n-1} P(A_i \cap A_j)\right. \\
&\quad \left. + \sum_{i<j<l}^{n-1} P(A_i \cap A_j \cap A_l) - \ldots + (-1)^{n-2} P\left(\bigcap_{i=1}^{n-1} A_i\right)\right\} \\
&= P(A_n) - \sum_{i=1}^{n-1} P(A_i \cap A_n) + \sum_{i<j}^{n-1} P(A_i \cap A_j \cap A_n) - \ldots \\
&\quad + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right) \\
&= P(A_n) - \sum_{i=1}^{n-1} P(A_i) P(A_n|A_i) + \sum_{i<j}^{n-1} P(A_i) P(A_j|A_i) P(A_n|A_i \cap A_j) \\
&\quad - \ldots + (-1)^{n-1} P(A_1) P(A_2|A_1) \\
&\quad \times P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).
\end{aligned} \tag{3}$$

Thus

$$\begin{aligned}
P(N = n) &= P(A)\left\{1 - \sum_{i=1}^{n-1} P(A_n|A_i) + \sum_{i<j}^{n-1} P(A_j|A_i) P(A_n|A_j)\right. \\
&\quad - \sum_{i<j<l}^{n-1} P(A_j|A_i) P(A_l|A_j) P(A_n|A_l) \\
&\quad \left. + \ldots + (-1)^{n-1} P(A_2|A_1) P(A_3|A_2) \ldots P(A_n|A_{n-1})\right\}, \quad n = 1, 2, \ldots.
\end{aligned} \tag{4}$$

For clarity of exposition we now confine ourselves to the case where $\theta = 1/2$, so that $P(A) = \left(\frac{1}{2}\right)^K$. (We remark that the general case is technically no more difficult, but the notation is more cumbersome, for it is only when $\theta = \frac{1}{2}$ that θ and $1-\theta$ become equal.)

Now for any $i \geq 1$, consider $P(A_{i+1}|A_i)$, that is the conditional probability that block $i+1$ is the given sequence knowing that block i is the given sequence $\underline{x}_K$. Some reflection shows that

$$P(A_{i+1}|A_i) = \begin{cases} \frac{1}{2}, & \text{if all elements of } \underline{x}_K \text{ are the same} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Some further reflection reveals that

$$P(A_{i+2}|A_i) = \begin{cases} \left(\frac{1}{2}\right)^K, & \text{if } K \leq 2 \\ \left(\frac{1}{2}\right)^2, & \text{if } K \geq 3 \text{ and every second element of } \underline{x}_K \text{ is the same} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In writing 'every second element of $\underline{x}_K$ is the same', we mean that $x_i = x_{i+2}$ for all $i = 1, 2, \ldots, K-2$. It is important to note that such a sequence may also have other subsequences of elements the same. Therefore, while the sequence $(1, 0, 1, 0)$ has only every second element the same, the sequence $(1, 0, 1, 0, 1, 0)$ has every second and every fourth element the same. In fact if a sequence has every second element the same, such as $(1, 0, 1, 0, 1, 0, 1)$, it will also have every fourth element the same (when $K \geq 5$) and every sixth element the same (when $K \geq 7$), and so on. Similarly, a sequence that has every third element the same will have every sixth and ninth element the same, and so on. (It will not affect our calculations if a statement like 'every uth element is the same' is vacuously true. Thus the sequence $(1, 0, 1, 0)$, which has every second element the same, may also be said to have every fourth and fifth element the same, even though this sequence does not have a fifth element.) The distribution of N, and hence the probability with which any given sequence will occur before any other given sequence of the same length, depends on the length K of the sequence, and on a property we call *periodicity.* This property relates to whether or not, for each fixed i, $\underline{x}_K$ has the property that 'every ith element of $\underline{x}_K$ is the same'. In fact each sequence can be classified according to the following events, some of which are not pairwise disjoint: every element of the sequence is the same; every second but not every third element is the same; every third but not every second element is the same; and so on. More precisely, sequences are classified by their length K, the smallest even integer u and the smallest odd integer u for which every uth element is the same. It is such a classification that enables us to distinguish, for example, the sequences HHH, HHT and HTH discussed in Section 1. The fact that $P(N=n)$ depends on the above factors is clear from (3) above and (7) below.

In general, we obtain

$$P(A_{i+u}|A_i) = \begin{cases} \left(\frac{1}{2}\right)^K, \text{ if } K \leq u \\ \left(\frac{1}{2}\right)^u, \text{ if } K > u \text{ and every } u\text{th element of } \underline{x}_K \text{ is the same} \\ 0, \quad \text{otherwise.} \end{cases} \tag{7}$$

It is important to observe from (7) that $P(A_j|A_i) = P(A_{i+(j-i)}|A_i)$ depends on i and j only through the difference $u = j - i$. Thus we have

$$g_u := P(A_{i+u}|A_i) \text{ is independent of } i. \tag{8}$$

Using this fact, and recalling that $P(A) = \left(\frac{1}{2}\right)^K$, we find that (3) may be re-written as

$$P(N = n) = \left(\tfrac{1}{2}\right)^K \left\{ 1 - \sum_{i=1}^{n-1} g_{n-i} + \sum_{i<j}^{n-1} g_{j-i}g_{n-j} - \sum_{i<j<l}^{n-1} g_{j-i}g_{l-j}g_{n-l} + \ldots + (-1)^{n-1}g_1^{n-1} \right\}. \tag{9}$$

3. The distribution of first occurrence

We are now in a position to give the main theorem that provides an explicit expression for the distribution of N. We use the notation $\pi(m)$ for a partition $(1^{r_1}, 2^{r_2}, \ldots, m^{r_m})$ of a positive integer m; thus the non-negative integers r_i satisfy $\sum_{i=1}^m ir_i = m$. Also let $\sum_{\pi(m)}$ denote summation over all partitions of m. For non-negative integers $j_1, j_2, \ldots, j_m$ satisfying $\sum_{i=1}^m j_i = j$, write $\binom{j}{j_1, j_2, \ldots, j_m}$ for the corresponding multinomial coefficient $\frac{j!}{j_1!\, j_2! \ldots j_m!}$.

Introduce the generating function $G(x) = \sum_{u=1}^\infty g_u x^u$, a formal power series in x. Let $G_m(x) = \sum_{u=1}^m g_u x^u$, put $H(x) = [1 + G(x)]^{-1}$, and denote by $H_m(x) := \{[1 + G(x)]^{-1}\}_m = \{[1 + G_m(x)]^{-1}\}_m$ the terms in the expansion of H as a power series in x as far as the term x^m.

Theorem 3.1. *The density function of N, the number of the block at which any given vector $\underline{x}_K$ of ones and zeros occurs for the first time in a sequence of Bernoulli$\left(\frac{1}{2}\right)$ random variables, is given by each of the following*:

$$P(N = n) = \left(\tfrac{1}{2}\right)^K H_{n-1}(1), \quad n = 1, 2, \ldots \tag{10}$$

and

$$P(N = n) = \left(\tfrac{1}{2}\right)^K \sum_{j=0}^{n-1} (-1)^j \sum_{m=j}^{n-1} B_{m,j}(g_1, g_2, \ldots), \quad n = 1, 2, \ldots, \tag{11}$$

where $B_{m,j}(g_1, g_2, \ldots) = \sum_{\pi(m)} \binom{j}{j_1, j_2, \ldots, j_m} g_1^{j_1} g_2^{j_2} \ldots g_m^{j_m}, \sum_{i=1}^m j_i = j$, $\pi(m)$ *denotes the partition* $(1^{j_1}, 2^{j_2}, \ldots, m^{j_m})$ *of* m, *and* $g_u = P(A_{i+u}|A_i)$ *is given by* (7).

Proof. Observe first that by direct calculation (see also [5, pp. 189–91]) the coefficient of x^m in the power series expansion of $[G(x)]^j = [\sum_{u=1}^\infty g_u x^u]^j$ is routinely seen to be $B_{m,j}(g_1, g_2, \ldots)$. It follows that

$$\begin{aligned}\{[1+G(x)]^{-1}\}_{n-1} &= 1-\{G(x)\}_{n-1}+\{[G(x)]^2\}_{n-1}-\{[G(x)]^3\}_{n-1}+\ldots \\ &= 1-\sum_{m=1}^{n-1}B_{m,1}(g_1,g_2,\ldots)x^m+\sum_{m=1}^{n-1}B_{m,2}(g_1,g_2,\ldots)x^m \\ &\quad -\ldots+(-1)^{n-1}\sum_{m=1}^{n-1}B_{m,n-1}(g_1,g_2,\ldots)x^m. \qquad (12)\end{aligned}$$

Since $B_{m,j}(g_1, g_2, \ldots) = 0$ for each m and j satisfying $m < j$, we see from (12) and the definition of $H_{n-1}(x)$ that

$$\begin{aligned}\left(\tfrac{1}{2}\right)^K H_{n-1}(1) = \left(\tfrac{1}{2}\right)^K \Bigg[1 &- \sum_{m=1}^{n-1}B_{m,1}(g_1,g_2,\ldots)+\sum_{m=2}^{n-1}B_{m,2}(g_1,g_2,\ldots) \\ &-\ldots+(-1)^{n-1}B_{n-1,n-1}(g_1,g_2,\ldots)\Bigg]. \qquad (13)\end{aligned}$$

It follows that the right hand sides of (10) and (11) are equal, since we also have

$$\sum_{m=0}^{n-1}B_{m,0}(g_1,g_2,\ldots)=B_{0,0}(g_1,g_2,\ldots)=1.$$

From (9), (10) and (13) the proof is thus complete if we can show that

$$\begin{aligned}&1-\sum_{m=1}^{n-1}B_{m,1}(g_1,g_2,\ldots)+\sum_{m=2}^{n-1}B_{m,2}(g_1,g_2,\ldots)-\ldots+(-1)^{n-1}B_{n-1,n-1}(g_1,g_2,\ldots) \\ &=1-\sum_{i=1}^{n-1}g_{n-i}+\sum_{i<j}^{n-1}g_{j-i}g_{n-j}-\sum_{i<j<l}^{n-1}g_{j-i}g_{l-j}g_{n-l}+\ldots+(-1)^{n-1}g_1^{n-1}. \qquad (14)\end{aligned}$$

The sum that appears in the second term on the left side of (14) is

$$\begin{aligned}\sum_{m=1}^{n-1}B_{m,1}(g_1,g_2,\ldots) &= \sum_{m=1}^{n-1}\sum_{\pi(m)}\binom{1}{j_1,j_2,\ldots,j_m}g_1^{j_1}g_2^{j_2}\ldots g_m^{j_m} \\ &= \binom{1}{1}g_1^1+\binom{1}{0,1}g_1^0g_2^1+\binom{1}{0,0,1}g_1^0g_2^0g_3^1 \\ &\quad +\ldots+\binom{1}{0,0,\ldots,0,1}g_1^0g_2^0\ldots g_{n-2}^0g_{n-1}^1=\sum_{i=1}^{n-1}g_i,\end{aligned}$$

and this is the sum in the second term on the right side of (14).

Next, the sum appearing in the third term on the left side of (14) is

$$\begin{aligned}
\sum_{m=2}^{n-1} B_{m,2}(g_1, g_2, \dots) &= \sum_{\pi(2)} \binom{2}{j_1, j_2} g_1^{j_1} g_2^{j_2} \\
&\quad + \dots + \sum_{\pi(n-1)} \binom{2}{j_1, j_2, \dots, j_{n-1}} g_1^{j_1} g_2^{j_2} \dots g_{n-1}^{j_{n-1}} \\
&= \binom{2}{2,0} g_1^2 g_2^0 + \binom{2}{1,1,0} g_1^1 g_2^1 g_3^0 \\
&\quad + \left[\binom{2}{1,0,1,0} g_1^1 g_2^0 g_3^1 g_4^0 + \binom{2}{0,2,0,0} g_1^0 g_2^2 g_3^0 g_4^0\right] \\
&\quad + \left[\binom{2}{1,0,0,1,0} g_1^1 g_2^0 g_3^0 g_4^1 g_5^0 + \binom{2}{0,1,1,0,0} g_1^0 g_2^1 g_3^1 g_4^0 g_5^0\right] \\
&\quad + \left[\binom{2}{1,0,0,0,1,0} g_1^1 g_2^0 g_3^0 g_4^0 g_5^1 g_6^0 \right.\\
&\quad + \binom{2}{0,1,0,1,0,0} g_1^0 g_2^1 g_3^0 g_4^1 g_5^0 g_6^0 \\
&\quad \left. + \binom{2}{0,0,2,0,0,0} g_1^0 g_2^0 g_3^2 g_4^0 g_5^0 g_6^0 \right] + \dots \\
&\quad + \left[\binom{2}{1,0,\dots,0,1,0} g_1^1 g_2^0 \dots g_{n-2}^1 g_{n-1}^0 + \dots \right] \\
&= g_1^2 + 2g_1g_2 + [2g_1g_3 + g_2^2] + [2g_1g_4 + 2g_2g_3] \\
&\quad + [2g_1g_5 + 2g_2g_4 + g_3^2] + \dots + 2g_1g_{n-2}. \qquad (15)
\end{aligned}$$

On the other hand, the sum appearing in the third term on the right side of (14) is

$$\begin{aligned}
& g_1 \sum_{i=1}^{n-2} g_{n-i-1} + g_2 \sum_{i=1}^{n-3} g_{n-i-2} + \dots + g_{n-2} g_1 \\
&= g_1 \sum_{i=1}^{n-2} g_i + g_2 \sum_{i=1}^{n-3} g_i + \dots + g_{n-2} g_1 = \sum_{i=1}^{n-2} \sum_{j=1}^{n-i-1} g_i g_j,
\end{aligned}$$

which is (15).

We continue now using induction. (This method may seem rather prosaic, but we feel it is instructive in providing insight into the equivalence of (9), (10) and (11). In addition, some of the algebra we exhibit will be used again in our proof of Theorem 3.2 below.)

Suppose that

$$\sum_{m=w-1}^{n-1} B_{m,w-1}(g_1, g_2, \dots) = \sum_{i_1 < i_2 < \dots < i_{w-1}}^{n-1} g_{i_2 - i_1} g_{i_3 - i_2} \dots g_{i_{w-1} - i_{w-2}} g_{n - i_{w-1}}. \qquad (16)$$

We will show that

$$\sum_{m=w}^{n-1} B_{m,w}(g_1, g_2, \ldots) = \sum_{i_1<i_2<\ldots<i_w}^{n-1} g_{i_2-i_1} g_{i_3-i_2} \cdots g_{i_w-i_{w-1}} g_{n-i_w}. \tag{17}$$

From [5, p. 190], the $B_{m,r}(g_1, g_2, \ldots)$ satisfy the recurrence relation

$$\begin{aligned} B_{m,r+1}(g_1, g_2, \ldots) &= g_1 B_{m-1,r}(g_1, g_2, \ldots) + g_2 B_{m-2,r}(g_1, g_2, \ldots) \\ &\quad + \ldots + g_{m-r} B_{r,r}(g_1, g_2, \ldots). \end{aligned} \tag{18}$$

It follows from (18) and an interchange of the order of summations (and replacing $m-r$ by t) that the left hand side of (17) can be expressed as

$$\begin{aligned} \sum_{m=w}^{n-1} B_{m,w}(g_1, g_2, \ldots) &= \sum_{r=1}^{n-w} g_r \sum_{m=w+r-1}^{n-1} B_{m-r,w-1}(g_1, g_2, \ldots) \\ &= \sum_{r=1}^{n-w} g_r \sum_{t=w-1}^{n-r-1} B_{t,w-1}(g_1, g_2, \ldots). \end{aligned} \tag{19}$$

Inserting (16) into (19), with n in (16) replaced by $n-r$, we see that

$$\sum_{m=w}^{n-1} B_{m,w}(g_1, g_2, \ldots) = \sum_{r=1}^{n-w} g_r \sum_{i_1<i_2<\ldots<i_{w-1}}^{n-r-1} g_{i_2-i_1} g_{i_3-i_2} \cdots g_{i_{w-1}-i_{w-2}} g_{n-r-i_{w-1}}.$$

With $i_w = n-r$, this agrees with the right hand side of (17), and the proof is complete. ■

Theorem 3.2. *The density function of N, the number of the block at which any given vector $\underline{x}_K$ of ones and zeros occurs for the first time in a sequence of Bernoulli$\left(\frac{1}{2}\right)$ random variables satisfies the recursion*

$$P(N=n) = \left(\tfrac{1}{2}\right)^K - \sum_{j=1}^{n-1} g_{n-j} P(N=j), \quad n = 1, 2, \ldots. \tag{20}$$

Proof. Write (11) as $P(N=n) = \left(\frac{1}{2}\right)^K \{1 + \sum_{w=1}^{n-1} (-1)^w \sum_{m=w}^{n-1} B_{m,w}(g_1, g_2, \ldots)\}$. Applying (19) gives

$$P(N=n) = \left(\tfrac{1}{2}\right)^K \left\{1 + \sum_{w=1}^{n-1} (-1)^w \sum_{r=1}^{n-w} g_r \sum_{t=w-1}^{n-r-1} B_{t,w-1}(g_1, g_2, \ldots)\right\}. \tag{21}$$

If we interchange the first two summations that appear in (21) we find that

$$P(N=n) = \left(\tfrac{1}{2}\right)^K \left\{1 + \sum_{r=1}^{n-1} g_r \sum_{w=1}^{n-r} (-1)^w \sum_{t=w-1}^{n-r-1} B_{t,w-1}(g_1, g_2, \ldots)\right\}.$$

After an application of (11), again with j and n replaced by $w-1$ and $n-r$, respectively, we see that

$$P(N=n)=\left(\tfrac{1}{2}\right)^K-\sum_{r=1}^{n-1}g_rP(N=n-r),\quad n=1,2,\ldots.$$

Since this is the same as (20), the proof is complete. ■

4. Examples

We now present some examples to illustrate the theorems. These examples are designed primarily to show the techniques involved in obtaining more explicit expressions when the values of the g_u are specified. While Section 3 provides general formulae and a simple recursion, the computations required in practice may be carried out using expressions that have inserted into the general formulae the specific values for the g_u appropriate for the pattern being studied.

Example 4.1. (The aperiodic case). Consider any vector $\underline{x}_K$ for which there does not exist any u ($u=1,2,\ldots,K-1$) where the statement 'every uth element of $\underline{x}_K$ is the same' holds. Examples of such vectors when $K=5$ are $(1,1,1,0,0)$ and $(1,1,0,0,0)$. We call such a vector *aperiodic* and from (7) and (8) it is clear that for such vectors we have $g_u=\left(\tfrac{1}{2}\right)^K$, for $u\geq K$ and $g_u=0$ for $1\leq u<K$.

For illustrative purposes, we shall derive $P(N=n)$ for such aperiodic sequences by using each of (10) and (11) in turn.

Using (10) first, note that for aperiodic sequences we have

$$G(x)=\left(\tfrac{1}{2}\right)^K\sum_{u=K}^{\infty}x^u=\left(\tfrac{1}{2}\right)^Kx^K(1-x)^{-1}$$

and hence

$$H(x)=[1+G(x)]^{-1}=1+\textstyle\sum_{j=1}^{\infty}(-1)^j\left[\left(\tfrac{1}{2}\right)^Kx^K(1-x)^{-1}\right]^j.$$

To obtain $H_{n-1}(x)$, we require all terms up to x^{n-1} in $H(x)$.

Observe that, for $j\geq 1$,

$$\begin{aligned}\left\{\left(\tfrac{1}{2}\right)^{Kj}x^{Kj}(1-x)^{-j}\right\}_{n-1}&=\left\{\left(\tfrac{1}{2}\right)^{Kj}x^{Kj}\sum_{t=j}^{\infty}\binom{t-1}{j-1}x^{t-j}\right\}_{n-1}\\&=\left(\tfrac{1}{2}\right)^{Kj}x^{Kj}\sum_{t=j}^{n-1+j-Kj}\binom{t-1}{j-1}x^{t-j}.\end{aligned}$$

Hence

$$H_{n-1}(1)=1+\sum_{j=1}^{n-1}(-1)^j\left(\tfrac{1}{2}\right)^{Kj}\sum_{t=j}^{n-1+j-Kj}\binom{t-1}{j-1}.$$

Since for each integer $j\geq 1$, $\sum_{t=j}^{n-1+j-Kj}\binom{t-1}{j-1}=\binom{n-1-j(K-1)}{j}$ by the upper summation identity in [2, p. 174], it follows that

$$H_{n-1}(1) = 1 + \sum_{j=1}^{n-1}(-1)^j \left(\tfrac{1}{2}\right)^{Kj} \binom{n-1-j(K-1)}{j}.$$

Putting this into (10) gives

$$P(N = n \mid \underline{x}_K \text{ is aperiodic}) = \left(\tfrac{1}{2}\right)^K \sum_{j=0}^{n-1}(-1)^j \left(\tfrac{1}{2}\right)^{Kj} \binom{n-1-j(K-1)}{j}, \quad n = 1, 2, \ldots. \tag{22}$$

To derive this distribution using (11) instead, note that for aperiodic sequences we have

$$\begin{aligned} B_{m,j}(g_1, g_2, \ldots) &= B_{m,j}(\underbrace{0, 0, \ldots 0}_{K \text{ terms}}, \underbrace{\left(\tfrac{1}{2}\right)^K, \left(\tfrac{1}{2}\right)^K, \ldots, \left(\tfrac{1}{2}\right)^K}_{m-K \text{ terms}}) \\ &= \sum_{\pi(m)} \binom{j}{j_1, j_2, \ldots, j_m} 0^{j_1} 0^{j_2} \ldots 0^{j_{K-1}} \left(\tfrac{1}{2}\right)^{Kj_K} \left(\tfrac{1}{2}\right)^{Kj_{K+1}} \cdots \left(\tfrac{1}{2}\right)^{Kj_m}. \end{aligned}$$

Notice that unless all $j_1, j_2, \ldots, j_{K-1}$ are 0, we will have $0^{j_1} 0^{j_2} \ldots 0^{j_{K-1}} = 0$. Hence

$$B_{m,j}(\underbrace{0, 0, \ldots 0}_{K \text{ terms}}, \underbrace{\left(\tfrac{1}{2}\right)^K, \left(\tfrac{1}{2}\right)^K, \ldots, \left(\tfrac{1}{2}\right)^K}_{m-K \text{ terms}}) = \left(\tfrac{1}{2}\right)^{Kj} \sum_{\pi^*(m)} \binom{j}{j_1, j_2, \ldots, j_m},$$

where $\sum_{\pi^*(m)}$ denotes summation over all partitions $(1^{j_1}, 2^{j_2}, \ldots, m^{j_m})$ of m satisfying $j_1 + j_2 + \ \ldots + j_m = j$. Clearly $\sum_{\pi^*(m)} \binom{j}{j_1, j_2, \ldots, j_m}$ is given by the coefficient of x^m in the expansion of $(x^K + x^{K+1} + \ \ldots)^j$. Proceeding as in the derivation of $H_{n-1}(x)$ above, this coefficient is $\binom{m-1-j(K-1)}{j-1}$.
Consequently,

$$\begin{aligned} P(N = n \mid \underline{x}_K \text{ is aperiodic}) &= \left(\tfrac{1}{2}\right)^K \sum_{j=0}^{n-1}(-1)^j \left(\tfrac{1}{2}\right)^{Kj} \sum_{m=j}^{n-1} \binom{m-1-j(K-1)}{j-1} \\ &= \left(\tfrac{1}{2}\right)^K \sum_{j=0}^{n-1}(-1)^j \left(\tfrac{1}{2}\right)^{Kj} \sum_{u=j-1-j(K-1)}^{n-2-j(K-1)} \binom{u}{j-1} \\ &= \left(\tfrac{1}{2}\right)^K \sum_{j=0}^{n-1}(-1)^j \left(\tfrac{1}{2}\right)^{Kj} \sum_{u=j-1}^{n-2-j(K-1)} \binom{u}{j-1}, \end{aligned}$$

and this agrees with (22) after again applying the 'summation on the upper index' identity.

For further illustration of (22), we look at the special cases $K = 1$ and $K = 2$.

For $K = 1$, $\underline{x}_K$ is either (0) or (1). From (22), the distribution of the first occurrence of either of these is $P(\ N = n \mid \underline{x}_K = (1) \textit{ or } (0)) = \frac{1}{2}\sum_{l=0}^{n-1}\left(-\frac{1}{2}\right)^l \binom{n-1}{l} = \left(\frac{1}{2}\right)^n$, $n = 1, 2, \ldots$. This result is obvious since N has the geometric distribution with parameter $\theta = 1/2$.

For $K=2$, the possible aperiodic vectors are $(1,0)$ and $(0,1)$. Then (22) yields

$$P(N=n \mid \underline{x}_2=(1,0) \text{ or } (0,1)) = \left(\tfrac{1}{2}\right)^2 \sum_{l=0}^{n-1} (-1)^l \left(\tfrac{1}{2}\right)^{2l} \binom{n-1-l}{l} = n\left(\tfrac{1}{2}\right)^{n+1}, \quad n=1,2,\ldots.$$

The last equality here follows from the fact that $P_n := P(N=n \mid \underline{x}_2=(1,\ 0) \text{ or } (0,1))$ satisfies the recursion $P_n = P_{n-1} - \frac{1}{4}P_{n-2}$, and subject to the initial conditions $P_1 = P_2 = \frac{1}{4}$, this has solution $n\left(\frac{1}{2}\right)^{n+1}$, $n \geq 1$. (This recursive formula is easily derived directly, but it also follows from the general recursion (20).)

We remark that this distribution also follows from the following simple argument (which we illustrate for the sequence $(0,1)$). Of the 2^{n+1} possible vectors that can occur up to the occurrence of block n, the n that are of the form $\underbrace{1,1,\ldots,1}_{i \text{ term}},\underbrace{0,0,\ldots 0}_{n-i \text{ terms}},\ 1$ $(i=0,1,\ldots,n-1)$ lead to $(0,1)$ occurring for the first time at block n.

Example 4.2. (Sequences with period one). For the sequences $\underline{1}_K := (\underbrace{1,1,\ldots,1}_{K \text{ terms}})$ and $\underline{0}_K = (\underbrace{0,0,\ldots,0}_{K \text{ terms}})$, we have from (7),

$$g_u = \begin{cases} \left(\tfrac{1}{2}\right)^K, & u \geq K \\ \left(\tfrac{1}{2}\right)^u, & u < K. \end{cases}$$

To obtain a simplified formula for the distribution of N in this case, it seems easier to apply Theorem 3.2 rather than either of the expressions in Theorem 3.1. By (20) and the present g_u, we have

$$P(N=n) = \begin{cases} \left(\tfrac{1}{2}\right)^K - \displaystyle\sum_{j=1}^{n-1} \left(\tfrac{1}{2}\right)^{n-j} P(N=j), & n \leq K \\ \left(\tfrac{1}{2}\right)^K - \left(\tfrac{1}{2}\right)^K \displaystyle\sum_{j=1}^{n-K} P(N=j) - \sum_{j=n-K+1}^{n-1} \left(\tfrac{1}{2}\right)^{n-j} P(N=j), & n > K. \end{cases} \tag{23}$$

For $K=1$, we see also from (23) that $P(N=n) = \left(\frac{1}{2}\right)^1 - \frac{1}{2}\sum_{j=1}^{n-1} P(N=j)$, so that

$$P(N=n) = \left(\tfrac{1}{2}\right)^n, \quad n \geq 1, K=1. \tag{24}$$

Next, the portion of (23) that refers to the case $n \leq K$ is easily seen to have the solution given by the next expression below. Since both expressions in (23) are equal for $n=K+1$, this solution applies for $n=K+1$ also. It is

$$P(N=n) = \begin{cases} \left(\tfrac{1}{2}\right)^K, & n=1, K \geq 1 \\ \left(\tfrac{1}{2}\right)^{K+1}, & 2 \leq n \leq K+1, K \geq 1. \end{cases} \tag{25}$$

For $n \geq K+2$ and $K \geq 2$, non-trivial manipulation of (23) leads to

$$P(N=n)=\sum_{j=1}^{K}\left(\tfrac{1}{2}\right)^{j}P(N=n-j),$$

from which we obtain the more convenient form

$$P(N=n)-P(N=n-1)=-\left(\tfrac{1}{2}\right)^{K+1}P(N=n-K-1),\quad n\geq K+2,\ K\geq 2. \tag{26}$$

The formulae (24) and (25) with the recursion (26) can be viewed as major simplifications of the recursion (20) in the case of our present sequences $\underline{1}_K$ and $\underline{0}_K$, but we can go further and obtain an explicit solution. To solve (26), it is convenient to partition the set of positive integers as

$$\{1,2,\ldots\}=\{1,2,\ldots,K+1\}\cup\left[\bigcup_{m=1}^{\infty}\{mK+2,\ mK+3,\ldots,(m+1)K+1\}\right].$$

The solution for $P(N=n)$ over the first set $\{1,2,\ldots,K+1\}$ of this partition has been given in (25) above. To obtain $P(N=n)$ over each of the remaining sets, we unfold the recurrence (26) so that for n in any given set of the partition, $P(N=n)$ is expressed in a suitable form in terms of the $P(N=j)$ with j belonging to the previous set. One convenient representation is

$$\begin{aligned}P(N=n)&=P(N=mK+1)-\left(\tfrac{1}{2}\right)^{K+1}\sum_{j=0}^{n-mK-2}P(N=j+[m-1]K+1),\\ &\qquad mK+2\leq n\leq[m+1]K+1,\quad K\geq 2, m\geq 1.\end{aligned} \tag{27}$$

Now in (27) we first put $m=1$, and allow n to range over $\{K+2,K+3,\ldots,2K+1\}$. Then employing (25) we see that

$$\begin{aligned}P(N=n)&=P(N=K+1)-\left(\tfrac{1}{2}\right)^{K+1}\left\{P(N=1)+\sum_{j=1}^{n-K-2}P(N=j+1)\right\}\\ &=\left(\tfrac{1}{2}\right)^{K+1}-\left(\tfrac{1}{2}\right)^{K+1}\left\{\left(\tfrac{1}{2}\right)^{K}+(n-K-2)\left(\tfrac{1}{2}\right)^{K+1}\right\}.\end{aligned}$$

That is,

$$P(N=n)=\left(\tfrac{1}{2}\right)^{K+1}-\left(\tfrac{1}{2}\right)^{2K+2}(n-K),\quad K+2\leq n\leq 2K+1,\ K\geq 2. \tag{28}$$

Next, in (27) we put $m=2$, allow n to range over $\{2K+2,2K+3,\ldots,3K+1\}$ and use (28) and (25) to find that

$$\begin{aligned}P(N=n)&=P(N=2K+1)-\left(\tfrac{1}{2}\right)^{K+1}\left\{P(N=K+1)+\sum_{j=1}^{n-2K-2}P(N=j+K+1)\right\}\\ &=\left(\tfrac{1}{2}\right)^{K+1}-(K+1)\left(\tfrac{1}{2}\right)^{2K+2}-\left(\tfrac{1}{2}\right)^{K+1}\left\{\left(\tfrac{1}{2}\right)^{K+1}+\sum_{j=1}^{n-2K-2}\left[\left(\tfrac{1}{2}\right)^{K+1}-(j+1)\left(\tfrac{1}{2}\right)^{2K+2}\right]\right\}.\end{aligned}$$

Simplifying gives

$$P(N=n)=\left(\tfrac{1}{2}\right)^{K+1}-\left(\tfrac{1}{2}\right)^{2K+2}(n-K)+\left(\tfrac{1}{2}\right)^{3K+3}\sum_{j=1}^{n-2K-2}(j+1),$$

$$2K+2\le n\le 3K+1,\ K\ge 2. \tag{29}$$

In general, using similar but tedious calculations, we find that

$$\begin{aligned}P(N=n)&=\left(\tfrac{1}{2}\right)^{K+1}-\left(\tfrac{1}{2}\right)^{2(K+1)}(n-K)+\left(\tfrac{1}{2}\right)^{3(K+1)}\sum_{j=1}^{n-2K-2}(j+1)\\&\quad-\left(\tfrac{1}{2}\right)^{4(K+1)}\sum_{j=1}^{n-3K-2}\sum_{i=1}^{j-1}(i+1)+\left(\tfrac{1}{2}\right)^{5(K+1)}\sum_{K=1}^{n-4K-2}\sum_{j=1}^{K-1}\sum_{i=1}^{j-1}(i+1)\\&\quad+\ldots+(-1)^m\left(\tfrac{1}{2}\right)^{(m+1)(K+1)}\sum_{i_{m-1}=1}^{n-mK-2}\sum_{i_{m-2}=1}^{i_{m-1}-1}\sum_{i_{m-3}=1}^{i_{m-2}-1}\ldots\sum_{i_1=1}^{i_2-1}(i_1+1),\end{aligned}$$

$$mK+2\le n\le(m+1)K+1,\ m\ge 1, K\ge 2. \tag{30}$$

Using the identity $\sum_{i=1}^{a}i(i+1)(i+2)\ldots(i+K-1)=\frac{1}{K+1}a(a+1)(a+2)\ldots(a+K)$, repeatedly, it is easy to show that (30) reduces to

$$\begin{aligned}P(N=n)&=\left(\tfrac{1}{2}\right)^{K+1}-\left(\tfrac{1}{2}\right)^{2(K+1)}(n-K)+\left(\tfrac{1}{2}\right)^{3(K+1)}\left(\tfrac{1}{2}\right)(n-2K-2)(n-2K+1)\\&\quad-\left(\tfrac{1}{2}\right)^{4(K+1)}\left(\tfrac{1}{3!}\right)(n-3K-3)(n-3K-2)(n-3K+2)\\&\quad+\left(\tfrac{1}{2}\right)^{5(K+1)}\left(\tfrac{1}{4!}\right)(n-4K-4)(n-4K-3)(n-4K-2)(n-4K+3)\\&\quad-\left(\tfrac{1}{2}\right)^{6(K+1)}\left(\tfrac{1}{5!}\right)(n-5K-5)(n-5K-4)(n-5K-3)(n-5K-2)(n-5K+4)\\&\quad\vdots\\&\quad+(-1)^m\left(\tfrac{1}{2}\right)^{(m+1)(K+1)}\left(\frac{1}{m!}\right)\left\{\prod_{i=0}^{m-2}(n-m[K+1]+i)\right\}\\&\quad\times(n-mK+m-1), mK+2\le n\le(m+1)K+1,\ m\ge 1, K\ge 2.\end{aligned} \tag{31}$$

Combining (24), (25) and (31) we have finally

$$P(N = n \,|\, \underline{x}_K = \underline{1}_K \text{ or } \underline{0}_K) =$$

$$= \begin{cases} \left(\frac{1}{2}\right)^n \text{ for } n \geq 1, \;\; K = 1, \\ \left(\frac{1}{2}\right)^K \text{ for } n = 1, \;\; K \geq 1, \\ \left(\frac{1}{2}\right)^{K+1} \text{ for } 2 \leq n \leq K+1, \;\; K \geq 1, \\ \left(\frac{1}{2}\right)^{K+1} - \left(\frac{1}{2}\right)^{2(K+1)}(n-K) \text{ for } K+2 \leq n \leq 2K+1, \;\; K \geq 1, \\ \left(\frac{1}{2}\right)^{K+1} + \sum_{j=2}^{m}(-1)^j \left(\frac{1}{2}\right)^{(j+1)(K+1)} \frac{1}{j!}\left\{\prod_{i=0}^{j-2}(n - j[K+1] + i)\right\}(n - j[K-1] - 1) \\ \qquad \text{for } mK + 2 \leq n \leq (m+1)K + 1, \; m \geq 2, \; K \geq 2. \end{cases} \tag{32}$$

The derivation of formula (32) suggests that some complexity exists even in the apparently simple case of the two patterns all of whose elements are the same (i.e. $\underline{1}_K$ and $\underline{0}_K$). Nevertheless, we note parenthetically that the form of (32) permits a clear explanation of the logic we actually employed in the derivation of formulae (3) and (9), which provided the basis for the general Theorems 3.1 and 3.2. To clarify the reasoning employed, suppose that our pattern of interest is $(1, 1, 1, 1, 1)$ so that $K = 5$. We will justify the appearance of the term $\left(\frac{1}{2}\right)^{K+1} - \left(\frac{1}{2}\right)^{2(K+1)}(n-K)$ in (32) for this case when $n = 7$. Now if the pattern $(1, 1, 1, 1, 1)$ is to appear for the first time at block 7, then the first 11 variables $X_1, X_2, \ldots, X_{11}$ must satisfy the following restrictions: we must have $X_7 = X_8 = X_9 = X_{10} = X_{11} = 1$ (to ensure that the pattern occurs at block 7); simultaneously we must have $X_6 = 0$ (to ensure that the pattern has not occurred at block 6); and at the same time we must ensure that the vector $(X_1, X_2, X_3, X_4, X_5)$ does not have value $(1, 1, 1, 1, 1)$ (for otherwise the pattern will have appeared at the first block). The probability that the first 11 random variables satisfy these restrictions is clearly $\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^1\left(1 - \left(\frac{1}{2}\right)^5\right)$, and this is precisely $\left(\frac{1}{2}\right)^{K+1} - \left(\frac{1}{2}\right)^{2(K+1)}(n - K)$ with $K = 5$ and $n = 7$.

5. Comments

We described the factors that determine the distribution of the first occurrence of any given pattern of ones and zeros in a sequence of independent Bernoulli variables. Mathematically, these factors are expressed in terms of the g_u of (8) and (7). Two expressions for the distribution of the first occurrence of the pattern are given in Theorem 3.1, and a recurrence relation for the distribution is provided in Theorem 3.2. The choice of which of the three expressions to use for any particular sequence is to some extent a function of whether or not a computer is being used to perform the calculations. The distribution we derived provides an alternative approach to discussing the distribution's moments, the mean and variance of which are discussed elsewhere in the literature without knowledge of the distribution (see [6]). Using the results of the present work, one can write an expression for the probability with which one sequence will occur before another. However, we will not discuss this here. Nor will we address the question of the asymptotic distribution of N, though a study of it is obviously suggested by the complexity of the special case of our general distribution, which we discussed in Example 4.2. We close by reiterating that the condition $\theta = 1/2$ was imposed throughout the paper only for convenience of notation. In fact, the results of the paper can be extended to more general distributions than Bernoulli, though the presentation is more tedious.

References

[1] A. Engel, The probability abacus. *Educational Studies in Mathematics* **6** (1975), 1–22.

[2] R.L. Graham, D.E. Knuth and O. Patashnik, *Concrete mathematics*. Addison-Wesley, New York, 1989.

[3] V.C. Hombas, Waiting time and expected waiting time—paradoxical situations. *The American Statistician* **51**(2) (1997), 130–33.

[4] C. Konold, Confessions of a coin flipper and would-be instructor, *The American Statistician* **49**(2) (1995), 203–9.

[5] J. Riordan, *An introduction to combinatorial analysis*, Wiley, New York, 1958.

[6] S.M. Ross, *Introduction to probability models*, Academic Press, Orlando, 1997.