Count of Runs of Length Two in Binary Sequences

Gadre M.P.
Department of Statistics,
University of Pune, Pune, INDIA.
e-mail: mpg28@rediffmail.com

Rattihalli R.N.
Department of Statistics,
Shivaji University, Kolhapur, INDIA.
e-mail:rnr5@rediffmail.com

Abstract

In this note we derive some results related to the number of counts of different types of runs of length two in binary sequences. These results are useful to develop attribute control charts based on a single attribute under Markovian dependence.

1. Introduction

Consider a sequence of 0's and 1's of length (n+1). This can be considered as a realization on a 'Markov Chain' (MC) with state space {0, 1}, and a one step 'transition probability matrix' (tpm) **P**. A subsequence of length two indicates a typical transition. The counts N_{ij}, the number of transitions from state i to j, (i, j = 0, 1) play an important role in inference related to an MC setup. There are $2^{(n+1)}$ possible sequences of length (n+1). Let $\underline{N} = \begin{pmatrix} N_{00} & N_{01} \\ N_{10} & N_{11} \end{pmatrix}$ denote the transition count matrix' (tcm). For details one may refer to Prakasa Rao and Basawa (1980) and Whittle's formula given in Iosifescu (1980, pp. 178).

In this note, we find the number of sequences yielding a specified tcm and the number of distinct tcms. These are very useful in developing attribute control charts under Markovian dependence.

Remainder of the paper is organized as follows. Section-2 contains the notations and a simple illustration. The next section consists of the main result and an application of the results is given in Section-4.

2. Notations

Let $\{X_0, X_1, \ldots, X_n\}$ be a sequence of 0's and 1's and $\underline{X} = (X_0, X_1, \ldots, X_n)$. We know that there are $2^{(n+1)}$ such sequences. Let Ω be the set of such sequences. Define a function $\underline{N} : \Omega \to \Re^4$ defined by

$$\underline{N} = \begin{pmatrix} N_{00} & N_{01} \\ N_{10} & N_{11} \end{pmatrix},$$

where, $N_{ij} = \sum_{r=0}^{n-1} k(x_r, x_{r+1})$ and

$$k(x_r, x_{r+1}) = \begin{cases} 1, \text{ if } (x_r, x_{r+1}) = (i, j) \\ 0, \text{ if } (x_r, x_{r+1}) \neq (i, j). \end{cases}$$

In other words, N_{ij} is the number of runs of length two for which (x_r, x_{r+1}) = (i, j) for r = 0, 1, …, (n-1).

For example, let n = 7 and the sequence be 01011000. This sequence starts, ends with 0 and yields the tcm **N** = $\begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}$. In this realization there are 5 (= n_{00} + n_{01} + 1) zeros and 3 (= $n_{01} + n_{11}$) 1's.

We note the following and use it to prove the result. The above sequence can be viewed as a specific realization of inserting five 0's into 3 cells and further inserting three 1's into 2 cells with no cells empty in either case.

We note that the number of ways in which 5 0's can be inserted into 3 cells with no cells empty is $\binom{4}{2}$. As such the number of sequences yielding the tcm $\begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}$ are $\binom{4}{2}\binom{2}{1}$ = 12. These sequences are 01011000, 01001100, 00101100, 00100110, 01000110, 00010110, 01101000, 01100100, 00110100, 00110010, 01100010 and 00011010.

In general we have the following.

3. Main Result

Theorem: Let n be fixed.

A. The number of distinct sequences x in Ω yielding a specific N are as follows.

a. The number of sequences with $x_0 = 0$, yielding the matrix $\begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix}$ is

i) 1 if $n_{00} = n$,

ii) $\binom{n_{00}+n_{01}}{n_{01}}\binom{n_{01}+n_{11}-1}{n_{01}-1}$ if $n_{01} > 0$ and the sequence ends with 0

iii) $\binom{n_{00}+n_{01}-1}{n_{01}-1}\binom{n_{01}+n_{11}-1}{n_{01}-1}$ if $n_{01} > 0$ and the sequence ends with 1.

b. The number of sequences with $x_0 = 1$, yielding the matrix $\begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix}$ is

iv) 1 if $n_{11} = n$,

v) $\binom{n_{11}+n_{10}}{n_{10}}\binom{n_{10}+n_{00}-1}{n_{10}-1}$ if $n_{10} > 0$ and the sequence ends with 1,

vi) $\binom{n_{00}+n_{01}-1}{n_{01}-1}\binom{n_{01}+n_{11}-1}{n_{01}-1}$ if $n_{10} > 0$ and the sequence ends with 0.

B. The number of distinct values of N are $\frac{n(n+1)}{2}+1$.

Proof: **A**. Below we consider two cases, namely $x_0 = 0$ and $x_0 = 1$.

Case-a: Let X_0 be 0.

i) The proof is trivial.

ii) In this case, $n_{01} = n_{10}$. This situation is similar to that of inserting $(n_{00} + n_{01} + 1)$ balls (zeros) into $(n_{01} + 1)$ cells with no cell being empty and it can be done in $\binom{n_{00} + n_{01}}{n_{01}}$ ways.

Further the total number of 1's is $(n_{01} + n_{11})$. These have to be put in n_{01} cells, with no cell empty, and it can be done in $\binom{n_{01} + n_{11} - 1}{n_{01} - 1}$ ways.

Thus the total number of ways is. $\binom{n_{00} + n_{01}}{n_{01}}\binom{n_{01} + n_{11} - 1}{n_{01} - 1}$.

iii) Consider a specific choice of n_{00} and n_{11}. In this case there are $(n_{00} + n_{01})$ zeros. We have to put these 0's in n_{01} cells with no cell empty. This can be done in $\binom{n_{00} + n_{01} - 1}{n_{01} - 1}$ ways.

Further the total no. of 1's is $(n_{01} + n_{11})$. These have to be put in n_{01} cells with no cell empty. This can be done in $\binom{n_{01} + n_{11} - 1}{n_{01} - 1}$ ways.

Thus the total number of ways is $\binom{n_{00} + n_{01} - 1}{n_{01} - 1}\binom{n_{01} + n_{11} - 1}{n_{01} - 1}$.

<u>Case-b</u>: Proofs are as in Case-a above.

B. <u>Case-1</u>: Let $x_0 = 0$. In this case note that,

(a) If $(N_{01} + N_{10})$ is even then $N_{01} = N_{10}$.

(b) If $(N_{01} + N_{10})$ is odd then $N_{01} = (N_{10} + 1)$.

That is in either case one of N_{01}, N_{10} determines the other.

I. If $N_{01} = 0$ then $N_{00} = n$.

II. Let $N_{01} > 0$. Define $g = \{n-(N_{01} + N_{10})\}$. As $g < n$ and since N_{00} and N_{11} can take any value from {0, 1, …, (n-1)} and as there is only one choice to the values of N_{01} and N_{10}; for given (n, g), in this case, the number of distinct values of $\underline{N}$ are $\sum_{g=0}^{n-1}(g+1) = \frac{n(n+1)}{2}$.

Hence for given (n, g) the total number of distinct values of $\underline{N}$ are $\frac{n(n+1)}{2}+1$.

<u>Case-2</u>: Let x_0 be 1. Proof is similar to case 1, just by interchanging the role of N_{00} and N_{11} as well that of N_{01} and N_{10}.

4. An Application

To compute optimal values of the design parameters of the 'Single Attribute Control Charts for Markovian Dependent Process' we need to compute the probability distribution of the chart statistic $T = \sum_{i=0}^{1}\sum_{j=0}^{1} a_{ij} N_{ij}$, which is a function of $\underline{N}$. Here a_{ij} are some specific real constants. The distribution of T can be obtained by considering all 2^n sequences of 0's 1's (if x_0 is assumed to be zero). However by considering the number of distinct tcms and the number of sequences yielding the same tcm the computations reduce very significantly.

References

1. B. L. S. Prakasa Rao and I. V. Basawa (1980): *Statistical Inference for Stochastic Processes,* Academic Press, London.

2. M. Iosifescu (1980): Finite Markov processes and Their Applications, John Wiley & Sons, Ltd.