Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Computational Statistics & Data Analysis 50 (2006) 604 – 610

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

# A recursive algorithm for computing the distribution of the number of successes in higher-order Markovian trials

Donald E.K. Martin*

*Department of Mathematics, Howard University, USA Statistical Research Division, US Census Bureau, Washington, DC 20233, USA*

Received 17 November 2003; accepted 18 September 2004
Available online 26 October 2004

**Abstract**

This paper presents a recursive method of computing the distribution of the number of successes in a sequence of binary trials that are Markovian of a general order. Waiting-time distributions are also obtained. Recurrence relations among probabilities of partitioned events are used to compute the desired probabilities. An advantage of computing the probabilities in the manner given is that the algorithm may be easily programmed on a computer, requiring no computer algebra system, as in computation based on conditional probability generating functions. The difference between calculated probabilities for various model orders emphasizes the importance of selecting a proper model order for a Markovian data set.


*Keywords:* Binary trials; Markovian sequences; Probability

## 1. Introduction

For a sequence of $n$ binary trials, the distribution of the number of successes $S_n$ and the distribution of the waiting time to a specified number of successes are important in many practical applications. In the case of independent trials, distributional theory for $S_n$ and

* Corresponding author. 8052 Quill Point Drive, Bowie, MD 20720, USA. Tel.: +1 301 763 3689; fax: +1 301 457 2299.

*E-mail address:* donald.e.martin@census.gov (D.E.K. Martin).

0167-9473/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2004.09.005

related quantities are well known, however when dependence is present, the distribution of $S_n$ is no longer binomial, and computing distributions of related quantities is not as straightforward.

A sequence $\{Y_t,\ t=0, \pm 1, \pm 2, \ldots\}$ having a finite or denumerable state space is called *Markovian of order m* if

$$P(Y_t{=}y_t \mid Y_{t-1}{=}y_{t-1}, Y_{t-2}{=}y_{t-2}, \ldots, Y_{t-m}{=}y_{t-m}, \ldots, Y_1{=}y_1, Y_0{=}y_0, \ldots) \\ = P(Y_t = y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \ldots, Y_{t-m} = y_{t-m}),$$

that is the outcome of each trial depends on the outcome of the $m$ directly preceding trials, and depends only on these. Markov dependence has been used to model data in such diverse fields as work force planning, production, finance, accounting, education, marketing, and health services, typically with the assumption that $m = 1$ (see, for example, Isaacson and Madsen, 1976). However, at times, higher-order dependence is required. For example, though several authors have fit Markov chain models to data on the presence or absence of rain and/or the length of dry and wet spells, Chin (1977) noted that the proper Markov order of the daily precipitation process depends on the season of the year and also on geographic location. Indeed, in one of the examples in Chin's paper, a third-order model was required to describe daily precipitation occurrence in winter.

Gabriel (1959) gave formulas for computing the distribution of $S_n$ for a Markov chain. Ladd (1975) gave a recursive algorithm for computing this distribution in the case of a stationary Markov chain. In Kedem (1980), a recursive algorithm was given to compute the distribution of $S_n$ for stationary Markovian sequences of order two, and Martin (2000) extended Kedem's results to stationary fourth-order Markovian trials. Uchida (1998) used conditional probability generating functions to obtain the distribution of $S_n$ for Markovian sequences of a general order $m$. Recurrence relations among the conditional probability generating functions of the desired random variable were developed by considering one-step ahead from every condition. The resulting linear system of equations was solved by a computer algebra system for the conditional probability generating functions. The desired probabilities were then obtained by expanding the unconditional probability generating functions.

In this paper, we give a simpler method for computing the distribution of the number of successes in finite Markovian sequences of a general order. We use recurrence relations among probabilities of partitioned events to compute probabilities in a straightforward manner. The recurrence relations among these probabilities yield equations that facilitate easy programming on a computer, without the use of a computer algebra system. The algorithm is a generalization of one given in Martin (2000), both because stationarity is not required, and because the model order is not limited. The unconditional distribution of the waiting time to the $k$th success, for different values of $k$, is obtained as a by-product.

In Section 2, the steps of the algorithm are given. Section 3 contains an example of computed probabilities. Section 4 is a summary.

## 2. The algorithm

Let $X_{-m+1}, X_{-m+2}, \ldots, X_0, X_1, X_2, \ldots$ denote a $\{0, 1\}$-valued (failure-success) $m$th order Markovian sequence with time-invariant transition probabilities. Denote the initial