Journal of Statistical Planning and Inference 143 (2013) 1233–1243

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Distributions associated with (k_1,k_2) events on semi-Markov binary trials

Valeri T. Stefanov [a,*], Raimondo Manca [b,1]

[a] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[b] Department of Methods and Models for Economics, Finance and Territory, University of Rome "La Sapienza", 00161 Rome, Italy

ARTICLE INFO

Article history:
Received 20 December 2011
Received in revised form
12 September 2012
Accepted 10 December 2012
Available online 25 December 2012

Keywords:
Binary trial
Generating function
Semi-Markov process
Stopping time

ABSTRACT

We derive neat expressions for the probability generating functions of relevant waiting times associated with (k_1,k_2) events on semi-Markov binary trials. These lead to evaluation of relevant probabilities associated with numbers of occurrence of such events on a string of a fixed length. Our methodology is general enough and provides a template for treating more general events than those of type (k_1,k_2). Also, the same template is extendable to semi-Markov trials with more than two outcomes.

1. Introduction

Each sequence of binary trials on {0,1} can be viewed as a sequence of alternating runs of 0's and 1's (0 and 1 are conventionally called 'failure' and 'success', respectively). The model of binary trials considered in this paper is a semi-Markov one. More specifically, the binary trials are generated by a two-state semi-Markov process, X_n say, with a discrete-time parameter, whose embedded discrete-time Markov chain, $\hat{X}_k$ say, has the following one-step transition probability matrix:

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

In the context of {0,1} binary trials state 1 marks an occurrence of a run of 0's and state 2 marks an occurrence of a run of 1's. The holding time distributions at the two states are identified by random variables, ν_1 and ν_2 say, and their corresponding probability generating functions (pgf's) are denoted by $h_1(s) = E(s^{\nu_1})$ and $h_2(s) = E(s^{\nu_2})$, respectively. In other words, the lengths of the runs of 0's are generated by independent copies of ν_1 and the lengths of the runs of 1's—by independent copies of ν_2. The initial probabilities are $\pi_1 = P(\hat{X}_0 = 1)$ and $\pi_2 = P(\hat{X}_0 = 2)$.

Note that this semi-Markov binary model embraces as special cases the Bernoulli model, the Markov-dependent model and the general two-state semi-Markov model.

* Corresponding author. Tel.: +61 83881712.
E-mail addresses: valery.stefanov@yahoo.com (V.T. Stefanov), raimondo.manca@uniroma1.it (R. Manca).
[1] Tel.: +39 064 9766507; fax: +39 064 9766765.

0378-3758/$ - see front matter
http://dx.doi.org/10.1016/j.jspi.2012.12.003

Example 1 (*Bernoulli model*). Assume the binary sequences are generated by Bernoulli trials with 'success' and 'failure' probabilities p and q, respectively. Of course, such trials are embraced by the aforementioned model as follows. Identify 'failure' with state 1 and 'success' with state 2 and assume v_1 is geometrically distributed on $\{1,2,\ldots\}$ with parameter p, that is $h_1(s)=ps/(1-qs)$, and v_2 is geometrically distributed with parameter q, that is $h_2(s)=qs/(1-ps)$.

Example 2 (*Markov-dependent model*). Assume the binary trials are generated by a two-state Markov chain with transition probabilities $p_{i,j}$, $i,j=1,2$. Similarly to the preceding example, the time spent in a state before entering the other state is geometrically distributed. Take $h_1(s)=p_{1,2}s/(1-p_{1,1}s)$ and $h_2(s)=p_{2,1}s/(1-p_{2,2}s)$ and identify state 1 with 'failure' and state 2 with 'success' to see that this model is a special case of the aforementioned semi-Markov model.

Example 3 (*general semi-Markov model*). Consider now a two-state semi-Markov process with discrete-time parameter, whose one-step transition probabilities for its embedded discrete-time Markov chain are $p_{i,j}$, $i,j=1,2$. Denote the pgf of the holding time in state i, given the next state to be visited is j, by $w_{i,j}(s)$. Recall that the holding times are conditionally independent given a trajectory of the embedded discrete-time Markov chain (Asmussen, 2003, p. 206). Therefore, the time spent in a state, i say, before entering the other state, j ($j\neq i$) say, is a random sum (the summation index is geometrically distributed with parameter $p_{i,j}$) of independent random variables, $Z_1,Z_2,\ldots$, say, where Z_1 has a pgf given by $w_{i,j}(s)$ and the remaining Z_m are identically distributed with a pgf given by $w_{i,i}(s)$. From well-known properties of random sums and their pgf's one gets that the pgf of this random sum is equal to $p_{i,j}w_{i,j}(s)/(1-p_{i,i}w_{i,i}(s))$. Therefore, this general semi-Markov model is embraced by the aforementioned semi-Markov one by taking $h_1(s)=p_{1,2}w_{1,2}(s)/(1-p_{1,1}w_{1,1}(s))$ and $h_2(s)=p_{2,1}w_{2,1}(s)/(1-p_{2,2}w_{2,2}(s))$.

Similarly to Dafnis et al. (2010), for positive integers k_1 and k_2 define the events E_1, E_2, E_3 as follows:

- E_1: at least k_1 consecutive 0's are followed by at least k_2 consecutive 1's.
- E_2: exactly k_1 consecutive 0's are followed by exactly k_2 consecutive 1's.
- E_3: at most k_1 (and at least 1) consecutive 0's are followed by at most k_2 (and at least 1) consecutive 1's.

The events E_i have been of interest both in distribution theory (see Balakrishnan and Koutras, 2002) and in applications (Dafnis et al., 2010). The Bernoulli case has been covered in the literature satisfactorily with neat explicit results for the relevant exact distributions (cf. Dafnis et al., 2010; Huang and Tsai, 1991). On the other hand, from the point of view of applications dependent binary trials constitute a more realistic model. Note that for some dependent binary trials there are some exact results, as well as some approximations, on distributions associated with the events E_i (see Makri, 2010; Sen et al., 2006; Vellaisamy, 2004).

The paper is organised as follows. The main results are stated in Section 2. Their proofs are found in Section 4. The special cases of Bernoulli trials and Markov dependent trials are covered in Section 3. Section 5 provides relevant comments about the applicability of our methodology to more general events and semi-Markov trials with more than two outcomes. The last section contains a few relevant remarks.

2. Main results

Recall from Section 1 that the holding times for the two states of our model are identified by the random variables v_1 and v_2 whose pgf's were denoted by $h_1(s)$ and $h_2(s)$, respectively. For $j=1,2$, denote by $v_j^{(<k_j)}$ a random variable whose distribution is equal to the conditional distribution of v_j given $(v_j<k_j)$, and by $v_j^{(\geq k_1)}$ a random variable whose distribution is equal to the conditional distribution of v_j given $(v_j\geq k_j)$. Likewise, one introduces the random variables $v_j^{(=k_j)}$, $v_j^{(\neq k_j)}$, $v_j^{(\leq k_j)}$, and $v_j^{(>k_j)}$. Furthermore, for $j=1,2$, denote by $h_j^{(<)}(s)$ the pgf of $v_j^{(<k_j)}$ and likewise denote by $h_j^{(\cdot)}(s)$ the pgf of $v_j^{(\cdot\ k_j)}$ where $(\cdot)$ stands for either $(\geq)$, or $(=)$, or $(\leq)$, etc. For $j=1,2$, denote

$$r_j^{(<)}=\sum_{i=1}^{k_j-1}P(v_j=i),\quad r_j^{(\geq)}=1-r_j^{(<)},\quad r_j^{(=)}=P(v_j=k_j),$$

$$r_j^{(\neq)}=1-r_j^{(=)},\quad r_j^{(\leq)}=\sum_{i=1}^{k_j}P(v_j=i),\quad r_j^{(>)}=1-r_j^{(\leq)},$$

$$g_j^{(<)}(s)=\sum_{i=1}^{k_j-1}s^iP(v_j=i),\quad g_j^{(\geq)}(s)=h_j(s)-g_j^{(<)}(s),$$

$$g_j^{(\leq)}(s)=\sum_{i=1}^{k_j}s^iP(v_j=i),\quad g_j^{(>)}(s)=h_j(s)-g_j^{(\leq)}(s),$$

$$g_j^{(=)}(s)=s^{k_j}P(v_j=k_j),\quad g_j^{(\neq)}(s)=h_j(s)-g_j^{(=)}(s).$$