Statistics and Probability Letters 119 (2016) 1–10

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Counts of Bernoulli success strings in a multivariate framework

Djilali Ait Aoudia [a], Éric Marchand [b,*], François Perron [c]

[a] École des HEC, Département de sciences de la décision, Montréal, Qc, Canada
[b] Université de Sherbrooke, Département de mathématiques, Sherbrooke, Qc, Canada
[c] Université de Montréal, Département de mathématiques et de statistique, Montréal, Qc, Canada

ARTICLE INFO

Article history:
Received 5 November 2015
Received in revised form 8 July 2016
Accepted 8 July 2016
Available online 16 July 2016

MSC:
60C05
60E05
62E15

Keywords:
Arrays
Binomial moments
Dirichlet
Poisson distribution
Poisson mixtures
Runs

ABSTRACT

Distributional results are obtained for runs of length 2 in Bernoulli arrays $X_{k,j}$ with multinomial distributed rows. These include a multivariate Poisson mixture representation with Dirichlet mixing for the joint distribution of the number of runs in each column.

© 2016 Published by Elsevier B.V.

1. Introduction

In many studies of data, one is concerned with the occurrence or non-occurrence of a special event over time. In the counting of species, one wishes to know whether a new catch is from a new species; in the study of records, one wishes to assess the likelihood of a new record or not; in classification, one wishes to know whether a current observation is from a different population than past observations. Distributional properties relative to such Bernoulli sequences of 0's and 1's, as well as to runs of consecutive 1's, have generated much interest over the years.

Multivariate versions are also of interest, namely for situations with r dependent sequences and where a motivating question is to describe the nature of the dependence. Bernoulli arrays were introduced by Ait Aoudia and Marchand (2010) and consist of entries $\{X_{k,j}, k \geqslant 1, j = 1, \ldots, r+1\}$, where the rows $\underline{X}_k = (X_{k,1}, \ldots, X_{k,r+1})$ are independently distributed as Multinomial$(1, p_{k,1}, \ldots, p_{k,r+1})$, $k \geqslant 1$. Such arrays arise in sampling one object at a time from an urn with $r+1$ colors, with replacement and reinforcement (i.e., adding balls as in a Polyà urn). Quantities of interest include the number of runs $S_{n,j} = \sum_{k=1}^{n} X_{k,j}X_{k+1,j}$ of length 2 in the jth column, their limits $S_j = \lim_{n\to\infty} S_{n,j}$, and the random vectors

$$\underline{S}_n = (S_{n,1}, \ldots, S_{n,r}), \quad \text{and} \quad \underline{S} = \lim_{n\to\infty} \underline{S}_n. \tag{1.1}$$

* Corresponding author.
E-mail address: eric.marchand@usherbrooke.ca (É. Marchand).

http://dx.doi.org/10.1016/j.spl.2016.07.009
0167-7152/© 2016 Published by Elsevier B.V.

For the univariate case ($r = 1$), an elegant result consists of the Poisson mixture representation

$$S_1|U \sim Poisson(aU) \text{ with } U \sim Beta(a, b), \quad \text{for } p_{k,1} = \frac{a}{a+b+k-1},\ a > 0,\ b \geqslant 0, \tag{1.2}$$

as obtained by Mori (2001), as well as Holst (2008). This class of solutions contains degenerate cases with $b = 0$, i.e., $\mathbb{P}(U = 1) = 1$, and where S_1 is Poisson distributed with mean parameter a. Moreover, the $a = 1, b = 0$ $Poisson(1)$ solution was recognized in the mid 1990s by Persi Diaconis, as well as Hahlin (1995), and earlier versions are due to Arratia et al. (1992), Kolchin (1971), and Goncharov (1944). This elegant result has inspired much interest and lead to various findings relative to the distributions of $S_{n,j}$ and S_j for various other configurations of the $\{p_{k,j}\}$'s, as well as relationships and implications for Pólya urns, records, matching problems, marked Poisson processes, etc., as witnessed by the work of Chern et al. (2000), Csörgó and Wu (2000), Holst (2007, 2008), Huffer and Sethuraman (2012), Huffer et al. (2009), Joffe et al. (2004, 2000), Mori (2001) and Sethuraman and Sethuraman (2004), among others.

With such Poisson distributions and Poisson mixtures arising naturally in these univariate ($r = 1$) situations, it seems natural to investigate multivariate versions of such models. The findings of Ait Aoudia et al. (2014) are contributions in this direction with analysis for the totals $\sum_{j=1}^{r} S_{n,j}$ and $T = \sum_{j=1}^{r} S_j$, and representations for the distributions of $\underline{S}_n$ and $\underline{S}$ in the bivariate case ($r = 2$). For instance, they obtain the following multivariate extension of (1.2):

$$T|U \sim Poisson\left(\frac{aU}{r}\right) \text{ with } U \sim Beta(a, b), \quad \text{for } p_{k,l} = \frac{a}{r(a+b+k-1)},\ a > 0,\ b \geqslant 0, \tag{1.3}$$

which yields $T \sim \text{Poisson}(\lambda)$ for the degenerate case $a = r\lambda$, $b = 0$.

The main findings here are presented in Section 3, are based on the assumption

$$p_{k,1} = \cdots = p_{k,r} = \frac{1}{b+k}, \quad \text{with } b \geqslant r-1, \tag{1.4}$$

and include: (i) explicit expressions (Theorems 3.3 and 3.4) for the mixed binomial moments of $S_{n,1}, S_{n,2}, \ldots, S_{n,r}$; for all n, r; and of $S_1, S_2, \ldots, S_r$, (ii) an explicit expression for the distribution function of $\underline{S}$ (Theorem 3.4), and (iii) two mixture representations (Theorem 3.5) for the distribution of $\underline{S}$.

As recalled in Section 2, mixed binomial moments yield expressions for probability mass and generating functions. The most intricate and challenging technical achievement is the binomial moment result of Theorem 3.3, which solves a recursive system of equations involving binomial moments related to the $S_{n,j}$'s and auxiliary variables $W_{n,j}$, defined below in Section 3.1. Solutions to such a system of equations were provided by Ait Aoudia et al. (2014) in the bivariate case $r = 2$, but our result is unified and general for $r \geqslant 2$. The mixture representations are quite interesting. In the first mixture representation, the components S_i of $\underline{S}$ are distributed as independent Poisson(V_i) conditional on $V_1, \ldots, V_r$ which follows a Dirichlet distribution. This neatly summarizes the dependence structure of the S_j's, as well as replicating the marginal Beta mixture representation (1.2). The second mixture representation is quite different as the mixing parameter is univariate Beta and the dependence is reflected otherwise through a multivariate discrete distribution with Poisson and non-independent marginals, which connects for instance to the Schur-constant distributions of Castañer et al. (2015).

2. Multivariate Poisson mixtures and mixed binomial moments

For non-negative integer valued random variables $Z_1, Z_2, \ldots, Z_r$ and constants $k_i \in \mathbb{N}$, $i = 1, \ldots, r$, the mixed binomial moments are defined as $\mathbb{E}\left(\prod_{i=1}^{r} \binom{Z_i}{k_i}\right)$, where $\binom{z}{u}$ is taken to be equal to 0 for $u > z$. Our distributional findings involve the determination of mixed binomial moments, which imply associated expressions for the probability generating and mass functions. Indeed, the Taylor series expansion about 1 of the probability generating function ψ_{Z_1} having radius of convergence greater than 1 can be expressed as

$$\psi_{Z_1}(t) = \mathbb{E}\left[t^{Z_1}\right] = \sum_{k \geqslant 0} \mathbb{E}\binom{Z_1}{k}(t-1)^k, \quad t \in [0, 1], \tag{2.1}$$

and the probability mass function (pmf) of Z can be written as

$$\mathbb{P}(Z_1 = i) = \sum_{k=i}^{\infty} (-1)^{k-i} \binom{k}{i} \mathbb{E}\binom{Z_1}{k}. \tag{2.2}$$

For the multivariate case, corresponding expressions are given by

$$\mathbb{E}\left[\prod_{i=1}^{r} t_i^{Z_i}\right] = \sum_{k_1 \geqslant 0, \ldots, k_r \geqslant 0} \mathbb{E}\left(\prod_{i=1}^{r} \binom{Z_i}{k_i}\right) \prod_{j=1}^{r} (t_j - 1)^{k_i}, \quad (t_1, t_2, \ldots, t_r) \in [0, 1]^r, \tag{2.3}$$

Download English Version:

https://daneshyari.com/en/article/1151453

Download Persian Version:

https://daneshyari.com/article/1151453

Daneshyari.com