Statistical Methodology 25 (2015) 36–50

Contents lists available at ScienceDirect

Statistical Methodology

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

On some distributions arising from a generalized trivariate reduction scheme

Christophe Chesneau [a], Maher Kachour [b], Dimitris Karlis [c,*]

[a] Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie, Campus II, Science 3, 14032 Caen, France

[b] Ecole supérieure de commerce IDRAC, 47, rue Sergent Michel Berthet CP 607, 69258 Lyon Cedex 09, France

[c] Department of Statistics, Athens University of Economics and Business, 76 Patission str, 10434, Athens, Greece

ARTICLE INFO

Article history:
Received 13 January 2014
Received in revised form 5 January 2015
Accepted 15 January 2015
Available online 22 January 2015

MSC:
62M10
62M20

Keywords:
Skellam distribution
Trivariate reduction
Rademacher distribution

ABSTRACT

In this article we construct bivariate discrete distributions in $\mathbb{Z}^2$. We make use of a generalized trivariate reduction technique. The special case leading to a generalization of a bivariate Skellam distribution is studied in detail. Properties of the derived models as well as estimation are examined. Real data application is provided. Discussion of extensions to different models is also mentioned.

1. Introduction

Discrete valued data have found a huge number of real applications. Noticeably, it is more common to work with positive counts rather than integers that can take values in $\mathbb{Z}$. In recent days there has been renewed interest for discrete valued models defined in $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, i.e. in both the positive and negative integers. Such data occur naturally in several circumstances in diverse scientific fields.

* Corresponding author.
E-mail addresses: chesneau@math.unicaen.fr (C. Chesneau), maher.kachour@idraclyon.com (M. Kachour), karlis@aueb.gr (D. Karlis).

http://dx.doi.org/10.1016/j.stamet.2015.01.001
1572-3127/

Examples of such data refer to financial applications where price changes for assets are given in certain ranges, known as ticks. Hence, when the number of ticks for an asset during trading needs to be modeled, this can go upwards or downward a specific number of ticks and thus it takes values in $\mathbb{Z}$. In sports applications and in particular in soccer, the interest lies on modeling the score as the difference of the goals scored by each team. Such a bivariate distribution can be of special interest for betting purposes since for example composite bets are used like the result in the first half and at the end of the game. In biostatistics interest lies on modeling outcomes before and after a treatment is applied, given that the outcome is a discrete random variable like the number of coughs or the number of epileptic seizures. Then the before–after outcome is defined on $\mathbb{Z}$. Furthermore in image analysis, the intensity in each pixel can take discrete values so the difference of the intensities between adjacent pixels takes values in $\mathbb{Z}$. Finally in discrete time series analysis in order to achieve stationarity we may need to take first order differences, leading again to data in $\mathbb{Z}$. In all the above examples clearly one needs to develop appropriate models in order to make inference.

Recently there are papers that deal with such data defined in $\mathbb{Z}$ but in almost all of them the interest lies on the univariate case. The aim of the present paper is to define and propose models in the bivariate (multivariate) case and, hence, to define interesting bivariate (multivariate) distributions in $\mathbb{Z}^d$, $d \geq 2$.

To work towards this direction we will make use of a generalized trivariate (multivariate) reduction technique. Trivariate reduction has been widely used to define bivariate models (see, e.g. [25]). The central idea is to start by independent random variables and mix them using particular functional forms so as to end up with marginal distribution with specific properties and some correlation structure. Of course this is not the only way to define bivariate (multivariate) models but trivariate reduction can produce easy to interpret models with useful and flexible properties. In particular in this paper we extend the idea so as to be able to define more flexible models.

In particular we make use of Rademacher distribution, which is a natural extension of the Bernoulli distribution from $\{0, 1\}$ to $\{-1, 1\}$. This allows to naturally define models in $\mathbb{Z}$ and hence create flexible models based on standard approaches. The models proposed here can have interesting interpretation as mixtures of simpler ones, taking very flexible shapes and hence being useful for real data modeling. We investigate this potential with a particular model in Section 4.

The remaining of the paper is organized as follows. In Section 2 we present some useful results that we will use later on. We also review existing univariate models and provide the reader the necessary information to follow our derivations. Section 3 presents our findings. New models are developed and their properties are examined. In particular we investigate estimation for a particular model and comparison between different estimates. In Section 4 one can find a real data application on soccer data. Finally discussions on possible extensions can be found in Section 5.

2. Useful results

In this section we briefly review existing results and provide the necessary definitions so as to derive in the next section our main findings.

The literature on discrete distributions defined on $\mathbb{Z}$ is increasing. Such models are derived mainly via two avenues: the first one is by considering discretized versions of existing continuous distributions, as for example the discrete normal [12,24] or some other continuous distributions [7,15]. The second avenue derives the distributions as the difference of two positive discrete random variables. An important member of this class, both from historical perspective but also for its application potential, is the so called Skellam distribution introduced by [26]. This distribution has found some applications recently, see [9,10,2]. Note that some other distributions can be also derived as the difference of two discrete variables, see for example the papers of [23,27]. We will focus mainly to the second idea as differentiating can be related to the trivariate reduction technique we plan to apply.

2.1. Skellam distribution

In this subsection we briefly review the Skellam distribution. Let us consider two variables X and Y in $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$ and their difference $Z = X - Y$. The probability function of the difference Z is a discrete distribution defined on the set of integer numbers $\mathbb{Z}$.