



**HAL**  
open science

## On some distributions arising from a generalized trivariate reduction scheme

Christophe Chesneau, Maher Kachour, Dimitris Karlis

► **To cite this version:**

Christophe Chesneau, Maher Kachour, Dimitris Karlis. On some distributions arising from a generalized trivariate reduction scheme. 2013. hal-00861100

**HAL Id: hal-00861100**

**<https://hal.science/hal-00861100>**

Preprint submitted on 11 Sep 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## On some distributions arising from a generalized trivariate reduction scheme

Christophe Chesneau · Maher Kachour ·  
Dimitris Karlis

Received:

**Abstract** In this article we construct bivariate discrete distribution in  $\mathbb{Z}^2$ . We make use of a generalized trivariate reduction technique. The special case leading to bivariate Skellam distributions 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.

**Keywords** Skellam distribution; trivariate reduction; Rademacher distribution.

**2000 Mathematics Subject Classification** 62M10, 62M20.

### 1 Introduction

Working with discrete valued data has 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  $\mathbb{Z}$ , i.e. in both the positive and negative integers. Such data occur naturally in several circumstances in diverse range of different scientific field.

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

---

Christophe Chesneau

Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie, Campus II, Science 3, 14032 Caen, France. E-mail: chesneau@math.unicaen.fr

Maher Kachour

Ecole supérieure de commerce IDRAC, 47, rue Sergent Michel Berthet CP 607, 69258 Lyon Cedex 09, France. E-mail: maher.kachour@idraclyon.com

Dimitris Karlis

Department of Statistics, Athens University of Economics and Business, 76 Patission str, 10434, Athens, Greece. E-mail: karlis@aueb.gr

of ticks for an asset during trading is need 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 modelling the score as the difference of the goals scored by each team. This 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 modelling 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  $\mathbb{Z}^2$  and above.

To work to this direction we will make use of a generalized trivariate (multivariate) reduction technique. Trivariate reduction have been widely used to define bivariate models (see, e.g. Sarabia and Gomez-Deniz (2008)). 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 modelling. We investigate this potential with a particular model in section 4.

The remaining of the paper is 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 limited. 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 Kemp (1997); Roy (2003) or some other continuous distributions, (see, e.g. Inusah and Kozubowski (2006); Kozubowski and Inusah (2006)). 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 Skellam (1946). This distribution has found some applications recently, (see Karlis and Ntzoufras 2006, 2009; Alzaid and Omair 2010). Note that some other distributions can be also derived as the difference of two discrete variables are given in Ong et al. (2008) and Szablowski (2001). 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  $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$  and their difference  $Z = X - Y$ . The probability function of the difference  $Z$  is a discrete distribution defined on the set of integer numbers  $\mathbb{Z}$ .

If  $X$  and  $Y$  follow independently Poisson distributions with parameters  $\lambda_1 \geq 0$  and  $\lambda_2 \geq 0$  respectively (note that a Poisson variate with  $\lambda = 0$  implies a degenerate at 0 distribution) then the random variable  $Z = X - Y$  has probability function given by

$$f_{PD}(z|\lambda_1, \lambda_2) = P(Z = z|\lambda_1, \lambda_2) = e^{-(\lambda_1 + \lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{z/2} I_{|z|}\left(2\sqrt{\lambda_1\lambda_2}\right)$$

for all  $z \in \mathbb{Z}$ ,  $\lambda_1, \lambda_2 \geq 0$  where  $I_r(x)$  is the modified Bessel function of order  $r$  (see Abramowitz and Stegun 1974, pp. 375) defined by

$$I_r(x) = \left(\frac{x}{2}\right)^r \sum_{k=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^k}{k!\Gamma(r+k+1)}.$$

The mean and the variance are  $E(Z) = \lambda_1 - \lambda_2$  and  $Var(Z) = \lambda_1 + \lambda_2$ , respectively. This can be used in order to reparameterize the distribution with a mean variance parametrization. We will denote the distribution as  $Sk(\lambda_1, \lambda_2)$

Clearly,  $Var(Z) \geq |E(Z)|$ . The skewness is determined by the sign of  $\lambda_1 - \lambda_2$ . The distribution is symmetric only when  $\lambda_1 = \lambda_2$ . For large values of the  $\lambda_1 + \lambda_2$  the distribution can be sufficiently approximated by the normal

distribution. If  $\lambda_2$  is 0, then the distribution tends to a Poisson distribution. If the parameter  $\lambda_1$  is 0, then the distribution is the negative of a Poisson distribution. The Skellam distribution is unimodal.

Let us consider two random variables say  $Z_1 \sim Sk(\lambda_1, \lambda_2)$  and  $Z_2 \sim Sk(\lambda_3, \lambda_4)$ . Then the sum  $S_2 = Z_1 + Z_2$  follows a  $Sk(\lambda_1 + \lambda_3, \lambda_2 + \lambda_4)$  distribution, while the difference  $D_2 = Z_1 - Z_2$  follows a Skellam distribution with parameters  $\lambda_1 + \lambda_4$  and  $\lambda_2 + \lambda_3$ .

Note that Skellam is not necessarily the distribution of the difference of two uncorrelated Poisson random variables (see, Karlis and Ntzoufras (2006)). This implies that we can derive the Skellam distribution as the difference of other distributions as well which motivates its usage in various cases.

Skellam distribution also arises in applying wavelets to discrete data. Haar wavelet and filterbank transform coefficients of Poisson data are distributed as sums and differences of Poisson counts the sums being again Poisson, and the differences taking the Skellam distribution (Hirakawa and Wolfe (2012)).

To our knowledge extension to more dimensions are limited. Bulla et al. (2012a) used a trivariate reduction approach, while Karlis and Ntzoufras (2012) defined a bivariate distribution using copulas. A bivariate Skellam is also used in a time series model in Bulla et al. (2012b). Our models derived in the next section generalize some of the models in a certain extend.

## 2.2 Rademacher distribution

We will make use of the so called Rademacher distribution. This is a generalization of simple Bernoulli distribution by allowing the random variable to take values in  $\{-1, 1\}$  instead of  $\{0, 1\}$ . Indeed, the Rademacher distribution is a recoding of the Bernoulli distribution, where 1 still indicates success, but failure is coded as  $-1$ . Namely we will denote as  $R \sim \mathcal{R}(\alpha)$  if the for the random variable  $R$  holds that  $P(R = 1) = \alpha = 1 - P(R = -1)$ . In other words, if  $B$  is a Bernoulli( $\alpha$ ) random variable, then  $R = 2B - 1$  is a Rademacher( $\alpha$ ) random variable. Such variables will be important for our derivations since they allow to define random variable in  $\mathbb{Z}$ . Note that Rademacher distribution was often used in the literature under the assumption  $\alpha = \frac{1}{2}$ . Especially, the Rademacher( $\frac{1}{2}$ ) is an essential tool for the symmetrization technique used in empirical process and statistical learning theory, see ,e.g., Koltchinskii (2001) and Massart (2003). The Rademacher( $\frac{1}{2}$ ) distribution has been also used in bootstrapping, see, e.g., Liu (1988). Moreover, the simple random walk on  $\mathbb{Z}$  can be defined by using the Rademacher( $\frac{1}{2}$ ). Explicitly, if  $R_i$  are i.i.d. Rademacher( $\frac{1}{2}$ ) random variables, the sequence  $S_1, S_2, \dots$  is a random walk, where  $S_n = X_1 + \dots + X_n$ . In this paper we make use of a more general case by allowing  $\alpha \in (0, 1)$ . It can be easily checked that  $E(R) = 2\alpha - 1$  and  $Var(R) = 4\alpha(1 - \alpha)$ . A useful result for the sequel is the following:

Let  $R \sim \mathcal{R}(\alpha)$  and  $Z \sim Sk(\lambda_1, \lambda_2)$  then the random variable  $Z^* = RZ$  is a mixture of two Skellam random variables, of the form  $\alpha Sk(\lambda_1, \lambda_2) + (1 - \alpha)Sk(\lambda_2, \lambda_1)$ .

### 2.3 Trivariate Reduction

This *trivariate reduction* method is a popular and old technique used for constructing dependent variables. It has been used for both continuous and discrete cases. The method consists of building a pair of dependent random variables starting from three (or more) independent (usually) random variables. The functions that connect initial variables are generally elementary functions, or are given by the structure of the variables that we want to generate. A general approach is the following: Consider random variables  $X_1, X_2$  and  $X_3$ . Then we may define a new pair of variables using  $Y_1 = g_1(X_1, X_3)$  and  $Y_2 = g_2(X_2, X_3)$  where  $g_i(\cdot)$ ,  $i = 1, 2$  are some functions. The central idea is that since  $Y$ 's share the common  $X_3$  they are correlated, the correlation structure is determined by the function(s)  $g$ . Typical choices are for example  $g(X, Y) = X + Y$  which is used to derive the bivariate Poisson distribution if the  $X$ 's follow Poisson distributions, and  $g(X, Y) = \min(X, Y)$  for bivariate exponential distributions if the  $X$ 's follow exponential distributions. The method can be very flexible, while the choice of the functions  $g$  determine the correlation properties and perhaps put restrictions on them. For example the usage of  $g_1 = g_2 = X + Y$  leads to necessarily positive correlation. Note that such reduction schemes can have other interesting interpretations as discussed in Lai (1995).

In a recent article Bulla et al. (2012a) constructed bivariate Skellam distributions by a simple trivariate reduction scheme with additive functions. Here we expand this by considering a more general scheme involving also Rademacher random variables which makes the functions  $g$  to be more general and hence to allow for more flexible correlation structure.

## 3 The proposed models

In this section we define our new bivariate models based on a generalized trivariate reduction technique.

### 3.1 Definition

**Definition 1 ( $\mathbb{Z}^2$ -distribution)** Let  $R_1, R_2, R_3, R_4, U_0, U_1$  and  $U_2$  be independent random variables such that

- $R_i \sim \mathcal{R}(\alpha_i)$ , i.e.,

$$\mathbb{P}(R_i = -1) = 1 - \alpha_i, \quad \mathbb{P}(R_i = 1) = \alpha_i,$$

where  $\alpha_i \in (0, 1)$  for any  $i \in \{1, 2, 3, 4\}$ ,

- $\text{supp}(U_j) \subseteq \mathbb{Z}$  for any  $j \in \{0, 1, 2\}$ .

We say that  $(X, Y)$  follows the  $\mathbb{Z}^2$ -distribution if and only if

$$X \stackrel{d}{=} R_1U_1 + R_2U_0, \quad Y \stackrel{d}{=} R_3U_2 + R_4U_0.$$

### 3.2 Properties

In this section, we study some properties of the  $\mathbb{Z}^2$ -distribution. We set  $\mu_j = \mathbb{E}(U_j)$  and  $\sigma_j^2 = \mathbb{V}(U_j)$  assuming that they exist.

**Expectations.** We have

$$\mathbb{E}(X) = (2\alpha_1 - 1)\mu_1 + (2\alpha_2 - 1)\mu_0, \quad \mathbb{E}(Y) = (2\alpha_3 - 1)\mu_2 + (2\alpha_4 - 1)\mu_0.$$

**Variations.** We have

$$\mathbb{V}(X) = \mathbb{V}(R_1U_1) + \mathbb{V}(R_2U_0) = \sigma_1^2 - 4\alpha_1(\alpha_1 - 1)\mu_1^2 + \sigma_0^2 - 4\alpha_2(\alpha_2 - 1)\mu_0^2$$

and

$$\mathbb{V}(Y) = \mathbb{V}(R_3U_2) + \mathbb{V}(R_4U_0) = \sigma_2^2 - 4\alpha_3(\alpha_3 - 1)\mu_2^2 + \sigma_0^2 - 4\alpha_4(\alpha_4 - 1)\mu_0^2.$$

**Covariance.**

$$\mathbb{C}_{ov}(X, Y) = \mathbb{C}_{ov}(R_2U_0, R_4U_0) = (2\alpha_2 - 1)(2\alpha_4 - 1)\sigma_0^2.$$

**Moments.** We have

$$\mathbb{E}(X^n) = \sum_{k=0}^n \binom{n}{k} \mathbb{E}(R_1^k) \mathbb{E}(R_2^{n-k}) \mathbb{E}(U_1^k) \mathbb{E}(U_0^{n-k})$$

and

$$\mathbb{E}(Y^n) = \sum_{k=0}^n \binom{n}{k} \mathbb{E}(R_3^k) \mathbb{E}(R_4^{n-k}) \mathbb{E}(U_2^k) \mathbb{E}(U_0^{n-k}).$$

**Distribution of  $X$ .** We have

$$\begin{aligned} \mathbb{P}(X = x) &= \alpha_1\alpha_2\mathbb{P}(U_1 + U_0 = x) + (1 - \alpha_1)\alpha_2\mathbb{P}(-U_1 + U_0 = x) \\ &+ \alpha_1(1 - \alpha_2)\mathbb{P}(U_1 - U_0 = x) + (1 - \alpha_1)(1 - \alpha_2)\mathbb{P}(-U_1 - U_0 = x). \end{aligned}$$

**Distribution of  $Y$ .** We have

$$\begin{aligned} \mathbb{P}(Y = y) &= \alpha_3\alpha_4\mathbb{P}(U_2 + U_0 = y) + (1 - \alpha_3)\alpha_4\mathbb{P}(-U_2 + U_0 = y) \\ &+ \alpha_3(1 - \alpha_4)\mathbb{P}(U_2 - U_0 = y) + (1 - \alpha_3)(1 - \alpha_4)\mathbb{P}(-U_2 - U_0 = y). \end{aligned}$$

**Characteristic function of  $(X, Y)$ .** We have

$$\begin{aligned} \varphi(s, t) &= \mathbb{E}(e^{isX + itY}) = \mathbb{E}\left(e^{isR_1U_1 + itR_3U_2 + iU_0(sR_2 + tR_4)}\right) \\ &= \mathbb{E}(e^{isR_1U_1}) \mathbb{E}(e^{itR_3U_2}) \mathbb{E}\left(e^{iU_0(sR_2 + tR_4)}\right) \\ &= ((1 - \alpha_1)\varphi_{U_1}(-s) + \alpha_1\varphi_{U_1}(s)) ((1 - \alpha_3)\varphi_{U_2}(-t) + \alpha_3\varphi_{U_2}(t)) \times \\ &\quad ((1 - \alpha_2)(1 - \alpha_4)\varphi_{U_0}(-s - t) + \alpha_2(1 - \alpha_4)\varphi_{U_0}(s - t) + (1 - \alpha_2)\alpha_4\varphi_{U_0}(-s + t) + \alpha_2\alpha_4\varphi_{U_0}(s + t)) \end{aligned}$$

**Joint Probability mass function of  $(X, Y)$ :** sum version. We have that according to the combinations of signs of the  $R$ 's we can derive the joint probability function as a mixture of several components (in fact up to 16 components). This can be simplified upon restrictions in the  $\alpha$ 's. The general form is given by:

$$\begin{aligned}
& \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\
&= \alpha_1 \alpha_2 \alpha_3 \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x - k) \mathbb{P}(U_2 = y - k) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) \alpha_2 \alpha_3 \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = k - x) \mathbb{P}(U_2 = y - k) \mathbb{P}(U_0 = k) \\
&+ \alpha_1 (1 - \alpha_2) \alpha_3 \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x + k) \mathbb{P}(U_2 = y - k) \mathbb{P}(U_0 = k) \\
&+ \alpha_1 \alpha_2 (1 - \alpha_3) \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x - k) \mathbb{P}(U_2 = k - y) \mathbb{P}(U_0 = k) \\
&+ \alpha_1 \alpha_2 \alpha_3 (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x - k) \mathbb{P}(U_2 = y + k) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) (1 - \alpha_2) \alpha_3 \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = -x - k) \mathbb{P}(U_2 = y - k) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) \alpha_2 (1 - \alpha_3) \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = k - x) \mathbb{P}(U_2 = k - y) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) \alpha_2 \alpha_3 (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = k - x) \mathbb{P}(U_2 = y + k) \mathbb{P}(U_0 = k) \\
&+ \alpha_1 (1 - \alpha_2) (1 - \alpha_3) \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x + k) \mathbb{P}(U_2 = k - y) \mathbb{P}(U_0 = k) \\
&+ \alpha_1 (1 - \alpha_2) \alpha_3 (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x + k) \mathbb{P}(U_2 = y + k) \mathbb{P}(U_0 = k) \\
&+ \alpha_1 \alpha_2 (1 - \alpha_3) (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x - k) \mathbb{P}(U_2 = -y - k) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) (1 - \alpha_2) (1 - \alpha_3) \alpha_4 \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = -x - k) \mathbb{P}(U_2 = k - y) \mathbb{P}(U_0 = k) \\
&+ \alpha_1 (1 - \alpha_2) (1 - \alpha_3) (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = x + k) \mathbb{P}(U_2 = -y - k) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) \alpha_2 (1 - \alpha_3) (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = k - x) \mathbb{P}(U_2 = -y - k) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) (1 - \alpha_2) \alpha_3 (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = -x - k) \mathbb{P}(U_2 = y + k) \mathbb{P}(U_0 = k) \\
&+ (1 - \alpha_1) (1 - \alpha_2) (1 - \alpha_3) (1 - \alpha_4) \sum_{k \in \mathbb{Z}} \mathbb{P}(U_1 = -x - k) \mathbb{P}(U_2 = -y - k) \mathbb{P}(U_0 = k).
\end{aligned}$$

This equation can be used to determine the likelihood function and estimate unknown parameters.

### 3.3 Study of a particular case

In this section we study the particular case where

- $R_1, R_2, R_3, R_4$  are i.i.d. with  $R_i \sim \mathcal{R}(\alpha)$  where  $\alpha \in (0, 1)$  for any  $i \in \{1, 2, 3, 4\}$
- $U_j \sim \mathcal{P}(\lambda_j)$  where  $\lambda_j > 0$ , i.e.,

$$\mathbb{P}(U_j = k) = e^{-\lambda_j} \frac{\lambda_j^k}{k!}, \quad k \in \mathbb{N},$$

where  $\lambda_j > 0$  for any  $j \in \{0, 1, 2\}$ .

We recall the  $\mathbb{Z}^2$ -distribution of  $(X, Y)$  is defined by

$$X \stackrel{d}{=} R_1 U_1 + R_2 U_0, \quad Y \stackrel{d}{=} R_3 U_2 + R_4 U_0.$$

(In the sequel, we adopt the convention  $\sum_{k=i}^j a_k = 0$  for  $j < i$ ).

Expectations. We have

$$\mathbb{E}(X) = (2\alpha - 1)(\lambda_1 + \lambda_0), \quad \mathbb{E}(Y) = (2\alpha - 1)(\lambda_2 + \lambda_0).$$

Variances. We have

$$\mathbb{V}(X) = \lambda_1 + \lambda_0 - 4\alpha(\alpha - 1)(\lambda_1^2 + \lambda_0^2)$$

$$\mathbb{V}(Y) = \lambda_2 + \lambda_0 - 4\alpha(\alpha - 1)(\lambda_2^2 + \lambda_0^2).$$

Covariance.

$$\mathbb{C}_{ov}(X, Y) = (2\alpha - 1)^2 \lambda_0.$$

Distribution of  $X$ . We have

$$\begin{aligned} \mathbb{P}(X = x) &= \alpha^2 e^{-(\lambda_0 + \lambda_1)} \frac{(\lambda_0 + \lambda_1)^x}{x!} \\ &+ (1 - \alpha) \alpha e^{-(\lambda_0 + \lambda_1)} \left( \sum_{k=\max(x, 0)}^{\infty} \frac{\lambda_1^{k-x}}{(k-x)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(-x, 0)}^{\infty} \frac{\lambda_1^{x+k}}{(x+k)!} \frac{\lambda_0^k}{k!} \right) \\ &+ (1 - \alpha)^2 e^{-(\lambda_0 + \lambda_1)} \frac{(\lambda_0 + \lambda_1)^{-x}}{(-x)!}. \end{aligned}$$

Distribution of  $Y$ . We have

$$\begin{aligned} \mathbb{P}(Y = y) &= \alpha^2 e^{-(\lambda_0 + \lambda_2)} \frac{(\lambda_0 + \lambda_2)^y}{y!} \\ &+ (1 - \alpha) \alpha e^{-(\lambda_0 + \lambda_2)} \left( \sum_{k=\max(y, 0)}^{\infty} \frac{\lambda_2^{k-y}}{(k-y)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(-y, 0)}^{\infty} \frac{\lambda_2^{y+k}}{(y+k)!} \frac{\lambda_0^k}{k!} \right) \\ &+ (1 - \alpha)^2 e^{-(\lambda_0 + \lambda_2)} \frac{(\lambda_0 + \lambda_2)^{-y}}{(-y)!}. \end{aligned}$$

*Remark 1* One can recognize that the marginal distributions are a mixture of Skellam distributions. Namely it has four components: two of them are Skellam, one of them is a simple Poisson and one negative Poisson. Since Poisson and negative Poisson can be considered as special cases of Skellam with one parameter equal to 0, the marginals are finite mixtures of Skellam. This gives a better insight as several results from finite mixtures are applicable. In figure 1 we have plotted the marginal distribution for  $X$  for a combination of parameters. One can see that the distribution can have as many as three modes. When  $\alpha = 0.5$  the distribution is symmetric.

Characteristic function of  $(X, Y)$ . We have

$$\begin{aligned}
\varphi(s, t) &= \mathbb{E} \left( e^{isX+itY} \right) = \mathbb{E} \left( e^{isR_1U_1+itR_3U_2+iU_0(sR_2+tR_4)} \right) \\
&= \mathbb{E} \left( e^{isR_1U_1} \right) \mathbb{E} \left( e^{itR_3U_2} \right) \mathbb{E} \left( e^{iU_0(sR_2+tR_4)} \right) \\
&= \left( (1 - \alpha)e^{-\lambda_1(s+1)} + \alpha e^{\lambda_1(s-1)} \right) \left( (1 - \alpha)e^{-\lambda_2(t+1)} + \alpha e^{\lambda_2(t-1)} \right) \times \\
&\quad \left( (1 - \alpha)^2 e^{-\lambda_0(s+t+1)} + \alpha(1 - \alpha)e^{\lambda_0(s-t-1)} + (1 - \alpha)\alpha e^{\lambda_0(-s+t-1)} + \alpha^2 e^{\lambda_0(s+t-1)} \right).
\end{aligned}$$

Distribution of  $(X, Y)$ . We have

$$\begin{aligned}
& \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\
&= \alpha^4 e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \sum_{k=0}^{\min(x,y)} \frac{\lambda_1^{x-k}}{(x-k)!} \frac{\lambda_2^{y-k}}{(y-k)!} \frac{\lambda_0^k}{k!} \\
&+ (1-\alpha)\alpha^3 e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \times \\
&\quad \left( \sum_{k=\max(x,0)}^y \frac{\lambda_1^{k-x}}{(k-x)!} \frac{\lambda_2^{y-k}}{(y-k)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(-x,0)}^y \frac{\lambda_1^{x+k}}{(x+k)!} \frac{\lambda_2^{y-k}}{(y-k)!} \frac{\lambda_0^k}{k!} \right. \\
&+ \sum_{k=\max(y,0)}^x \frac{\lambda_1^{x-k}}{(x-k)!} \frac{\lambda_2^{k-y}}{(k-y)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(-y,0)}^x \frac{\lambda_1^{x-k}}{(x-k)!} \frac{\lambda_2^{y+k}}{(y+k)!} \frac{\lambda_0^k}{k!} \left. \right) \\
&+ (1-\alpha)^2 \alpha^2 e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \times \\
&\quad \left( \sum_{k=0}^{\min(-x,y)} \frac{\lambda_1^{-x-k}}{(-x-k)!} \frac{\lambda_2^{y-k}}{(y-k)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(x,y,0)}^{\infty} \frac{\lambda_1^{k-x}}{(k-x)!} \frac{\lambda_2^{k-y}}{(k-y)!} \frac{\lambda_0^k}{k!} \right. \\
&+ \sum_{k=\max(-y,x,0)}^{\infty} \frac{\lambda_1^{k-x}}{(k-x)!} \frac{\lambda_2^{y+k}}{(y+k)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(y,-x,0)}^{\infty} \frac{\lambda_1^{x+k}}{(x+k)!} \frac{\lambda_2^{k-y}}{(k-y)!} \frac{\lambda_0^k}{k!} \\
&+ \sum_{k=\max(-y,-x,0)}^{\infty} \frac{\lambda_1^{x+k}}{(x+k)!} \frac{\lambda_2^{y+k}}{(y+k)!} \frac{\lambda_0^k}{k!} + \sum_{k=0}^{\min(x,-y)} \frac{\lambda_1^{x-k}}{(x-k)!} \frac{\lambda_2^{-y-k}}{(-y-k)!} \frac{\lambda_0^k}{k!} \left. \right) \\
&+ (1-\alpha)^3 \alpha e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \times \\
&\quad \left( \sum_{k=\max(y,0)}^{-x} \frac{\lambda_1^{-x-k}}{(-x-k)!} \frac{\lambda_2^{k-y}}{(k-y)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(-x,0)}^{-y} \frac{\lambda_1^{x+k}}{(x+k)!} \frac{\lambda_2^{-y-k}}{(-y-k)!} \frac{\lambda_0^k}{k!} \right. \\
&+ \sum_{k=\max(x,0)}^{-y} \frac{\lambda_1^{k-x}}{(k-x)!} \frac{\lambda_2^{-y-k}}{(-y-k)!} \frac{\lambda_0^k}{k!} + \sum_{k=\max(-y,0)}^{-x} \frac{\lambda_1^{-x-k}}{(-x-k)!} \frac{\lambda_2^{y+k}}{(y+k)!} \frac{\lambda_0^k}{k!} \left. \right) \\
&+ (1-\alpha)^4 e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \sum_{k=0}^{\min(-x,-y)} \frac{\lambda_1^{-x-k}}{(-x-k)!} \frac{\lambda_2^{-y-k}}{(-y-k)!} \frac{\lambda_0^k}{k!}. \tag{1}
\end{aligned}$$

*Remark 2* In Figure 2 we have depicted some joint distributions for this family to show the wide range of shapes it can take. The plots are image plots, the darker the color the larger the probability. We have used various values of  $\alpha$  namely 0.5, 0.3, 0.8, 0.97 and  $\lambda_0 = 2, \lambda_1 = 0.5, \lambda_2 = 1$ . Since  $\lambda_0$  determines the covariance but also relates to the marginal means we used a higher value for this than the other  $\lambda$ 's. Plots are very interesting showing the potential. For the down-right  $\alpha$  is close to 1, which means that with high probability we get positive  $R$ 's the joint density is very close to a standard bivariate Poisson one, that is why it is concentrated at the positive orthant. When  $\alpha = 0.5$  then we got 1 and -1 with equal chance leading to a multimodal distribution with a let say  $X$  shape. More-

over, in this case, we have zero covariance/correlation but one can observe that there exists clearly a dependence between the observations. Now, for  $\alpha = 0.3$  (resp. 0.8), we see higher concentrations in the negative (resp. positive) orthants. Finally the conditional expectations are linear due to the generalized trivariate reduction used.

### 3.4 Parametric estimation

Using the setting described in the section above, we investigate the estimation of the unknown parameters  $(\alpha, \lambda_0, \lambda_1, \lambda_2)$  from a sample of size  $n$ ,  $(X_1, Y_1), \dots, (X_n, Y_n)$  of  $(X, Y)$ .

Method of Moments. We can express  $(\lambda_0, \lambda_1, \lambda_2)$  in terms of  $\alpha$ ,  $\mathbb{C}_{ov}(X, Y)$ ,  $\mathbb{E}(X)$  and  $\mathbb{E}(Y)$  as

$$\begin{cases} \lambda_0 = \frac{\mathbb{C}_{ov}(X, Y)}{(2\alpha - 1)^2}, \\ \lambda_1 = \frac{\mathbb{E}(X)}{2\alpha - 1} - \frac{\mathbb{C}_{ov}(X, Y)}{(2\alpha - 1)^2}, \\ \lambda_2 = \frac{\mathbb{E}(Y)}{2\alpha - 1} - \frac{\mathbb{C}_{ov}(X, Y)}{(2\alpha - 1)^2}. \end{cases}$$

Therefore, using the expression  $\mathbb{V}(X) = \lambda_1 + \lambda_0 - 4\alpha(\alpha - 1)(\lambda_1^2 + \lambda_0^2)$  (resp.  $\mathbb{V}(Y)$ ) and the above equations, one can write  $\alpha$  in terms of  $\mathbb{E}(X)$ ,  $\mathbb{V}(X)$  (resp.  $\mathbb{E}(Y)$ ,  $\mathbb{V}(Y)$ ) and  $\mathbb{C}_{ov}(X, Y)$ . Furthermore, by replacing the empirical estimators of these previous quantities, we end up with the following equation which has to be solved:

$$f(\alpha) = \frac{\bar{x}}{(2\alpha - 1)} - 4\alpha(\alpha - 1) \left( \frac{\bar{x}^2}{(2\alpha - 1)^2} - 2\frac{\bar{x} C_{X,Y}}{(2\alpha - 1)^3} + 2\frac{C_{X,Y}^2}{(\alpha - 1)^4} - S_X^2 \right) = 0$$

where,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad C_{X,Y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$

This is a nonlinear equation for  $\alpha$  with an asymptote at  $\alpha = 0.5$ . For the sake of clarity, in Figure 3 we plot the  $f(\alpha)$  function, based on the real data example used in section 4. Clearly there is an asymptote in  $\alpha = 0.5$  which may create difficulties when solving the equation. Also note that perhaps there exists several solutions. An easy way to select the solution but also the interval to search for the solution is to select the interval so as to match the sign of the covariance.

Solving this one can obtain a consistent moment estimate  $\hat{\alpha}$ . Estimations of the other parameters follow from a plug-in method:

$$\begin{cases} \hat{\lambda}_0 = \frac{\mathbb{C}_{ov}(X, Y)}{(2\hat{\alpha} - 1)^2}, \\ \hat{\lambda}_1 = \frac{\mathbb{E}(X)}{2\hat{\alpha} - 1} - \frac{\mathbb{C}_{ov}(X, Y)}{(2\hat{\alpha} - 1)^2}, \\ \hat{\lambda}_2 = \frac{\mathbb{E}(Y)}{2\hat{\alpha} - 1} - \frac{\mathbb{C}_{ov}(X, Y)}{(2\hat{\alpha} - 1)^2}. \end{cases}$$

Standard errors can be derived using standard asymptotic results based on the delta method (see, e.g. Lehmann and Casella (1998)), however in practice bootstrap standard errors are more easy to derive. Note that the estimate near 0.5 can have very inefficient behavior since the equation is too steep and hence small changes in the moments can lead to large changes in the estimated values.

Maximum likelihood estimation. This method considers the estimators:

$$(\hat{\alpha}, \hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2) = \operatorname{argmax} \prod_{i=1}^n \mathbb{P}(\{X = X_i\} \cap \{Y = Y_i\}),$$

where  $\mathbb{P}(\{X = X_i\} \cap \{Y = Y_i\})$  is given by (1).

Standard numerical methods can be used to derive the estimates. We have used simple code in R which worked without problems. Standard errors can be derived from the Hessian as usual (see, e.g. Lehmann and Casella (1998)).

### 3.5 Some simulation comparison

In order to gain an idea on the estimation methods, we have conducted a small simulation experiment for the two methods. We have used two configurations ( $\alpha = 0.2, \lambda_0 = 1, \lambda_1 = 0.5, \lambda_2 = 1$ ) and ( $\alpha = 0.4, \lambda_0 = 0.5, \lambda_1 = 0.5, \lambda_2 = 1$ ). For each one, we run 5000 replications. Simulation is straightforward based on the definition, as Rademacher's random variables simulation is simplistic. We have used three sample sizes namely  $n = 100, 500, 1000$ . For each sample we estimated the parameters with moment and ML methods. Note that for moment method a large number of times, we derived inadmissible estimates (e.g. negative  $\lambda$ 's). Such samples were excluded, however for small sample sizes this can be as large as 30% of the replications. The reported values are based on 5000 replication where the estimates existed for both methods. Figures 4 and 5 depict boxplots of both methods (dark grey is the moment method, light grey the ML method) for all sample sizes. The horizontal line is the true values used for the simulations. Boxplots obviously show the superiority of the ML methods. Moment estimates while as mentioned above can fail for a large number of times, they lead to biased estimates especially for small sample sizes but also with much larger variances as the length of the boxplot show. On the contrary ML estimates are unbiased with much less variability, leading to much better behavior.

## 4 Application

In this section, we use the football data from the Italian Series A championship. There are 20 teams playing matches in both home and guest. In total 380 matches were played. We have data for 379 of them as one match never finished. Our data consist of the scores in the final ( $X$ ) and the half time ( $Y$ ). We take the difference between the home team and the guest team, positive scores imply the win of the home team while negative the win of the guest team. Data were downloaded from [www.stat-football.com](http://www.stat-football.com). We have that  $\bar{x} = 0.358$ ,  $\bar{y} = 0.2031$  while  $S_X^2 = 2.807$ ,  $S_Y^2 = 1.172$ ,  $C_{X,Y} = 1.151$ , implying strong positive correlation ( $\tau = 0.634$ ) as expected.

Data are depicted in Figure 6. The size of the circle is proportional to the frequency of this. It is clear a positive correlation between the half time and the final scores.

In the sequel we fitted the bivariate distribution defined in 1, as well as other candidates like the distribution in Bulla et al. (2012a) and the one in Karlis and Ntzoufras (2012). For our model moment estimates failed leading to negative estimates. ML estimates can be seen in the table below, where the last line also show the log-likelihood from the competing models. The new model provides much better fit judged by the smaller log-likelihood. It is also interesting that it has less parameters but also better fit.

	parameter	std
	$\alpha$	0.6158
	$\lambda_0$	0.6884
	$\lambda_1$	0.9200
	$\lambda_2$	0.0882
	Log-likelihood	
Model	Log-Lik	Parameters
New	-1247.995	4
Bulla	-1273.889	3
Copula	-1254.919	5

Table 1 shows the observed and fitted frequencies of the Italian soccer data using the estimated parameters. The resemblance in the center of the data is very good. Also the marginals are very close, indicating that the proposed model can capture the general characteristics.

		Y										
		-4	-3	-2	-1	0	1	2	3	4	Total	
X	-5	obs	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
		expec	0.04	0.10	0.13	0.10	0.16	0.15	0.21	0.16	1.10	
		-4	obs	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	3.00
			expec	0.10	0.22	0.42	0.43	0.84	0.67	0.68	0.36	3.89
		-3	obs	0.00	1.00	5.00	1.00	3.00	0.00	0.00	0.00	10.00
			expec	0.07	0.59	0.95	1.45	3.61	2.29	1.54	0.98	11.61
		-2	obs	0.00	1.00	6.00	10.00	17.00	7.00	0.00	0.00	41.00
			expec	0.04	0.41	2.61	3.58	11.64	5.66	4.28	0.74	29.04
		-1	obs	0.00	0.00	3.00	17.00	23.00	6.00	2.00	0.00	51.00
			expec	0.02	0.22	1.92	9.84	25.53	15.70	3.36	0.42	57.04
		0	obs	0.00	0.00	1.00	20.00	61.00	14.00	0.00	0.00	96.00
			expec	0.01	0.15	1.65	10.44	70.35	16.68	2.90	0.29	102.48
		1	obs	0.00	0.00	0.00	7.00	45.00	28.00	8.00	0.00	88.00
			expec	0.02	0.23	2.08	14.61	40.80	23.34	3.77	0.44	85.31
		2	obs	0.00	0.00	0.00	1.00	19.00	21.00	10.00	1.00	52.00
			expec	0.04	0.46	4.22	7.82	18.88	12.40	6.98	0.88	51.77
		3	obs	0.00	0.00	0.00	0.00	6.00	8.00	11.00	3.00	28.00
			expec	0.09	0.98	2.38	3.40	5.90	5.37	3.88	1.64	23.80
		4	obs	0.00	0.00	0.00	0.00	0.00	2.00	3.00	2.00	7.00
			expec	0.17	0.55	1.08	1.03	1.38	1.62	1.74	0.91	8.78
	5	obs	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
		expec	0.10	0.25	0.33	0.24	0.26	0.37	0.53	0.41	2.64	
	6	obs	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	2.00	
		expec	0.04	0.08	0.08	0.04	0.04	0.07	0.12	0.12	0.67	
	Total	obs	1.00	3.00	16.00	56.00	175.00	86.00	35.00	6.00	379.00	
		expec	0.74	4.23	17.84	52.96	179.37	84.31	29.99	7.35	378.13	

**Table 1** Observed and fitted frequencies of the Italian data. The resemblance in the center of the data is very good. Also the marginals are very close

## 5 Discussion

There are some natural extensions of the work presented in the current paper. First of all, the generalized reduction method can be used to derive multivariate analogues. We did not pursue this in the present paper to avoid complicating notations and expositions. Also, a slight reparameterization could allow for possible covariates in the mean of the models and hence to allow more flexibility for modelling purposes.

More importantly, in Definition 1, we introduce a new distribution on  $\mathbb{Z}^2$  based on generalized trivariate (multivariate) reduction technique and the Rademacher distribution. One of possible extensions of this distribution, can be achieved by replacing the multiplication in the equations of  $X$  and  $Y$ , by a thinning operator. Explicitly,

$$X \stackrel{d}{=} R_1 \odot U_1 + R_2 \odot U_0, \quad Y \stackrel{d}{=} R_3 \odot U_2 + R_4 \odot U_0,$$

with  $U_0, U_1$  and  $U_2$  are non-negative integer-valued i.i.d. random variables, and for all  $i \in \{1, 2, 3, 4\}$ ,  $j \in \{0, 1, 2\}$ ,

$$R_i \odot U_j = \sum_{k=1}^{U_j} \xi_{k,i}$$

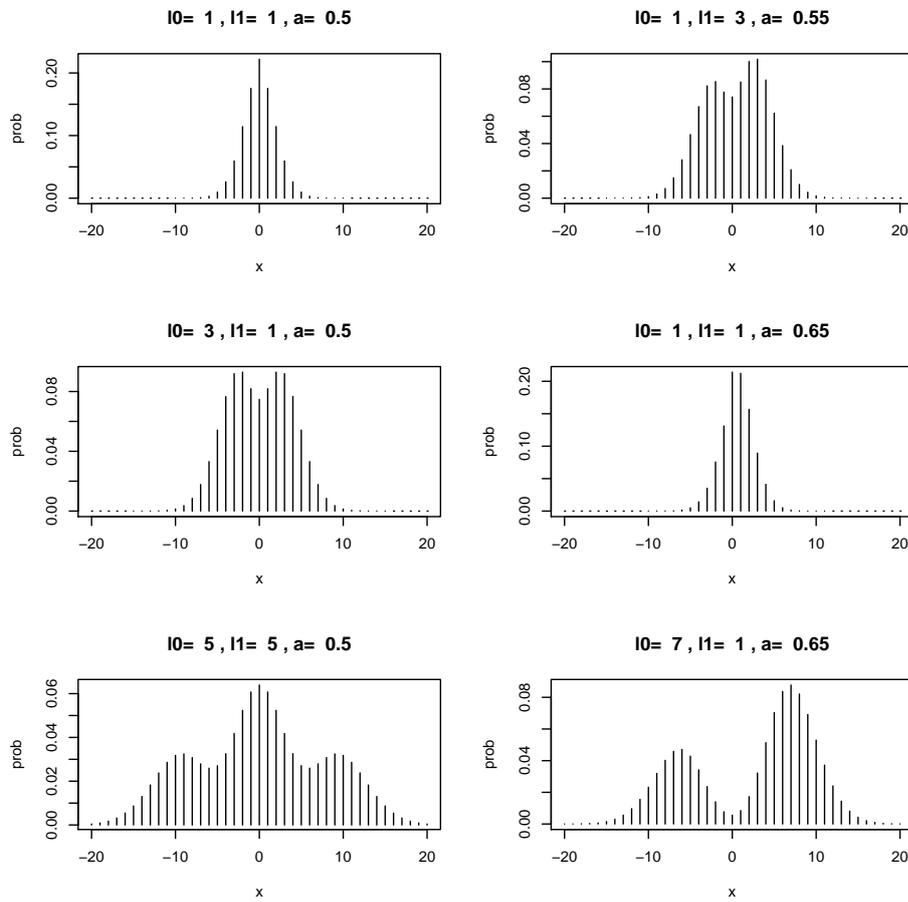
where  $(\xi_{k,i})$  is a sequence (called counting sequence) of i.i.d. Rademacher( $\alpha_i$ ) and independent of  $U_j$ . It is understood that the counting sequences associated to each operator are mutually independent. Note that the above operator is none other than a particular case of the signed thinning operator, for details please see Latour and Truquet (2008) and Kachour and Truquet (2011). On the other hand, one can see the  $\odot$  operator as a generalization of the random walk on  $\mathbb{Z}$ , where the number of steps is also an non-negative integer-valued random variable.

## References

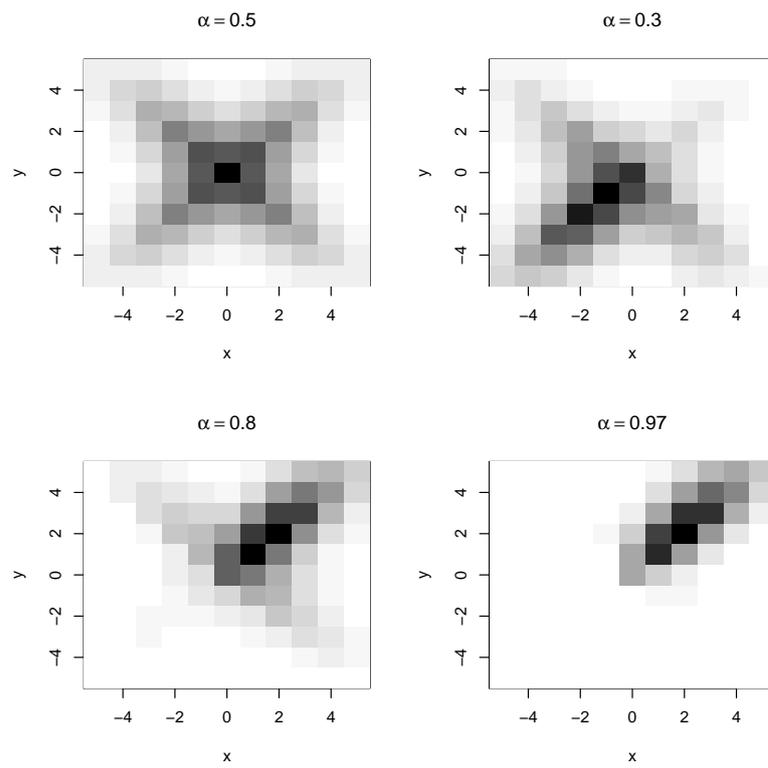
- M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1974.
- A.A. Alzaid and M.A. Omaid. On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society*, 8(33):17–45, 2010.
- J. Bulla, C. Chesneau and M. Kachour. On the bivariate Skellam distribution. *submitted*, 2012.
- J. Bulla, C. Chesneau and M. Kachour. A bivariate first-order signed integer-valued autoregressive process. *submitted*, 2012.
- K. Hirakawa and P.J. Wolfe. Skellam shrinkage: Wavelet-based intensity estimation for inhomogeneous poisson data. *IEEE Transactions on Information Theory*, 58(2):1080–1093, 2012.
- S. Inusah and T.J. Kozubowski, A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, 136(3):1090–1102, 2006.
- M. Kachour, and L. Truquet. A p-order signed integer-valued autoregressive (SINAR(1)) model. *Journal of Time Series Analysis*, 32(4):, 223–236, 2011.
- D. Karlis and I. Ntzoufras. Bayesian analysis of the differences of count data. *Statistics in medicine*, 25(11):1885–1905, 2006.
- D. Karlis and I. Ntzoufras. Bayesian modelling of football outcomes: Using the skellam's distribution for the goal difference. *IMA Journal Management Mathematics*, 20(2): 133–145, 2009.
- D. Karlis and I. Ntzoufras. Bivariate Skellam distributions with applications to soccer modelling and betting. *submitted*, 2012.
- A.W. Kemp. Characterizations of a discrete normal distribution. *Journal of Statistical Planning and Inference*, 63(2):223–229, 1997
- V. Koltchinskii. Rademacher penalties and structural risk minimization. *IEEE Trans. Inform. Theory* 47 :1902–1914, 2001.
- T.J. Kozubowski and S. Inusah. A skew laplace distribution on integers. *Annals of the Institute of Statistical Mathematics*, 58(3):555–571, 2006.
- C.D. Lai, Construction of bivariate distributions by a generalized trivariate reduction technique. *Statistics and Probability Letters*, 25(3):265–270, 1995.
- A. Latour, and L. Truquet. An integer-valued bilinear type model. Research report, available on <http://hal.archives-ouvertes.fr/hal-00373409>, 2008.
- E.L. Lehmann, and G. Casella, *Theory of Point Estimation* (2nd ed.), Springer, 1998.
- R.Y. Liu, Bootstrap procedures under some non i.i.d. models. *Annals of Statistics*, 16(3): 1696–1708, 1988.

- 
- P. Massart. Concentration Inequalities and Model Selection *Ecole d'été de Probabilités de Saint-Flour 23*. Springer, 2003
- S.H. Ong, K. Shimizu, and C. Min Ng. A class of discrete distributions arising from difference of two random variables. *Computational Statistics and Data Analysis*, 52(3):1490–1499, 2008.
- D. Roy. The discrete normal distribution. *Communications in Statistics - Theory and Methods*, 32(10):1871–1883, 2003.
- J.M. Sarabia and E. Gomez-Deniz. Construction of multivariate distributions: a review of some recent results. *Statistics and Operations Research Transactions (SORT)*, 32(1): 3–36, 2008.
- J.G. Skellam. The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society: Series A*, 109:296, 1946.
- P.J. Szablowski. Discrete normal distribution and its relationship with jacobi theta functions. *Statistics and Probability Letters*, 52(3):289–299, 2001.

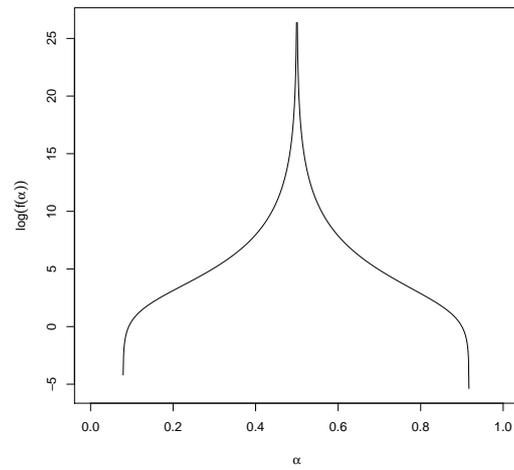
## Figures



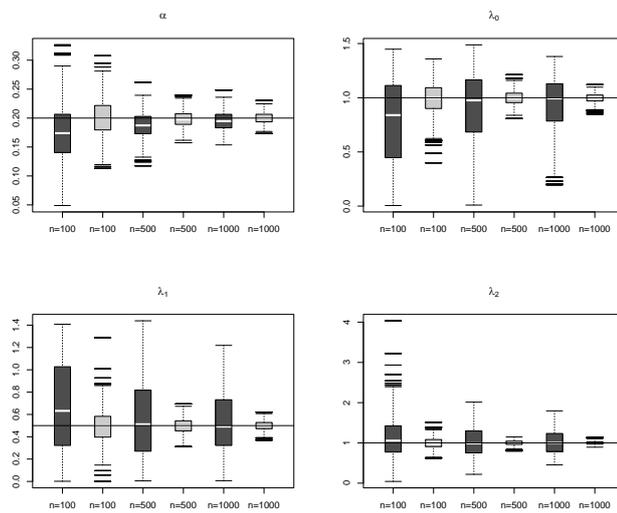
**Fig. 1** Plot of the marginal distribution for several parameter combinations. It is interesting the wide range of shapes of the distributions derived.



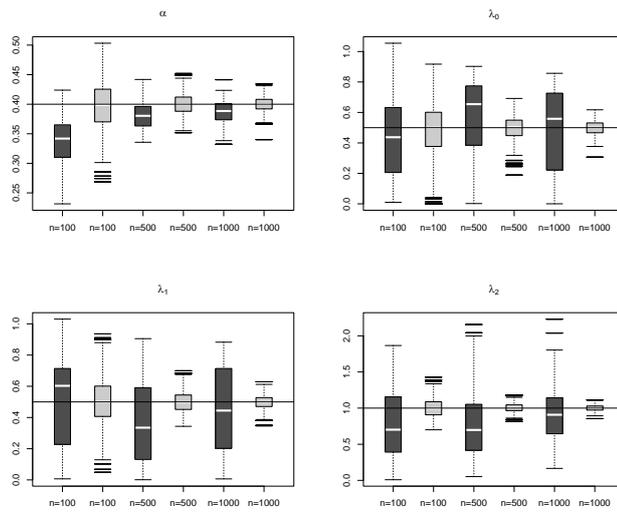
**Fig. 2** The joint probability mass function for various values of  $\alpha$  and  $\lambda_0 = 2, \lambda_1 = 0.5, \lambda_2 = 1$ .



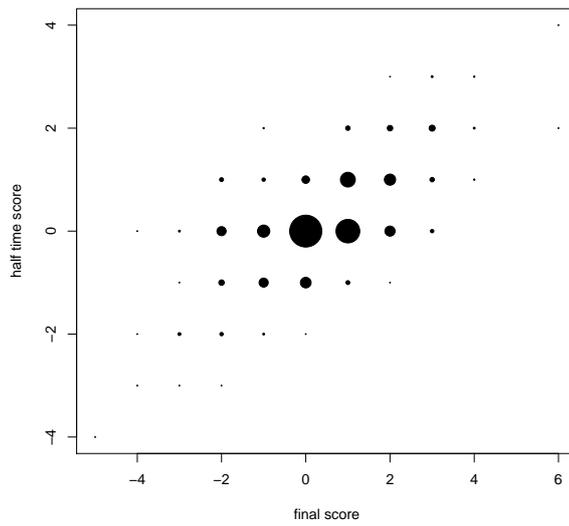
**Fig. 3** The plot of function  $f(\alpha)$ , which is used to calculate the moments estimators, for the data on Italian football.



**Fig. 4** Boxplots from the simulations using  $(\alpha = 0.2, \lambda_0 = 1, \lambda_1 = 0.5, \lambda_2 = 1)$ . 1000 replications were used, cases when the moment estimates failed were excluded.



**Fig. 5** Boxplots from the simulations using  $(\alpha = 0.4, \lambda_0 = 0.5, \lambda_1 = 0.5, \lambda_2 = 1)$ . 1000 replications were used, cases when the moment estimates failed were excluded.



**Fig. 6** Plots of Italian data used. The size of the point relates to the observed frequency.