



HAL
open science

On the bivariate Skellam distribution

Jan Bulla, Christophe Chesneau, Maher Kachour

► **To cite this version:**

Jan Bulla, Christophe Chesneau, Maher Kachour. On the bivariate Skellam distribution. 2013. hal-00744355v2

HAL Id: hal-00744355

<https://hal.science/hal-00744355v2>

Preprint submitted on 23 Oct 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 the bivariate Skellam distribution

Jan Bulla · Christophe Chesneau ·
Maher Kachour

accepted at *Communications in Statistics - Theory and Methods*

Abstract In this paper, we introduce a new distribution on \mathbb{Z}^2 , which can be viewed as a natural bivariate extension of the Skellam distribution. The main feature of this distribution is a possible dependence of the univariate components, both following univariate Skellam distributions. We explore various properties of the distribution and investigate the estimation of the unknown parameters via the method of moments and maximum likelihood. In the experimental section, we illustrate our theory. First, we compare the performance of the estimators by means of a simulation study. In the second part, we present two applications to a real data set and show how an improved fit can be achieved by estimating mixture distributions.

Keywords Skellam distribution · Bivariate Skellam distribution · Maximum likelihood estimate.

2000 Mathematics Subject Classification 62G30, 62E10.

1 Introduction

In recent years, there has been a growing interest in studying bivariate discrete random variables, in an effort to explain phenomena in various areas of application. More specifically, bivariate discrete random variables are appropriate for modelling paired count data exhibiting correlation. Paired count data occur in many contexts such as marketing (joint purchases of two products), econometrics (number of voluntary and involuntary job changes), insurance

Jan Bulla and Christophe Chesneau
Université de Caen, Laboratoire de Mathématiques Nicolas Oresme, CNRS UMR 6139,
14032 Caen Cedex, France. E-mail: bulla@math.unicaen.fr, 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: kachour.maher@gmail.com

(number of accidents in a site before and after infrastructure changes), engineering (faults due to different causes), epidemiology (joint concurrence of two different diseases), and many others.

The Bivariate Poisson Distribution (BPD), originally derived by McKendrick (1926) as a solution to a differential equation arising in a biological application, is probably the best known bivariate discrete distribution. For a comprehensive treatment of the bivariate Poisson distribution and its multivariate extensions, the reader can refer to the books of Kocherlakota and Kocherlakota (1992) and Johnson et al. (1997), and the review articles by Papageorgiou (1997) and Kocherlakota and Kocherlakota (1997). The BPD has found many applications recently (e.g., Karlis and Ntzoufras (2003) have used it for modelling sports data). Many other bivariate discrete distributions have been introduced, for details we refer to Marshall and Olkin (1990), Boucher et al. (2008), and Cheon et al. (2009). In particular, we cite the Bivariate Negative Binomial Distribution (BNBD), which was used in insurance theory for describing of the number of accidents in transportation, and the Bivariate Geometric Distribution (BGD), which is considered plausible for reliability modelling. Note that the literature on bivariate count distributions with negative correlation is limited. One of the reasons is that negative correlation in bivariate counts occurs rather infrequently. However there are such models in the literature, as for example the bivariate Poisson-lognormal model of Aitchinson and Ho (1989), the finite mixture model developed in Karlis and Meligkotsidou (2007) and models based on copulas see, e.g. Nikoloulopoulos and Karlis (2009) and the references therein.

In contrast to the well-known situation when the paired data are counts, i.e., observed on \mathbb{N}^2 , sometimes the data take values in \mathbb{Z}^2 . For example, when analysing intra-daily stock prices the changes take both positive and negative integer values, known also as ticks (the price can go up or down on certain predefined ranges of value). The price change is therefore characterized by discrete jumps. While bivariate discrete distribution for non-negative paired data are now abundant, to our best knowledge there is a shortage of bivariate discrete distribution defined on \mathbb{Z}^2 . We intend to contribute to this literature by presenting a bivariate discrete distribution based on the Skellam distribution.

The remainder of the paper proceeds as follows. Firstly, we present the considered bivariate Skellam distribution and investigate some of its properties in Section 2. Section 3 describes two possible approaches for estimating the unknown parameters efficiently. Section 4 presents results from a simulation study and a real data set, respectively. Last, Section 5 is devoted to the proofs.

2 The BSkellam($\lambda_0, \lambda_1, \lambda_2$) distribution

Before analyzing the bivariate setting, we recall the definition of the univariate Skellam distribution.

Definition 1 (*Skellam(λ_1, λ_2) distribution*) Let $\lambda_1 > 0$ and $\lambda_2 > 0$. We say that the random variable X has the Skellam distribution, denoted by

Skellam(λ_1, λ_2), if and only if

$$X \stackrel{d}{=} U_1 - U_2,$$

where U_1 and U_2 are two independent random variables such that $U_i \sim \text{Poisson}(\lambda_i)$ for any $i \in \{1, 2\}$. The probability mass function of X is given by

$$\mathbb{P}(X = x) = e^{-(\lambda_1 + \lambda_2)} \lambda_1^x \sum_{i=\max(0, -x)}^{\infty} \frac{(\lambda_1 \lambda_2)^i}{(x+i)!i!}, \quad x \in \mathbb{Z}.$$

The distribution of the difference between two independent Poisson random variables was derived by Irwin (1937) for the case of equal parameters. Skellam (1946) and Prékopa (1953) discussed the case of unequal parameters. Further details can be found in e.g. Karlis and Ntzoufras (2008) and Al-Zaid and Omair (2010).

Remark 1 The definition of the Skellam distribution can be extended to more than the simple difference of two independent Poisson distributions. Indeed, let X_1 and X_2 be two independent Poisson random variables with parameters θ_1 and θ_2 respectively. Let $Y_i = X_i + W$, for $i = 1, 2$, where W is a random variable independent of X_1 and X_2 . Thus, $Z = Y_1 - Y_2 = X_1 - X_2$ also follows a Skellam(θ_1, θ_2) distribution.

Inspired by the trivariate reduction method derived by Johnson et Kotz (1969), the purpose of this paper is to study a natural bivariate extension of the Skellam(λ_1, λ_2) distribution, including dependence between the components. It is described below.

Definition 2 (*BSkellam*($\lambda_0, \lambda_1, \lambda_2$) *distribution*) Let $\lambda_0 \geq 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$. We say that the bivariate random variable (X_1, X_2) follows a bivariate Skellam distribution, denoted by BSkellam($\lambda_0, \lambda_1, \lambda_2$), if and only if

$$X_1 \sim \text{Skellam}(\lambda_1, \lambda_0), \quad \text{and} \quad X_2 \sim \text{Skellam}(\lambda_2, \lambda_0).$$

The probability mass function of (X_1, X_2) is given by

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = e^{-(\lambda_1 + \lambda_2 + \lambda_0)} \lambda_1^{x_1} \lambda_2^{x_2} \sum_{i=\max(0, -x_1, -x_2)}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x_1+i)!(x_2+i)!i!}$$

for all $(x_1, x_2) \in \mathbb{Z}^2$.

Remark 2 Note that $Z = X_1 - X_2 \sim \text{Skellam}(\lambda_1, \lambda_2)$. Hence, the probability distribution of Z is independent of λ_0 . Moreover, for $\lambda_0 = 0$, the bivariate Skellam distribution reduces to the product of two independent Poisson distributions.

Lemma 1 below investigates some of the basic properties of the BSkellam($\lambda_0, \lambda_1, \lambda_2$) distribution.

Lemma 1 Let $\lambda_0 \geq 0$, $\lambda_1 > 0$, $\lambda_2 > 0$ and $(X_1, X_2) \sim \text{BSkellam}(\lambda_0, \lambda_1, \lambda_2)$. Then

– the mean of (X_1, X_2) is $(\lambda_1 - \lambda_0, \lambda_2 - \lambda_0)$ and the covariance matrix equals

$$\Sigma = \begin{pmatrix} \lambda_1 + \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_2 + \lambda_0 \end{pmatrix}.$$

– the correlation coefficient of (X_1, X_2) is

$$\text{cor}(X_1, X_2) = \frac{\lambda_0}{\sqrt{(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)}}.$$

Once again, one can see that $\lambda_0 > 0$ implies that X_1 and X_2 are dependent random variables.

– the characteristic function of (X_1, X_2) can be calculated as

$$\phi_{(X_1, X_2)}(t_1, t_2) = e^{\lambda_1(e^{it_1} - 1)} e^{\lambda_2(e^{it_2} - 1)} e^{\lambda_0(e^{-it_1 - it_2} - 1)} \quad \forall (t_1, t_2) \in \mathbb{Z}^2.$$

Remark 3 Using the characteristic function of $(X_1, X_2) \sim \text{BSkellam}(\lambda_0, \lambda_1, \lambda_2)$, the probability mass function of (X_1, X_2) can be expressed by

$$P(X_1 = x_1, X_2 = x_2) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi_{(X_1, X_2)}(t_1, t_2) e^{-ix_1 t_1} e^{-ix_2 t_2} dt_1 dt_2 \quad \forall (x_1, x_2) \in \mathbb{Z}^2. \quad (1)$$

In particular, this yields the following equality

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{\lambda_1(e^{it_1} - 1)} e^{\lambda_2(e^{it_2} - 1)} e^{\lambda_0(e^{-it_1 - it_2} - 1)} e^{-ix_1 t_1} e^{-iy t_2} dt_1 dt_2 = (2\pi)^2 e^{-(\lambda_1 + \lambda_2 + \lambda_0)} \lambda_1^x \lambda_2^y \sum_{i=0}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)!i!} \quad \forall (x_1, x_2) \in \mathbb{Z}^2. \quad (2)$$

Lemma 2 below presents a Gaussian approximation result for the $\text{BSkellam}(\lambda_0, \lambda_1, \lambda_2)$ distribution.

Lemma 2 Let $\lambda_0 \geq 0$, $\lambda_1 > \lambda_0$ and $\lambda_2 > \lambda_0$ and suppose that λ_1 and λ_2 are large enough. Then, the distribution of (X_1, X_2) can be approximate by the following bivariate gaussian distribution

$$\mathcal{N}_2 \left(\begin{pmatrix} \lambda_1 - \lambda_0 \\ \lambda_2 - \lambda_0 \end{pmatrix}, \begin{pmatrix} \lambda_1 - \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_2 - \lambda_0 \end{pmatrix} \right)$$

Remark 4 Recall that the distribution of X_1 and X_2 can be approximated by a Gaussian distribution for large values of $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$, respectively.

3 Parameter estimation

This section is devoted to the estimation of the parameters of a BSkellam($\lambda_0, \lambda_1, \lambda_2$) distribution via the method of moments and the method of the maximum likelihood, respectively. Let $\lambda_0 \geq 0$, $\lambda_1 > 0$, and $\lambda_2 > 0$ be unknown parameters and $(X_{1,1}, X_{2,1}), \dots, (X_{1,n}, X_{2,n})$ be n i.i.d. bivariate random variables with the common distribution BSkellam($\lambda_0, \lambda_1, \lambda_2$).

3.1 The method of moments

Set, for any $j \in \{1, 2\}$,

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{j,i}, \text{ and } C_{1,2} = \frac{1}{n-1} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2).$$

In view of Lemma 1, the method of moment provides

- the estimator $\hat{\lambda}_0 = C_{1,2}$ for λ_0 ,
- the estimator

$$\hat{\lambda}_j = \bar{X}_j + \hat{\lambda}_0$$

for λ_j , $j \in \{1, 2\}$.

Remark 5 Thanks to the strong law of large numbers, note that $\hat{\lambda}_j$ converge a.s. to λ_j for $j \in \{0, 1, 2\}$.

Remark 6 The moment estimates of λ_0 do not exist if $C_{1,2} < 0$. To solve this problem, a modification is done such that the negative estimate of λ_0 is set to zero and the other estimate is set to equal the absolute value of the respective sample mean.

3.2 Maximum likelihood estimation

For any $z = (x, y) \in \mathbb{Z}^2$ and $\lambda = (\lambda_0, \lambda_1, \lambda_2) \in [0, \infty[^3$, set

$$G(\lambda, z) = \lambda_1^x \lambda_2^y \sum_{i=\max(0, -x, -y)}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)!i!}.$$

Lemma 3 For any $(x, y) \in \mathbb{Z}^2$ holds

$$\frac{\partial G(\lambda, z)}{\partial \lambda_1} = \frac{x}{\lambda_1} G(\lambda, z) + \frac{\lambda_0}{\lambda_1} G(\lambda, (x+1, y+1)),$$

$$\frac{\partial G(\lambda, z)}{\partial \lambda_2} = \frac{y}{\lambda_2} G(\lambda, z) + \frac{\lambda_0}{\lambda_2} G(\lambda, (x+1, y+1)),$$

and

$$\frac{\partial G(\lambda, z)}{\partial \lambda_0} = G(\lambda, (x+1, y+1)).$$

For any $i \in \{1, \dots, n\}$, set $z_i = (x_{1,i}, x_{2,i}) \in \mathbb{Z}^2$, $z = (z_1, \dots, z_n)$, and $\lambda = (\lambda_0, \lambda_1, \lambda_2) \in [0, \infty]^3$. Then, the likelihood function is given by

$$L(\lambda, z) = \prod_{i=1}^n \mathbb{P}(X_1 = x_{1,i}, X_2 = x_{2,i}) = e^{-n(\lambda_0 + \lambda_1 + \lambda_2)} \prod_{i=1}^n G(\lambda, z_i).$$

Therefore, the j -th partial derivative of the log-likelihood function equals

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_j} = -n + \sum_{i=1}^n \frac{\frac{\partial G(\lambda, z_i)}{\partial \lambda_j}}{G(\lambda, z_i)}.$$

for any $j \in \{0, 1, 2\}$. From Lemma 3 follows

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_1} = -n + \frac{1}{\lambda_1} \sum_{i=1}^n x_{1,i} + \frac{\lambda_0}{\lambda_1} \sum_{i=1}^n \frac{G(\lambda, z_i + 1)}{G(\lambda, z_i)},$$

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_2} = -n + \frac{1}{\lambda_2} \sum_{i=1}^n x_{2,i} + \frac{\lambda_0}{\lambda_2} \sum_{i=1}^n \frac{G(\lambda, z_i + 1)}{G(\lambda, z_i)},$$

and

$$\frac{\partial \ln L(\lambda, z)}{\partial \lambda_0} = -n + \sum_{i=1}^n \frac{G(\lambda, z_i + 1)}{G(\lambda, z_i)}.$$

Let $Z_i = (X_{1,i}, X_{2,i})$ for any $i \in \{1, \dots, n\}$. The maximum likelihood estimate $\hat{\lambda} = (\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2)$ satisfies

$$\begin{cases} 0 = -n + \frac{1}{\hat{\lambda}_1} \sum_{i=1}^n X_{1,i} + \frac{\hat{\lambda}_0}{\hat{\lambda}_1} \sum_{i=1}^n \frac{G(\hat{\lambda}, Z_i + 1)}{G(\hat{\lambda}, Z_i)}, \\ 0 = -n + \frac{1}{\hat{\lambda}_2} \sum_{i=1}^n X_{2,i} + \frac{\hat{\lambda}_0}{\hat{\lambda}_2} \sum_{i=1}^n \frac{G(\hat{\lambda}, Z_i + 1)}{G(\hat{\lambda}, Z_i)}, \\ 0 = -n + \sum_{i=1}^n \frac{G(\hat{\lambda}, Z_i + 1)}{G(\hat{\lambda}, Z_i)}. \end{cases} \quad (3)$$

Therefore, $\sum_{i=1}^n G(\hat{\lambda}, Z_i + 1)/G(\hat{\lambda}, Z_i) = n$ and, a fortiori,

$$\hat{\lambda}_1 = \bar{X}_1 + \hat{\lambda}_0, \quad \hat{\lambda}_2 = \bar{X}_2 + \hat{\lambda}_0. \quad (4)$$

Substituting (4) into the last equation of the system of equations (3), we obtain the following equation for $\hat{\lambda}_0$:

$$n = \sum_{i=1}^n \frac{G((\bar{X}_1 + \hat{\lambda}_0, \bar{X}_2 + \hat{\lambda}_0, \hat{\lambda}_0), Z_i + 1)}{G((\bar{X}_1 + \hat{\lambda}_0, \bar{X}_2 + \hat{\lambda}_0, \hat{\lambda}_0), Z_i)}.$$

Then, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ follow directly from $\hat{\lambda}_0$ and (4), and the maximum likelihood estimates $\hat{\lambda}$ are complete. Nevertheless, for practical purposes a direct numerical maximization of expression 1 and 2, respectively, is also a straightforward approach, and applied in the what follows.

Remark 7 The maximum likelihood estimator $\hat{\lambda}$ is asymptotically normally distributed $\mathcal{N}_3(\lambda, I^{-1}(\lambda))$, where $I(\lambda)$ is the Fisher information matrix with entries

$$I_{i,j}(\lambda) = -\mathbb{E} \left(\frac{\partial \ln L(\lambda, (X_1, X_2))}{\partial \lambda_{i-1} \partial \lambda_{j-1}} \right) \quad \text{with } (i, j) \in \{1, 2, 3\}.$$

Traditionally, this matrix can be estimated by $I(\hat{\lambda})$, which allows the constructions of confidence intervals or hypothesis tests.

3.3 Extensions of the bivariate Skellam

Having only three parameters, the bivariate Skellam distribution is a rather parsimonious distribution for bivariate data. However, along with the parsimony naturally come certain drawbacks. For example, mean and skewness of the marginal distributions have identical sign, and the standard deviation depends on the same parameters as the mean. In order to increase the flexibility of the bivariate Skellam, we propose two extensions: Firstly, the inclusion of a shift parameter and, secondly, mixtures of bivariate Skellam distributions.

Shifted BSkellam distribution. The inclusion of a shift parameter $k = (k_1, k_2)$ in the bivariate Skellam is straightforward. Then, we obtain a new bivariate random variable $(X_{1*}, X_{2*}) = (X_1 + k_1, X_2 + k_2)$ with probability mass

$$\mathbb{P}(X_{1*} = x_1, X_{2*} = x_2) = \mathbb{P}(X_1 = x_1 - k_1, X_2 = x_2 - k_2)$$

where $(X_1, X_2) \sim \text{BSkellam}(\lambda_0, \lambda_1, \lambda_2)$.

Basic properties of this distribution are $(\mathbb{E}(X_{1*}), \mathbb{E}(X_{2*})) = (\mathbb{E}(X_1) + k_1, \mathbb{E}(X_2) + k_2) = (\lambda_1 - \lambda_0 + k_1, \lambda_2 - \lambda_0 + k_2)$, and the covariance matrix identical to that of (X_1, X_2) .

Mixture of BSkellam distributions. Distributions based on the mixture of several bivariate Skellam distributions can be characterized by the probability mass

$$\mathbb{P}(X_{1\bullet} = x_1, X_{2\bullet} = x_2) = \sum_{m=1}^M p_m \mathbb{P}(X_{1m} = x_1, X_{2m} = x_2),$$

where $\sum_{m=1}^M p_m = 1$, $p_m \geq 0$ and $(X_{1m}, X_{2m}) \sim \text{BSkellam}(\lambda_{0m}, \lambda_{1m}, \lambda_{2m})$ for all $m \in \{1, \dots, M\}$. Let us observe that, for all $i \in \mathbb{Z}$, we have

$$\mathbb{P}(X_{1\bullet} = i) = \sum p_m \mathbb{P}(X_{1m} = i) \quad \text{and} \quad \mathbb{P}(X_{2\bullet} = i) = \sum p_m \mathbb{P}(X_{2m} = i).$$

It follows that

$$\begin{aligned} (\mathbb{E}(X_{1\bullet}), \mathbb{E}(X_{2\bullet})) &= \left(\sum_{m=1}^M p_m(\lambda_{1m} - \lambda_{0m}), \sum_{m=1}^M p_m(\lambda_{2m} - \lambda_{0m}) \right), \\ \text{cov}(X_{1\bullet}, X_{2\bullet}) &= \sum_{m=1}^M p_m (\lambda_{0m} + (\lambda_{1m} - \lambda_{0m})(\lambda_{2m} - \lambda_{0m})) - \\ &\quad \sum_{m=1}^M p_m(\lambda_{1m} - \lambda_{0m}) \sum_{m=1}^M p_m(\lambda_{2m} - \lambda_{0m}), \\ \mathbb{V}(X_{1\bullet}) &= \sum_{m=1}^M p_m(\lambda_{1m} + \lambda_{0m} + (\lambda_{1m} - \lambda_{0m})^2) - \left(\sum_{m=1}^M p_m(\lambda_{1m} - \lambda_{0m}) \right)^2, \end{aligned}$$

and

$$\mathbb{V}(X_{2\bullet}) = \sum_{m=1}^M p_m(\lambda_{2m} + \lambda_{0m} + (\lambda_{2m} - \lambda_{0m})^2) - \left(\sum_{m=1}^M p_m(\lambda_{2m} - \lambda_{0m}) \right)^2.$$

Moments of higher order for this distribution cannot be derived easily, and may thus be approximated by Monte Carlo methods for practical applications.

Naturally, both extensions can also be combined by mixtures of shifted bivariate Skellam distributions. Parameter estimation for a shifted BSkellam distribution, a mixture of BSkellam distributions, and mixtures of shifted BSkellam distribution may be carried out by numerical maximization of the log likelihood. However, it may be noted that this procedure can become computationally demanding, because it requires optimization over a set of parameters with continuous and discrete-values entries.

4 Applications

In this application section, we present a Monte Carlo study in the next Section 4.1, followed by an application to real data in Section 4.2 and 4.3.

4.1 Monte Carlo study

In this section we briefly demonstrate some properties of the bivariate Skellam distribution and the parameter estimation procedures. More precisely, we consider two settings (termed Setting 1 and 2 in the following), dealing with a bivariate Skellam distribution with parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 2, 3)$ and $(\lambda_0, \lambda_1, \lambda_2) = (4, 35, 40)$, respectively. Figure 1 shows random samples with 500 observations each from Setting 1 in the left and Setting 2 in the right panel. The two settings are subject to different correlation, which equals 0.617

and 0.0966, respectively. For the two samples displayed, the estimated correlation equal 0.628 ($p < 2.2e-16$) and 0.0975 ($p = 0.0292$), respectively.

In a first Monte Carlo experiment, we investigate the departure from normality in the two settings. For this purpose, we generated 1000 samples having 1000 observations for each of the settings and performed the Shapiro-Wilk test for multivariate normality. In Setting 1, having comparably small values of λ_i , $i = 1, 2, 3$, normality is rejected at 5%-level for 87.2% of the samples. However, the corresponding proportion in the second setting reduces to 30.1%. This indicates that the bivariate Skellam can be approximated by a Gaussian distribution for large values of λ_i , $i = 1, 2, 3$, which supports the results of Lemma 2.

In a second Monte Carlo experiment, we investigate the properties of the method of moments and maximum likelihood estimators. To this end, the two parameter sets described above were used for generating 1000 samples of observations of size $n = 50, 100, \text{ and } 200$, respectively. Subsequently, the parameters were estimated by the method of moments and by maximization of the likelihood. Figure 2 shows Box plots of the estimated parameters. The method of moments estimates have white background and the boxes corresponding to the maximum likelihood estimator are coloured grey. The dashed black lines represent the true parameter values. Several phenomena are visible: First, both estimators seem to be consistent. Second, both estimators seem unbiased. Third, the maximum likelihood estimator has a lower variability than the method of moments estimator. Finally, it may be noted that normality of the estimators cannot be rejected for the large majority of samples at 5%-level (30 of 36, Shapiro-Wilk).

4.2 Soccer results

The data analyzed in this section are soccer results from the qualifications matches to the soccer world championship 2014 that took place in June and September 2012. The data are available for download on the FIFA web site (<http://www.fifa.com/worldcup/preliminaries/matches/date=092012.html>). More precisely, we consider the difference of goals scored by the home and the guest team during the first and second half, respectively, of the match. Figure 3 shows the 239 observations. The average difference in the first half equals 0.155 (s.d.: 1.19) and 0.172 (s.d.: 1.54) in the second half. Moreover, the estimated correlation of the differences calculated from the first and second half takes the value 0.412 ($p < 0.001$). This indicates that an independent modelling of the variables would not be appropriate.

We fitted four different models to the data: first, a bivariate Skellam and a shifted bivariate Skellam distribution. Secondly, mixtures of bivariate Skellam and shifted bivariate Skellam distributions. These four models are denoted BSk, sBSk, mix-BSk, and mix-sBSk, respectively, in the following. The parameters of all models were estimated by maximization of the log-likelihood. Table

1 displays the maximized values of the log-likelihood and the model selection criteria AIC and BIC. Both mixture models have two components, because models with a higher number of components are rejected by both selection criteria.

The AIC shows a preference for the mix-sBSk, while the BIC prefers the more parsimonious mix-BSk without shift. The estimated parameters of the mix-BSk take the values

$$\begin{aligned}(\hat{\lambda}_{01}, \hat{\lambda}_{11}, \hat{\lambda}_{21}) &= (0.383, 0.561, 0.955), \\(\hat{\lambda}_{02}, \hat{\lambda}_{12}, \hat{\lambda}_{22}) &= (2.190, 2.300, 1.490), \\(\hat{p}_1, \hat{p}_2) &= (0.685, 0.315),\end{aligned}$$

and the respective estimates for mix-sBSk equal

$$\begin{aligned}(\hat{\lambda}_{01}, \hat{\lambda}_{11}, \hat{\lambda}_{21}) &= (0.272, 0.463, 0.836), \\(\hat{\lambda}_{02}, \hat{\lambda}_{12}, \hat{\lambda}_{22}) &= (1.000, 1.120, 2.780), \\(\hat{p}_1, \hat{p}_2) &= (0.503, 0.497), \\(\hat{k}_{11}, \hat{k}_{21}) &= (0, -2), \\(\hat{k}_{12}, \hat{k}_{22}) &= (0, 0).\end{aligned}$$

The simpler models with only one component are less favoured, which is underlined by the statistics presented in Table 2. This table shows the first three moments of the marginal distributions and the correlation for both the sample data and the estimated models.

The marginal means reproduced well by all models. As to standard deviation, skewness and correlation values, the mixture models show their superiority to the simpler models with only one component. Each of the two mixture models has its small advantages and disadvantages: While the model without shift comes closer to the correlation and skewness values of the data, the model incorporation a shift better captures the standard deviation of the sample.

The better fit of the mixture models also becomes visible in Figure 4, which shows histograms of the marginal distribution of the data with the four fitted models. From a contextual point of view, the application of a mixture model may be motivated by the comparably large heterogeneity of the matches carried out for a the qualification of the soccer world championship. For example, matches of more or less equally strong teams can be expected to have a distribution different to that of a matches of teams with large performance differences. However, further investigation of such hypothesis goes beyond the scope of this paper.

4.3 Analysis of EURO-Bund and Schatz future

In this section, we analyze data from the EURO-Bund and Schatz future, respectively, recorded on November 29th 2011. The maturity of both futures was

December 2011, and the smallest possible price change (tick) equals 0.01 and 0.005 for Bund and Schatz future, respectively. Prices of the two futures are recorded for intervals of two minute length, where the last trade in each interval determines the corresponding price. If no trades took place in an interval, the price was supposed to be unchanged. In order to avoid non-representative outliers, we restrict our observation period to the main trading hours from 9h15 to 17h15, resulting in 240 observations. During this period, the Bund future dropped slightly from 134.28 to 133.4, whereas the Schatz future remained almost unchanged (109.98 vs. 109.985). Figure 5 displays the price changes of the two futures in ticks, as well as their sample ACF, cross-correlation function, and a bivariate scatter plot. The two data series are time series, however, neither the ACF nor the cross-correlation function indicate the presence of any temporally lagged dependencies. Nevertheless, the two series are correlated at lag zero, the estimated correlation equals 0.664 ($p < 0.001$).

As in the previous sections, we fitted different models to the data. Here, we restrict ourselves to the simple bivariate Skellam (BSk) model and mixtures of bivariate Skellam distributions (mix-BSk), because preliminary analyses showed that a substantial increase in the log-likelihood could be easily achieved by increasing the number of mixture components, whereas shifted distributions only had a comparably small effect. Table 3 displays the maximized values of the log-likelihood and the model selection criteria AIC and BIC. Both AIC and BIC prefer the mixture model with 4 states.

The better fit of the mixture models also becomes visible in Figure 6, which shows histograms of the marginal distribution of the data with the two fitted models. The simple BSk fails to reproduce in particular the marginal of the Schatz future, for which the mix-BSk provides a more satisfactory image.

The estimated parameters of BSk take the values

$$(\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2) = (6.051, 5.672, 6.051),$$

and the respective estimates for mix-BSk equal

$$\begin{aligned} (\hat{\lambda}_{01}, \hat{\lambda}_{11}, \hat{\lambda}_{21}) &= (0.570, 1.375, 0.651), \\ (\hat{\lambda}_{02}, \hat{\lambda}_{12}, \hat{\lambda}_{22}) &= (3.042, 0.000, 2.657), \\ (\hat{\lambda}_{03}, \hat{\lambda}_{13}, \hat{\lambda}_{23}) &= (0.000, 5.873, 1.300), \\ (\hat{\lambda}_{04}, \hat{\lambda}_{14}, \hat{\lambda}_{24}) &= (7.406, 0.000, 6.106), \\ (\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4) &= (0.410, 0.283, 0.183, 0.125). \end{aligned}$$

5 Proofs

Proof of Lemma 1.

- For any $i \in \{1, 2\}$, we have $\mathbb{E}(X_i) = \lambda_i - \lambda_0$ and $\mathbb{V}(X_i) = \lambda_i + \lambda_0$. We have $\mathbb{C}_{ov}(X_1, X_2) = \lambda_0$. From these equalities follows the covariance matrix Σ .
- Using the first point, we obtain $\text{cor}(X_1, X_2) = \mathbb{C}_{ov}(X_1, X_2) / \sqrt{\mathbb{V}(X_1)\mathbb{V}(X_2)} = \lambda_0 / \sqrt{(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)}$.
- For any $(t_1, t_2) \in \mathbb{Z}^2$, the characteristic function of (X_1, X_2) is,

$$\begin{aligned} \phi_{(X_1, X_2)}(t_1, t_2) &= \mathbb{E}(e^{it_1 X_1 + it_2 X_2}) = \mathbb{E}(e^{it_1(U_1 - U_0) + it_2(U_2 - U_0)}) \\ &= \mathbb{E}(e^{it_1 U_1 + it_2 U_2 - i(t_1 + t_2)U_0}) = \phi_{U_1}(t_1) \phi_{U_2}(t_2) \phi_{U_0}(-t_1 - t_2) \\ &= e^{\lambda_1(e^{it_1} - 1)} e^{\lambda_2(e^{it_2} - 1)} e^{\lambda_0(e^{-it_1 - it_2} - 1)}. \end{aligned}$$

This ends the proof of Lemma 1. ■

Proof of Lemma 2. Set $W_1 = 1/(\lambda_1 - \lambda_0)X_1$ and $W_2 = 1/(\lambda_2 - \lambda_0)X_2$. For λ_2 and λ_1 large enough, using the third point of Lemma 1 and $e^x \underset{0}{\sim} 1 + x + x^2/2$ follows

$$\begin{aligned} \phi_{(W_1, W_2)}(t_1, t_2) &= \phi_{(X_1, X_2)}(t_1/(\lambda_1 - \lambda_0), t_2/(\lambda_2 - \lambda_0)) \\ &= e^{\lambda_1(e^{it_1/(\lambda_1 - \lambda_0)} - 1)} e^{\lambda_2(e^{it_2/(\lambda_2 - \lambda_0)} - 1)} e^{\lambda_0(e^{-it_1/(\lambda_1 - \lambda_0) - it_2/(\lambda_2 - \lambda_0)} - 1)} \\ &\approx e^{\lambda_1(it_1/(\lambda_1 - \lambda_0) - (1/2)t_1^2/(\lambda_1 - \lambda_0)^2)} e^{\lambda_2(it_2/(\lambda_2 - \lambda_0) - (1/2)t_2^2/(\lambda_2 - \lambda_0)^2)} \\ &\quad e^{\lambda_0(-it_1/(\lambda_1 - \lambda_0) - it_2/(\lambda_2 - \lambda_0) - (1/2)(t_1/(\lambda_1 - \lambda_0) + t_2/(\lambda_2 - \lambda_0))^2)} \\ &= e^{it_1} e^{it_2} e^{-(1/2)(t_1^2/(\lambda_1 - \lambda_0) + t_2^2/(\lambda_2 - \lambda_0) + 2t_1 t_2/(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0))} \end{aligned}$$

for any $(t_1, t_2) \in \mathbb{Z}^2$. The last term is the characteristic function of a bivariate random variable

$$Z \sim \mathcal{N}_2 \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1/(\lambda_1 - \lambda_0) & \lambda_0/(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0) \\ \lambda_0/(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0) & 1/(\lambda_2 - \lambda_0) \end{pmatrix} \right).$$

Therefore (W_1, W_2) can be approximated by this distribution and, a fortiori,

$$(X_1, X_2) \approx \mathcal{N}_2 \left(\begin{pmatrix} \lambda_1 - \lambda_0 \\ \lambda_2 - \lambda_0 \end{pmatrix}, \begin{pmatrix} \lambda_1 - \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_2 - \lambda_0 \end{pmatrix} \right).$$

Lemma 2 is proved. ■

Proof of Lemma 3. For any $(x, y) \in \mathbb{Z}^2$ holds

$$\begin{aligned} \frac{\partial G(\lambda, z)}{\partial \lambda_1} &= \lambda_1^x \lambda_2^y \sum_{i=0}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)!i!} \\ &= x \lambda_1^{x-1} \lambda_2^y \sum_{i=0}^{\infty} \frac{(\lambda_0 \lambda_1 \lambda_2)^i}{(x+i)!(y+i)!i!} + \lambda_1^x \lambda_2^y \sum_{i=1}^{\infty} i \frac{(\lambda_0 \lambda_2)^i \lambda_1^{i-1}}{(x+i)!(y+i)!i!} \\ &= \frac{x}{\lambda_1} G(\lambda, z) + \lambda_1^x \lambda_2^y \sum_{i=0}^{\infty} \frac{(\lambda_0 \lambda_2)^{i+1} \lambda_1^i}{(x+i+1)!(y+i+1)!i!} \\ &= \frac{x}{\lambda_1} G(\lambda, z) + \frac{\lambda_0}{\lambda_1} G(\lambda, (x+1, y+1)). \end{aligned}$$

The other equalities can be proved similarly, and the proof of Lemma 3 is complete.



Acknowledgements

We would like to thank the Chair of Quantitative Finance of the Ecole Centrale Paris for granting us access to their data base. Special thanks go to F. Abergel for his suggestion to closer investigating futures on government bonds and R. Zaatour for his continuous support with the data preparation process. Furthermore, we are would like to render thanks to two anonymous reviewers for their helpful comments.

References

- Aitchinson, J. and Ho, C. The multivariate Poisson-log normal distribution. *Biometrika*, 75:621629, 1989.
- Al-Zaid, A.A. and Omair, M.A. (2010). On the poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society*, 8, (33), 17-45.
- Boucher, J. P., Michel, D. and Montserrat, G. Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, 2(1):135162, 2008.
- Cheon, S., Song, S.H. and Jung, B.C. Tests for independence in a bivariate negative binomial model. *Journal of the Korean Statistical Society*, 38: 185190, 2009.
- Irwin, J. O. The frequency distribution of the difference between two independent variates following the same Poisson distribution, *Journal of the Royal Statistical Society Series A* 100, pp. 415, 1937.
- Johnson, N.L. and Kotz, S. *Distributions in statistics: Discrete distribution*, New York, Wiley and Sons, 1969.
- Johnson, N., Kotz, S. and Balakrishnan, N. *Multivariate Discrete Distributions*. Wiley, New York, 1997.
- Karlis, D. and Ntzoufras, I. Analysis of Sports Data by Using Bivariate Poisson Models. *Journal of the Royal Statistical Society D (The Statistician)*, 52, 381–393, 2003.
- Karlis, D. and Meligkotsidou, L. Finite multivariate Poisson mixtures with applications. *Journal of Statistical Planning and Inference*, 137: 19421960, 2007.
- Karlis, D. and Ntzoufras, I. (2008). Bayesian modeling of football outcomes: Using the skellam distribution for the goal difference. *Journal Management Mathematics*, 20, (2), 133-145.
- Nikoloulopoulos, A.K. and Karlis, D. Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference*, Volume 139, Issue 11, 3878-3890, 2009.
- Kocherlakota, S. and Kocherlakota, K. *Bivariate Discrete Distributions, Statistics: textbooks and monographs*, volume 132. Markel Dekker, New York, 1992.
- Kocherlakota, S. and Kocherlakota, K. Bivariate discrete distributions, in S. Kotz, C. B. Read & D. L. Banks, eds, *Encyclopedia of Statistical Sciences- Update*, Vol. 2, Wiley, New York, pp. 6883, 1997.
- Marshall, A. W. and Olkin, I. Multivariate distribution generated from mixtures of convolution and product families, pages 371393. *Topics in Statistical Dependence*, Block, Sampson y Sanits (Eds), Institute of Mathematical Statistics, 1990.
- McKendric, A. G. Applications of mathematics to medical problems, *Proc. Edinburgh Math. Soc.* 44, 98–130, 1926.
- Papageorgiou, H. Multivariate discrete distributions, in C. B. Kotz, S. Read & D. L. Banks, eds, *Encyclopedia of Statistical Sciences-Updat*, Vol. 1, Wiley, New York, pp. 408419, 1997.
- Prékopa, A. On composed Poisson distributions. IV. Remarks on the theory of differential processes, *Acta Math. Acad. Sci. Hungar.* 3, pp. 317–325, 1953.
- Skellam, J.G. (1946). The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society, Series A*, 109-296.

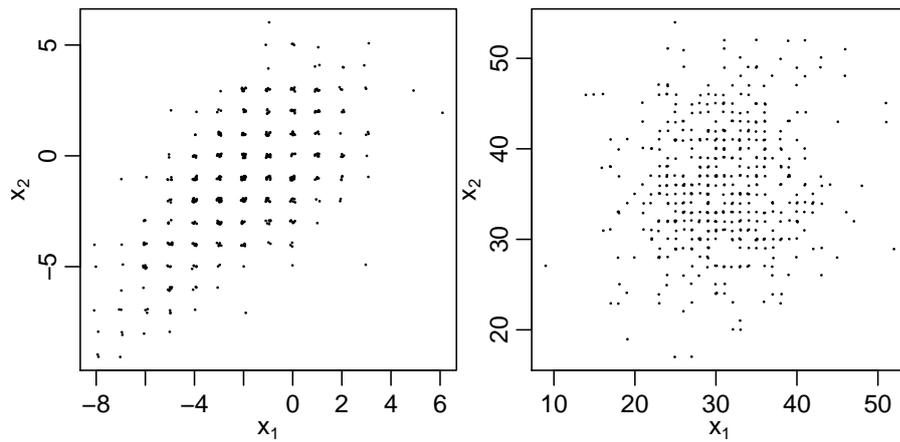


Fig. 1 Two samples with 500 observations each, generated from a bivariate Skellam distribution with true parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 2, 3)$ and $(\lambda_0, \lambda_1, \lambda_2) = (4, 35, 40)$ in the left (respectively right) panel. To improve the visualization, Gaussian jitter was added to the observations.

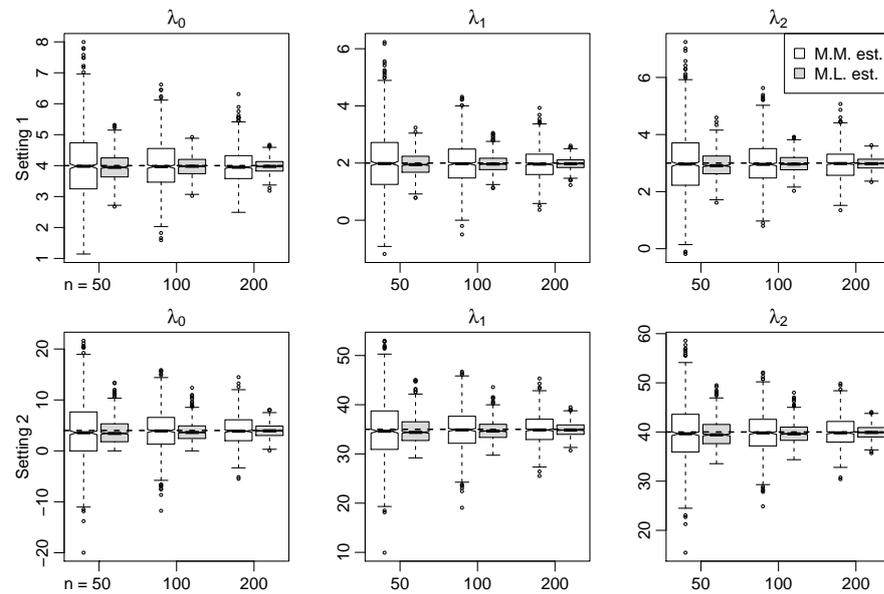


Fig. 2 Parameter estimates from 100 from simulated sequences of length $n = 50, 100,$ and 200 . The upper three panels were simulated from a bivariate Skellam with parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 2, 3)$, the lower three panels result from the parameters $(\lambda_0, \lambda_1, \lambda_2) = (4, 35, 40)$ in the left and right panel, respectively. Box plots corresponding to the method of moments (M.M.) and maximum likelihood (M.L.) estimator have white (respectively grey) background. The dashed black lines represent the true parameter values.

Fig. 3 Goal difference between home and guest team at half-time and at the end of the match. For improved visualization, Gaussian jitter was added to the observations.

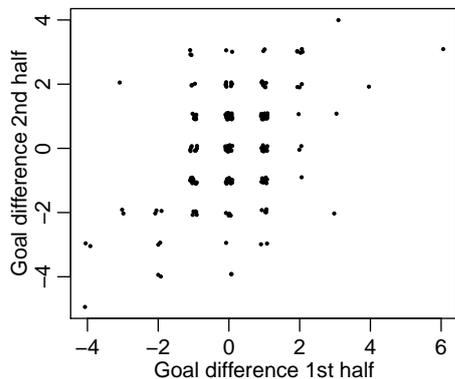


Table 1 Model selection

This table shows log-likelihood, AIC, and BIC values for the four estimated models.

	BSk	sBSk	mix-BSk	mix-sBSk
logL	-814	-802	-794	-787
AIC	1634	1614	1600	1595
BIC	1645	1632	1621	1630

Table 2 Moments and correlation

This table summarizes the first three moments of the marginal distributions and the correlation of the data and models.

	Mean		S.d.		Skew.		Corr.
Data	0.155	0.172	1.19	1.54	0.00101	-0.439	0.412
BSk	0.155	0.172	2.1	2.12	0.0507	0.0555	0.461
sBSk	0.155	0.172	1.49	2.51	0.0845	0.292	0.346
mix-BSk	0.155	0.172	1.43	1.56	-0.00228	-0.503	0.434
mix-sBSk	0.155	0.172	1.19	1.61	0.0422	-0.102	0.338

Table 3 Model selection

This table shows log-likelihood, AIC, and BIC values for the four estimated models.

no. states	1	2	3	4	5
logL	-1075	-1149	-1096	-1075	-1073
AIC	2173	2310	2211	2173	2176
BIC	2215	2331	2242	2215	2228

Fig. 4 Histogram of the marginal distributions with fitted models.

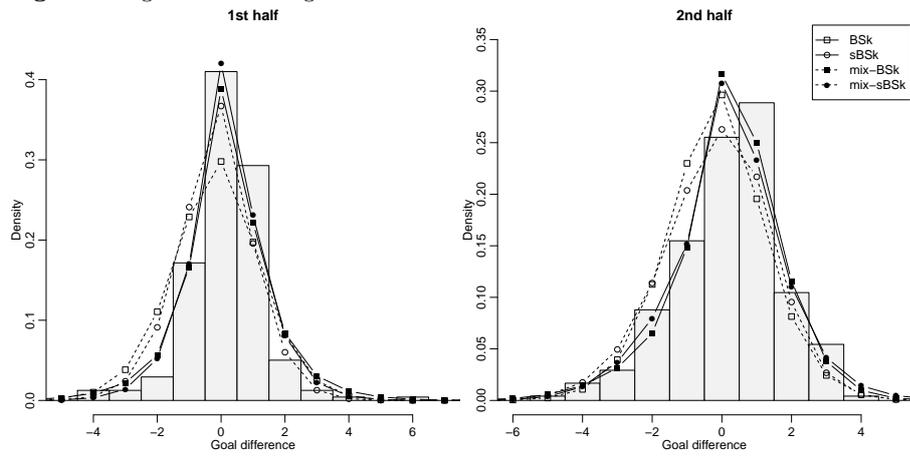


Fig. 5 Price changes of Bund and Schatz future in the top panels. The middle panels show the sample ACF of the two data series. The lower two panels show a bivariate scatter plot of the data and the empirical cross-correlation. For improved visualization, Gaussian jitter was added to the observations in the scatter plot.

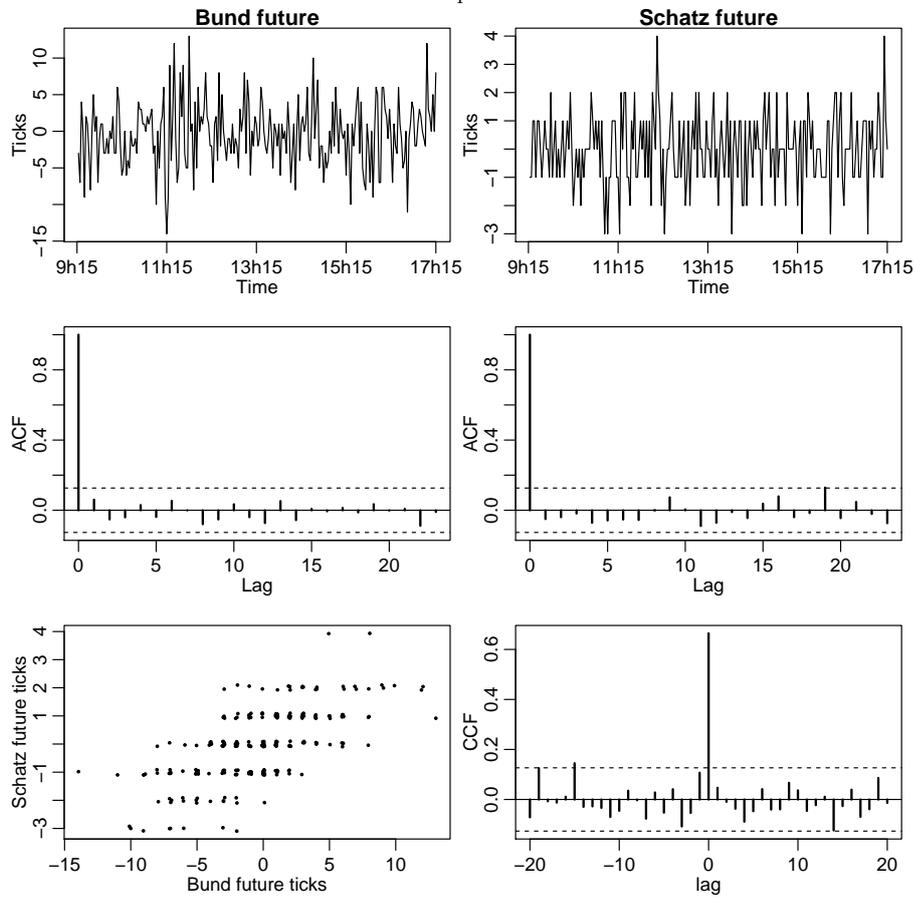


Fig. 6 Histogram of the marginal distributions with fitted models.

