ORIGINAL ARTICLE

Bivariate Conway–Maxwell Poisson Distributions with Given Marginals and Correlation

Seng Huat Ong[1] ⓘ **· Ramesh C. Gupta[2] · Tiefeng Ma[3] · Shin Zhu Sim[4]**

Abstract

The Conway–Maxwell Poisson (CMP) distribution is a popular model for analyzing data that exhibit under or over dispersion. In this article, we construct bivariate CMP distributions with given marginal CMP distributions and range of correlation coefficient over $(-1, 1)$ based on the Sarmanov family of bivariate distributions. One of the constructions is based on a general method for weighted distributions. The dependence property is examined. Parameter estimation, tests of independence and adequacy of model and a Monte Carlo power study are discussed. A real data set is used to exemplify its usefulness with comparison to other bivariate models.

Keywords Conway–Maxwell Poisson · Negative and positive correlation · Positive likelihood ratio dependence · Test of independence · Likelihood ratio test · Score test · Weighted distribution

✉ Seng Huat Ong
  ongsh@ucsiuniversity.edu.my

  Ramesh C. Gupta
  rcgupta@maine.edu

  Tiefeng Ma
  matiefeng@swufe.edu.cn

  Shin Zhu Sim
  simsz@utar.edu.my

[1] Department of Actuarial Science and Applied Statistics, UCSI University, 56000 Kuala Lumpur, Malaysia

[2] Department of Mathematics and Statistics, University of Maine Orono, Maine 04469-5752, USA

[3] School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, Sichuan, People's Republic of China

[4] Department of Mathematical and Actuarial Sciences, Universiti Tunku Abdul Rahman, 43000 Kajang, Selangor, Malaysia

1 Introduction

The construction of a bivariate distribution with specified marginals and correlation has been a challenging problem since the early twentieth century. There is much interest in this problem because of its wide ranging applications. For instance, Ong [22] considered computer sampling of some bivariate discrete distributions with given marginals and correlation. Recently, Lin et al. [21] gave a good survey on this topic and this complements previous surveys. A comprehensive overview is found in the monograph by Balakrishnan and Lai [2].

A simple method to formulate a bivariate distribution with fixed marginals and varying correlation is the well-known Farlie–Gumbel–Morgenstern (FGM) family of distributions defined by

$$H(x, y) = F(x)G(y)\left\{1 + \beta \bar{F}\bar{G}\right\} \tag{1.1}$$

where $F(x)$ and $G(y)$ are the cumulative distribution functions, $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$ and $\beta \varepsilon [-1, 1]$ is the parameter that controls the correlation. Schucany et al. [25] had shown that the FGM family (1.1) has correlation coefficient restricted to the interval $(-1/3, 1/3)$. Various researchers ([14–16, 20]; see also, Balakrishnan and Lai [2]) have advocated methods to overcome this drawback of the FGM family.

Sarmanov [23] introduced a family of distributions with better flexibility, and this family includes the FGM family as a particular case. The Sarmanov family is defined by

$$h(x, y) = f(x)g(y)\left\{1 + \beta \phi_1(x)\phi_2(y)\right\}, \quad x, y \varepsilon R. \tag{1.2}$$

where $f(x)$ and $g(y)$ are the probability density functions (pdf) of $F(x)$ and $G(y)$, respectively. $\phi_1(x)$ and $\phi_2(y)$ are measurable functions [21] (also known as mixing functions) satisfying the conditions

$$E\left[\phi_1(X)\right] = 0, \quad E\left[\phi_2(Y)\right] = 0 \quad \text{and} \quad 1 + \phi_1(x)\phi_2(y) \geq 0.$$

If $\phi_1(x) = 1 - 2F$ and $\phi_2(y) = 1 - 2G$, then (1.2) reduces to (1.1). Lee [19] and Shubina and Lee [28] have made a detailed study of the Sarmanov family. In the literature, the family (1.2) is referred to as the Sarmanov-Lee family. There is vast improvement in the range of correlation [21]. For example, the maximum correlation for the bivariate distribution with uniform marginals is 3/4 as opposed to 1/3 for the FGM distribution (1.1). Different bivariate distributions with given marginals are constructed by choosing different mixing functions $\phi_1(x)$ and $\phi_2(y)$. Lee [19] has given some examples of the choice of $\phi_1(x)$ and $\phi_2(y)$. In this paper, we give a general method of constructing bivariate generalizations of weighted discrete distributions by considering a particular simple choice of $\phi_1(x)$ and $\phi_2(y)$.

The objective of this paper is to propose bivariate extensions of a univariate CMP distribution where marginals are CMP. Recently Sellers et al. [24] considered a bivariate CMP (BCMP) distribution where marginal distributions are not CMP.

We construct BCMP distributions with CMP marginals and range of correlation over $[-1, 1]$ by using two instances of (1.2). The first is a general method for weighted distributions. The second bivariate distribution is constructed by using $\phi_1(x)$ and $\phi_2(y)$ based on the probability generating function. The proposed BCMP distributions includes as a special case a bivariate Poisson distribution which has correlation in $[-1, 1]$. Holgate's [12] bivariate Poisson distribution constructed by random element in common attains positive correlation only.

The CMP distribution generalizes the Poisson distribution by allowing for over-dispersion ($v < 1$) or under-dispersion ($v > 1$). Its probability mass function (pmf) is given by

$$P(X = x) = \frac{\lambda^x}{(x!)^v} \frac{1}{Z(\lambda, v)}, \ x = 0, 1, 2, \ldots \tag{1.3}$$

where

$$Z(\lambda, v) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^v}, \ \lambda > 0, \ v \geq 0 \tag{1.4}$$

is the normalizing constant. The mean and variance of the CMP distribution have the following approximations [27]

$$E[X] \approx \lambda^{1/v} - \frac{v-1}{2v}, \text{var}(X) \approx \frac{1}{v}\lambda^{1/v} \tag{1.5}$$

The CMP distribution may be regarded as a weighted Poisson distribution with pmf

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \frac{(x!)^{1-v}}{W(\lambda, v)}, \ x = 0, 1, 2, \ldots \tag{1.6}$$

where $W(\lambda, v)$ is the normalizing constant.

The CMP distribution is also appealing from a theoretical point of view because it belongs to the class of two parameters power series distribution [27]. As a consequence, sufficient statistics and other elegant properties may be derived. Some of these have recently been investigated by Gupta et al. [10]. Kadane et al. [17] investigated the number of solutions which give rise to the same sufficient statistics.

The paper is organized as follows. Section 2 defines the proposed BCMP distributions and Sect. 3 discusses some dependence properties. The statistical analyses concerning parameter estimation, tests of hypotheses of independence and adequacy are presented in Sect. 4. Section 4 also contains simulation studies to study the power of Rao score and likelihood ratio tests. Section 5 illustrates an application to a real data set. Some concluding remarks are given in Sect. 6.

2 BCMP Distribution and Properties

2.1 Bivariate Discrete Distributions Based on Sarmanov–Lee Family

In this section, we present two bivariate distributions by choosing different mixing functions $\phi_1(x)$ and $\phi_2(y)$.

2.1.1 Bivariate Weighted Discrete Distributions

We propose a general method for constructing Sarmanov-type bivariate distributions for weighted distributions. Consider a discrete distribution with pmf $p(x)$ and mean μ. The weighted distribution for $p(x)$ has pmf

$$P(X = x) = p(x)\frac{w(x)}{W}, \ x = 0, 1, 2, \dots$$

where $w(x)$ is the weight and W is the normalizing constant. We make use of the simpler $p(x)$ to construct the mixing functions. Let α be a positive real number,

$$\phi_1(x) = p^\alpha(x) - E\big[p^\alpha(X)\big]$$

and

$$\phi_2(y) = p^\alpha(y) - E\big[p^\alpha(Y)\big].$$

The expectation $E\big[p^\alpha(X)\big]$ is taken with respect to the weighted pmf $P(X = x)$. A bivariate distribution with joint pmf based on (1.2) is given by

$$
\begin{aligned}
&P(X = x, Y = y) \\
&= P(X = x)P(Y = y)\big\{1 + \beta\big(p^\alpha(x) - E\big[p^\alpha(X)\big]\big)\big(p^\alpha(y) - E\big[p^\alpha(Y)\big]\big)\big\}, \ \ \beta\varepsilon[-1, 1]
\end{aligned}
\tag{2.1}
$$

2.1.2 Bivariate Discrete Distributions Based on Probability Generating Functions

Let $\phi_1(x) = \theta^x - G(\theta)$ and $\phi_2(y) = \theta^y - G(\theta)$, where $0 < \theta < 1$ and $G(\theta)$ is the probability generating function of the marginal distribution. A bivariate distribution is defined as follows with joint pmf

$$P(X = x, Y = y) = P(X = x)P(Y = y)\{1 + \beta(\theta^x - G(\theta))(\theta^y - G(\theta))\} \tag{2.2}$$

Note that the functions $\phi_i(t), i = 1, 2$ are bounded and $\sum \phi_i(t)P(t) = 0, \ t = 0, 12, \dots$
.

Let $\xi_i = \sum t\phi_i(t)P(t)$. From Theorem 2 [19], the correlation coefficient ρ of (2.1) and (2.2) is given by

$$\rho = \frac{\beta\xi_1\xi_2}{\sigma_1\sigma_2}. \tag{2.3}$$

where σ_1^2, σ_2^2 are the variances.

For model (2.1) the function ξ_i is given by

$$\xi_i = \sum x\phi_i(x)P(x) = \sum \left(xp(x)P(x) - xP(x)E[p(X)] \right) = \delta_i - \mu_i\gamma_i, i = 1, 2$$

where $\delta = E[Xp(X)]$, μ is the mean for $P(x)$ and $\gamma = E[p(X)]$. Thus,

$$\rho = \frac{\beta\left(\delta_1 - \mu_1\gamma_1\right)\left(\delta_2 - \mu_2\gamma_2\right)}{\sigma_1\sigma_2}$$

2.2 BCMP Distributions

Based on (2.1) and (2.2) we have the following BCMP distributions.

The bivariate distribution with joint pmf based on (2.1) is given by

$$P(X = x, Y = y) = P(X = x)P(Y = y)\left\{1 + \beta\left[p^\alpha(x) - E[p^\alpha(X)]\right]\left[p^\alpha(y) - E[p^\alpha(Y)]\right]\right\} \tag{2.4}$$

where the CMP marginals $P(X = x)$ and $P(Y = y)$ are given by (1.3) with parameters $\left(\lambda_1, \nu_1\right)$ and $\left(\lambda_2, \nu_2\right)$, respectively. We consider computation of $E[p^\alpha(Y)]$ by writing it in terms of a CMP pmf. For simplicity, we suppress the subscripts and write the parameters as (λ, ν).

Since $Z(\lambda, \nu) = \sum_{j=0}^\infty \frac{\lambda^j}{(j!)^\nu}$, we express

$$E[p^\alpha(X)] = \sum_{x=0}^\infty e^{-\lambda\alpha} \frac{\lambda^{x\alpha}}{(x!)^\alpha} P(x) = \frac{e^{-\lambda\alpha}Z\left(\lambda^2, \nu + 1\right)}{Z(\lambda, \nu)} \sum_{x=0}^\infty \frac{\lambda^{x(\alpha+1)}}{(x!)^{\nu+\alpha}Z\left(\lambda^2, \nu + 1\right)}$$

That is,

$$E[p^\alpha(X)] = \frac{e^{-\lambda\alpha}Z\left(\lambda^2, \nu + 1\right)}{Z(\lambda, \nu)} P(R = x)$$

where $P(R = x) = \frac{1}{Z(\lambda^2, \nu+1)} \sum_{x=0}^\infty \frac{\lambda^{x(\alpha+1)}}{(x!)^{\nu+\alpha}}$ is a CMP pmf with parameters $\left(\lambda^2, \nu + 1\right)$. Note that $E[p^\alpha(X)]$ is an infinite sum of product of Poisson and CMP probabilities and it is easy to see that $E[p^\alpha(X)] \leq 1$. Hence, $\phi_1(x)$ and $\phi_2(y)$ are bounded.

Let $\phi_1(x) = \theta^x - g\left(\theta; \lambda_1, \nu_1\right)$ and $\phi_2(y) = \theta^y - g\left(\theta; \lambda_2, \nu_2\right)$, where $0 < \theta < 1$ and $g(\theta; \lambda, \nu)$ is the probability generating function of the CMP distribution given by

$$g(\theta; \lambda, \nu) = \frac{Z(\lambda\theta, \nu)}{Z(\lambda, \nu)}.$$

The BCMP distribution corresponding to (2.2) has joint pmf defined as follows:

$$P(X = x, Y = y) = P(X = x)P(Y = y)\left\{1 + \beta\phi_1(x)\phi_2(y)\right\}, \tag{2.5}$$

where the CMP marginals $P(X = x)$ and $P(Y = y)$ are given by (1.3) with parameters (λ_1, ν_1) and (λ_2, ν_2), respectively.

If $\nu_1 = \nu_2 = 1$, then (2.5) is the joint pmf of the bivariate Poisson distribution given in Section 6.3 of Lee [19] with $\theta = e^{-1}$.

2.3 Correlation Coefficient

Let μ and σ^2 denote the mean and variance of the CMP distribution. For BCMP distributions (2.4) and (2.5), the variances σ_1^2, σ_2^2 may be approximated by (1.5).

For BCMP distribution (2.4),

$$\xi_i = \sum x\phi_i(x)P(x) = \sum \left(x \frac{e^{-\lambda_i}\lambda_i^x}{x!} - xP(x) \right) = \lambda_i - \mu_i, i = 1, 2$$

Thus (2.4) has a very simple expression for the correlation coefficient.

$$\rho = \frac{\beta(\lambda_1 - \mu_1)(\lambda_2 - \mu_2)}{\sigma_1 \sigma_2}$$

For (2.5),

$$\xi_i = \theta \frac{\partial g_i(\theta)}{\partial \theta} - \mu_i g_i(\theta), i = 1, 2$$

where $g_i(\theta) = g(\theta; \lambda_i, \nu_i), i = 1, 2$. The correlation coefficient is given by

$$\rho = \frac{\beta\left(\theta \frac{\partial g_1(\theta)}{\partial \theta} - \mu_1 g_1(\theta)\right)\left(\theta \frac{\partial g_2(\theta)}{\partial \theta} - \mu_2 g_2(\theta)\right)}{\sigma_1 \sigma_2}$$

For the computation of $Z(\theta, \nu)$ given by (1.4) and related quantities such as derivatives see Section 4 of Gupta et al. [10].

3 Dependence Properties of the BCMP Distribution

The most common measures for determining the relationship between two variables are the Pearson correlation coefficient, Kendall's tau and Spearman's rho. As a generalization of Pearson's correlation coefficient, Bjerve and Doksum [3], Doksum et al. [7] and Blyth [4–6] introduced and discussed correlation curve. The correlation curve is a local measure of the strength of association between the two variables X and Y. The correlation curve $\rho(x)$, a function of x, describes the amount of variance explained by a regression curve varies locally. However, $\rho(x)$ does not treat X and Y on equal footing, but needs Y to be a response and X, a predictor variable. The correlation curve is a regression concept.

A local dependence function, a function of x and y, should measure the strength and direction of association locally treating both variables symmetrically. For a bivariate distribution, it is defined as follows:

An $r \times c$ contingency table with cell probabilities $p_{i,j}$ specify the joint distribution for two discrete random variables X and Y as

$$p_{i,j} = P(X = i; Y = j); 1 \leq i \leq r; 1 \leq j \leq c \tag{3.1}$$

The two marginal distributions for X and Y are

$$P(X = i) = \sum_{j=1}^{c} p_{i,j} = p_{i+} \text{ and } P(Y = j) = \sum_{i=1}^{r} p_{i,j} = p_{+j}$$

respectively. Yule and Kendall [30] and Goodman [9] suggested the following set of local cross product ratios

$$\alpha_{i,j} = \frac{p_{i,j} p_{i+1,j+1}}{p_{i,j+1} p_{i+1,j}}, 1 \leq i \leq r, 1 \leq j \leq c \tag{3.2}$$

Equation (3.2) defines the local dependence function. Also, let $\gamma_{i,j} = ln\alpha_{i,j}$. Both $\alpha_{i,j}$ and $\gamma_{i,j}$ measure the association of the 2×2 tables found by adjacent rows and adjacent columns. It is known that the set $\{\alpha_{i,j}\}$ or equivalently $\{\gamma_{i,j}\}$ together with marginal probability distributions uniquely determine the bivariate distribution. For more explanation, see Wang [29] and Holland and Wang [13].

We now present a very important property of the local dependence function. It is in terms of the totally positive of order 2, TP$_2$ (reverse regular of order 2, RR$_2$), property defined below.

Definition A discrete bivariate distribution $P(X = i; Y = j)$ is said to be TP$_2$(RR$_2$) if for $a_1 < b_1$, $a_2 < b_2$

$$\begin{vmatrix} p(a_1, b_1) & p(a_1, b_2) \\ p(a_2, b_1) & p(a_2, b_2) \end{vmatrix} \geq (\leq)0 \tag{3.3}$$

where $p(a_i, b_j) = P(X = i; Y = j)$. It can be easily verified that the TP$_2$(RR$_2$) property is equivalent to $\alpha_{i,j} \geq (\leq)1$; where $\alpha_{i,j}$ is the local dependence function. This is also equivalent to $\gamma_{i,j} \geq (\leq)0$.

We now obtain the local dependence function for the Sarmanov family of discrete distributions.

3.1 Local Dependence Function for Sarmanov Family of Discrete Distributions

For this family

$$\alpha_{ij} = \frac{[1 + \omega\phi_1(x)\phi_2(y)][1 + \omega\phi_1(x+1)\phi_2(y+1)]}{[1 + \omega\phi_1(x)\phi_2(y+1)][1 + \omega\phi_1(x+1)\phi_2(y)]} \tag{3.4}$$

So $\alpha_{i,j} \geq (\leq)1$ is equivalent to

$$\omega\left[\phi_2(y+1) - \phi_2(y)\right]\left[\phi_1(x+1) - \phi_1(x)\right] \geq (\leq)0 \tag{3.5}$$

Thus, Sarmanov family is TP$_2$ if

(a) $\phi_1(x)$ and $\phi_2(y)$ are both increasing or both decreasing and
(b) $\omega > 0$.

Similarly, the Sarmanov family is RR$_2$ if

(a) $\phi_1(x)$ and $\phi_2 y$ are both increasing or both decreasing and
(b) $\omega < 0$.

Let us now obtain such conditions for the BCMP distribution.
We have

$$\phi_1(x) = \theta^x - g_1(\theta), 0 < \theta < 1.$$

This gives

$$\begin{aligned}\phi_1(x+1) - \phi_1(x) &= \left[\theta^{x+1} - g_1(\theta)\right]\left[\theta^x - g_1(\theta)\right]\\ &= \theta^x(\theta - 1) < 0, 0 < \theta < 1\end{aligned} \tag{3.6}$$

Hence, $\phi_1(x)$ is decreasing. Similarly, $\phi_2(y)$ is decreasing.

Hence, BCMP is TP$_2$ if and only if $\omega > 0$.

Note that the TP$_2$ condition is the same as the condition for positive dependence notion that Lehmann [18] called positive likelihood ratio dependence. This notion leads naturally to the order described below.

3.2 Positively Likelihood Ratio Dependent Ordering

Let (X_1, X_2) and (Y_1, Y_2) be two bivariate random vectors having the same marginals. Then, we say that (X_1, X_2) is smaller than (Y_1, Y_2) in the positively likelihood ratio dependent (PLRD) order denoted by

$$\left(X_1, X_2\right) \leq_{\text{PLRD}} \left(Y_1, Y_2\right)$$

if

$$f\left(x_1, y_1\right)f\left(x_2, y_2\right)g\left(x_1, y_2\right)g\left(x_2, y_1\right) \leq f\left(x_1, y_2\right)f\left(x_2, y_1\right)g\left(x_1, y_1\right)g\left(x_2, y_2\right), x_1 \leq x_2, y_1 \leq y_2$$

where F and G have (continuous or discrete) densities f and g; see [26].

It can be easily seen that

$$\left(X_1, X_2\right) \leq_{PLRD} \left(Y_1, Y_2\right)$$

is equivalent to

$$\alpha_{(X_1,X_2)} \le \alpha_{(Y_1,Y_2)}$$

where $\alpha_{(X_1,X_2)}$ and $\alpha_{(Y_1,Y_2)}$ are the local dependence functions for (X_1, X_2) and (Y_1, Y_2), respectively.

We shall now compare two Sarmanov families with parameters w_1 and w_2:

After tedious algebra, it can be verified that

$$\alpha_{(X_1,X_2)} - \alpha_{(Y_1,Y_2)}$$
$$= (\omega_1 - \omega_2)\left[\phi_1(x+1) - \phi_1(x)\right]\left[\phi_1(y+1) - \phi_1(y)\right] \times \left[1 - \omega_1\omega_2\phi_1(x)\phi_2(x)\phi_1(x+1)\phi_2(y+1)\right]$$

Now assume

(a) $\phi_1(x)$ and $\phi_2(y)$ are both increasing or both decreasing and
(b) $\omega_1\omega_2\phi_1(x)\phi_2(x)\phi_1(x+1)\phi_2(y+1) \le 1$

Then, $\alpha_{(X_1,X_2)} \le \alpha_{(Y_1,Y_2)}$ if $\omega_1 < \omega_2$.

The alternative conditions are

(c) $\phi_1(x)$ and $\phi_2(y)$ are both increasing or both decreasing and
(d) $\omega_1\omega_2\phi_1(x)\phi_2(x)\phi_1(x+1)\phi_2(y+1) \ge 1$

Then, $\alpha_{(X_1,X_2)} \le \alpha_{(Y_1,Y_2)}$ if $\omega_1 > \omega_2$.

We now investigate PLRD ordering for BCMP distribution. In this case,

$$\omega_1\phi_1(x)\phi_2(y) \ge -1 \quad \text{and} \quad \omega_2\phi_1(x+1)\phi_2(y+1) \ge -1.$$

This does not imply that $\omega_1\omega_2\phi_1(x)\phi_2(x)\phi_1(x+1)\phi_2(y+1) \ge 1$. Hence, we cannot compare the two vectors of Sarmanov family according to the PLRD ordering.

Remark. The PLRD ordering is difficult to check in most cases. By contrast finding families of distributions that are not ordered by this relation is relatively easy. For more explanation and comments, see [8].

4 Statistical Analysis

In this section, we examine statistical inference for the BCMP distribution given by (2.2).

4.1 Parameter Estimation

The method of moments estimation for estimating the parameters is conducted as follows:

The marginal parameters (λ, v) are estimated by equating the first and second marginal sample moments by using the approximations in (1.5). The estimate for β is obtained from (2.3) by equating with the sample correlation coefficient.

For maximum likelihood estimation (MLE), simulated annealing (SA) algorithm is used to determine the estimates corresponding to the global optimum. It is a popular algorithm for searching global extremum in non-smooth functions with a large number of local extrema. Henderson et al. [11] have discussed the convergence of SA and presented practical guidelines for the implementation of SA algorithm, especially its control parameters, to ensure good performance.

The log-likelihood function of the model is given by

$$\ln L = \sum_{i=1}^{n} \left[x_i \ln \lambda_1 + y_i \ln \lambda_2 + \ln \left\{ 1 + \beta \phi_1(x_i) \phi_2(y_i) \right\} - v_1 (\ln x_i!) - v_2 (\ln y_i!) - \ln Z_1 - \ln Z_2 \right]$$

(4.1)

where $Z_1 = Z(\lambda_1, v_1), Z_2 = Z(\lambda_2, v_2)$. For the following sections, we consider $\theta = e^{-1}$.

4.2 Test of Hypotheses

In this subsection, the tests for the two hypotheses of interest, that is, test of independence and test of adequacy of the proposed BCMP distribution are discussed. To compare the null model (restricted model) against the alternative model (unrestricted model), the score and the likelihood ratio (LR) tests are chosen and their test statistics are summarized as follows.

Let $H_0: \theta = \theta^*$ versus $H_1: \theta \neq \theta^*$. The score test statistic is $T = S'V^{-1}S$ where

$$S' = \left. \frac{\partial}{\partial \theta} \ln L(\theta) \right|_{\theta = \hat{\theta}^*}$$

is the score function and

$$V = E\left[-\frac{\partial^2 \ln L(\theta)}{\partial \theta} \right]\Bigg|_{\theta = \hat{\theta}^*}$$

is the information matrix.

The test statistic based on the LR test is defined as $\mathrm{LR} = -2 \log L(\hat{\theta}^*)/L(\hat{\theta})$ where $\hat{\theta}^*$ and $\hat{\theta}$ are the restricted and unrestricted maximum likelihood estimates.

4.2.1 Test of Independence

The random variables X and Y are independent if $\beta = 0$. The hypotheses to be tested are $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$.

The score functions evaluated under the null hypothesis are needed and they can be easily obtained from the log-likelihood equation in (4.1).

$$\left.\frac{\partial \ln L}{\partial \beta}\right|_{\beta=0} = \sum_{i=1}^{n} \phi_1(x_i)\phi_2(y_i)$$

$$\left.\frac{\partial \ln L}{\partial \lambda_1}\right|_{\beta=0} = \sum_{i=1}^{n} \frac{x_i}{\lambda_1} - \frac{1}{Z_1}\left(\frac{\partial Z_1}{\partial \lambda_1}\right), \quad \left.\frac{\partial \ln L}{\partial \lambda_2}\right|_{\beta=0} = \sum_{i=1}^{n} \frac{y_i}{\lambda_2} - \frac{1}{Z_2}\left(\frac{\partial Z_2}{\partial \lambda_2}\right)$$

$$\left.\frac{\partial \ln L}{\partial v_1}\right|_{\beta=0} = \sum_{i=1}^{n} -\ln(x_i!) - \frac{1}{Z_1}\left(\frac{\partial Z_1}{\partial v_1}\right), \quad \left.\frac{\partial \ln L}{\partial v_2}\right|_{\beta=0} = \sum_{i=1}^{n} -\ln(y_i!) - \frac{1}{Z_2}\left(\frac{\partial Z_2}{\partial v_2}\right)$$

The elements of the information matrix corresponding to $\beta = 0$ are

$$\left.\frac{\partial^2 \ln L}{\partial \beta^2}\right|_{\beta=0} = \sum_{i=1}^{n} -(\phi_1(x_i)\phi_2(y_i))^2$$

$$\left.\frac{\partial^2 \ln L}{\partial \beta \partial \lambda_1}\right|_{\beta=0} = \sum_{i=1}^{n} \phi_2(y_i)\left(\frac{\partial \phi_1(x_i)}{\partial \lambda_1}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \beta \partial \lambda_2}\right|_{\beta=0} = \sum_{i=1}^{n} \phi_1(x_i)\left(\frac{\partial \phi_2(y_i)}{\partial \lambda_2}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \beta \partial v_1}\right|_{\beta=0} = \sum_{i=1}^{n} \phi_2(y_i)\left(\frac{\partial \phi_1(x_i)}{\partial v_1}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \beta \partial v_2}\right|_{\beta=0} = \sum_{i=1}^{n} \phi_1(x_i)\left(\frac{\partial \phi_2(y_i)}{\partial v_2}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_1^2}\right|_{\beta=0} = \sum_{i=1}^{n} -\frac{x_i}{\lambda_1^2} + \left(\frac{1}{Z_1}\frac{\partial Z_1}{\partial \lambda_1}\right)^2 - \frac{1}{Z_1}\left(\frac{\partial^2 Z_1}{\partial \lambda_1^2}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_2^2}\right|_{\beta=0} = \sum_{i=1}^{n} -\frac{y_i}{\lambda_2^2} + \left(\frac{1}{Z_2}\frac{\partial Z_2}{\partial \lambda_2}\right)^2 - \frac{1}{Z_2}\left(\frac{\partial^2 Z_2}{\partial \lambda_2^2}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial v_j^2}\right|_{\beta=0} = \sum_{i=1}^{n} \frac{1}{Z_j^2}\left[\left(\frac{\partial Z_j}{\partial v_j}\right)^2 - Z_j\frac{\partial^2 Z_j}{\partial v_j^2}\right], j = 1, 2$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_j \partial v_j}\right|_{\beta=0} = \sum_{i=1}^{n} \frac{1}{Z_j^2}\left(\frac{\partial Z_j}{\partial \lambda_j}\frac{\partial Z_j}{\partial v_j} - Z_j\frac{\partial^2 Z_j}{\partial \lambda_j \partial v_j}\right), j = 1, 2$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_1 \partial \lambda_2}\right|_{\beta=0} = \left.\frac{\partial^2 \ln L}{\partial \lambda_1 \partial v_2}\right|_{\beta=0} = \left.\frac{\partial^2 \ln L}{\partial \lambda_2 \partial v_1}\right|_{\beta=0} = \left.\frac{\partial^2 \ln L}{\partial v_1 \partial v_2}\right|_{\beta=0} = 0$$

Both the score and LR tests have an approximate Chi-square distribution with 1 degree of freedom.

4.2.2 Test of BCMP

To test the bivariate Poisson distribution against a BCMP distribution, that is, if the bivariate Poisson is adequate, the proposed hypotheses are

$$H_0 : v_1 = v_2 = 1 \text{ versus } H_0 \text{ is not true.}$$

From Eqs. (2.5) and (4.1), the score functions are found to be

$$\frac{\partial \ln L}{\partial \beta}\bigg|_{v_1=v_2=1} = \sum_{i=1}^{n} \frac{a_1 a_2}{f}$$

$$\frac{\partial \ln L}{\partial \lambda_1}\bigg|_{v_1=v_2=1} = \sum_{i=1}^{n} -1 + \frac{x_i}{\lambda_1} + \frac{\beta a_1 e^{-\lambda_1 + \lambda_1 \theta}(1-\theta)}{f}$$

$$\frac{\partial \ln L}{\partial \lambda_2}\bigg|_{v_1=v_2=1} = \sum_{i=1}^{n} -1 + \frac{y_i}{\lambda_2} + \frac{\beta a_2 e^{-\lambda_2 + \lambda_2 \theta}(1-\theta)}{f}$$

$$\frac{\partial \ln L}{\partial v_1}\bigg|_{v_1=v_2=1} = \sum_{i=1}^{n} -\ln(x_i!) - e^{-\lambda_1}\left(\frac{\partial Z_1}{\partial v_1}\right) + \frac{\beta a_1}{f}\left[e^{-2\lambda_1 + \lambda_1 \theta}\frac{\partial Z_1}{\partial v_1} - e^{-\lambda_1}\frac{\partial Z_1^*}{\partial v_1}\right]$$

$$\frac{\partial \ln L}{\partial v_2}\bigg|_{v_1=v_2=1} = \sum_{i=1}^{n} -\ln(y_i!) - e^{-\lambda_2}\frac{\partial Z_2}{\partial v_2} + \frac{\beta a_2}{f}\left[e^{-2\lambda_2 + \lambda_2 \theta}\frac{\partial Z_2}{\partial v_2} - e^{-\lambda_2}\frac{\partial Z_2^*}{\partial v_2}\right]$$

Table 1 Simulated power of Score and LR tests for test of independence ($\lambda_1 = 0.9, v_1 = 2, \lambda_2 = 0.7, v_2 = 3$)

		$n=100$		$n=500$		$n=1000$	
β	α	T	LR	T	LR	T	LR
-1	0.05	0.174	0.114	0.678	0.678	0.926	0.955
	0.10	0.264	0.226	0.768	0.765	0.969	0.975
-0.8	0.05	0.134	0.089	0.511	0.510	0.786	0.787
	0.10	0.209	0.184	0.618	0.618	0.860	0.859
-0.5	0.05	0.088	0.052	0.224	0.206	0.415	0.415
	0.10	0.146	0.123	0.340	0.322	0.539	0.540
0	0.05	0.053	0.035	0.051	0.052	0.056	0.057
	0.10	0.110	0.102	0.109	0.109	0.112	0.112
0.5	0.05	0.100	0.068	0.276	0.278	0.455	0.455
	0.10	0.169	0.160	0.384	0.383	0.569	0.57
0.8	0.05	0.163	0.115	0.548	0.547	0.826	0.826
	0.10	0.260	0.231	0.668	0.669	0.897	0.898
1.0	0.05	0.229	0.144	0.737	0.737	0.952	0.955
	0.10	0.337	0.298	0.835	0.835	0.974	0.975

where $a_1 = \left(-e^{-\lambda_2 + \lambda_2\theta} + \theta^{y_i}\right)$, $a_2 = \left(-e^{-\lambda_1 + \lambda_1\theta} + \theta^{x_i}\right)$, $f = 1 + \beta\left(-e^{-\lambda_1 + \lambda_1\theta} + \theta^{x_i}\right)$ $\left(-e^{-\lambda_2 + \lambda_2\theta} + \theta^{y_i}\right)$, $Z_1^* = Z(\theta\lambda_1, v_1)$ and $Z_2^* = Z(\theta\lambda_2, v_2)$.

The elements of information matrix evaluated under the null hypothesis are

$$\left.\frac{\partial^2 \ln L}{\partial \beta^2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} -\left(\frac{a_1 a_2}{f}\right)^2$$

$$\left.\frac{\partial^2 \ln L}{\partial \beta \partial \lambda_j}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} -\frac{b_j}{g^2}\left(e^{\lambda_1 + \lambda_2 + \lambda_j \theta}\right)(\theta - 1), \; j = 1, 2$$

$$\left.\frac{\partial^2 \ln L}{\partial \beta \partial v_1}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} \frac{(e^{\lambda_2} b_1)}{g^2}\left(e^{\lambda_1 \theta}\frac{\partial Z_1}{\partial v_1} - e^{\lambda_1}\frac{\partial Z_1^*}{\partial v_1}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \beta \partial v_2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} \frac{(e^{\lambda_1} b_2)}{g^2}\left(e^{\lambda_2 \theta}\frac{\partial Z_2}{\partial v_2} - e^{\lambda_2}\frac{\partial Z_2^*}{\partial v_2}\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_1^2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} -\frac{x_i}{\lambda_1^2} - \frac{\beta b_1 e^{\lambda_1 \theta}}{g}(\theta - 1)^2\left(\frac{\beta b_1 e^{\lambda_1 \theta}}{g} + 1\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_1 \partial \lambda_2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} \left(\frac{\theta - 1}{g}\right)^2 \beta e^{(\lambda_1 + \lambda_2)(1 + \theta)}$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_2^2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} -\frac{y_i}{\lambda_2^2} - \frac{\beta b_2 e^{\lambda_2 \theta}(\theta - 1)^2}{g}\left(\frac{\beta b_2 e^{\lambda_2 \theta}}{g} + 1\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_j \partial v_j}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} \begin{array}{l} e^{-\lambda_j}\left(\frac{\partial Z_j}{\partial v_j} - \frac{\partial^2 Z_j}{\partial \lambda_j \partial v_j}\right) + \left(\frac{\beta b_j}{g}\right)^2\left(e^{\lambda_j(\theta-1)}(\theta - 1)\left(e^{\lambda_j \theta}\frac{\partial Z_j}{\partial v_j} - e^{\lambda_j}\frac{\partial Z_j^*}{\partial v_j}\right)\right) \\ + \frac{1}{g}\left[\beta b_j e^{-\lambda_j}\left(e^{\lambda_j \theta}\frac{\partial^2 Z_j}{\partial \lambda_j \partial v_j} + e^{\lambda_j \theta}(\theta - 2)\frac{\partial Z_j}{\partial v_j} + e^{\lambda_j}\left(\frac{\partial Z_j^*}{\partial v_j} - \frac{\partial^2 Z_j^*}{\partial \lambda_j \partial v_j}\right)\right)\right] \end{array}, \; j = 1, 2$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_1 \partial v_2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} \frac{1}{g^2}\left(\beta e^{\lambda_1 + \lambda_1 \theta}(\theta - 1)\left(-e^{\lambda_2 \theta}\frac{\partial Z_2}{\partial v_2} + e^{\lambda_2}\frac{\partial Z_2^*}{\partial v_2}\right)\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial \lambda_2 \partial v_1}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} \frac{1}{g^2}\left(\beta e^{\lambda_2 + \lambda_2 \theta}(\theta - 1)\left(-e^{\lambda_1 \theta}\frac{\partial Z_1}{\partial v_1} + e^{\lambda_1}\frac{\partial Z_1^*}{\partial v_1}\right)\right)$$

$$\left.\frac{\partial^2 \ln L}{\partial v_j^2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} e^{-4\lambda_j} \left\{ \begin{array}{l} e^{2\lambda_j}\left(\frac{\partial Z_j}{\partial v_j}\right)^2 - e^{3\lambda_j}\frac{\partial^2 Z_j}{\partial v_j^2} - \left(\frac{\beta a_j}{f}\left(e^{\lambda_j \theta}\frac{\partial Z_j}{\partial v_j} - e^{\lambda_j}\frac{\partial Z_j^*}{\partial v_j}\right)\right)^2 \\ + \frac{\beta a_j e^{\lambda_j}}{f} \begin{array}{l} -2e^{\lambda_j \theta}\left(\frac{\partial Z_j}{\partial v_j}\right)^2 + e^{\lambda_j + \lambda_j \theta}\frac{\partial^2 Z_j}{\partial v_j^2} + \\ 2e^{\lambda_j}\left(\frac{\partial Z_j}{\partial v_j}\frac{\partial Z_j^*}{\partial v_j}\right) - e^{2\lambda_j}\frac{\partial^2 Z_j^*}{\partial v_j^2} \end{array} \end{array} \right\}, \; j = 1, 2$$

$$\left.\frac{\partial^2 \ln L}{\partial v_1 \partial v_2}\right|_{v_1 = v_2 = 1} = \sum_{i=1}^{n} \frac{\beta}{g^2}\left(e^{\lambda_1 \theta}\frac{\partial Z_1}{\partial v_1} - e^{\lambda_1}\frac{\partial Z_1^*}{\partial v_1}\right)\left(e^{\lambda_2 \theta}\frac{\partial Z_2}{\partial v_2} - e^{\lambda_2}\frac{\partial Z_2^*}{\partial v_2}\right)$$

where $b_1 = e^{\lambda_2}\theta^{y_i} - e^{\lambda_2\theta}$, $b_2 = e^{\lambda_1}\theta^{x_i} - e^{\lambda_1\theta}$ and $g = \beta e^{(\lambda_1 + \lambda_2)\theta} - \beta e^{\lambda_1 + \lambda_2\theta}\theta^{x_i} - \beta e^{\lambda_2 + \lambda_1\theta}\theta^{y_i} + e^{\lambda_1 + \lambda_2}\left(1 + \beta\theta^{x_i + y_i}\right)$.

Both the score and LR tests have an approximate Chi-square distribution with 2 degrees of freedom.

Table 2 Simulated power of Score and LR tests for test of adequacy of BCMP $(\lambda_1 = 0.9, v_1 = 2, \lambda_2 = 0.7, v_2 = 3)$

			$n=100$		$n=500$		$n=1000$	
v_1	v_2	α	T	LR	T	α	T	LR
1.0	1.0	0.05	0.056	0.129	0.047	0.060	0.049	0.047
		0.10	0.109	0.172	0.100	0.106	0.093	0.088
0.5	0.5	0.05	0.760	0.744	1.000	1.000	1.000	1.000
		0.10	0.829	0.814	1.000	1.000	1.000	1.000
2.0	2.0	0.05	0.660	0.739	1.000	1.000	1.000	1.000
		0.10	0.778	0.837	1.000	1.000	1.000	1.000
1.0	2.0	0.05	0.279	0.404	0.968	0.976	1.000	1.000
		0.10	0.431	0.526	0.982	0.986	1.000	1.000
1.0	0.5	0.05	0.404	0.374	0.963	0.958	0.999	0.999
		0.10	0.516	0.481	0.982	0.976	0.999	1.000
2.0	1.0	0.05	0.378	0.484	0.990	0.992	1.000	1.000
		0.10	0.539	0.622	0.996	0.997	1.000	1.000
0.5	1.0	0.05	0.594	0.577	0.997	0.996	1.000	1.000
		0.10	0.701	0.687	0.998	0.998	1.000	1.000

4.3 Monte Carlo Simulation Study

The performance of the proposed score and LR tests is compared in order to check the efficiencies of the tests. A simulation study of 1000 replications has been carried out by considering small ($n=100$), medium ($n=500$) and large ($n=1000$) sample sizes for the bivariate discrete distributions based on probability generating functions. The nominal significant level, α is taken as 5% and 10%. For the test of independence, different values of β, ranging from -1 to 1 are considered with a few combinations of v_1 and v_2. Tables 1 and 2 display, respectively, the simulated power of the tests under the test of independence and the test of adequacy of BCMP distribution.

From Table 1, the score test performs better in maintaining the nominal significance level of 5% ($\beta=0$) if compared to the LR test for small sample sizes, but vice versa for nominal level of 10%. A weak detection is achieved when β is 0.5 away from zero and when the sample size is small. The power of detection is greatly improved when the sample size increases. When β is 0.8 or 1.0 away from zero, the score test outperforms the LR test when the sample size is small, but the values are very close to each other when the sample size increases.

The powers of the proposed tests increase when β diverges further from the value 0 and when the sample size increases. Besides that the powers of both of the tests are very close to each other when the sample size increases to 500 and 1000 regardless of the value of β and the nominal levels.

As shown in Table 2, the score test outperforms the LR test in maintaining the nominal levels of 5% and 10% regardless of the sample size. The LR test overestimates the nominal significance level for small sample sizes. In addition, when the parameters v_1 and v_2 are smaller than 1.0, the score and LR tests are comparable to

each other. However, if both of the parameters are higher than 1.0, LR test is able to achieve a higher detection compared to the score test. If v_1 is fixed as 1.0, the LR test outperforms the score test when v_2 is larger than 1.0 but vice versa when v_2 is smaller than 1.0. The same result applies when v_2 is fixed as 1.0 and v_1 is set as larger or smaller than 1.0.

Overall, both score and LR tests are powerful when the sample size is equal or larger than 500 as almost 96% detection can be achieved.

5 Example of Application to Real Data

As an illustration of application we consider the data set of the number of accidents sustained by 122 experienced shunters over 2 successive periods of time [1] which was also used by Sellers et al. [24].

In this section, all the summation series involved are computed by recursion with double-precision accuracy and truncation approach is applied to the normalizing constant, $Z(\lambda, v)$. It is set as $Z(\lambda, v) \leq 1 \times 10^{200}$. To compute the correlation coefficient $\rho = \frac{\alpha \xi_1 \xi_2}{\sigma_1 \sigma_2}$, the calculation of mean μ_1, μ_2 and variances σ_1^2, σ_2^2 from marginals $P(X = x)$ and $P(Y = y)$ are required. They are computed by using the following equations as the accuracy of approximation Eqs. (1.5) hold when $\lambda > 10^v$. The equations fail when the value of the parameter v is close to zero. For example, when $v = 0.01$, we have to make sure that $\lambda > 10^v = 1.0233$.

$$\mu_i = \lambda_i \frac{\partial \ln Z(\lambda_i, v_i)}{\partial \lambda_i} = \lambda_i \sum_j \frac{j \lambda_i^{j-1}}{(j!)^v} \Big/ \sum_j \frac{\lambda_i^j}{(j!)^v}, \quad i = 1, 2$$

$$\sigma_i^2 = \lambda_i \frac{\partial \mu_i}{\partial \lambda_i} = \mu_i + \lambda_i^2 \left[-\left(\sum_j \frac{j \lambda_i^{j-1}}{(j!)^v} \right)^2 + \left(\sum_j \frac{(j-1) j \lambda_i^{j-2}}{(j!)^v} \right) \left(\sum_j \frac{\lambda_i^j}{(j!)^v} \right) \right] \left(\sum_j \frac{\lambda_i^j}{(j!)^v} \right)^{-2}, \quad i = 1, 2$$

For the data set on the number of accidents sustained by 122 experienced shunters over 2 successive periods of time. [1], the sample moments are:

<div style="text-align:center">

Sample mean: $\hat{\mu}_1 = 0.9754$, $\hat{\mu}_2 = 1.2705$

Variance: $\hat{\sigma}_1^2 = 1.2969$, $\hat{\sigma}_2^2 = 1.6535$.

Sample correlation, $\hat{\rho} = 0.2585$.

</div>

This data set has also been fitted by the bivariate negative binomial (BNB) distribution as a comparison. See Table 3.

To confirm that the MLE's for the data in Table 3 really give a maximum, some likelihood function values are computed for points around the ML estimates. Based on the likelihood function values, the ML estimates do correspond to the maximum (Table 4).

It is observed that in the summary statistics presented in Table 5, the fits by both proposed BCMP are significantly better than the BNB based upon the χ^2 values and the log-likelihood values. For this data set, Sellers et al. [24] gave a log-likelihood value of -341.704 for their BCMP distribution which is about the same as BNB.

Table 3 Observed and expected number of accidents sustained by 122 experienced shunters over 2 successive periods of time

	0	1	2	3	4	5	6	Total
0	21	18	8	2	1	0	0	50
BNB	21.90	16.67	7.98	3.07	1.04	0.32	0.09	51.07
BCMPI	18.08	16.60	10.02	4.82	1.99	0.73	0.24	52.47
BCMPII	21.09	15.44	8.63	4.05	1.65	0.59	0.19	51.65
1	13	14	10	1	4	1	0	43
BNB	12.52	13.18	8.06	3.77	1.50	0.53	0.18	39.74
BCMPI	12.98	11.95	7.38	3.62	1.51	0.56	0.18	38.18
BCMPII	11.56	12.43	8.12	4.05	1.69	0.61	0.20	38.66
2	4	5	4	2	1	0	1	17
BNB	4.50	6.06	4.54	2.52	1.16	0.47	0.17	19.42
BCMPI	6.36	5.90	3.82	1.95	0.83	0.31	0.10	19.28
BCMPII	5.08	6.48	4.43	2.25	0.94	0.34	0.11	19.63
3	2	1	3	2	0	1	0	9
BNB	1.30	2.13	1.90	1.23	0.65	0.30	0.12	7.63
BCMPI	2.54	2.37	1.59	0.83	0.36	0.14	0.05	7.87
BCMPII	1.94	2.65	1.84	0.94	0.39	0.14	0.05	7.95
4	0	0	1	1	0	0	0	2
BNB	0.33	0.64	0.66	0.49	0.29	0.15	0.07	2.63
BCMPI	0.88	0.82	0.56	0.30	0.13	0.05	0.02	2.76
BCMPII	0.65	0.91	0.64	0.33	0.14	0.05	0.02	2.74
5	0	0	0	0	0	0	0	0
BNB	0.08	0.17	0.20	0.17	0.11	0.06	0.03	0.82
BCMPI	0.27	0.26	0.18	0.09	0.04	0.02	0.01	0.86
BCMPII	0.20	0.28	0.20	0.10	0.04	0.02	0.01	0.83
6	0	0	0	0	0	0	0	0
BNB	0.02	0.04	0.06	0.05	0.04	0.02	0.01	0.24
BCMPI	0.08	0.07	0.05	0.03	0.01	0.00	0.00	0.24
BCMPII	0.05	0.08	0.05	0.03	0.01	0.00	0.00	0.23
7	0	1	0	0	0	0	0	1
BNB	0.00	0.01	0.01	0.01	0.01	0.01	0.38	0.43
BCMPI	0.02	0.02	0.01	0.01	0.00	0.00	0.27	0.34
BCMPII	0.01	0.02	0.01	0.01	0.00	0.00	0.25	0.31

BCMPI is BCMP distribution based on weighted distribution; BCMPII is BCMP distribution based on probability generating function

ML estimates:

BNB: $\hat{p}_1 = 0.20, \hat{p}_2 = 0.15, \hat{p}_3 = 0.01, \hat{v} = 3.88$

BCMPI: $\hat{\lambda}_1 = 0.92, \hat{v}_1 = 0.57, \hat{\lambda}_2 = 0.73, \hat{v}_2 = 0.53, \hat{\alpha} = 0.58, \hat{\beta} = 1.00$

BCMPII: $\hat{\lambda}_1 = 0.94, \hat{v}_1 = 0.59, \hat{\lambda}_2 = 0.75, \hat{v}_2 = 0.56, \hat{\beta} = 1.00$

Table 4 Likelihood function values for points around the ML estimates

Parameters (BCMPI) $(\hat{\lambda}_1, \hat{v}_1, \hat{\lambda}_2, \hat{v}_2, \hat{\alpha}, \hat{\beta})$	Likelihood values	Parameters (BCMPII) $(\hat{\lambda}_1, \hat{v}_1, \hat{\lambda}_2, \hat{v}_2, \hat{\beta})$	Likelihood values
(0.92, 0.57, 0.73, 0.53, 0.58, 1.00) (ML estimates)	− 345.5534	(0.94, 0.59, 0.75, 0.56, 1.00) (ML estimates)	− 343.5029
(0.90, 0.50, 0.70, 0.50, 0.50, 0.90)	− 345.8766	(0.90, 0.50, 0.70, 0.50, 1.00)	− 343.8142
(0.90, 0.50, 0.70, 0.50, 0.50, 0.80)	− 345.9731	(0.90, 0.50, 0.70, 0.50, 0.80)	− 344.0005
(0.85, 0.45, 0.60, 0.40, 0.45, 0.85)	− 347.0905	(0.80, 0.40, 0.60, 0.40, 0.80)	− 345.6643
(1.00, 1.45, 1.20, 1.40, 0.79, 0.85)	− 375.7474	(1.00, 0.60, 0.80, 0.60, 1.00)	− 343.8895
(1.00, 1.45, 1.20, 1.40, 0.79, 1.00)	− 375.6931	(1.20, 0.80, 0.90, 1.00, 1.00)	− 346.8308
(1.00, 1.45, 1.20, 1.40, 0.79, 1.00)	− 382.6309	(2.20, 0.80, 2.90, 1.00, 1.00)	− 495.3810

Table 5 Summary statistics

Model	χ^2	Log-likelihood value	P value
BNB	121.70	− 341.61	0.000
BCMPI	86.99	− 345.55	0.001
BCMPII	80.49	− 343.50	0.004

Thus, based on the log-likelihood values, the proposed BCMP distributions fit better than the BCMP distribution of Sellers et al. [24].

6 Concluding Remarks

BCMP distributions with marginal distributions which are CMP distributions and range of correlation coefficient over $(-1, 1)$ have been proposed. This is based on the Sarmanov family of bivariate distributions which is a simple and elegant approach in constructing bivariate distributions. A method is proposed for constructing bivariate generalizations of weighted distributions. The univariate CMP distribution is a very popular distribution for applications since it has the flexibility to analyze data that exhibit under or over dispersion. The BCMP distribution proposed by Sellers et al. [24] does not have the marginal distributions which are CMP distributions. It is shown in this article that the proposed BCMP distributions fit much better than the BCMP of Sellers et al. [24]. Thus, the proposed BCMP distributions will be of great utility to the data analysts.

Compliance with Ethical Standards

Conflict of interest There is no conflict of interest concerning this manuscript

Appendix

Derivatives of the BCMP normalizing constant, $Z(\lambda\theta, v)$ are shown as follows

$$\frac{\partial Z_1^*}{\partial \lambda_1} = \theta \sum_{j=1}^{\infty} \frac{j(\lambda_1\theta)^{j-1}}{(j!)^{v_1}}, \quad \frac{\partial Z_1^*}{\partial v_1} = -\sum_{j=1}^{\infty} \frac{\ln(j!)(\lambda_1\theta)^{j}}{(j!)^{v_1}}$$

$$\frac{\partial^2 Z_1^*}{\partial \lambda_1 \partial v_1} = -\theta \sum_{j=1}^{\infty} \frac{j\ln(j!)(\lambda_1\theta)^{j-1}}{(j!)^{v_1}}, \quad \frac{\partial^2 Z_1^*}{\partial v_1^2} = -\sum_{j=1}^{\infty} \frac{\ln(j!)^2(\lambda_1\theta)^{j}}{(j!)^{v_1}}.$$

References

1. Arbous AG, Kerrich JE (1951) Accident statistics and the concept of accident proneness. Biometrics 7:340–342
2. Balakrishnan N, Lai CD (2009) Continuous bivariate distributions, 2nd edn. Springer, New York
3. Bjerve S, Doksum K (1993) Correlation curves: measures of association as functions of covariate values. Ann Stat 21:890–902
4. Blyth S (1993) A note on correlation curves and Chernoff's inequality. Scand J Stat 20(4):375–377
5. Blyth S (1994) Karl Pearson and correlation curve. Int Stat Rev 62(3):393–403
6. Blyth S (1994) Measuring local association: an introduction to the correlation curve. Sociol Methods 24:171–197
7. Doksum K, Blyth S, Bradlow E, Meng XL, Zhao H (1994) Correlation curves as local measures of variance explained by regression. J Am Stat Assoc 89(426):571–582
8. Genest C, Verret F (2002) The TP$_2$ ordering of Kimeldorf and Sampson has the normal-agreeing property. Stat Probab Lett 57:387–391
9. Goodman LA (1969) How to ransack social mobility tables and other kinds of cross-classification tables. Am J Sociol 75:1–39
10. Gupta RC, Sim SZ, Ong SH (2014) Analysis of discrete data by Conway–Maxwell Poisson distribution. AStA Adv Stat Anal 98:327–343
11. Henderson D, Jacobson SH, Johnson AW (2003) The theory and practice of simulated annealing. In: Handbook of metaheuristics, vol 57, pp 287–319
12. Holgate P (1964) Estimation for the bivariate Poisson distribution. Biometrika 51:241–245
13. Holland PW, Wang YJ (1987) Dependence function for continuous bivariate densities. Commun Stat Theory Methods 16:863–876
14. Huang JS, Kotz S (1984) Correlation structure in iterated Farlie–Gumbel–Morgenstern distributions. Biometrika 71:633–636
15. Huang JS, Kotz S (1999) Modifications of the Farlie–Gumbel–Morgenstern distributions: a tough hill to climb. Metrika 49:307–323
16. Johnson NL, Kotz S (1975) On some generalized Farlie–Gumbel– Morgenstern distributions. Commun Stat Theory Methods 4:415–427
17. Kadane JB, Krishnan R, Shmeuli G (2006) A data disclosure policy for count data based on the COM-Poisson distribution. Manag Sci 52(10):1610–1617
18. Lehmann EL (1966) Some concepts of dependence. Ann Math Stat 37:1137–1153
19. Lee M-LT (1996) Properties and applications of the Sarmanov family of bivariate distributions. Commun Stat Theory Methods 25:1207–1222
20. Lin GD (1987) Relationships between two extensions of Farlie–Gumbel–Morgenstern distribution. Ann Inst Stat Math 39:129–140
21. Lin GD, Dou X, Kuriki S, Huang JS (2014) Recent developments on the construction of bivariate distributions with fixed marginals. J Stat Distrib Appl 1:14

22. Ong SH (1992) The computer generation of bivariate binomial variables with given marginals and correlation. Commun Stat Simul Comput 21:285–299
23. Sarmanov IO (1966) Generalized normal correlation and two dimensional Frechet classes. Soviet Math Dokl **7**:596–599 [English translation; Russian original in Dokl. Akad. Nauk. SSSR 168, 32–35 (1966)]
24. Sellers KF, Morris DS, Balakrishnan N (2016) Bivariate Conway–Maxwell–Poisson distribution: formulation, properties, and inference. J Multivar Anal 150:152–168
25. Schucany WR, Parr WC, Boyer JE (1978) Correlation structure in Farlie-Gumbel-Morgenstern distributions. Biometrika 65(3):650–653
26. Shaked M, Shanthikumar JG (2007) Stochastic orders. Springer, New York
27. Shmueli G, Minka TP, Kadane JB, Borle S, Boatwright S (2005) A useful distribution for fitting discrete data-revival of the Conway–Maxwell–Poisson distribution. J R Stat Soc Ser C (Appl Stat) 54(1):127–142
28. Shubina M, Lee M-LT (2004) On maximum attainable correlation and other measures of dependence for the Sarmanov family of bivariate distributions. Commun Stat Theory Methods 33:1031–1052
29. Wang YJ (1993) Construction of continuous bivariate density functions. Stat Sin 3:173–187
30. Yule GU, Kendall MG (1950) An introduction to the theory of statistics, 14th edn. Charles Griffith & Company Limited, London