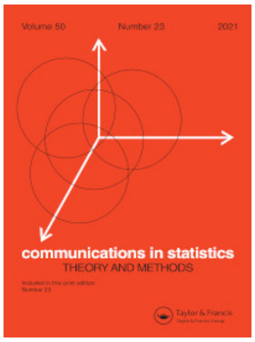

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

A flexible bivariate distribution for count data expressing data dispersion

Kimberly S. Weems, Kimberly F. Sellers & Tong Li

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2021.1999474

A flexible bivariate distribution for count data expressing data dispersion

Kimberly S. Weems[a], Kimberly F. Sellers[b,c], and Tong Li[b]

[a]Department of Mathematics and Physics, North Carolina Central University, Durham, North Carolina, USA; [b]Department of Mathematics and Statistics, Georgetown University, Washington, DC, USA; [c]Center for Statistical Research and Methodology, U. S. Census Bureau, Washington, DC, USA

ABSTRACT
The bivariate Poisson distribution is a natural choice for modeling bivariate count data. Its constraining assumption, however, limits model flexibility in some contexts. This work considers the trivariate reduction method to construct a Bivariate Conway-Maxwell-Poisson (BCMP) distribution, which accommodates over- and under-dispersed data. The approach produces marginals that have a flexible form which includes several special case distributions for certain parameters. Moreover, this BCMP model performs well relative to other bivariate models for count data, including BCMP models based on different methods of construction. As a result, the trivariate-reduced BCMP distribution is a flexible alternative for modeling bivariate count data containing data dispersion.

ARTICLE HISTORY
Received 14 September 2020
Accepted 22 October 2021

KEYWORDS
Bivariate count data; Conway-Maxwell-Poisson distribution; trivariate reduction

1. Introduction

The trivariate reduction method is a popular approach for constructing dependent random variables, whether continuous or discrete (Chesneau, Kachour, and Karlis 2015). For three random variables, X_i, $i = 1, 2, 3$, the idea behind the trivariate reduction method is to define new random variables, say $X = h_1(X_1, X_3)$ and $Y = h_2(X_2, X_3)$ for functions $h_i(X_i, X_3)$, $i = 1, 2$ thus clearly capturing some measure of interdependence through X_3. In order to construct bivariate discrete distributions, a popular choice is to let $h_i(X_i, X_3) = X_i + X_3$, $i = 1, 2$ define the dependent discrete variables, given three discrete random variables X_i, $i = 1, 2, 3$. Also known as the "variables in common" method, the trivariate reduction method can be generalized to allow for three or more random variables that may or may not themselves be independent (Lai 2006). Here, we consider independent X_i, $i = 1, 2, 3$, and let $X = X_1 + X_3$ and $Y = X_2 + X_3$.

The trivariate reduction method is a particularly appealing means by which to establish the bivariate Poisson (BP) distribution, which is a popular model for count data (M'Kendrick 1926; Maritz 1952; Teicher 1954; Holgate 1964; Marshall and Olkin 1985;

CONTACT Kimberly S. Weems ksweems@nccu.edu Department of Mathematics and Physics, North Carolina Central University, Durham, NC, USA.

Supplemental data for this article is available online at https://doi.org/10.1080/03610926.2021.1999474.
This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Kocherlakota and Kocherlakota 1992; Johnson, Kotz, and Balakrishnan 1997). Letting X_i, $i = 1, 2, 3$ be independent Poisson(λ_i) random variables, the joint probability mass function (pmf) for the BP distribution is

$$P(X, Y) = \exp\{-(\lambda_1 + \lambda_2 + \lambda_3)\} \frac{\lambda_1^x \lambda_2^y}{x!\, y!} \sum_{k=0}^{\min(x,y)} \binom{x}{k} \binom{y}{k} k! \left(\frac{\lambda_3}{\lambda_1 \lambda_2}\right)^k \tag{1}$$

with marginal probability functions for X and Y taking the form of univariate Poisson pmfs with respective rate parameters, $\lambda_1 + \lambda_3$ and $\lambda_2 + \lambda_3$. Marshall and Olkin (1985) and Kocherlakota and Kocherlakota (1992) note the following properties of this bivariate distribution:

1. the probability generating function (pgf) is
$$\begin{aligned} \Pi(t_1, t_2) &= \exp[\lambda_1(t_1 - 1) + \lambda_2(t_2 - 1) + \lambda_3(t_1 t_2 - 1)] \\ &= \exp\left((\lambda_1 + \lambda_3)(t_1 - 1) + (\lambda_2 + \lambda_3)(t_2 - 1) + \lambda_3(t_1 - 1)(t_2 - 1)\right); \end{aligned} \tag{2}$$
2. the covariance is $\text{Cov}(X, Y) = \lambda_3$;
3. the correlation, $\text{Corr}(X, Y) = \frac{\lambda_3}{\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}}$, is non-negative; and
4. the conditional mean, $E(X|Y = y) = \lambda_1 + y\left(\frac{\lambda_3}{\lambda_2+\lambda_3}\right)$, shows the linear regression of X on Y.

Kokonendji and Puig (2018) introduce a generalized dispersion index (GDI),

$$GDI(X, Y) = \frac{\begin{pmatrix} \sqrt{E(X)} & \sqrt{E(Y)} \end{pmatrix} \begin{pmatrix} Var(X) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) \end{pmatrix} \begin{pmatrix} \sqrt{E(X)} \\ \sqrt{E(Y)} \end{pmatrix}}{\begin{pmatrix} E(X) & E(Y) \end{pmatrix} \begin{pmatrix} E(X) \\ E(Y) \end{pmatrix}}, \tag{3}$$

which can assess data dispersion relative to the uncorrelated Poisson distribution. The GDI of a bivariate model is equi-dispersed relative to the uncorrelated BP distribution if $GDI = 1$; alternately, $GDI > (<)1$ indicates over-dispersion (under-dispersion) relative to the uncorrelated BP distribution. For example, the trivariate-reduced BP distribution has

$$GDI(X, Y) = 1 + \frac{2\lambda_3 \sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}{(\lambda_1 + \lambda_3)^2 + (\lambda_2 + \lambda_3)^2} \geq 1 \quad \Longleftrightarrow \quad \lambda_3 \geq 0,$$

thus the trivariate-reduced BP distribution is over-dispersed (equi-dispersed) relative to the uncorrelated BP distribution when $\lambda_3 > (=)0$.

The Poisson distribution is known to be constrained by data equi-dispersion (i.e., where the associated mean and variance equal); accordingly, associated limiting characteristics hold true even in the bivariate case. Stein and Juritz (1987) instead utilize the trivariate reduction method with independent negative binomial random variables[1], $X_i \sim \text{NB}(\alpha_i, \beta)$, $i = 1, 2, 3$, to produce a bivariate negative binomial (BNB) distribution whose joint pmf is

$$p(X = x, y = y) = \left(\frac{\beta}{1+\beta}\right)^{\sum_i^3 \alpha_i} \left(\frac{1}{1+\beta}\right)^{x+y} \times \sum_{i=0}^{\min(x,y)} \binom{\alpha_1 + x - i - 1}{x - i} \binom{\alpha_3 + y - i - 1}{y - i} \binom{\alpha_3 + i + 1}{i} (1+\beta)^i,$$

with respective NB($\alpha_1 + \alpha_3, \beta$) and NB($\alpha_2 + \alpha_3, \beta$) marginal distributions for X and Y. Other properties of this BNB distribution include the conditional probability,

$$P(X = x | Y = y) = \left(\frac{\beta}{1+\beta}\right)^{\alpha_1} \left(\frac{1}{1+\beta}\right)^{x} \Big/ \binom{y + \alpha_2 + \alpha_3 - 1}{y} \times \sum_{i=0}^{\min(x,y)} \binom{\alpha_1 + x - i - 1}{x - i} \binom{\alpha_2 + y - i - 1}{y - i} \binom{\alpha_3 + i - 1}{i} (1+\beta)^i,$$

and the conditional mean and correlation, respectively, namely

$$E(X|Y = y) = \frac{\alpha_1}{\beta} + \left(\frac{\alpha_3}{\alpha_2 + \alpha_3}\right) y,$$
$$\rho(X, Y) = \frac{\alpha_3}{\sqrt{(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_3)}}.$$

While the BNB can address data over-dispersion (where the variance is greater than the mean), it is not able to accommodate data under-dispersion. The bivariate generalized Poisson (BGP) distribution, introduced by Famoye and Consul (1995), instead allows for either form of dispersion by using trivariate reduction with independent GP(θ_i, λ_i) random variables X_i, $i = 1, 2, 3$. The resulting distribution has the pmf

$$P(X = x, Y = y) = \theta_1 \theta_2 \theta_3 e^{-\theta_1 - \theta_2 - \theta_3 - x\lambda_1 - y\lambda_2} \cdot \sum_{u=0}^{\min(x,y)} k(u), \tag{4}$$

where

$$k(u) = \frac{[\theta_1 + (x-u)\lambda_1]^{x-u-1}}{(x-u)!} \frac{[\theta_2 + (y-u)\lambda_2]^{y-u-1}}{(y-u)!} \frac{[\theta_3 + u\lambda_3]^{u-1}}{u!} e^{u(\lambda_1 + \lambda_2 - \lambda_3)}.$$

The BGP distribution reduces to the Holgate (1964) BP distribution when $\lambda_i = 0$ for all $i = 1, 2, 3$, and has the following properties:

$$\begin{aligned}
E(X) &= \theta_1 (1 - \lambda_1)^{-1} + \theta_3 (1 - \lambda_3)^{-1}, \\
E(Y) &= \theta_2 (1 - \lambda_2)^{-1} + \theta_3 (1 - \lambda_3)^{-1}, \\
E(X^2) &= \theta_1 (1 - \lambda_1)^{-3} + \theta_3 (1 - \lambda_3)^{-3}, \\
E(Y^2) &= \theta_2 (1 - \lambda_2)^{-3} + \theta_3 (1 - \lambda_3)^{-3}, \\
E(XY) &= \theta_3 (1 - \lambda_3)^{-3}, \\
\rho(X, Y) &= \theta_3 \{ [\theta_1 (1 - \lambda_1)^{-3} (1 - \lambda_3)^3 + \theta_3] [\theta_2 (1 - \lambda_2)^{-3} (1 - \lambda_3)^3 + \theta_3] \}^{-1/2}.
\end{aligned} \tag{5}$$

Equation (5) shows that, for $\theta_3 = 0$, X and Y are uncorrelated. Meanwhile, for $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$, the correlation structure provided in Equation (5) simplifies to $\rho =$

$\theta_3[(\theta_1+\theta_3)(\theta_2+\theta_3)]^{-1/2}$, and the marginal distributions of X and Y are univariate GP distributions with respective parameters, $(\theta_1+\theta_3,\lambda)$ and $(\theta_2+\theta_3,\lambda)$.

The conditional distribution of X given a value for Y is

$$P(X=x|Y=y)=\frac{\theta_3}{\theta_2+\theta_3}\sum_{u=0}^{\min(x,y)}\binom{y}{u}\frac{\theta_1[\theta_1+\lambda(x-u)]^{x-u-1}}{(x-u)!} \times\frac{\theta_2[\theta_2+\lambda(y-u)]^{y-u-1}}{(y-u)!}\frac{(\theta_3+\lambda u)^{u-1}}{(\theta_2+\theta_3+\lambda y)^{y-1}}e^{-\theta_1-\lambda(x-u)}, \tag{6}$$

from which it can be shown that the conditional expectation of X given Y is

$$E(X|Y=y)=\theta_1(1-\lambda)^{-1}+\theta_3(\theta_2+\theta_3)^{-1}y;$$

likewise, the conditional mean of Y given X is

$$E(Y|X=x)=\theta_2(1-\lambda)^{-1}+\theta_3(\theta_1+\theta_3)^{-1}x$$

(Famoye and Consul 1995). While the BGP distribution can accommodate data over- or under-dispersion, it is limited in the extent to which it can handle data under-dispersion (Famoye 1993).

Each of the above bivariate models strives to sufficiently describe correlated count data, yet each suffers from some limitation in flexibility as it relates to data dispersion. To summarize, the BP distribution is a natural choice for modeling count data stemming from two correlated random variables; however, this construct is limited by the underlying model assumption that the data are equi-dispersed. The BNB and the BGP distributions are welcomed alternatives to the BP; however, they likewise suffer from their own respective limitations with regard to data dispersion. This work uses the univariate Conway-Maxwell-Poisson (CMP) distribution to construct a bivariate CMP (BCMP) model that allows for under- or over-dispersion. Section 2 introduces the reader to this univariate model and related structures that offer flexibility in the face of data dispersion.

The trivariate reduction approach is one way by which to create a bivariate distribution. Various other construction techniques can likewise be used where the univariate CMP distribution serves as motivation. Sellers, Morris, and Balakrishnan (2016) use the compounding method to obtain a BCMP distribution (BCMP_C). Ong et al. (2021) use an approach based on the Sarmanov family of distributions to construct two additional BCMP distributions (BCMP_{S1} and BCMP_{S2}, respectively). This work, instead, utilizes the trivariate reduction approach via sums of independent CMP random variables, thus deriving yet another BCMP distribution (trivariate-reduced BCMP or BCMP_T). Section 3 describes these bivariate models, outlining their associated statistical properties. Section 4 addresses matters of statistical inference and computation for the trivariate-reduced BCMP, including parameter estimation and hypothesis testing. Section 5 provides simulated and real data examples illustrating the flexibility of the trivariate-reduced BCMP distribution for bivariate dispersed count data. Lastly, Section 6 concludes with remarks and some generalizations.

2. The Conway-Maxwell-Poisson and related distributions

The CMP distribution (introduced by Conway and Maxwell (1962), and revived by Shmueli et al. (2005)) is a flexible count distribution whose pmf has the form

$$\Pr(W = w|\lambda, \nu) = \frac{\lambda^w}{(w!)^\nu Z(\lambda, \nu)}, \quad w = 0, 1, 2, \ldots$$

for a random variable W, where $\lambda = \mathrm{E}(W^\nu) > 0$ is a location parameter and $\nu \geq 0$ is a dispersion parameter such that $\nu = 1$ denotes equi-dispersion, while $\nu > (<)\ 1$ signifies under-dispersion (over-dispersion) (Shmueli et al. 2005). Meanwhile, $Z(\lambda, \nu) = \sum_{s=0}^{\infty} \frac{\lambda^s}{(s!)^\nu}$ is the normalizing constant that has been well-studied, particularly with varying suggestions on how to approximate the infinite sum (Minka et al. 2003; Gillispie and Green 2015; Şimşek and Iyengar 2016; Gaunt et al. 2019). The CMP distribution includes three well-known distributions as special cases: Poisson with rate parameter λ (for $\nu = 1$), Bernoulli with success probability $\frac{\lambda}{1+\lambda}$ (for $\nu \to \infty$), and geometric with success probability $1 - \lambda$ (for $\nu = 0, \lambda < 1$).

The probability generating function (pgf) for the CMP distribution is $\Pi_W(t) = \frac{Z(\lambda t, \nu)}{Z(\lambda, \nu)}$, and the moment generating function (mgf) is $M_W(t) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}$ (Sellers, Shmueli, and Borle 2012), from which the associated moments can be derived. Shmueli et al. (2005) meanwhile report the CMP moments via the recursion,

$$\mathrm{E}(W^{r+1}) = \begin{cases} \lambda\{\mathrm{E}(W+1)\}^{1-\nu}, & r = 0 \\ \lambda \dfrac{\partial}{\partial \lambda} \mathrm{E}(W^r) + \mathrm{E}(W)\mathrm{E}(W^r), & r > 0. \end{cases} \tag{7}$$

In particular, the expected value and variance can be written in the form and approximated, respectively, as

$$\mathrm{E}(W) = \frac{\partial \ln Z(\lambda, \nu)}{\partial \ln \lambda} \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \quad \text{and} \tag{8}$$

$$\mathrm{Var}(W) = \frac{\partial \mathrm{E}(W)}{\partial \ln \lambda} \approx \frac{1}{\nu} \lambda^{1/\nu}, \tag{9}$$

where the approximations are especially good for $\nu \leq 1$ or $\lambda > 10^\nu$ (Shmueli et al. 2005); these results and the associated constraints stem from the aforementioned approximations to the normalizing constant.

Sellers, Swift, and Weems (2017) introduce the sum-of-CMP (sCMP) as a generalization of the CMP distribution. Letting $\mathcal{W} = \sum_{i=1}^{s} W_i$, where $W_i \sim \mathrm{CMP}(\lambda, \nu), i = 1, \ldots, s$ are independent and identically distributed (iid) random variables, we say that $\mathcal{W}$ is an sCMP(λ, ν, s) variable, and has the pmf for $w = 0, 1, 2, \ldots,$ namely

$$P(\mathcal{W} = w) = P\left(\sum_{i=1}^{s} W_i = w\right) = \frac{\lambda^w}{(w!)^\nu Z^s(\lambda, \nu)} \sum_{\substack{a_1, \ldots, a_s = 0 \\ a_1 + \ldots + a_s = w}}^{w} \binom{w}{a_1, \ldots, a_s}^\nu, \tag{10}$$

where $\binom{w}{a_1, \ldots, a_s} = \frac{w!}{a_1! \cdots a_s!}$ is a multinomial coefficient, and $Z^s(\lambda, \nu)$ denotes the sth power of the CMP normalizing constant. The sCMP(λ, ν, s) distribution contains the

Poisson distribution with rate parameter $s\lambda$ (for $\nu=1$), negative binomial$(s, 1-\lambda)$ distribution (for $\nu=0$ and $\lambda<1$), and Binomial(s, p) distribution (as $\nu\to\infty$ with success probability $p=\frac{\lambda}{\lambda+1}$) as special cases; for $s=1$, the sCMP$(\lambda,\nu,s=1)$ is the CMP(λ,ν) distribution (Sellers, Swift, and Weems 2017). The sCMP pgf has the form $\Pi_W(t)=\left(\frac{Z(\lambda t,\nu)}{Z(\lambda,\nu)}\right)^s$ where this pgf form implies that, while it is easy to interpret the parameter s assuming a discrete form, s can actually be a continuous parameter. Under appropriate conditions, the sCMP distribution is closed under addition, i.e., sums of independent sCMP random variables (with the same λ and ν) produce sCMP distributed random variables. These models motivate discussion regarding the bivariate structures discussed in Section 3.

3. Bivariate CMP distributions

Several bivariate distributions based on the CMP distribution have been constructed. Sections 3.1 and 3.2 describe some recent developments, while Section 3.3 introduces the trivariate-reduced BCMP distribution, which is the main focus of this paper.

3.1. Using the compounding method

Sellers, Morris, and Balakrishnan (2016) generalize the compounding approach for the BP distribution (Kocherlakota and Kocherlakota 1992) by "compounding" a bivariate binomial with a CMP distribution, i.e., letting the joint conditional distribution of $\{(X,Y)|n\}$ have a bivariate binomial distribution, where the number of trials n is a CMP(λ,ν) random variable. The resulting BCMP model (called BCMP$_C$) for (X, Y) has the joint pgf

$$\Pi(t_1^*,t_2^*)=\frac{Z\big(\lambda\big[1+p_{1+}(t_1^*-1)+p_{+1}(t_2^*-1)+p_{11}(t_1^*-1)(t_2^*-1)\big],\nu\big)}{Z(\lambda,\nu)}. \quad (11)$$

This construct for the pgf of (X, Y) yields three special cases that we desire of a BCMP distribution: for $\nu=1$, the BCMP$_C$ distribution reduces to the BP described in Section 1 where $\lambda_1+\lambda_3=\lambda_0p_{1+}$, $\lambda_2+\lambda_3=\lambda_0p_{+1}$, and $\lambda_3=\lambda_0p_{11}$; when $\nu=0$, it reduces to the Marshall and Olkin (1985) bivariate geometric model; and, for $\nu\to\infty$, we obtain the bivariate Bernoulli distribution described in Marshall and Olkin (1985) with the form,

		Y: 0	Y: 1	
X	0	$\tilde{p}_{00}\doteq 1-\frac{\lambda}{\lambda+1}(p_{01}+p_{10}+p_{11})$	$\tilde{p}_{01}\doteq\frac{\lambda}{\lambda+1}p_{01}$	$\tilde{p}_{0+}\doteq 1-\frac{\lambda}{\lambda+1}p_{1+}$
	1	$\tilde{p}_{10}\doteq\frac{\lambda}{\lambda+1}p_{10}$	$\tilde{p}_{11}\doteq\frac{\lambda}{\lambda+1}p_{11}$	$\tilde{p}_{1+}\doteq\frac{\lambda}{\lambda+1}p_{1+}$
		$\tilde{p}_{+0}\doteq 1-\frac{\lambda}{\lambda+1}p_{+1}$	$\tilde{p}_{+1}\doteq\frac{\lambda}{\lambda+1}p_{+1}$	1

For this special case, $GDI(X,Y)\in(0,1]$ implies that the bivariate Bernoulli distribution is under- or equi-dispersed relative to the uncorrelated BP distribution (Kokonendji and

Puig 2018). Further, the recurring term in the above table, $\frac{\lambda}{\lambda+1}$, is the success probability in the univariate CMP case when $\nu \to \infty$, i.e., the univariate Bernoulli $\left(\frac{\lambda}{\lambda+1}\right)$ case.

Equation (11) is used to derive the joint pmf of (X, Y) as

$$P(X = x,\ Y = y) = \frac{1}{Z(\lambda, \nu)} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n!)^{\nu}} \sum_{a=n-x-y}^{n} \binom{n}{a,\ n-a-y,\ n-a-x,\ x+y+a-n} \times p_{00}^{a} p_{10}^{n-a-y} p_{01}^{n-a-x} p_{11}^{x+y+a-n}, \tag{12}$$

for some parameters, λ, ν, and probabilities $p_{00}, p_{10}, p_{01}, p_{11}$ such that $p_{00} + p_{10} + p_{01} + p_{11} = 1$, $p_{i+} = p_{i0} + p_{i1}$ for $i = 0,\ 1$, and $p_{+j} = p_{0j} + p_{1j}$ for $j = 0,\ 1$. Moments and product moments via the joint factorial mgf, as well as the regression of Y on X, and the conditional pgf are likewise obtained (Sellers, Morris, and Balakrishnan 2016). While the marginal distributions of X and Y are not themselves CMP distributed, Poisson marginals serve as a special case.

3.2. Using the Sarmanov family of distributions

A drawback to the compounding approach toward constructing a BCMP distribution is that its correlation cannot attain the full range; it is constrained to only be greater than or equal to 0. Ong et al. (2021) instead construct two versions of a BCMP distribution based on the Sarmanov family of distributions (Sarmanov 1966). For random variables X and Y and mixing functions $\phi_1(X)$ and $\phi_2(Y)$, such that $E[\phi_1(X)] = E[\phi_2(Y)] = 0$ and $1 + \phi_1(x)\phi_2(y) \geq 0$, the Sarmanov family of distributions is defined as

$$P(X = x, Y = y) = P(X = x)P(Y = y)\left[1 + \tau\phi_1(x)\phi_2(y)\right], \quad x, y \in \mathbb{R},$$

for $\tau \in [-1, 1]$. This construction is argued as more desirable for a BCMP because the respective marginals for X and Y are CMP distributed, and the resulting models allow for a [-1, 1] range in correlation.

The first BCMP distribution (BCMP_{S1}) utilizes weighted Poisson distributions (Kokonendji, Mizere, and Balakrishnan 2008). For $\alpha > 0$, let

$$\phi^1(x) = p^{\alpha}(x) - E(p^{\alpha}(X)) \quad \text{and} \quad \phi^2(y) = p^{\alpha}(y) - E(p^{\alpha}(Y)), \tag{13}$$

which yield the joint pmf

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

where $P(X = x)$ and $P(Y = y)$ denote marginal pmfs $\text{CMP}(\lambda_1, \nu_1)$ and $\text{CMP}(\lambda_2, \nu_2)$, respectively. Note that

$$E\left[p^{\alpha}(X)\right] = \frac{e^{-\lambda_1\alpha} Z(\lambda_1^2, \nu_1 + 1)}{Z(\lambda_1, \nu_1)} \sum_{x=0}^{\infty} \frac{\lambda_1^{x(\alpha+1)}}{(x!)^{\nu_1+\alpha} Z(\lambda_1^2, \nu_1 + 1)} \leq 1, \tag{14}$$

and similarly, $E[p^{\alpha}(Y)]$ is defined as in Equation (14) with $(\lambda_2,\ \nu_2)$ replacing $(\lambda_1,\ \nu_1)$. The correlation coefficient between X and Y is expressed as

$$\rho = \frac{\tau(\lambda_1 - \mu_1)(\lambda_2 - \mu_2)}{\sigma_1 \sigma_2}, \tag{15}$$

where (μ_1, σ_1) and (μ_2, σ_2) denote the marginal mean and standard deviation for X and Y, respectively.

The second BCMP distribution (BCMP_{S2}) relies on the pgfs of the marginal CMP distributions. Letting $\Pi_X(\theta) = \frac{Z(\lambda_1\theta, \nu_1)}{Z(\lambda_1, \nu_1)}$ and $\Pi_Y(\theta) = \frac{Z(\lambda_2\theta, \nu_2)}{Z(\lambda_2, \nu_2)}$ denote the respective pgfs of X and Y, Ong et al. (2021) consider

$$\phi_1(x) = \theta^x - \Pi_X(\theta) \quad \text{and} \quad \phi_2(y) = \theta^y - \Pi_Y(\theta), \tag{16}$$

where $0 < \theta < 1$. The joint pmf for this construction is given by

$$P(X = x, Y = y) = P(X = x)P(Y = y)\left[1 + \tau\phi_1(x)\phi_2(y)\right] \tag{17}$$

with correlation

$$\rho = \frac{\tau\left(\theta\frac{\partial \Pi_X(\theta)}{\partial\theta} - \mu_1\Pi_X(\theta)\right)\left(\theta\frac{\partial \Pi_Y(\theta)}{\partial\theta} - \mu_2\Pi_Y(\theta)\right)}{\sigma_1\sigma_2}. \tag{18}$$

For $\nu_1 = \nu_2 = 1$ and $\theta = e^{-1}$, the BCMP_{S2} distribution corresponds to the Lee (1996) BP distribution.

3.3. Using trivariate reduction

Here, we introduce a simple BCMP model that extends the flexibility and utility of the univariate CMP to a bivariate form via trivariate reduction (henceforth, referred to as trivariate-reduced BCMP or BCMP_{T}) and relates to models described in Section 1. Let W_i, $i = 1, 2, 3$, be independently distributed CMP(λ_i, ν) random variables such that $X = W_1 + W_3$ and $Y = W_2 + W_3$. Clearly, X and Y are correlated via W_3 and thus have a joint distribution whose pmf is

$$\begin{aligned}
P(X = x, Y = y) &= P(W_1 + W_3 = x, W_2 + W_3 = y) \\
&= \sum_{w_3=0}^{\infty} P(W_1 + W_3 = x, W_2 + W_3 = y | W_3 = w_3)P(W_3 = w_3) \\
&= \sum_{w_3=0}^{\infty} P(W_1 = x - w_3)P(W_2 = y - w_3)P(W_3 = w_3) \\
&\quad \text{by independence of } W_i, i = 1, 2, 3 \\
&= \sum_{w_3=0}^{\min(x,y)} \frac{\lambda_1^{x-w_3}}{[(x-w_3)!]^{\nu} Z(\lambda_1, \nu)} \frac{\lambda_2^{y-w_3}}{[(y-w_3)!]^{\nu} Z(\lambda_2, \nu)} \frac{\lambda_3^{w_3}}{[(w_3)!]^{\nu} Z(\lambda_3, \nu)} \\
&= \frac{\lambda_1^x \lambda_2^y}{Z(\lambda_1, \nu)Z(\lambda_2, \nu)Z(\lambda_3, \nu)} \sum_{w_3=0}^{\min(x,y)} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^{w_3} \frac{1}{[(x-w_3)!(y-w_3)!w_3!]^{\nu}} \\
&= \frac{\lambda_1^x \lambda_2^y}{Z(\lambda_1, \nu)Z(\lambda_2, \nu)Z(\lambda_3, \nu)(x!y!)^{\nu}} \\
&\quad \times \sum_{w_3=0}^{\min(x,y)} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^{w_3} \left[\binom{x}{w_3}\binom{y}{w_3} w_3!\right]^{\nu}
\end{aligned} \tag{19}$$

and joint pgf

$$\begin{aligned}\prod(t_1,t_2) &= E(t_1^X t_2^Y) = E(t_1^{W_1+W_3} t_2^{W_2+W_3}) \\ &= E(t_1^{W_1})E(t_2^{W_2})E((t_1t_2)^{W_3})(\text{by independence of } W_i, i=1,2,3) \\ &= \frac{Z(\lambda_1 t_1,\nu)}{Z(\lambda_1,\nu)} \frac{Z(\lambda_2 t_2,\nu)}{Z(\lambda_2,\nu)} \frac{Z(\lambda_3 t_1 t_2,\nu)}{Z(\lambda_3,\nu)}.\end{aligned} \tag{20}$$

For the special case where $\nu=1$, the BCMP_T model reduces to the Holgate (1964) BP distribution described in Section 1 with joint pgf, $\Pi(t_1,t_2;\nu=1) = \exp(\lambda_1(t_1-1)+\lambda_2(t_2-1)+\lambda_3(t_1t_2-1))$.[2] Meanwhile when $\nu=0$,

$$\Pi(t_1,t_2;\nu=0) = \frac{(1-\lambda_1)(1-\lambda_2)(1-\lambda_3)}{(1-\lambda_1 t_1)(1-\lambda_2 t_2)(1-\lambda_3 t_1 t_2)} \tag{21}$$

is the joint pgf of a bivariate geometric distribution of the type obtained via trivariate reduction for $\lambda_i < 1$, $i=1, 2, 3$, $\lambda_j t_j < 1$, $j=1, 2$, and $\lambda_3 t_1 t_2 < 1$. This is a special case of the Stein and Juritz (1987) BNB distribution where $\alpha_i = 1$ for $i=1, 2, 3$ and $\lambda_1 = \lambda_2 = \lambda_3 = \lambda = \frac{1}{1+\beta}$ produces

$$P(X=x, Y=y) = \left(\frac{\beta}{1+\beta}\right)^3 (1+\beta)^{-x-y} \sum_{w_3=0}^{\min(x,y)} (1+\beta)^{w_3}.$$

Finally, when $\nu \to \infty$, the BCMP_T distribution reduces to the trivariate-reduced bivariate Bernoulli distribution with joint pgf

$$\begin{aligned}\Pi(t_1,t_2;\nu\to\infty) &= \frac{(1+\lambda_1 t_1)(1+\lambda_2 t_2)(1+\lambda_3 t_1 t_2)}{(1+\lambda_1)(1+\lambda_2)(1+\lambda_3)} \\ &= (q_1+p_1t_1)(q_2+p_2t_2)(q_3+p_3t_1t_2),\end{aligned} \tag{22}$$

where $p_i = \frac{\lambda_i}{1+\lambda_i}$ and $q_i = 1-p_i$.

The BCMP_T has marginal distributions that are convolutions of CMP distributions; the marginal pgfs have the form

$$\Pi_X(t) = \Pi(t,1) = \frac{Z(\lambda_1 t,\nu)Z(\lambda_3 t,\nu)}{Z(\lambda_1,\nu)Z(\lambda_3,\nu)}, \text{ and} \tag{23}$$

$$\Pi_Y(t) = \Pi(1,t) = \frac{Z(\lambda_2 t,\nu)Z(\lambda_3 t,\nu)}{Z(\lambda_2,\nu)Z(\lambda_3,\nu)}. \tag{24}$$

For $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$, the marginal pgfs simplify to $\Pi_X(t) = \Pi_Y(t) = \left(\frac{Z(\lambda t,\nu)}{Z(\lambda,\nu)}\right)^2$, i.e., the form of a $\text{sCMP}(\lambda,\nu,2)$ random variable (see Section 2), which includes the Poisson(2λ), negative binomial($2, 1-\lambda$), and binomial$\left(2, p=\frac{\lambda}{1+\lambda}\right)$ distributions as special cases under certain conditions (Sellers, Swift, and Weems 2017). As such, the marginals are likewise equi-, over-, or under-dispersed depending on the value of ν ($=, <, > 1$). The marginal pgfs, along with the analogous mgfs, aid in determining the following distributional properties, where the approximations hold for $\nu \le 1$ or $\lambda > 10^\nu$:

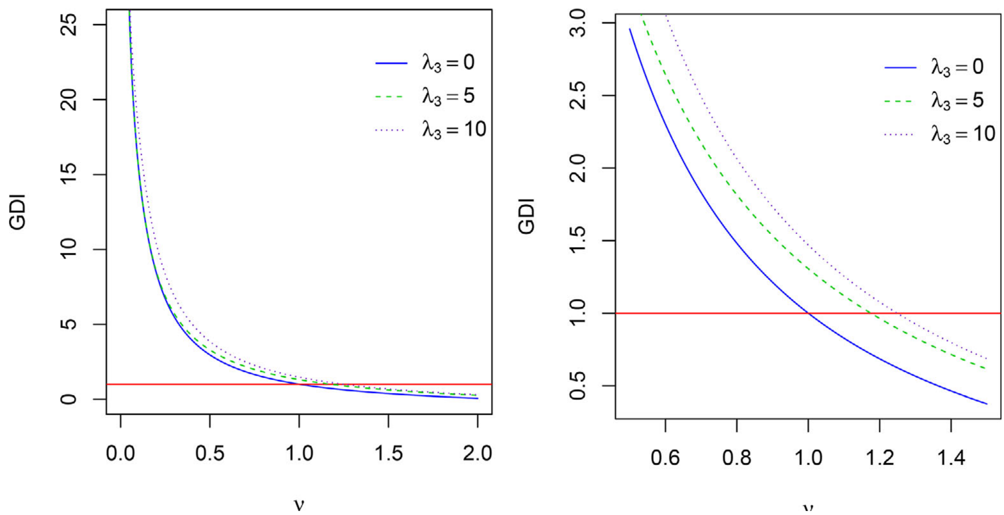

Figure 1. Plots of GDI versus ν for $\lambda_1 = 10.5$, $\lambda_2 = 12$, and $\lambda_3 = (0, 5, 10)$. The line at GDI = 1 indicates equi-dispersion. Figure 1a gives plots when $\nu \in [0, 2]$; Figure 1b gives a magnified view when $\nu \in [0.5, 1.5]$ for clarity.

1. $E(X) = E(W_1) + E(W_3) \approx \lambda_1^{1/\nu} + \lambda_3^{1/\nu} - \frac{\nu-1}{\nu}$ by independence, where $E(W_i)$ is determined via Equation (8); similarly, we determine $E(Y) = E(W_2) + E(W_3) \approx \lambda_2^{1/\nu} + \lambda_3^{1/\nu} - \frac{\nu-1}{\nu}$;
2. $Var(X) = Var(W_1) + Var(W_3) \approx (\lambda_1^{1/\nu} + \lambda_3^{1/\nu})/\nu$ by independence, where $Var(W_i)$ is determined via Equation (9); similarly, we determine $Var(Y) = Var(W_2) + Var(W_3) \approx (\lambda_2^{1/\nu} + \lambda_3^{1/\nu})/\nu$;
3. $E(XY) = E\{(W_1 + W_3)(W_2 + W_3)\} = E\{(W_1W_2) + (W_1W_3) + (W_2W_3) + W_3^2\}$, where $E(W_iW_j) = E(W_i)E(W_j)$ for $i \neq j$, and $E(W_3^2) = Var(W_3) + E^2(W_3)$, where $E(W_i)$ and $Var(W_i)$ can be determined via Equations (8)-(9), respectively;
4. $Cov(X, Y) = E(XY) - (E(X) \cdot E(Y)) \approx \frac{\nu-1}{2\nu}\left[3\left(\frac{\nu-1}{2\nu}\right) - \lambda_1^{1/\nu} - \lambda_2^{1/\nu} - 2\lambda_3^{1/\nu}\right] + \frac{1}{\nu}\lambda_3^{1/\nu}$;
5. $Corr(X, Y) = Cov(X, Y)/\sqrt{Var(X) \cdot Var(Y)}$, where the covariance and respective variances are determined above. While it is not immediately obvious given the approximations presented above, we know that $0 \leq \rho = Corr(X, Y) \leq 1$ since X and Y share a common non-negative component.

The statistical measures associated with the BCMP_T do not have a closed form; however, their approximations aid in determining an approximate GDI. The special case where $\nu = 1$ reduces to the GDI for the Holgate (1964) BP distribution, hence we again find that the distribution is over- or equi-dispersed relative to the uncorrelated BP distribution when $\lambda_3 > (=)\ 0$. For other values of ν, the form of the approximate GDI is not as easily interpreted; therefore, we investigate its behavior computationally. Figure 1 presents approximate GDI plots for certain values of ν when $\lambda_1 = 10.5$, $\lambda_2 = 12$, and $\lambda_3 = (0, 5, 10)$; Figure 1a provides the approximate GDI for $\nu \in [0, 2]$ while Figure 1b focuses on the reduced range $\nu \in [0.5, 1.5]$ to magnify the different functions of GDI with respect to λ_3. These plots illustrate that the BCMP_T is over-dispersed relative to the uncorrelated BP distribution ($GDI > 1$) when $\nu < 1$; this is consistent with our univariate interpretation of ν. At $\nu = 1$, we confirm that $BCMP_T$ is equi- or over-dispersed relative to the uncorrelated BP distribution ($GDI \geq 1$). Finally, for $\nu > 1$, the GDI reveals over-, under-, or equi-dispersion, depending on the combination of λ_3 and ν yet, as ν increases, we more consistently determine $GDI < 1$, hence $BCMP_T$ becomes more under-dispersed relative to the uncorrelated BP distribution as $\nu > 1$ increases. Thus, these GDI values provide evidence of the flexibility of the $BCMP_T$ distribution that demonstrates (to a great extent) an analogous interpretation of dispersion to that regarding the univariate CMP model, modified to account for the dependence intensity λ_3.

4. Statistical inference

This manuscript focuses on the development and statistical inference tools associated with the $BCMP_T$ model. This section particularly discusses parameter estimation details via the method of moments (MOM) and maximum likelihood (ML), respectively, to estimate λ_1, λ_2, λ_3, and ν along with hypothesis testing and statistical computing. The interested reader is referred to Sellers, Morris, and Balakrishnan (2016) or Ong et al. (2021) which discuss these matters as they relate to the $BCMP_C$, and $BCMP_{S1}$ and $BCMP_{S2}$ distributions, respectively.

Conducting parameter estimation via the MOM method, we consider the true $BCMP_T$ moment equations,

$$\mu_x = \lambda_1 \frac{\partial \ln Z(\lambda_1, \nu)}{\partial \lambda_1} + \lambda_3 \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3}, \tag{25}$$

$$\mu_y = \lambda_2 \frac{\partial \ln Z(\lambda_2, \nu)}{\partial \lambda_2} + \lambda_3 \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3}, \tag{26}$$

$$\begin{aligned} \mu_{x^2} = \lambda_1 \frac{\partial \ln Z(\lambda_1, \nu)}{\partial \lambda_1} + \frac{\lambda_1^2}{Z(\lambda_1, \nu)} \frac{\partial^2 Z(\lambda_1, \nu)}{\partial \lambda_1^2} + \lambda_3 \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3} + \frac{\lambda_3^2}{Z(\lambda_3, \nu)} \frac{\partial^2 Z(\lambda_3, \nu)}{\partial \lambda_3^2} \\ + \lambda_1 \lambda_3 \frac{\partial \ln Z(\lambda_1, \nu)}{\partial \lambda_1} \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3}, \end{aligned} \tag{27}$$

$$\begin{aligned} \mu_{y^2} &= \lambda_2 \frac{\partial \ln Z(\lambda_2, \nu)}{\partial \lambda_2} + \frac{\lambda_2^2}{Z(\lambda_2, \nu)} \frac{\partial^2 Z(\lambda_2, \nu)}{\partial \lambda_2^2} + \lambda_3 \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3} + \frac{\lambda_3^2}{Z(\lambda_3, \nu)} \frac{\partial^2 Z(\lambda_3, \nu)}{\partial \lambda_3^2} \\ &\quad + \lambda_2 \lambda_3 \frac{\partial \ln Z(\lambda_2, \nu)}{\partial \lambda_2} \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3}, \end{aligned} \tag{28}$$

$$\begin{aligned} \mu_{xy} &= \lambda_1 \lambda_2 \frac{\partial \ln Z(\lambda_1, \nu)}{\partial \lambda_1} \frac{\partial \ln Z(\lambda_2, \nu)}{\partial \lambda_2} + \lambda_1 \lambda_3 \frac{\partial \ln Z(\lambda_1, \nu)}{\partial \lambda_1} \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3} \\ &\quad + \lambda_2 \lambda_3 \frac{\partial \ln Z(\lambda_2, \nu)}{\partial \lambda_2} \frac{\partial \ln(\lambda_3, \nu)}{\partial \lambda_3} + \lambda_3 \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3} + \frac{\lambda_3^2}{Z(\lambda_3, \nu)} \frac{\partial^2 Z(\lambda_3, \nu)}{\partial \lambda_3^2}, \end{aligned} \tag{29}$$

where $\mu_{g(X,Y)} \doteq E(g(X, Y))$ denotes the generalized expectation for some function $g(X, Y)$, and compare them with their corresponding sampling estimators, $\bar{X}$, $\bar{Y}$, $\overline{X^2}$, $\overline{Y^2}$, $\overline{XY}$.

Closed form solutions do not exist when equating the above respective forms; thus, we instead determine the MOM estimators, $\tilde{\lambda}_i,\ i=1,\ 2,\ 3,$ and $\tilde{\nu}$, by minimizing the squared-error loss function,

$$l(\lambda_1, \lambda_2, \lambda_3, \nu; (\boldsymbol{x}, \boldsymbol{y})) = \left(\mu_x - \bar{X}\right)^2 + \left(\mu_y - \bar{Y}\right)^2 + \left(\mu_{x^2} - \overline{X^2}\right)^2 + \left(\mu_{y^2} - \overline{Y^2}\right)^2 + \left(\mu_{xy} - \overline{XY}\right)^2, \quad (30)$$

where $\mu_x, \mu_y, \mu_{x^2}, \mu_{y^2}, \mu_{xy}$ are defined in Equations (25)-(29). We meanwhile utilize the Delta method to obtain standard errors associated with the MOM estimators; see Appendix A for details. No closed form for the MOM estimates exists; thus, the associated statistical computations are obtained via the `optim` function in R. We use `optim` to optimally solve the loss function (Equation (30)) in order to obtain the MOM estimates, and we use the Delta method approach detailed in Appendix A to calculate the standard errors, where the approximation to the Hessian matrix in Equation (36) is provided in the `optim` output.

Various approaches have been proposed to evaluate the normalizing constant Z (Sellers and Shmueli 2010; Gillispie and Green 2015; Şimşek and Iyengar 2016; Gaunt et al. 2019); for our computations, however, we find summing the first 101 terms of the infinite series to generally be sufficient. To compute the moments of the BCMP_T distribution, we need to evaluate $\frac{\partial Z}{\partial \lambda}\frac{1}{Z}$; however, Z and $\frac{\partial Z}{\partial \lambda}$ become numerically unstable when λ is large or ν is close to zero. To handle this, notice that Z and $\frac{\partial Z}{\partial \lambda}$ are, respectively, the sum of sequences $\{x_n\}_{n\in\mathbb{N}} = \left\{\frac{\lambda^n}{(n!)^\nu}\right\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}} = \left\{\frac{n\lambda^{n-1}}{(n!)^\nu}\right\}_{n\in\mathbb{N}}$. Therefore, we have

$$\frac{\partial Z}{\partial \lambda}\frac{1}{Z} = \frac{\sum y_n}{\sum x_n} = \frac{\sum e^{\ln y_n}}{\sum e^{\ln x_n}} = \frac{\sum e^{\ln y_n - C}}{\sum e^{\ln x_n - C}},$$

where C is a constant that sufficiently shrinks both the numerator and denominator; usually, $C = \max_{n\in\mathbb{N}}\{\ln x_n\}$ suffices.[3] In practice, we first calculate the sequences, $\{\ln x_n\}_{n\in\mathbb{N}}$ and $\{\ln y_n\}_{n\in\mathbb{N}}$; then, we determine C and evaluate $\frac{\partial Z}{\partial \lambda}\frac{1}{Z}$. Additionally, since we approximate $\sum_{n=1}^{\infty} e^{\ln x_n - C}$ using $\sum_{n=1}^{K} e^{\ln x_n - C}$ for some large K, it is important to choose an appropriate K which provides sufficient accuracy to our approximation while maintaining computational speed. Analysts should bear in mind the following situations:

1. for fixed $\nu > 0.01$ and $\lambda \geq 0$, $\frac{\lambda^n}{(n!)^\nu}$ achieves its approximate maximum at $n = \lambda^{1/\nu}$ and it decreases for $n > \lambda^{1/\nu}$. Accordingly, this quantity becomes negligible for $n \gg \lambda^{1/\nu}$, e.g. $n > 2\lambda^{1/\nu}$; $K = \max(\lceil 2\lambda^{1/\nu}\rceil, 100)$ is often a reasonable choice.
2. when $\lambda < 0.99$, we choose K such that λ^K vanishes. Generally, we find that $K \in [100, 100000]$ works well.

A similar approach is used to evaluate $\frac{\partial^2 Z}{\partial \lambda^2}\frac{1}{Z}$.

For $\nu > 0$, the BCMP_T distribution is defined for all $\lambda_i \geq 0,\ i=1,\ 2,\ 3$; for $\nu = 0$, however, $\lambda_i < 1,\ i=1,\ 2,\ 3$. This constraint introduces potential optimization issues when trying to determine the MOM estimators; for example, we do not want the optimization algorithm to enter the parameter space where $\lambda_i > 1$ and $\nu = 0$ for any i. To avoid this, we divide our optimization region into two spaces: (1) $\nu > 0.01$, $\lambda_i \geq 0$, and (2) $0 \leq \nu < 0.2$, $0 \leq \lambda_i < 1$. While the latter constraint is less useful than the former, we maintain its use for completeness. Accordingly, we first obtain the respective MOM estimators under the respective constraints, and then compare their respective loss function values to determine which result is optimal.

To determine the ML estimators (MLEs) of $\lambda_1, \lambda_2, \lambda_3, \nu$, we consider the log-likelihood

$$\ln L(\lambda_1, \lambda_2, \lambda_3, \nu; (\boldsymbol{x}, \boldsymbol{y})) = \ln \prod_{i=1}^{n} p(x_i, y_i) = \sum_{i=1}^{n} \ln p(x_i, y_i), \tag{31}$$

where

$$\ln p(x_i, y_i) = x_i \ln \lambda_1 + y_i \ln \lambda_2 - \ln Z(\lambda_1, \nu) - \ln Z(\lambda_2, \nu) - \ln Z(\lambda_3, \nu) - \nu\left[\ln(x_i!) + \ln(y_i!)\right] + \ln\left(\sum_{w=0}^{\min(x_i, y_i)} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right).$$

The corresponding normal equations are

$$\frac{\partial \ln L}{\partial \lambda_1} = \sum_{i=1}^{n} \frac{\partial \ln p(x_i, y_i)}{\partial \lambda_1} = \sum_{i=1}^{n} \left(\frac{x_i}{\lambda_1} - \frac{\partial \ln Z(\lambda_1, \nu)}{\partial \lambda_1} - \frac{1}{\left(\sum_{w=0}^{\min(x_i, y_i)} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right)} \left(\sum_{w=0}^{\min(x_i, y_i)} \frac{w}{\lambda_1} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right)\right)$$

$$\frac{\partial \ln L}{\partial \lambda_2} = \sum_{i=1}^{n} \frac{\partial \ln p(x_i, y_i)}{\partial \lambda_2} = \sum_{i=1}^{n} \left(\frac{y_i}{\lambda_2} - \frac{\partial \ln Z(\lambda_2, \nu)}{\partial \lambda_2} - \frac{1}{\left(\sum_{w=0}^{\min(x_i, y_i)} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right)} \left(\sum_{w=0}^{\min(x_i, y_i)} \frac{w}{\lambda_2} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right)\right)$$

$$\frac{\partial \ln L}{\partial \lambda_3} = \sum_{i=1}^{n} \frac{\partial \ln p(x_i, y_i)}{\partial \lambda_3} = \sum_{i=1}^{n} \left(-\frac{\partial \ln Z(\lambda_3, \nu)}{\partial \lambda_3} + \frac{1}{\left(\sum_{w=0}^{\min(x_i, y_i)} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right)} \left(\sum_{w=0}^{\min(x_i, y_i)} \frac{w}{\lambda_3} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right)\right)$$

$$\frac{\partial \ln L}{\partial \nu} = \sum_{i=1}^{n} \frac{\partial \ln p(x_i, y_i)}{\partial \nu} = \sum_{i=1}^{n} \left(-\frac{\partial \ln Z(\lambda_1, \nu)}{\partial \nu} - \frac{\partial \ln Z(\lambda_2, \nu)}{\partial \nu} - \frac{\partial \ln Z(\lambda_3, \nu)}{\partial \nu} - [\ln(x_i!) + \ln(y_i!)] + \frac{1}{\left(\sum_{w=0}^{\min(x_i, y_i)} \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w \left[\binom{x_i}{w}\binom{y_i}{w} w!\right]^\nu\right)}\right.$$

$$\times\left(\sum_{w=0}^{\min(x_i,y_i)}\left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^w\left[\binom{x_i}{w}\binom{y_i}{w}w!\right]^\nu \ln\left[\binom{x_i}{w}\binom{y_i}{w}w!\right]\right)\Bigg).$$

These score equations do not have a closed-form solution, thus we circumvent this issue by utilizing statistical computing in R. We use the `optim` function to optimize the negated log-likelihood, recognizing the constraint that $\nu \geq 0$ and $\lambda_i > 0$ for all $i = 1$, 2, 3. The corresponding standard errors associated with these MLEs $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\nu})$ may be determined by utilizing the approximate Hessian matrix provided with the `optim` output. Computational matters discussed above with regard to calculating Z for the MOM estimates are likewise utilized here to determine the MLEs.

4.1. Hypothesis testing

Two hypotheses tests are of interest, given this distributional development. The first inquiry regards detecting the existence of statistically significant data dispersion in the bivariate data set such that an assumed bivariate Poisson distribution would be an inappropriate model to describe the data. Accordingly, we conduct a hypothesis test where the null hypothesis assumes data equi-dispersion $(H_{0,\nu} : \nu = 1)$ such that the bivariate Poisson model is reasonable, and the alternative hypothesis (i.e., $H_{1,\nu} : \nu \neq 1$) suggests sufficient data over- or under-dispersion such that the trivariate-reduced BCMP model is more appropriate. We are not concerned with the direction of the data dispersion because our model can accommodate both forms. We use the likelihood ratio test (LRT) statistic,

$$\Lambda_\nu = \frac{\max_{\Omega_{0,\nu}} L(\lambda_1, \lambda_2, \lambda_3, \nu = 1)}{\max_\Omega L(\lambda_1, \lambda_2, \lambda_3, \nu)},$$

to draw inference, where Ω and $\Omega_{0,\nu}$ denote the respective parameter spaces in general and under the null hypothesis, $H_{0,\nu}$. Noting that the null hypothesis represents the special bivariate Poisson case, we use the `bivpois` package in R to conduct analyses (Karlis and Ntzoufras 2005). The second investigation seeks to determine if a statistically significant amount of dependence exists between the two random variables, X and Y. Given that X and Y were determined via trivariate reduction with a shared parameter λ_3, the question of independence reduces to testing whether $H_{0,\lambda} : \lambda_3 = 0$ or $H_{1,\lambda} : \lambda_3 > 0$. For this test, we have the LRT statistic,

$$\Lambda_\lambda = \frac{\max_{\Omega_{0,\lambda}} L(\lambda_1, \lambda_2, \lambda_3 = 0, \nu)}{\max_\Omega L(\lambda_1, \lambda_2, \lambda_3, \nu)},$$

where $\Omega_{0,\lambda}$ denotes the parameter spaces under $H_{0,\lambda}$.

The $\max_{\Omega_{0,\nu}} L(\lambda_1, \lambda_2, \lambda_3, \nu = 1)$ is obtained via `bivpois`, while $\max_\Omega L(\lambda_1, \lambda_2, \lambda_3, \nu)$ is provided among the `optim` output. Meanwhile, we obtain $\max_{\Omega_{0,\lambda}} L(\lambda_1, \lambda_2, \lambda_3 = 0, \nu)$ via an analogous `optim` computation that optimizes the resulting function under this constrained space. We utilize the distributional theory result that $-2\ln\Lambda_\nu$ converges to a chi-squared distribution with one degree of freedom and draw inference from the resulting p-value. Moreover, taking into account the boundary condition for the second test, the asymptotic null distribution of the LRT statistic $-2\ln\Lambda_\lambda$ is a mixture of chi-square distributions (Balakrishnan and Pal 2013).

Table 1. MOM and ML estimates with respective standard errors (in parentheses) associated with two bivariate Poisson data simulations stemming from generated data sets of size $n = \{100, 250, 500, 1000\}$ where (1) $\lambda_1 = 0.3$, $\lambda_2 = 1.6$, $\lambda_3 = 2.6$; and (2) $\lambda_1 = 2.3$, $\lambda_2 = 1.9$, $\lambda_3 = 0.5$. While the data simulated are from a bivariate Poisson model, we recognize that this distribution equals a trivariate-reduced BCMP model where $\nu = 1$, thus we compare the estimated dispersion values to this true value.

		Sample size			
Truth	Method	100	250	500	1000
$\lambda_1 = 0.3$	MLE	0.138 (0.083)	0.230 (0.061)	0.307 (0.055)	0.283 (0.037)
	MOM	0.216 (0.158)	0.269 (0.095)	0.260 (0.070)	0.344 (0.057)
$\lambda_2 = 1.6$	MLE	1.576 (0.254)	1.346 (0.141)	1.686 (0.142)	1.543 (0.088)
	MOM	1.894 (0.312)	1.554 (0.201)	1.653 (0.155)	1.721 (0.117)
$\lambda_3 = 2.6$	MLE	2.598 (0.468)	2.375 (0.273)	2.706 (0.231)	2.586 (0.153)
	MOM	2.790 (0.357)	2.712 (0.324)	2.702 (0.221)	2.782 (0.168)
$\nu = 1$	MLE	0.983 (0.147)	0.923 (0.096)	1.093 (0.075)	0.982 (0.049)
	MOM	1.088 (0.097)	1.049 (0.102)	1.082 (0.074)	1.062 (0.052)
$\lambda_1 = 2.3$	MLE	2.204 (0.550)	2.272 (0.367)	2.419 (0.257)	2.209 (0.166)
	MOM	2.100 (0.457)	2.268 (0.359)	2.413 (0.245)	2.190 (0.168)
$\lambda_2 = 1.9$	MLE	1.899 (0.476)	1.896 (0.308)	2.003 (0.213)	1.804 (0.136)
	MOM	1.826 (0.417)	1.912 (0.288)	1.992 (0.204)	1.794 (0.144)
$\lambda_3 = 0.5$	MLE	0.736 (0.260)	0.716 (0.168)	0.440 (0.117)	0.562 (0.078)
	MOM	0.717 (0.235)	0.716 (0.182)	0.436 (0.117)	0.558 (0.079)
$\nu = 1$	MLE	1.052 (0.171)	1.084 (0.111)	1.035 (0.072)	0.993 (0.051)
	MOM	1.004 (0.156)	1.086 (0.104)	1.030 (0.067)	0.984 (0.052)

5. Examples

5.1. Simulated data examples

We consider three data simulations to illustrate the flexibility of the BCMP_T distribution in its ability to represent three special cases: the bivariate Poisson (BP); and the bivariate Bernoulli and geometric distributions, each obtained via trivariate reduction. In all cases, we consider the sample sizes $n = \{100, 250, 500, 1000\}$ to study the parameter estimate accuracy as n increases, and we consider both the MOM and ML estimation approaches to compare results with respect to the considered estimation procedures.

For the BP example, we consider two data simulations: (1) $\lambda_1 = 0.3$, $\lambda_2 = 1.6$, $\lambda_3 = 2.6$; and (2) $\lambda_1 = 2.3$, $\lambda_2 = 1.9$, $\lambda_3 = 0.5$. In both cases, the generated data further set $\nu = 1$ to reflect the BCMP_T model constrained to reflect an assumed BP distribution. Table 1 provides the respective MLE and MOM estimates, along with the respective corresponding standard errors provided in parentheses. In both simulation examples, we see that both the MOM and MLE methods recognize the distributional form as BP (i.e., neither $\tilde{\nu}$ nor $\hat{\nu}$ is statistically different from 1 based on the respective 95% confidence intervals). Further, both models reasonably estimate the true λ_i for all i, obtaining estimates within one standard error of the true parameter. In particular, the amount of dependence represented via λ_3 is likewise estimated reasonably. While estimator accuracy does not necessarily converge to the true parameter as the sample size increases, we still see a decreasing trend in the standard errors as n gets large.

For the case of a trivariate-reduced bivariate Bernoulli distribution example, we again consider two data simulations, where $\nu \to \infty$ for the trivariate-reduced BCMP model: (1) $\lambda_1 = 3$, $\lambda_2 = 4$, $\lambda_3 = 0.5$; and (2) $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\lambda_3 = 3$. Table 2 provides the respective MLE and MOM estimates assuming a BCMP_T distribution, along with the respective corresponding standard errors provided in parentheses. In the first

Table 2. MOM and ML estimates with respective standard errors (in parentheses) associated with two trivariate-reduced bivariate Bernoulli data simulations stemming from generated data sets of size $n = \{100, 250, 500, 1000\}$ where (1) $\lambda_1 = 3$, $\lambda_2 = 4$, $\lambda_3 = 0.5$; and (2) $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\lambda_3 = 3$. While the data simulated are from a trivariate-reduced bivariate Bernoulli model, we recognize that this distribution equals a trivariate-reduced BCMP model where $\nu \rightarrow \infty$, thus we compare the estimated dispersion values to this truth.

		Sample size			
Truth	Method	100	250	500	1000
$\lambda_1 = 3$	MLE	2.661 (0.622)	3.531 (0.571)	2.900 (0.315)	2.959 (0.228)
	MOM	0.189 (0.053)	0.275 (0.048)	0.245 (0.030)	0.266 (0.023)
$\lambda_2 = 4$	MLE	5.529 (1.630)	3.225 (0.508)	4.144 (0.503)	4.220 (0.363)
	MOM	0.330 (0.093)	0.301 (0.051)	0.335 (0.041)	0.319 (0.028)
$\lambda_3 = 0.5$	MLE	0.415 (0.095)	0.481 (0.068)	0.573 (0.056)	0.574 (0.040)
	MOM	4.719 (1.329)	4.065 (0.715)	4.869 (0.647)	4.871 (0.462)
$\nu = \infty$	MLE	25.880 (723.300)	27.40 (727.400)	27.29 (484.300)	29.510 (730.100)
	MOM	6.466 (1.1610)	5.782 (0.556)	5.745 (0.311)	5.719 (0.216)
$\lambda_1 = 1$	MLE	0.889 (0.193)	1.480 (0.231)	1.219 (0.127)	0.947 (0.068)
	MOM	0.984 (0.242)	1.345 (0.219)	1.296 (0.181)	0.973 (0.088)
$\lambda_2 = 0.5$	MLE	0.564 (0.128)	0.636 (0.100)	0.551 (0.060)	0.537 (0.041)
	MOM	0.616 (0.159)	0.584 (0.110)	0.591 (0.084)	0.560 (0.053)
$\lambda_3 = 3$	MLE	3.746 (1.034)	2.197 (0.369)	3.076 (0.385)	2.839 (0.241)
	MOM	3.216 (1.057)	2.267 (0.476)	2.834 (0.473)	2.717 (0.287)
$\nu = \infty$	MLE	27.850 (1360.800)	26.570 (623.500)	29.510 (1142.000)	28.470 (625.300)
	MOM	29.230 (11.900)	7.864 (1.646)	22.680 (6.867)	21.930 (3.483)

simulation, the MOM approach produced heavily biased estimates, while the MLE method performed reasonably well, producing estimates ($\hat{\lambda}_i$ for all i) that are generally within one standard error of the true parameter. We believe these results occur because the true likelihood with these parameters has a near plateau or at least an approximate local minimum near the point $\{\tilde{\lambda}_1 = 0.25, \tilde{\lambda}_2 = 0.33, \tilde{\lambda}_3 = 4.6, \tilde{\nu} = 5\}$, i.e., the optimization solution may not be unique. This demonstrates the impact that the starting values for the optimization procedure have on the resulting MOM estimate, and that the MOM estimate may not be reliable. The discrepancy between the true values and MOM estimates regarding the first simulation appears to be further due to the dependence parameter, λ_3. While this true distribution does not assume a strong association between X and Y, the MOM approach appears to assume that the codependent intensity term is the major contributor to the distributional form. In contrast, the second simulation assumes a stronger codependent intensity $\lambda_3 = 3$, and the corresponding MOM estimates now appear more reasonable, capturing the true intensity within 1-2 standard errors of the associated estimates.

The MLE and MOM estimates for the true dispersion at first appear considerably small in relation to the theorized infinite value, with the MOM estimator $\tilde{\nu}$ substantially smaller than its MLE counterpart. The MLE results, however, speak to the computational impact of the dispersion parameter in the CMP models; these results are consistent with those obtained from conducting parameter estimation with underlying CMP structures (Sellers and Shmueli 2010; Sellers 2012; Sellers, Morris, and Balakrishnan 2016; Sellers and Raim 2016). As seen in these cases, ν at approximately 30 effectively represents a plateau in the log-likelihood. Meanwhile, the large standard error associated with $\hat{\nu}$ demonstrates the computational effort to estimate infinity. The second example, however, shows that the MOM and MLE methods both produce reasonable estimates of

Table 3. MOM and ML estimates with respective standard errors (in parentheses) associated with two trivariate-reduced bivariate geometric data simulations stemming from generated data sets of size $n = \{100, 250, 500, 1000\}$ where (1) $\lambda_1 = 0.8$, $\lambda_2 = 0.5$, $\lambda_3 = 0.2$; and (2) $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\lambda_3 = 0.7$. While the data simulated are from a trivariate-reduced bivariate geometric model, we recognize that this distribution equals a trivariate-reduced BCMP model where $\nu = 0$, thus we compare the estimated dispersion values to this true value.

		Sample size			
Truth	Method	100	250	500	1000
$\lambda_1 = 0.8$	MLE	0.902 (0.126)	0.886 (0.067)	0.791 (0.036)	0.795 (0.026)
	MOM	0.962 (0.132)	0.779 (0.007)	0.790 (0.005)	0.990 (0.047)
$\lambda_2 = 0.5$	MLE	0.506 (0.102)	0.531 (0.049)	0.474 (0.030)	0.486 (0.022)
	MOM	0.563 (0.087)	0.430 (0.077)	0.428 (0.071)	0.576 (0.039)
$\lambda_3 = 0.2$	MLE	0.325 (0.118)	0.249 (0.070)	0.272 (0.042)	0.265 (0.033)
	MOM	0.210 (0.167)	0.333 (0.100)	0.326 (0.089)	0.183 (0.082)
$\nu = 0$	MLE	0.079 (0.073)	0.063 (0.039)	0.000 (0.022)	0.000 (0.016)
	MOM	0.105 (0.050)	0.000 (0.017)	0.000 (0.011)	0.100 (0.016)
$\lambda_1 = 0.2$	MLE	0.248 (0.043)	0.220 (0.025)	0.195 (0.017)	0.200 (0.012)
	MOM	0.281 (0.058)	0.235 (0.039)	0.178 (0.028)	0.120 (0.054)
$\lambda_2 = 0.1$	MLE	0.117 (0.035)	0.124 (0.022)	0.112 (0.015)	0.105 (0.010)
	MOM	0.195 (0.045)	0.138 (0.044)	0.094 (0.033)	0.007 (0.061)
$\lambda_3 = 0.7$	MLE	0.837 (0.110)	0.741 (0.057)	0.697 (0.035)	0.703 (0.025)
	MOM	0.830 (0.102)	0.781 (0.028)	0.752 (0.041)	0.784 (0.021)
$\nu = 0$	MLE	0.140 (0.095)	0.050 (0.052)	0.000 (0.032)	0.000 (0.022)
	MOM	0.143 (0.083)	0.087 (0.031)	0.044 (0.023)	0.055 (0.010)

the true parameters, λ_i, $i = 1, 2, 3$. The difference in performance for the MOM estimators occurs because different sets of parameters can produce similar moments.

Whether or not the estimators reasonably approximate the true corresponding parameters, we still generally see a decreasing trend in the λ_i estimator standard errors as n gets large, while the dispersion estimator standard errors lack robustness due to the computational implications described above.

Finally, we consider two data simulations regarding a trivariate-reduced geometric (i.e., $\nu = 0$) distribution, where (1) $\lambda_1 = 0.8$, $\lambda_2 = 0.5$, $\lambda_3 = 0.2$; and (2) $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\lambda_3 = 0.7$. Table 3 provides the respective MLE and MOM estimates assuming a BCMP_T distribution, along with the respective corresponding standard errors provided in parentheses. Again, the ML estimation procedure seems to outperform the MOM approach; this makes sense because the MOM estimation requires less information than the MLEs. For instance, we see that, for the first example where $n = 1000$, the MOM estimates include $\tilde{\lambda}_1 = 0.99$ and $\tilde{\nu} = 0.1$ with standard errors that are sufficiently small such that these estimators are significantly biased. These results indicate that the MOM estimators are not robust, even with a large sample size. Both estimators appear to perform better in the latter example, with the MLE outperforming the MOM procedure. In particular, the ML estimation more accurately estimates the dispersion as $\hat{\nu} = 0.000$ while the MOM procedure produces $\tilde{\nu} = 0.055$ with a standard error that is too small for $\nu = 0$ to be considered as a possible value. Nonetheless, the size of the dependence intensity, λ_3, does not appear to influence the MLE and MOM approaches toward proper estimation. Both approaches reasonably estimate λ_i, $i = 1, 2, 3$ for both simulated exercises.

Table 4. Maximum likelihood estimates, along with the corresponding log-likelihood, Akaike Information Criterion (AIC), and χ^2 goodness-of-fit values for various bivariate models: bivariate Poisson (BP); bivariate negative binomial (BNB); bivariate generalized Poisson (BGP); and four bivariate CMP using the compounding, Sarmanov, and trivariate reduction approaches ($BCMP_C$, $BCMP_{S1}$, $BCMP_{S2}$, and $BCMP_T$), respectively on shunter accidents data set.

Model	Estimated Parameters			LogLik	AIC	χ^2	p-value
BP	$\hat{\lambda}_1 = 0.717$	$\hat{\lambda}_2 = 1.012$	$\hat{\lambda}_3 = 0.258$	−345.635	697.3	48.05	0.13
BNB	$\hat{m} = 0.891$	$\hat{r} = 3.876$	$\hat{\alpha}_1 = 1.331$	−341.610	691.2	21.92	0.97
	$\hat{\alpha}_2 = 0.095$						
BGP	$\hat{\theta}_1 = 0.560$	$\hat{\theta}_2 = 0.837$	$\hat{\theta}_3 = 0.305$	−341.513	695.0	23.59	0.93
	$\hat{\lambda}_1 = 0.151$	$\hat{\lambda}_2 = 0.123$	$\hat{\lambda}_3 = 0.031$				
$BCMP_C$	$\hat{\lambda} = 1.328$	$\hat{\nu} = 0.084$	$\hat{p}_{00} = 0.939$	−341.704	695.4	22.16	0.95
	$\hat{p}_{01} = 0.034$	$\hat{p}_{10} = 0.025$	$\hat{p}_{11} = 0.002$				
$BCMP_{S1}$	$\hat{\lambda}_1 = 0.92$	$\hat{\lambda}_2 = 0.73$	$\hat{\alpha} = 0.58$	−345.550	703.1	37.06	0.37
	$\hat{\nu}_1 = 0.57$	$\hat{\nu}_2 = 0.53$	$\hat{\tau} = 1.00$				
$BCMP_{S2}$	$\hat{\lambda}_1 = 0.94$	$\hat{\lambda}_2 = 0.75$	$\hat{\nu}_1 = 0.59$	−343.500	697.0	31.46	0.68
	$\hat{\nu}_2 = 0.56$	$\hat{\tau} = 1.00$					
$BCMP_T$	$\hat{\lambda}_1 = 0.517$	$\hat{\lambda}_2 = 0.684$	$\hat{\lambda}_3 = 0.270$	−342.009	692.0	23.36	0.96
	$\hat{\nu} = 0.438$						

5.2. Real data example: shunter accidents

Several works regarding bivariate discrete data analysis consider this data set that reports the number of accidents incurred by 122 shunters in two consecutive year periods (1937–1942 and 1943–1947); X and Y denote the number of shunter accidents between 1937–1942 and 1943–1947, respectively. This data set is recognized as being over-dispersed, thus numerous works argue against the use of the bivariate Poisson to analyze the data (Arbous and Kerrich 1951; Adelstein 1952; Famoye and Consul 1995; Sellers, Morris, and Balakrishnan 2016; Ong et al. 2021). Arbous and Kerrich (1951) and Adelstein (1952) use a BNB model to fit the data, while Famoye and Consul (1995) consider a BGP distribution. Sellers, Morris, and Balakrishnan (2016) utilize the $BCMP_C$ model to analyze the data, while Ong et al. (2021) examine two models motivated by the Sarmanov family, $BCMP_{S1}$ and $BCMP_{S2}$. Meanwhile, because the ML estimation outperforms the MOM procedure in simulated examples, we focus our attention on comparing the MLEs from the aforementioned models and those from the $BCMP_T$ model. Table 4 provides the MLEs for the respective models, along with their respective log-likelihood, Akaike Information Criterion (AIC), and chi-square goodness-of-fit (GoF) test statistic values. Adopting the Burnham and Anderson (2002) approach, we compare model performance via $\Delta_i = AIC_i - AIC_{\min}$, where AIC_i denotes the AIC associated with Model i, and $AIC_{\min}$ is the minimum AIC among the considered models. Table 5 supplies the levels of model support based on recommended Δ_i ranges.

While the BNB model is the optimal model based on AIC (691.2), the $BCMP_T$ model is the optimal BCMP representation based on AIC (692.0) and is the only model found to offer substantial empirical support in comparison to the BNB model ($\Delta = 0.8$). This occurs because the $BCMP_T$ distribution likewise only requires four parameters to model the data, and the resulting log-likelihood stemming from its MLEs is near optimal (LogLik $= -342.009$) in comparison with the other resulting log-likelihood values. Meanwhile, the other considered models (BP, BGP, $BCMP_C$, $BCMP_{S1}$, $BCMP_{S2}$) are found to have (considerably) less to essentially no empirical support in comparison to the BNB model ($\Delta \in (3.8, 11.9)$). This occurs because the other BCMP models require

Table 5. Levels of model support based on AIC difference values, $\Delta_i = AIC_i - AIC_{\min}$, for Model i (Burnham and Anderson 2002).

Δ_i	Empirical Support Level for Model i
$[0, 2]$	Substantial
$[4, 7]$	Considerably less
$(10, \infty)$	Essentially none

more parameters yet still produce log-likelihoods that are no better than the BCMP_T optimal log-likelihood. In particular, the MLEs associated with the BCMP models derived via the Sarmanov family produce log-likelihood values that align with the BP model, which has repeatedly been scrutinized for poor performance relative to this data set. Further, because the BCMP distributions derived via the Sarmanov family require more estimated parameters, these models produce two of the largest AIC values among the considered models. Thus, for this example, there does not appear to be any offered additional benefit to considering a model that allows for the full correlation range.

The BCMP_T model detects statistically significant data-overdispersion, estimating the associated dispersion parameter at $\hat{\nu} = 0.438$ ($-2\log\Lambda_\nu = 7.25$; p-value < 0.01); this agrees with the results from the other considered models that allow for data dispersion (Arbous and Kerrich 1951; Adelstein 1952; Famoye and Consul 1995; Sellers, Morris, and Balakrishnan 2016; Ong et al. 2021). Meanwhile, $\hat{\lambda}_3 = 0.27$ implies a relatively small level of positive association between the two random variables. The associated LRT statistic $-2\log\Lambda_\lambda = 9.0$ produces a small p-value < 0.01 confirming that statistically significant dependence exists.

The raw data, along with the estimated frequencies determined from the respective model MLEs are provided in Tables 6 and 7. The BCMP_T model appears to reasonably estimate the observed number of shunter accidents over the combination of respective time periods, thus providing comparable marginal distributions as well. In fact, its joint and marginal estimates appear to be approximately equal to those from the BNB, BCMP_C, and BGP models. In order to better assess GoF, approximate chi-square test statistics are also used to evaluate the models with degrees of freedom equal to $k - c - 1$, where k equals the number of bins and c is the number of estimated parameters. For these GoF tests, we maintain the bin structure with regard to Y while combining the bins for $X \geq 5$ in Tables 6 and 7 in order to account for small expected frequencies, thus we have $k = 42$. We consider these tests to have approximate chi-square distributions since, even after rebinning, some of the expected frequencies remain less than 0.25 (a threshold noted by Koehler and Larntz (1980)). Nevertheless, the results of these GoF tests are consistent with aforementioned interpretations of model adequacy based on the AIC; revisiting Table 4, we note that large p-values suggest stronger support for the associated model. For instance, the GoF tests support the BNB ($\chi^2 = 21.92$, p-value $= 0.97$) as the best fitting model followed closely by the BCMP_T ($\chi^2 = 23.36$, p-value $= 0.96$), BCMP_C ($\chi^2 = 22.16$, p-value $= 0.95$), and BGP models ($\chi^2 = 23.59$, p-value $= 0.93$).

5.3. Real data example: NBA data

This section considers an under-dispersed data set to further illustrate the flexibility of the BCMP_T model. This bivariate data set consists of the number of players selected for

Table 6. Observed accident counts for 122 shunters with associated estimated counts from various bivariate distributions: bivariate Poisson (BP); bivariate negative binomial (BNB); bivariate generalized Poisson (BGP); and four bivariate CMP distributions resulting from the compounding, Sarmanov, and trivariate reduction approaches, respectively ($BCMP_C$, $BCMP_{S1}$, $BCMP_{S2}$, and $BCMP_T$). Estimated counts are determined from MLEs for respective model parameters reported in Table 4.

x		$y=0$	$y=1$	$y=2$	$y=3$	$y=4$	$y=5$	$y=6+$	
0	OBS	21	18	8	2	1	–	–	50
	BP	16.72	16.92	8.56	2.89	0.73	0.15	0.02	46.00
	BNB	21.90	16.67	7.98	3.07	1.04	0.32	0.13	51.11
	BGP	22.21	16.44	7.88	3.12	1.11	0.37	0.17	51.32
	$BCMP_C$	22.48	16.10	7.88	3.14	1.09	0.34	0.13	51.16
	$BCMP_{S1}$	18.08	16.60	10.02	4.82	1.99	0.73	0.24	52.47
	$BCMP_{S2}$	21.09	15.44	8.63	4.05	1.65	0.59	0.19	51.65
	$BCMP_T$	22.52	15.40	7.77	3.28	1.22	0.41	0.18	50.78
1	OBS	13	14	10	1	4	1	–	43
	BP	11.99	16.45	10.51	4.28	1.27	0.29	0.07	44.87
	BNB	12.52	13.18	8.06	3.77	1.50	0.53	0.26	39.83
	BGP	10.70	14.51	8.67	3.84	1.46	0.51	0.25	39.93
	$BCMP_C$	12.11	12.94	8.14	3.90	1.57	0.55	0.25	39.46
	$BCMP_{S1}$	12.98	11.95	7.38	3.62	1.51	0.56	0.18	38.18
	$BCMP_{S2}$	11.56	12.43	8.12	4.05	1.69	0.61	0.20	38.66
	$BCMP_T$	11.64	14.04	8.17	3.79	1.52	0.54	0.26	39.96
2	OBS	4	5	4	2	1	0	1	17
	BP	4.30	7.45	5.89	2.89	1.01	0.27	0.07	21.88
	BNB	4.50	6.06	4.54	2.52	1.16	0.47	0.26	19.52
	BGP	3.97	6.11	4.93	2.55	1.06	0.39	0.19	19.20
	$BCMP_C$	4.46	6.12	4.68	2.62	1.20	0.47	0.23	19.78
	$BCMP_{S1}$	6.36	5.90	3.82	1.95	0.83	0.31	0.10	19.28
	$BCMP_{S2}$	5.08	6.48	4.43	2.25	0.94	0.34	0.11	19.63
	$BCMP_T$	4.44	6.18	4.89	2.56	1.12	0.43	0.22	19.83
3	OBS	2	1	3	2	0	1	0	9
	BP	1.03	2.15	2.05	1.20	0.49	0.15	0.05	7.11
	BNB	1.30	2.13	1.90	1.23	0.65	0.30	0.19	7.69
	BGP	1.35	2.18	1.92	1.19	0.56	0.22	0.12	7.53
	$BCMP_C$	1.34	2.21	1.97	1.26	0.64	0.28	0.16	7.85
	$BCMP_{S1}$	2.54	2.37	1.59	0.83	0.36	0.14	0.05	7.87
	$BCMP_{S2}$	1.94	2.65	1.84	0.94	0.39	0.14	0.05	7.95
	$BCMP_T$	1.42	2.17	1.93	1.25	0.61	0.25	0.14	7.77
4	OBS	0	0	1	1	–	–	–	2
	BP	0.18	0.45	0.51	0.35	0.17	0.06	0.02	1.73
	BNB	0.33	0.64	0.66	0.49	0.29	0.15	0.11	2.66
	BGP	0.45	0.73	0.67	0.44	0.23	0.10	0.06	2.67
	$BCMP_C$	0.35	0.67	0.68	0.48	0.27	0.13	0.08	2.66
	$BCMP_{S1}$	0.88	0.82	0.56	0.30	0.13	0.05	0.02	2.76
	$BCMP_{S2}$	0.65	0.91	0.64	0.33	0.14	0.05	0.02	2.74
	$BCMP_T$	0.40	0.66	0.64	0.46	0.26	0.12	0.08	2.60

the Center (C) and the Forward (F) positions from the All-Star game rosters of the 2000–2016 National Basketball Association (NBA) (National Basketball Association 2020). To conduct model comparisons for these bivariate data, we consider the bivariate Poisson (BP); bivariate negative binomial (BNB); bivariate generalized Poisson (BGP); and bivariate CMP models attained via compounding ($BCMP_C$) and trivariate-reduction ($BCMP_T$), respectively. Table 8 reports the MLEs, log-likelihood, number of parameters, and AIC values for the aforementioned models.

The $BCMP_T$ model produces the largest log-likelihood (-46.262) and smallest AIC, $AIC_{\min} = 100.521$, making it the optimal model among the considered models based on AIC. The BGP and $BCMP_C$ model estimates meanwhile produce AICs that are close in value with the BGP slightly outperforming the $BCMP_C$. The respective AICs are

Table 7. Continued: Observed accident counts for 122 shunters with associated estimated counts from various bivariate distributions: bivariate Poisson (BP); bivariate negative binomial (BNB); bivariate generalized Poisson (BGP); and four bivariate CMP distributions resulting from the compounding, Sarmanov, and trivariate reduction approaches, respectively ($BCMP_C$, $BCMP_{S1}$, $BCMP_{S2}$, and $BCMP_T$). Estimated counts are determined from MLEs for respective model parameters reported in Table 4.

x		$y=0$	$y=1$	$y=2$	$y=3$	$y=4$	$y=5$	$y=6+$	
5	OBS	–	–	–	–	–	–	–	0
	BP	0.03	0.07	0.10	0.08	0.04	0.02	0.01	0.34
	BNB	0.08	0.17	0.20	0.17	0.11	0.06	0.05	0.84
	BGP	0.15	0.24	0.22	0.15	0.08	0.04	0.03	0.90
	$BCMP_C$	0.08	0.18	0.20	0.16	0.10	0.05	0.03	0.80
	$BCMP_{S1}$	0.27	0.26	0.18	0.09	0.04	0.02	0.01	0.86
	$BCMP_{S2}$	0.20	0.28	0.20	0.10	0.04	0.02	0.01	0.83
	$BCMP_T$	0.10	0.18	0.18	0.14	0.09	0.05	0.03	0.78
6	OBS	–	–	–	–	–	–	–	0
	BP	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.06
	BNB	0.02	0.04	0.06	0.05	0.04	0.02	0.02	0.25
	BGP	0.05	0.08	0.07	0.05	0.03	0.01	0.01	0.30
	$BCMP_C$	0.02	0.04	0.05	0.05	0.03	0.02	0.01	0.22
	$BCMP_{S1}$	0.08	0.07	0.05	0.03	0.01	0.00	0.00	0.24
	$BCMP_{S2}$	0.05	0.08	0.05	0.03	0.01	0.00	0.00	0.23
	$BCMP_T$	0.02	0.04	0.05	0.04	0.03	0.02	0.01	0.21
7+	OBS	–	1	0	–	–	–	–	1
	BP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	BNB	0.00	0.01	0.02	0.02	0.02	0.01	0.02	0.10
	BGP	0.02	0.04	0.03	0.02	0.01	0.01	0.00	0.14
	$BCMP_C$	0.00	0.01	0.02	0.01	0.01	0.01	0.00	0.07
	$BCMP_{S1}$	0.02	0.02	0.01	0.01	0.00	0.00	0.27	0.34
	$BCMP_{S2}$	0.01	0.02	0.01	0.01	0.00	0.00	0.25	0.31
	$BCMP_T$	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.07
	OBS	40	39	26	8	6	2	1	122
	BP	34.24	43.51	27.64	11.70	3.72	0.94	0.24	121.98
	BNB	40.65	38.90	23.41	11.32	4.80	1.87	1.02	122.00
	BGP	38.90	40.32	24.38	11.36	4.55	1.65	0.83	122.00
	$BCMP_C$	40.84	38.27	23.62	11.62	4.91	1.84	0.90	122.00
	$BCMP_{S1}$	41.21	37.99	23.61	11.65	4.87	1.81	0.87	122.01
	$BCMP_{S2}$	40.58	38.29	23.92	11.76	4.86	1.75	0.83	121.99
	$BCMP_T$	40.54	38.67	23.66	11.54	4.85	1.82	0.91	121.96

$AIC_{\text{BGP}} = 105.322$ and $AIC_{\text{BCMP}_\text{C}} = 105.972$ with respective difference measures $\Delta_{\text{BGP}} = 4.801$ and $\Delta_{\text{BCMP}_\text{C}} = 5.451$ in relation to the optimal model with regard to AIC, BCMP_T. The BCMP_T model recognizes this data set as being statistically significantly under-dispersed ($\hat{\nu}_T = 3.515 > 1;\ -2\log\Lambda_\nu = 16,\ \text{p-value} < 0.0001$). Similarly, the BCMP_C and BGP models likewise detect data under-dispersion ($\hat{\nu}_C = 8.370 > 1$, and $\hat{\lambda}_2, \hat{\lambda}_3 < 0$). Thus, even though the BGP and BCMP_C models have "considerably less" empirical support than the BCMP_T (Burnham and Anderson 2002), they demonstrate themselves to be more effective than the BP and BNB distributions in modeling these data. As previously noted, the BP and BNB models cannot address data under-dispersion. This example illustrates that, under such circumstances, the (B)NB model can only perform at best as well as the (B)P model in estimating under-dispersed data because the (B)NB distribution only allows for data equi- or over-dispersion (Hilbe 2007). Accordingly, we see $\hat{r} \to \infty$, and the respective log-likelihood values approximately equaling each other, while the added parameter for the BNB produces a larger AIC than that for the BP model. Both the BP and BNB models produce AICs that show essentially no empirical support relative to the BCMP_T ($AIC_{\text{BP}} = 114.790$ and $AIC_{\text{BNB}} =$

Table 8. Respective maximum likelihood estimates, log-likelihood values, and Akaike Information Criterion (AIC) values for various bivariate models, namely the bivariate Poisson (BPD), bivariate negative binomial (BNB), bivariate generalized Poisson (BGP), and two bivariate COM-Poisson using the compounding, and trivariate reduction approaches, respectively ($BCMP_C$, and $BCMP_T$) on NBA data set described in National Basketball Association (2020).

Model	Estimated Parameters			Log Likelihood	No. Param.	AIC
BP	$\hat{\lambda}_1 = 2.941$	$\hat{\lambda}_2 = 2.647$	$\hat{\lambda}_3 = 0$	−54.395	3	114.790
BNB	$\hat{m} = 2.938$	$\hat{r} = 100000$	$\hat{\alpha}_1 = 0.899$	−54.397	4	116.794
	$\hat{\alpha}_2 = 0$					
BGP	$\hat{\theta}_1 = 0.560$	$\hat{\theta}_2 = 0.605$	$\hat{\theta}_3 = 4.048$	−46.661	6	105.322
	$\hat{\lambda}_1 = 0.324$	$\hat{\lambda}_2 = -0.133$	$\hat{\lambda}_3 = -1.000$			
$BCMP_C$	$\hat{\lambda} = 1,082,035$	$\hat{\nu} = 8.370$	$\hat{p}_{00} = 0$	−47.986	5	105.972
	$\hat{p}_{01} = 0.158$	$\hat{p}_{10} = 0.185$	$\hat{p}_{11} = 0.658$			
$BCMP_T$	$\hat{\lambda}_1 = 67.249$	$\hat{\lambda}_2 = 48.573$	$\hat{\lambda}_3 = 0$	−46.262	4	100.521
	$\hat{\nu} = 3.515$					

116.794, respectively; $\Delta_{BP} = 14.269$ and $\Delta_{BNB} = 16.273$, respectively). These results further demonstrate that the BP and BNB models are inappropriate for under-dispersed bivariate data because they cannot effectively model such constructs. The BGP and presented BCMP models, however, offer impressive results where the $BCMP_T$ proves itself to offer a simple yet most effective form.

6. Discussion

The $BCMP_T$ distribution is a flexible bivariate model for count data that can accommodate data dispersion. With the BP distribution as a special case, our proposed model likewise contains marginal forms including the Poisson and particular NB and binomial marginals as special cases. Simulated and real data examples demonstrate that this distribution performs at least comparably with other considered bivariate count models, outperforming the BP model because of its ability to accommodate data dispersion. We further have parameter estimation procedures discussed for MOM and ML estimation with hypothesis tests established to detect statistically significant data dispersion or dependence, respectively. While the $BCMP_T$ correlation structure is non-negative, future work seeks to develop a BCMP alternative that allows for positive and negative correlation while retaining strong model fitting capabilities.

Several BCMP models have already been developed; four of them are discussed in this paper. Another popular approach for constructing bivariate distributions is through the use of copulas, given its potential flexibility to allow for positive or negative correlations. This is one avenue for future study, particularly given the vast number of potential copulas for consideration and the need to better understand their resulting properties. Further research is also needed to suggest ways in which to circumvent identifiability concerns that surface when using copulas to create multivariate discrete distributions (Trivedi and Zimmer 2017). Such models can be further compared to other bivariate discrete models (e.g., the bivariate double Poisson model (Islam and Chowdhury 2017) which also accommodates over-, under-, and equi-dispersion).

The $BCMP_T$ distribution can be extended to more elaborate models. For instance, covariates can be incorporated (e.g., Jowaheer, Khan, and Sunecher (2018) and Sunecher, Khan, and Jowaheer (2020)). Here, we have refrained from this generalization as we want to focus attention on the distribution itself and the interpretation of its parameters; this helps ensure that readers can properly understand, interpret, and gain inference regarding parameter discussions (e.g., dispersion). Accordingly, while our model has allowed for varying location parameters, we have assumed a common ν for computational ease, which is consistent with the discussion in Section 2.4.2 of Shmueli et al. (2005). Varying dispersion levels, however, may also be considered.

The real data analyses provide the MLEs but not the corresponding standard errors; this work demonstrates that the issue of standard error computation remains an area of further study. For both examples, the approximated Hessian matrix for the BCMP_T model was not positive definite; investigations of these statistical computations showed that some of the eigenvalues were negative. This phenomenon is not unique to BCMP_T; similar issues can occur with BCMP_C, BCMP_{S1}, and BCMP_{S2} (Sellers, Morris, and Balakrishnan 2016; Ong et al. 2021). Future work will consider alternative procedures for approximating the Hessian matrix in order to ensure a positive definite structure, thus allowing for standard errors to be determined.

While this work focuses on the trivariate reduction method as a tool to develop a bivariate distribution, this approach can be generalized to higher dimensions thus establishing a multivariate analog. For $i = 1, 2, \ldots, m,$ let

$$X_i = W_i + W, \tag{32}$$

where W_i are $\text{CMP}(\lambda_i, \nu)$ distributed and W is a $\text{CMP}(\lambda, \nu)$ random variable such that all of these random variables are mutually independent with a common dispersion parameter, ν. Considering the case when $\nu = 1$, the resulting marginal distributions are Poisson with location parameters, $\lambda_i + \lambda$, $i = 1, 2, \ldots, m,$ respectively. Furthermore, when $\lambda_1 = \ldots = \lambda_m = \lambda$, each of the random variables $X_1, \ldots, X_m$ has an $\text{sCMP}(\lambda, \nu, 2)$ marginal distribution as discussed in Section 2. ML and MOM parameter estimation can likewise be generalized. See Johnson, Kotz, and Balakrishnan (1997) for additional discussion regarding the trivariate reduction method in a multivariate setting.

Notes

1. The Stein and Juritz (1987) parametrization for a $\text{NB}(r, \beta)$ distributed random variable X has the pmf $P(X = x) = \binom{r + x - 1}{x} (\frac{1}{1+\beta})^x (\frac{\beta}{1+\beta})^r$.
2. The trivariate-reduced BP model corresponds to the BP model derived via the compounding method in that $\lambda_1 + \lambda_3 = \lambda p_{1+}$, $\lambda_2 + \lambda_3 = \lambda p_{+1}$, and $\lambda_3 = \lambda p_{11}$ (Kocherlakota and Kocherlakota 1992).
3. Another option is to set $C = \max_{n \in \mathbb{N}} \{\ln y_n\}$ since the only requirement is for C to be sufficiently large.

Funding

Support for Kimberly S. Weems was provided by the National Science Foundation (Grant #1700235).

ORCID

Kimberly S. Weems http://orcid.org/0000-0002-2108-9257
Kimberly F. Sellers http://orcid.org/0000-0001-6516-0548

References

Adelstein, A. M. 1952. Accident proneness: A criticism of the concept based upon an analysis of shunters' accidents. *Journal of the Royal Statistical Society Series A* 115 (3):354–410. doi:10.2307/2980739.

Arbous, A. G., and J. E. Kerrich. 1951. Accident statistics and the concept of accident proneness. *Biometrics* 7 (4):340–432. doi:10.2307/3001656.

Balakrishnan, N., and S. Pal. 2013. Lognormal lifetimes and likelihood-based inference for flexible cure rate models based on COM-Poisson family. *Computational Statistics & Data Analysis* 67: 41–67. doi:10.1016/j.csda.2013.04.018.

Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference*. New York: Springer.

Chesneau, C.,. M. Kachour, and D. Karlis. 2015. On some distributions arising from a generalized trivariate reduction scheme. *Statistical Methodology* 25:36–50. doi:10.1016/j.stamet.2015.01.001.

Conway, R. W., and W. L. Maxwell. 1962. A queuing model with state dependent service rates. *Journal of Industrial Engineering* 12:132–6.

Famoye, F. 1993. Restricted generalized Poisson regression model. *Communications in Statistics - Theory and Methods* 22 (5):1335–54. doi:10.1080/03610929308831089.

Famoye, F., and P. C. Consul. 1995. Bivariate generalized Poisson distribution with some applications. *Metrika* 42 (1):127–38. doi:10.1007/BF01894293.

Gaunt, R. E., S. Iyengar, A. B. O. Daalhuis, and B. Simsek. 2019. An asymptotic expansion for the normalizing constant of the Conway-M axwell-Poisson distribution. *Annals of the Institute of Statistical Mathematics* 71 (1):163–80. doi:10.1007/s10463-017-0629-6.

Gillispie, S. B., and C. G. Green. 2015. Approximating the Conway-Maxwell-Poisson distribution normalization constant. *Statistics* 49 (5):1062–73. doi:10.1080/02331888.2014.896919.

Hilbe, J. M. 2007. *Negative binomial regression*. 5th ed. Cambridge: Cambridge University Press.

Holgate, P. 1964. Estimation for the bivariate Poisson distribution. *Biometrika* 51 (1–2):241–5. doi:10.1093/biomet/51.1-2.241.

Islam, M., and R. Chowdhury. 2017. Models for bivariate count data: Bivariate Poisson distribution. In *Analysis of repeated measures data*, 97–124. Singapore: Springer.

Johnson, N., S. Kotz, and N. Balakrishnan. 1997. *Discrete multivariate distributions*. New York: John Wiley & Sons.

Jowaheer, V., N. M. Khan, and Y. Sunecher. 2018. A BINAR(1) time-series model with cross-correlated COM-Poisson innovations. *Communications in Statistics - Theory and Methods* 47 (5): 1133–54. doi:10.1080/03610926.2017.1316400.

Karlis, D., and I. Ntzoufras. 2005. Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software* 14 (10):1–36. doi:10.18637/jss.v014.i10.

Kocherlakota, S., and K. Kocherlakota. 1992. *Bivariate discrete distributions*. New York: Marcel Dekker.

Koehler, K. J., and K. Larntz. 1980. An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association* 75 (370):336–44. doi:10.1080/01621459.1980.10477473.

Kokonendji, C. C., D. Mizere, and N. Balakrishnan. 2008. Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference* 138 (5):1287–96. doi:10.1016/j.jspi.2007.05.028.

Kokonendji, C. C., and P. Puig. 2018. Fisher dispersion index for multivariate count distributions: A review and a new proposal. *Journal of Multivariate Analysis* 165:180–93. doi:10.1016/j.jmva.2017.12.010.

Lai, C. D. 2006. Constructions of discrete bivariate distributions. In *Advances in distribution theory, order statistics, and inference*, ed. N. Balakrishnan, J. M. Sarabia, and E. Castillo, 29–58. Boston, MA: Birkhäuser Boston.

Lee, M.-L T. 1996. Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics - Theory and Methods* 25 (6):1207–22.

M'Kendrick, A. G. 1926. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130.
Maritz, J. S. 1952. Note on a certain family of discrete distribution. *Biometrika* 39 (1–2):196–8. doi:10.1093/biomet/39.1-2.196.
Marshall, A. W., and I. Olkin. 1985. A family of bivariate distributions generated by the bivariate Bernoulli distribution. *Journal of the American Statistical Association* 80 (390):332–8. doi:10.1080/01621459.1985.10478116.
Minka, T. P., G. Shmueli, J. B. Kadane, S. Borle, and P. Boatwright. 2003. Computing with the COM-Poisson distribution. Technical Report 776, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA.
National Basketball Association. 2020. NBA All-Star Game, 2000–2016. Accessed April 22, 2020. https://www.kaggle.com/fmejia21/nba-all-star-game-20002016.
Ong, S. H., R. C. Gupta, T. Ma, and S. Z. Sim. 2021. Bivariate Conway-Maxwell Poisson distributions with given marginals and correlation. *Journal of Statistical Theory and Practice* 15 (10): 1–19. doi:10.1007/s42519-020-00141-4.
Sarmanov, O. V. 1966. Generalized normal correlation and two-dimensional Fréchet classes. *Doklady Akademii Nauk* 168 (1):32–5.
Sellers, K. F. 2012. A generalized statistical control chart for over- or under-dispersed data. *Quality and Reliability Engineering International* 28 (1):59–65. doi:10.1002/qre.1215.
Sellers, K. F., D. S. Morris, and N. Balakrishnan. 2016. Bivariate Conway-Maxwell-Poisson distribution: Formulation, properties, and inference. *Journal of Multivariate Analysis* 150:152–68. doi:10.1016/j.jmva.2016.04.007.
Sellers, K. F., and A. Raim. 2016. A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis* 99:68–80. doi:10.1016/j.csda.2016.01.007.
Sellers, K. F., and G. Shmueli. 2010. A flexible regression model for count data. *Annals of Applied Statistics* 4 (2):943–61.
Sellers, K. F., G. Shmueli, and S. Borle. 2012. The COM-Poisson model for count data: A survey of methods and applications. *Applied Stochastic Models in Business and Industry* 28 (2):104–16. doi:10.1002/asmb.918.
Sellers, K. F., A. W. Swift, and K. S. Weems. 2017. A flexible distribution class for count data. *Journal of Statistical Distributions and Applications* 4 (1):1–21. doi:10.1186/s40488-017-0077-0.
Shmueli, G., T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright. 2005. A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54 (1):127–42. doi:10.1111/j.1467-9876.2005.00474.x.
Şimşek, B., and S. Iyengar. 2016. Approximating the Conway-Maxwell-Poisson normalizing constant. *Filomat* 30 (4):953–60. doi:10.2298/FIL1604953S.
Stein, G. Z., and J. M. Juritz. 1987. Bivariate compound Poisson distributions. *Communications in Statistics - Theory and Methods* 16 (12):3591–607. doi:10.1080/03610928708829593.
Sunecher, Y., N. Khan, and V. Jowaheer. 2020. BINMA(1) model with COM-Poisson innovations: Estimation and application. *Communications in Statistics - Simulation and Computation* 49 (6): 1631–52. doi:10.1080/03610918.2018.1502780.
Teicher, H. 1954. On the multivariate Poisson distribution. *Skandinavisk Aktuarietidskrift* 37:1–9.
Trivedi, P., and D. Zimmer. 2017. A note on identification of bivariate copulas for discrete count data. *Econometrics* 5 (1):10. doi:10.3390/econometrics5010010.

Appendix A. Approximating MOM estimator standard errors

As noted in Section 4, we determine the MOM estimators by minimizing the squared-error loss function,

$$l(\lambda_1, \lambda_2, \lambda_3, \nu; (\boldsymbol{x}, \boldsymbol{y})) = \left(\mu_x - \bar{X}\right)^2 + \left(\mu_y - \bar{Y}\right)^2 + \left(\mu_{x^2} - \overline{X^2}\right)^2 + \left(\mu_{y^2} - \overline{Y^2}\right)^2 + \left(\mu_{xy} - \overline{XY}\right)^2,$$

thus, the MOM estimators solve the system of equations,

$$\begin{cases} \frac{\partial l}{\partial \lambda_1} = f_1(\lambda_1, \lambda_2, \lambda_3, \nu; \bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) = 0 \\ \frac{\partial l}{\partial \lambda_2} = f_2(\lambda_1, \lambda_2, \lambda_3, \nu; \bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) = 0 \\ \frac{\partial l}{\partial \lambda_3} = f_3(\lambda_1, \lambda_2, \lambda_3, \nu; \bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) = 0 \\ \frac{\partial l}{\partial \nu} = f_4(\lambda_1, \lambda_2, \lambda_3, \nu; \bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) = 0 \end{cases} \tag{33}$$

from which we get implicit relations between the MOM estimators and data moments, namely

$$\begin{cases} \tilde{\lambda}_1 = F_1(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) \\ \tilde{\lambda}_2 = F_2(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) \\ \tilde{\lambda}_3 = F_3(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) \\ \tilde{\nu} = F_4(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}), \end{cases} \tag{34}$$

hence, computing the respective variances of these MOM estimators is equivalent to computing the variances of the functions, F_i, $i = 1, 2, 3, 4$.

We use the Delta method to accomplish this goal. Let μ_x, μ_y, μ_{x^2}, μ_{y^2}, μ_{xy} denote the respective BCMP_T moments. We know that, as the sample size n increases, the following results hold:

$$\begin{cases} \sqrt{n}\left(\bar{X} - \mu_x\right) \xrightarrow{d} N\left(0, \sigma_x^2\right) \\ \sqrt{n}\left(\bar{Y} - \mu_x\right) \xrightarrow{d} N\left(0, \sigma_y^2\right) \\ \sqrt{n}\left(\overline{X^2} - \mu_{x^2}\right) \xrightarrow{d} N\left(0, \sigma_{x^2}^2\right) \\ \sqrt{n}\left(\overline{Y^2} - \mu_{y^2}\right) \xrightarrow{d} N\left(0, \sigma_{y^2}^2\right) \\ \sqrt{n}\left(\overline{XY} - \mu_{xy}\right) \xrightarrow{d} N\left(0, \sigma_{xy}^2\right), \end{cases} \tag{35}$$

where $\sigma^2_{g(X,\,Y)}$ denotes the variance of a function $g(X, Y)$ associated with the BCMP_T distribution. Accordingly, a function of those moments $F_i(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY})$ can be represented via Taylor expansion as

$$\begin{aligned} F_i(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}) = F_i(\mu_x, \mu_y, \mu_{x^2}, \mu_{y^2}, \mu_{xy}) + \frac{\partial F_i}{\partial \mu_x} \cdot (\bar{X} - \mu_x) + \frac{\partial F_i}{\partial \mu_y} \cdot (\bar{Y} - \mu_y) \\ + \frac{\partial F_i}{\partial \mu_{x^2}} \cdot (\overline{X^2} - \mu_{x^2}) + \frac{\partial F_i}{\partial \mu_{y^2}} \cdot (\overline{Y^2} - \mu_{y^2}) + \frac{\partial F_i}{\partial \mu_{xy}} \cdot (\overline{XY} - \mu_{xy}) + R_2, \end{aligned}$$

where, as n gets large, the second order remainder of the Taylor expansion R_2 converges to 0. We use $\frac{\partial F_i}{\partial \bar{X}}, \frac{\partial F_i}{\partial \bar{Y}}, \frac{\partial F_i}{\partial \overline{X^2}}, \frac{\partial F_i}{\partial \overline{Y^2}}, \frac{\partial F_i}{\partial \overline{XY}}$ to approximate $\frac{\partial F_i}{\partial \mu_x}, \frac{\partial F_i}{\partial \mu_y}, \frac{\partial F_i}{\partial \mu_x^2}, \frac{\partial F_i}{\partial \mu_y^2}, \frac{\partial F_i}{\partial \mu_{xy}}$, recognizing the close proximity of the respective values when the sample size is sufficiently large. Thus, applying the Delta method, we find that

$$Var\left(\sqrt{n} \cdot F_i\left(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY}\right)\right) \to \left(\nabla F_i(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY})\right)^T \Sigma\left(\nabla F_i(\bar{X}, \bar{Y}, \overline{X^2}, \overline{Y^2}, \overline{XY})\right)$$

where Σ denotes the variance-covariance matrix of (X, Y, X^2, Y^2, XY). Thus, the estimated variance of the estimators is

$$Var\begin{pmatrix} \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \\ \tilde{\lambda}_3 \\ \tilde{\nu} \end{pmatrix} = Var\begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{pmatrix} = \frac{1}{n}\text{diag}\Big(\left(\nabla F_1 \, \nabla F_2 \, \nabla F_3 \, \nabla F_4\right)^T \Sigma \left(\nabla F_1 \, \nabla F_2 \, \nabla F_3 \, \nabla F_4\right)\Big), \tag{36}$$

where

$$\begin{pmatrix} \nabla F_1 \, \nabla F_2 \, \nabla F_3 \, \nabla F_4 \end{pmatrix}^T = -H^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial \bar{X}} & \frac{\partial f_1}{\partial \bar{Y}} & \frac{\partial f_1}{\partial \overline{X^2}} & \frac{\partial f_1}{\partial \overline{Y^2}} & \frac{\partial f_1}{\partial \overline{XY}} \\ \frac{\partial f_2}{\partial \bar{X}} & \frac{\partial f_2}{\partial \bar{Y}} & \frac{\partial f_2}{\partial \overline{X^2}} & \frac{\partial f_2}{\partial \overline{Y^2}} & \frac{\partial f_2}{\partial \overline{XY}} \\ \frac{\partial f_3}{\partial \bar{X}} & \frac{\partial f_3}{\partial \bar{Y}} & \frac{\partial f_3}{\partial \overline{X^2}} & \frac{\partial f_3}{\partial \overline{Y^2}} & \frac{\partial f_3}{\partial \overline{XY}} \\ \frac{\partial f_4}{\partial \bar{X}} & \frac{\partial f_4}{\partial \bar{Y}} & \frac{\partial f_4}{\partial \overline{X^2}} & \frac{\partial f_4}{\partial \overline{Y^2}} & \frac{\partial f_4}{\partial \overline{XY}} \end{pmatrix}$$

and H is the Hessian matrix of $l(\lambda_1, \lambda_2, \lambda_3, \nu; (\boldsymbol{x}, \boldsymbol{y}))$ over $\lambda_1, \lambda_2, \lambda_3,$ and ν. We recognize, however, that the standard error estimates are reasonable for MOM estimates close to their respective true values.