Statistical inference for distributions with one Poisson conditional

Barry C. Arnold\textsuperscript{a} and B. G. Manjunath\textsuperscript{a, b}

\textsuperscript{a}Department of Statistics, University of California, Riverside, CA, USA; \textsuperscript{b}School of Mathematics and Statistics, University of Hyderabad, Hyderabad, India

ABSTRACT

It will be recalled that the classical bivariate normal distributions have normal marginals and normal conditionals. It is natural to ask whether a similar phenomenon can be encountered involving Poisson marginals and conditionals. However, it is known, from research on conditionally specified models, that Poisson marginals will be encountered, together with both conditionals being of the Poisson form, only in the case in which the variables are independent. In order to have a flexible dependent bivariate model with some Poisson components, in the present article, we will be focusing on bivariate distributions with one marginal and the other family of conditionals being of the Poisson form. Such distributions are called Pseudo-Poisson distributions. We discuss distributional features of such models, explore inferential aspects and include an example of applications of the Pseudo-Poisson model to sets of over-dispersed data.

1. Introduction

The bivariate (and multivariate) normal model is a mathematically attractive instrument for analyzing vector-valued continuous data sets. A discrete distribution with equally attractive features has long been sought, but with, at best, only partial success. A particularly popular univariate discrete distribution is provided by the Poisson model. It would be desirable to identify a bivariate model which (similar to the normal case) has both Poisson marginal and Poisson conditional distributions. Some historical light on this search will be provided in Section 5. The search will be fruitless unless you will be satisfied with a bivariate density with independent Poisson distributed marginals (and, trivially, conditional densities of the same form). In the present paper, we focus on a less restrictive model, called a pseudo-Poisson model, which has one marginal density, say of \(X_1\), of the Poisson form and all of the conditional distributions of \(X_2\) given \(X_1\) also of the Poisson form. We will put particular emphasis on the case in which the regression of \(X_2\) on \(X_1\) is linear. In addition to discussion of relevant distribution theory, attention will be directed to questions of parameter estimation (point and interval) for the model and various sub-models.
A small simulation study is included. In addition, the models introduced in this paper are utilized to re-analyze a well-known bivariate data set dealing with traffic accident data. Extensions to higher dimensions are also discussed. We begin with a discussion of a general triangular model, which subsumes and extends the models with Poisson component distributions that are the focus of the present paper.

2. Triangular transformation models

We begin by reviewing a family of models called triangular transformation models which were introduced in Filus et al. [2] as follows:

Let \( \mathcal{F} = \{ F(x; \theta) : \theta = (\theta_1, \ldots, \theta_m)^T \in \Theta \subset \mathbb{R}^m \} \) be an \( m \)-parameter family of univariate distributions. A \( k \)-dimensional Pseudo-\( \mathcal{F} \) distribution can be constructed as follows:

\[
P(X_1 \leq x_1) = F(x_1; \theta_1) \tag{1}
\]

and for \( \ell = 2, 3, \ldots, k \)

\[
P(X_\ell \leq x_\ell \mid X_{(\ell-1)} = x_{(\ell-1)}) = F(x_\ell; \theta_\ell(x_{(\ell-1)})), \tag{2}
\]

where \( \theta_1 \in \Theta \) and, for each \( \ell, \theta_\ell : \mathbb{R}^{\ell-1} \to \Theta \). Note that we use the notational convention \( a(j) = (a_1, a_2, \ldots, a_j) \).

3. \( k \)-dimensional Pseudo-Poisson models

Within the general triangular transformation class of \( k \)-dimensional models can be found the class of \( k \)-dimensional Pseudo-Poisson models, which are more simply described in terms of discrete mass functions rather than distribution functions.

Using standard notation, we will write \( X \sim \mathcal{P}(\lambda) \) (Poisson distribution) if, for \( x \in \{0, 1, 2, \ldots\} \) we have

\[
P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}.
\]

**Definition 3.1:** A \( k \)-dimensional random variable \( X = (X_1, X_2, \ldots, X_k) \) is said to have a \( k \)-dimensional Pseudo-Poisson distribution if there exists a positive constant \( \lambda_1 \) such that

\[
X_1 \sim \mathcal{P}(\lambda_1)
\]

and \( k-1 \) functions \( \{\lambda_\ell : \ell = 2, 3, \ldots, k\} \) where, for each \( \ell, \lambda_\ell : \{0, 1, 2, \ldots\}^{(\ell-1)} \to (0, \infty) \) such that

\[
X_\ell \mid X_{(\ell-1)} = x_{(\ell-1)} \sim \mathcal{P}(\lambda_\ell(x_{(\ell-1)})).
\]

Note that there are no constraints on the forms of the functions \( \lambda_\ell, \ell = 2, 3, \ldots, k \) that appear in the definition, save for measurability.

In applications, it would typically be the case that the \( \lambda_\ell \)'s would be chosen to be relatively simple functions depending on a limited number of parameters.
4. Bivariate-Pseudo-Poisson models

We will consider in some detail the Pseudo-Poisson models in the case in which the dimension \(k\) is 2. In that case, we can describe the joint distribution as follows.

**Definition 4.1:** A 2-dimensional random variable \(X = (X_1, X_2)\) is said to have a bivariate Pseudo-Poisson distribution if there exists a positive constant \(\lambda_1\) such that

\[
X_1 \sim \mathcal{P}(\lambda_1)
\]

and a function \(\lambda_2 : \{0, 1, 2, \ldots\} \to (0, \infty)\) such that

\[
X_2 \mid X_1 = x_1 \sim \mathcal{P}(\lambda_2(x_1)).
\]

In this case also, it might be desirable to restrict the form of the function \(\lambda_2(x_1)\). For example, we might restrict it to be a polynomial with unknown coefficients.

5. A historical note

It will be recalled that the classical bivariate normal distributions have normal marginals AND normal conditionals. Thus, it has \(X_1 \sim \mathcal{N}\) (normal distribution), \(X_2 \sim \mathcal{N}\) and for each \(x_1 \in \mathbb{R}\), \(X_2 \mid X_1 = x_1 \sim \mathcal{N}\), while for each \(x_2 \in \mathbb{R}\), \(X_1 \mid X_2 = x_2 \sim \mathcal{N}\) also. It is natural to ask whether a similar phenomenon can be encountered with Poisson marginals and conditionals. A trivial example of this kind is one in which \(X_1\) and \(X_2\) are independent Poisson variables. In fact, no other examples exist. In an early paper, Seshadri and Patil [14] argued that one could not have a non-trivial distribution with \(X_1\) having a Poisson distribution and, for each \(x_2\), \(X_1 \mid X_2 = x_2\) also Poisson distributed. Arnold et al. [1] (for example) consider the case in which, for each \(x_1\), \(X_2 \mid X_1 = x_1 \sim \mathcal{P}\), while for each \(x_2\), \(X_1 \mid X_2 = x_2 \sim \mathcal{P}\). Such distributions are called Poisson-conditionals distributions. They only have Poisson marginals in the case of independence (which could be deduced using the Seshadri-Patil result). We refer to Ghosh et al. [3] for properties and application of the conditional Poisson model. However, if we are satisfied with having one marginal \((X_1)\) and the 'other' family of conditionals \((X_2 \mid X_1 = x_1)\) being of the Poisson form, then the Pseudo-Poisson distributions fill the bill precisely.

6. The bivariate Pseudo-Poisson model with a linear regression function

In this section, we will consider in some detail a particularly simple bivariate Pseudo-Poisson model. For it, we assume that

\[
X_1 \sim \mathcal{P}(\lambda_1)
\]

and

\[
X_2 \mid X_1 = x_1 \sim \mathcal{P}(\lambda_2 + \lambda_3 x_1).
\]

The natural parameter space for this model is \(\{(\lambda_1, \lambda_2, \lambda_3) : \lambda_1 > 0, \lambda_2 > 0, \lambda_3 \geq 0\}\). The case in which the variables are independent, corresponds to choice \(\lambda_3 = 0\). Note that in the
limiting case, \( \lambda_2 = 0 \) is a plausible value for the above model. However, in such a framework \( \lambda_3 > 0 \), i.e. independence of variables is forsaken. Subsequently, when \( \lambda_3 = 0 \) then \( \lambda_2 \) is necessarily takes value greater than zero. With this framework the plausible parameter space for the model is \( \{ (\lambda_1, \lambda_2, \lambda_3) : \lambda_1 > 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \} \).

In the following, we derive the joint probability generating function (p.g.f.) and marginal p.g.f. of the above bivariate Pseudo-Poisson distribution.

**Theorem 6.1:** The p.g.f. for the bivariate Pseudo-Poisson distribution is given by

\[
G(t_1, t_2) = e^{\lambda_2 (t_2 - 1)} e^{\lambda_1 [t_1 e^{\lambda_3 (t_2 - 1)} - 1]}, \quad t_1, t_2 \in \mathbb{R}. \tag{5}
\]

For the proof of theorem, see Appendix 1.

**Corollary 6.1:** The marginal mass function of \( X_2 \) is that of a Neyman Type A distribution, when \( \lambda_2 = 0 \).

For the proof of corollary, see Appendix 2.

The above p.g.f. is of the form of a Neyman Type A distribution with \( \lambda_3 \) being the index of clumping (see page 403 of Johnson et al. [6]) which can also be recognized as a Poisson mixture of Poisson distributions.

Now, the marginal mass function of \( X_2 \) is given by

\[
p_2(x_2) = P(X_2 = x_2) = \frac{e^{-\lambda_1} \lambda_3^{x_2}}{x_2!} \sum_{j=0}^{\infty} \frac{(\lambda_1 e^{-\lambda_3})^j f_j}{j!}; \quad x_2 = 0, 1, 2, \ldots, \tag{6}
\]

i.e. \( X_2 \) has a Poisson distribution with the parameter \( \lambda_3 \phi \) while \( \phi \) itself is a random variable with the Poisson distribution with the parameter \( \lambda_1 \).

In the following section, we derive moments of the bivariate Pseudo-Poisson distribution.

### 6.1. Moments

Now note that

\[
E(X_1) = \lambda_1, \tag{7}
\]

\[
E(X_2 \mid X_1 = x_1) = \lambda_2 + \lambda_3 x_1, \quad E(E(X_2 \mid X_1)) = \lambda_2 + \lambda_3 E(X_1), \quad E(X_2) = \lambda_2 + \lambda_3 \lambda_1, \tag{8}
\]

and

\[
\text{Var}(X_1) = \lambda_1, \tag{9}
\]

\[
\text{Var}(X_2 \mid X_1 = x_1) = \lambda_2 + \lambda_3 x_1, \quad \text{Var}(X_2) = E(\text{Var}(X_2 \mid X_1)) + \text{Var}(E(X_2 \mid X_1))
\]

\[
= E(\lambda_2 + \lambda_3 X_1) + \text{Var}(\lambda_2 + \lambda_3 X_1),
\]
\[ \text{Var}(X_2) = \lambda_2 + \lambda_3 \lambda_1 + \lambda_2^2 \lambda_1. \]  

Also

\[
E(X_1X_2) = E\{E(X_1X_2 \mid X_1)\} = E\{E(X_1)E(X_2 \mid X_1)\} = E(\lambda_2 X_1 + \lambda_3 X_1^2) = \lambda_2 \lambda_1 + \lambda_3 (\lambda_1 + \lambda_1^2). 
\]

The covariance between \(X_1\) and \(X_2\) is thus

\[
\text{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)
= \lambda_2 \lambda_1 + \lambda_3 (\lambda_1 + \lambda_1^2) - \lambda_1 (\lambda_2 + \lambda_1 \lambda_3),
\]

\[
\text{Cov}(X_1, X_2) = \lambda_1 \lambda_3, \quad (11)
\]

and the corresponding correlation is

\[
\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}} = \frac{\lambda_1 \lambda_3}{\sqrt{\lambda_1 (\lambda_2 + \lambda_3 \lambda_1 + \lambda_2^2 \lambda_1)}}. \quad (12)
\]

In the following, we note three special cases which merit consideration.

**Case I:** When \(\lambda_3 = 0\), it follows that \(\rho = 0\). In fact, in this case, \(X_1\) and \(X_2\) are independent.

**Case II:** (Sub-Model I) When \(\lambda_2 = \lambda_3\), the correlation is of the form

\[
\rho = \sqrt{\frac{\lambda_1 \lambda_3}{1 + \lambda_1 + \lambda_1 \lambda_3}}. \quad (13)
\]

**Case III:** (Sub-Model II) In the limiting case in which \(\lambda_2 = 0\), \(\rho\) simplifies to become

\[
\rho = \sqrt{\frac{\lambda_3}{1 + \lambda_3}}. \quad (14)
\]

Note that this correlation only depends on \(\lambda_3\), and not on \(\lambda_1\), when \(\lambda_2 = 0\).

We also remark that, for the case \(\lambda_2 = 0\), the bivariate Pseudo-Poisson distribution reduces to the bivariate Poisson-Poisson distribution. The Poisson-Poisson distribution was originally introduced by Leiter and Hamdani [10] in 1973 in analyzing traffic accidents and fatalities data. However, the two approaches leading to the same distribution are different. Nevertheless, one can consider the bivariate Pseudo-Poisson as a generalization of the Poisson-Poisson distribution.
6.2. Characterizing the Poisson-Poisson distribution

In the following, we state and prove a characterization of the bivariate Pseudo-Poisson distribution submodel (with $\lambda_2 = 0$) or the Poisson-Poisson distribution. A similar characterization for Power Series distributions is in Kyriakoussis and Papageorgiou [9].

**Theorem 6.2:**  If $E(X_2 \mid X_1 = x_1) = \lambda_3 x_1$ and the marginal distribution of $X_2$ is of the form in (6) (Neyman Type A Distribution), then the joint distribution of $(X_1, X_2)$ is necessarily that of the bivariate Poisson-Poisson distribution.

For the proof of above theorem, see Appendix 3.

6.3. Fisher dispersion index

In this section, we derive the Fisher dispersion index for the bivariate Pseudo-Poisson distribution. We use the definition of the multivariate Fisher dispersion index provided by Kokonendji and Puig [7] in Section 3 given as; for any $d$-dimensional discrete random variable $Y$ with mean vector $EY$ and covariance matrix $\text{Cov}(Y)$ the generalized dispersion index is

$$GDI(Y) = \frac{\sqrt{EY^T \text{Cov}(Y)} \sqrt{EY}}{EY^T EY}.$$  \hspace{1cm} (15)

The marginal dispersion indices are

$$DI(X_1) = \frac{\text{Var}(X_1)}{E(X_1)} = 1 \text{ (equi-dispersion)}, \hspace{1cm} (16)$$

$$DI(X_2) = \frac{\text{Var}(X_2)}{E(X_2)} = \frac{\lambda_2 + \lambda_3 \lambda_1 + \lambda_3^2 \lambda_1}{\lambda_2 + \lambda_3 \lambda_1}$$

$$= 1 + \frac{\lambda_3^2 \lambda_1}{\lambda_2 + \lambda_3 \lambda_1} \text{ (over-dispersion)}. \hspace{1cm} (17)$$

We state and prove the following theorem.

**Theorem 6.3:** The bivariate Pseudo-Poisson distribution is always over-dispersed.

For the proof of theorem, see Appendix 4.

Note that in a set of bivariate count data, if one marginal is equi-dispersed and other is over-dispersed, one can consider the bivariate Pseudo-Poisson distribution as a possible model.

7. Statistical Inference

In this section, we obtain moment and maximum likelihood estimators of parameters $\lambda_1$, $\lambda_2$ and $\lambda_3$. In addition, we construct the likelihood ratio test for the simpler submodel, i.e. for $\lambda_2 = \lambda_3$. Finally, we consider a simulation study and a real-life application of the bivariate Pseudo-Poisson distribution.
7.1. Moment estimators

Now suppose that we have data of the form $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ which are i.i.d. with common distribution (3)–(4). Note that, for each $i$, $X^{(i)} = (X_{1i}, X_{2i})^T$. Method of moments estimators of the parameters are readily derived. The respective simpler submodels and their statistical inference are also considered in the following section.

Now, if we equate the sample means and the sample covariance to their expectations, and if $M_1 > 0$, we obtain the following consistent asymptotically normal method of moments estimates.

\[
\tilde{\lambda}_1 = M_1, \quad (18)
\]
\[
\tilde{\lambda}_2 = M_2 - S_{12}, \quad (19)
\]
\[
\tilde{\lambda}_3 = \frac{S_{12}}{M_1}, \quad (20)
\]

where

\[
M_1 = \frac{1}{n} \sum_{i=1}^{n} X_{1i},
\]
\[
M_2 = \frac{1}{n} \sum_{i=1}^{n} X_{2i},
\]

and

\[
S_{12} = \frac{1}{n} \sum_{i=1}^{n} (X_{1i} - M_1)(X_{2i} - M_2).
\]

If we consider the simpler sub-model in which $\lambda_2 = 0$, then the method of moments estimates of the remaining two $\lambda$’s are even simpler. Thus, again provided that $M_1 > 0$,

\[
\tilde{\lambda}_1 = M_1, \quad (21)
\]
\[
\tilde{\lambda}_3 = \frac{M_2}{M_1}. \quad (22)
\]

For the sub-model in which $\lambda_2 = \lambda_3$, method of moments estimates of the parameters are given by

\[
\tilde{\lambda}_1 = M_1, \quad (23)
\]
\[
\tilde{\lambda}_3 = \frac{M_2}{1 + M_1}. \quad (24)
\]

7.2. Maximum likelihood estimators

In the two-parameter model (i.e. when $\lambda_2 = 0$ or $\lambda_2 = \lambda_3$), the maximum likelihood estimates can be verified to coincide with the method of moments estimates derived in the previous subsection, provided that $M_1 > 0$. 

Identifying the maximum likelihood estimators (m.l.e.’s) in the three parameter model is a little more challenging. For the given data of the form \( X^{(1)}, X^{(2)}, \ldots, X^{(n)} \) which are i.i.d. with common distribution (3)–(4) then the likelihood function is as follows:

\[
L(\theta) = \prod_{i=1}^{n} \left\{ \frac{e^{-\lambda_1 x_{1i}} \cdot e^{-(\lambda_2 + \lambda_3 x_{1i}) (\lambda_2 + \lambda_3 x_{1i}) x_{2i}}}{x_{1i}! x_{2i}!} \right\} \\
= \frac{e^{-n(\lambda_1 + \lambda_2)} \sum_{i=1}^{n} x_{1i} e^{-\lambda_3 \sum_{i=1}^{n} x_{1i} \prod_{i=1}^{n} (\lambda_2 + \lambda_3 x_{1i}) x_{2i}}}{\prod_{i=1}^{n} ((x_{1i})! (x_{2i})!)},
\]

where \( \theta = (\lambda_1, \lambda_2, \lambda_3)^T \).

The corresponding log-likelihood function is

\[
l = \log L(\theta) = -n(\lambda_1 + \lambda_2) + \log(\lambda_1) \sum_{i=1}^{n} x_{1i} - \lambda_3 \sum_{i=1}^{n} x_{1i} \\
+ \sum_{i=1}^{n} x_{2i} \log(\lambda_2 + \lambda_3 x_{1i}) + h(x_1, x_2),
\]

where \( h(x_1, x_2) = \log(\frac{1}{\prod_{i=1}^{n} ((x_{1i})! (x_{2i})!)}) \).

Now, differentiating with respect to the \( \lambda_i \)'s we get the following likelihood equations:

\[
-n + \frac{1}{\lambda_1} \sum_{i=1}^{n} X_{1i} = 0, \quad (27)
\]

\[
-n + \sum_{i=1}^{n} \frac{X_{2i}}{\lambda_2 + \lambda_3 X_{1i}} = 0, \quad (28)
\]

\[
- \sum_{i=1}^{n} \frac{X_{1i}}{\lambda_2 + \lambda_3 X_{1i}} + \sum_{i=1}^{n} \frac{X_{1i} X_{2i}}{\lambda_2 + \lambda_3 X_{1i}} = 0. \quad (29)
\]

If \( M_1 = (1/n) \sum_{i=1}^{n} X_{1i} = 0 \), then there is no solution to (27), otherwise Equation (27) is readily solved, yielding the m.l.e. for \( \lambda_1 \), namely

\[
\hat{\lambda}_1 = M_1. \quad (30)
\]

The remaining two equations must be solved numerically (provided that \( M_1 > 0 \)), to obtain \( \hat{\lambda}_2 \) and \( \hat{\lambda}_3 \), see Ypma [5] for available nonlinear optimization.

### 7.3. Likelihood ratio test

As usual, the general form of a generalized likelihood ratio test statistic is of the form

\[
\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}. \quad (31)
\]

Here, \( \Theta_0 \) is a subset of \( \Theta \) and we envision testing \( H_0 : \theta \in \Theta_0 \). We reject the null hypothesis for a small value of \( \Lambda \).

In the following subsection, we construct likelihood ratio tests for the simpler submodels.
7.3.1. Sub-Model I: for \( \lambda_2 = \lambda_3 \), equivalently, testing for \( H_0 : \lambda_2 = \lambda_3 \)

The natural parameter space under the full model is \( \Theta = (\lambda_1, \lambda_2, \lambda_3)^T : \lambda_1 > 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \). Besides, under the null hypothesis the natural parameter space is \( \Theta_0 = (\lambda_1, \lambda_3)^T : \lambda_1 > 0, \lambda_3 > 0 \).

Under \( H_0 \), Equation (26) will be

\[
l = -n(\lambda_1 + \lambda_3) + \log(\lambda_1) \sum_{i=1}^{n} x_{1i} - \lambda_3 \sum_{i=1}^{n} x_{1i} + \sum_{i=1}^{n} x_{2i} \log[\lambda_3(x_{1i} + 1)] + h(x_1, x_2). \tag{32}
\]

Now, taking partial derivatives with respect to \( \lambda_1 \) and \( \lambda_3 \) and equating them to zero, we get the following equations:

\[
-n + \frac{1}{\lambda_1} \sum_{i=1}^{n} X_{1i} = 0,
\]

\[
-n - \sum_{i=1}^{n} X_{1i} + \sum_{i=1}^{n} \frac{X_{2i}}{\lambda_3} = 0.
\]

Then, the m.l.e.’s of \( \lambda_1 \) and \( \lambda_3 \) are, provided that \( M_1 > 0 \), given by

\[
\hat{\lambda}_1^* = M_1,
\]

\[
\hat{\lambda}_3^* = \frac{M_2}{1 + M_1}.
\]

Note that these estimates agree with the method of moments estimates given in Section 6.1.

Now, in the unrestricted parameter space \( \Theta \), i.e. under the full model, the m.l.e.’s for \( \lambda_1 \), \( \lambda_2 \) and \( \lambda_3 \) are obtained from Equations (27)–(29).

Let \( \hat{\lambda}_1 \), \( \hat{\lambda}_2 \) and \( \hat{\lambda}_3 \) be the respective m.l.e.’s of \( \lambda_i \)’s then the generalized likelihood ratio test statistic defined in (31) will be

\[
\Lambda_1 = \frac{e^{-n(\hat{\lambda}_1^* + \hat{\lambda}_3^*)}(\hat{\lambda}_1^*)^\sum_{i=1}^{n} x_{1i} e^{-\hat{\lambda}_3^* \sum_{i=1}^{n} x_{1i}} \prod_{i=1}^{n} \left( \frac{\hat{\lambda}_1^* (1 + x_{1i})}{\lambda_2 + \hat{\lambda}_3 x_{1i}} \right)^{x_{2i}}}{e^{-n(\hat{\lambda}_1 + \hat{\lambda}_3)}(\hat{\lambda}_1)^\sum_{i=1}^{n} x_{1i} e^{-\hat{\lambda}_3 \sum_{i=1}^{n} x_{1i}} \prod_{i=1}^{n} \left( \frac{\hat{\lambda}_1 (1 + x_{1i})}{\lambda_2 + \hat{\lambda}_3 x_{1i}} \right)^{x_{2i}}}. \tag{33}
\]

Since \( \hat{\lambda}_1 = \hat{\lambda}_1^* \), the above test statistic simplifies to become

\[
\Lambda_1 = e^{n\hat{\lambda}_2} e^{-\left(\hat{\lambda}_3^* - \hat{\lambda}_3\right) \sum_{i=1}^{n} x_{1i} + \sum_{i=1}^{n} x_{2i} \log \left( \frac{\hat{\lambda}_3^* (1 + x_{1i})}{\lambda_2 + \hat{\lambda}_3 x_{1i}} \right)}. \tag{34}
\]

Now by taking the logarithm, we have

\[
\log \Lambda_1 = n\hat{\lambda}_2 - \left(\hat{\lambda}_3^* - \hat{\lambda}_3\right) \sum_{i=1}^{n} x_{1i} + \sum_{i=1}^{n} x_{2i} \log \left( \frac{\hat{\lambda}_3^* (1 + x_{1i})}{\lambda_2 + \hat{\lambda}_3 x_{1i}} \right). \tag{34}
\]

If \( n \) is large, then \(-2 \log \Lambda_1\) may be compared with a suitable \( \chi^2_1 \) percentile in order to decide whether \( H_0 \) should be accepted.
7.3.2. Sub-Model II: for $\lambda_2 = 0$, equivalently, testing for $H_0 : \lambda_2 = 0$

Under the null hypothesis the natural parameter space is $\Theta_0 = \{ (\lambda_1, \lambda_3)^T : \lambda_1 > 0, \lambda_3 > 0 \}$ and the m.l.e’s of $\lambda_1$ and $\lambda_3$ are, provided that $M_1 > 0$, given by

$$\hat{\lambda}_1^* = M_1,$$

$$\hat{\lambda}_3^* = \frac{M_2}{M_1}.$$ 

Also, these estimates coincide with the method of moments estimates given in Section 6.1.

The natural parameter space under the full model is $\Theta = \{ (\lambda_1, \lambda_2, \lambda_3)^T : \lambda_1 > 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \}$ and $\hat{\lambda}_1, \hat{\lambda}_2$ and $\hat{\lambda}_3$ are respective m.l.e’s obtained from Equations (27)–(29).

Therefore, the generalized likelihood ratio test statistic defined in (31) will be

$$\Lambda_2 = \frac{e^{-n(\hat{\lambda}_1 + \hat{\lambda}_3)} (\hat{\lambda}_1)^{\sum_{i=1}^n x_{1i}} e^{-\hat{\lambda}_3} (\sum_{i=1}^n x_{2i}) \prod_{i=1}^n (\hat{\lambda}_1 x_{1i})^{x_{2i}}}{\prod_{i=1}^n (x_{1i})! (x_{2i})!}.$$ 

Since $\hat{\lambda}_1 = \hat{\lambda}_1^*$, then the above test statistic becomes

$$\Lambda_2 = e^{n\hat{\lambda}_2} e^{-(\hat{\lambda}_3 - \hat{\lambda}_3)} \sum_{i=1}^n x_{1i} \prod_{i=1}^n \left( \frac{\hat{\lambda}_3 x_{1i}}{\hat{\lambda}_2 + \hat{\lambda}_3 x_{1i}} \right)^{x_{2i}}. \quad (35)$$ 

Taking logarithm, we have

$$\log \Lambda_2 = n\hat{\lambda}_2 - (\hat{\lambda}_3 - \hat{\lambda}_3) \sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i} \log \left( \frac{\hat{\lambda}_3 x_{1i}}{\hat{\lambda}_2 + \hat{\lambda}_3 x_{1i}} \right). \quad (36)$$ 

If $n$ is large, then $-2 \log \Lambda_2$ may be compared with a suitable $\chi^2_1$ percentile in order to decide whether $H_0$ should be accepted.

7.3.3. Testing for independence, i.e. $H_0 : \lambda_3 = 0$

Under the null hypothesis the natural parameter space is $\Theta_0 = \{ (\lambda_1, \lambda_2)^T : \lambda_1 > 0, \lambda_2 > 0 \}$ and the m.l.e’s of $\lambda_1$ and $\lambda_2$ are, provided that $M_1 > 0$ and $M_2 > 0$, given by

$$\hat{\lambda}_1^* = M_1,$$

$$\hat{\lambda}_2^* = M_2.$$ 

Also, these estimates coincide with the method of moments estimates as given in Section 6.1.

The natural parameter space under the full model is $\Theta = \{ (\lambda_1, \lambda_2, \lambda_3)^T : \lambda_1 > 0, \lambda_2 \geq 0, \lambda_3 \geq 0 \}$ and $\hat{\lambda}_1, \hat{\lambda}_2$ and $\hat{\lambda}_3$ are the respective m.l.e’s obtained from Equations (27)–(29).

Therefore, the generalized likelihood ratio test statistic defined in (31) will be

$$\Lambda_3 = \frac{e^{-n(\hat{\lambda}_1 + \hat{\lambda}_3)} (\hat{\lambda}_1)^{\sum_{i=1}^n x_{1i}} e^{-\hat{\lambda}_3} (\sum_{i=1}^n x_{2i}) \prod_{i=1}^n (\hat{\lambda}_1 x_{1i})^{x_{2i}}}{\prod_{i=1}^n (x_{1i})! (x_{2i})!}.$$ 

$$e^{-n(\hat{\lambda}_1 + \hat{\lambda}_3)} (\hat{\lambda}_1)^{\sum_{i=1}^n x_{1i}} e^{-\hat{\lambda}_3} (\sum_{i=1}^n x_{2i}) \prod_{i=1}^n (\hat{\lambda}_1 x_{1i})^{x_{2i}}.$$ 

$$\prod_{i=1}^n (x_{1i})! (x_{2i})!.$$ 

$$\Lambda_3 = \frac{e^{-n(\hat{\lambda}_1 + \hat{\lambda}_3)} (\hat{\lambda}_1)^{\sum_{i=1}^n x_{1i}} e^{-\hat{\lambda}_3} (\sum_{i=1}^n x_{2i}) \prod_{i=1}^n (\hat{\lambda}_1 x_{1i})^{x_{2i}}}{\prod_{i=1}^n (x_{1i})! (x_{2i})!}.$$
Since $\hat{\lambda}_1 = \hat{\lambda}_1^*$, then the above test statistic becomes

$$
\Lambda_3 = e^{n\hat{\lambda}_2} e^{-(\hat{\lambda}_2^* - \hat{\lambda}_2)} \sum_{i=1}^n x_{1i} \prod_{i=1}^n \left( \frac{\hat{\lambda}_2^*}{\hat{\lambda}_2 + \lambda_3 x_{1i}} \right)^{x_{2i}}.
$$

(37)

Taking logarithm, we have

$$
\log \Lambda_3 = n\hat{\lambda}_2 - (\hat{\lambda}_2^* - \hat{\lambda}_2) \sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i} \log \left( \frac{\hat{\lambda}_2^*}{\hat{\lambda}_2 + \lambda_3 x_{1i}} \right).
$$

(38)

If $n$ is large then $-2 \log \Lambda_3$ may be compared with a suitable $\chi^2$ percentile in order to decide whether $H_0$ should be accepted.

**Remark 7.1:** The two hypotheses mentioned in Sections 7.3.1 and 7.3.2 have corresponding likelihood ratio tests that involve parameters lying on the boundary. We refer to Kopylev and Sinha [8] and Molenberghs and Verbeke [12] for the asymptotic distributions of likelihood tests when parameters lie on the boundary. As mentioned in Molenberghs and Verbeke [12] page 24 and also the Theorem 2.1 of Kopylev and Sinha [8] the asymptotic distribution of likelihood ratio test will be a mixture of a $\chi^2_0$ (i.e. degenerate distribution) and a $\chi^2_1$ with mixing coefficient $1/2$.

### 7.4. Confidence intervals of the parameters

We know that even for the univariate Poisson distribution, it is difficult to construct a confidence interval for $\lambda$ with a specified confidence coefficient of $100(1 - \alpha)$%. Using the relationship between the Poisson and $\chi^2$-distribution, lower and upper limit for the confidence interval with specified confidence coefficient can be found by interpolation in tables of the central $\chi^2$ distribution. We refer to Johnson et al. [6, Section 4.7.3] for further discussion on constructing a confidence interval for the Poisson distribution.

In the present note, we will be focusing on the Wald method of constructing confidence intervals. In general, for any parameter $\theta$ and the corresponding point estimator $\hat{\theta}$ (say), then Wald confidence interval is given by

$$
\hat{\theta} \pm Z_{\alpha/2} S.E.(\hat{\theta}),
$$

(39)

where $Z_{\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal distribution and $S.E.(\hat{\theta})$ is standard error of the estimator $\hat{\theta}$. Also note that Wald confidence interval has poor coverage properties for small sample sizes.

Since the Pseudo-Poisson marginal distribution of $X_1$ is Poisson with parameter $\lambda_1$ one can use the existing approach (using the relationship between the Poisson and $\chi^2$-distribution) to obtain a confidence interval for $\lambda_1$. For $\lambda_2$ and $\lambda_3$ confidence intervals can be obtained using the aforementioned Wald method. In the following subsections, we will be analyzing the behavior of Wald confidence interval for the parameters for small and large samples.
7.5. Examples

In the following two subsections, we provide a simulation study and give examples of real-life applications of the bivariate Pseudo-Poisson distribution.

7.5.1. Simulation data

Simulating from Pseudo models is straightforward because of the marginal and conditional structure of the model. In the following, we give a simple simulation algorithm for the bivariate Pseudo-Poisson model with linear regression. For a given $\lambda_1$, $\lambda_2$ and $\lambda_3$.

**Step 1:** Simulate $x_1$ from $\mathcal{P}(\lambda_1)$.

**Step 2:** Simulate $x_2$ from $\mathcal{P}(\lambda_2 + \lambda_3 x_1)$.

Repeat the above two steps for the desired number of observations.

We have simulated 10,000 data sets of sample size $n = 20, 30, 50, 100, 500, 1000$ from the following full and sub-models:

**Full-Model:** for the parameter values $\lambda_1 = 1, \lambda_2 = 3$ and $\lambda_3 = 4$ (cf. Figure 1 for the probability mass function plot). The corresponding moment estimates (MM) and m.l.e’s (MLE) and also their bootstrapped standard errors and confidence intervals (CI) are displayed in Table 1.\(^1\)

**Sub-Model I (i.e. when $\lambda_2 = \lambda_3$):** for the parameter values $\lambda_1 = 1$ and $\lambda_3 = 3$ (Refer Figure 2 for the probability mass function plot). The corresponding MM (also equal to MLE), bootstrapped SE and CI are displayed in Table 2.

---

**Figure 1.** Bivariate Pseudo-Poisson (Full-model) mass function plot ($n = 1000$, $\lambda_1 = 1$, $\lambda_2 = 3$, $\lambda_3 = 4$).
Table 1. Simulation (Full Model).

| n   | λ₁     | λ₂     | λ₃     | ρ     | SE (MM) | SE (MLE) | 95% CI (MM) | 95% CI (MLE) | PC  |
|-----|--------|--------|--------|-------|---------|----------|-------------|-------------|-----|
| 20  | 0.996  | 3.301  | 3.715  | 0.782 | 0.221   | 1.242    | (0.533, 1.399) | (1.285, 5.869) | 0.826|
| 30  | 1.001  | 3.184  | 3.821  | 0.799 | 0.179   | 1.077    | (0.563, 1.429) | (1.326, 6.140) | 0.829|
| 50  | 1.000  | 3.086  | 3.911  | 0.814 | 0.142   | 0.896    | (0.563, 1.429) | (1.440, 5.140) | 0.831|
| 100 | 1.000  | 3.048  | 3.953  | 0.824 | 0.099   | 0.896    | (0.563, 1.429) | (1.373, 4.823) | 0.833|
| 500 | 1.000  | 3.009  | 3.997  | 0.832 | 0.044   | 0.896    | (0.563, 1.429) | (1.766, 5.906) | 0.834|
| 1000| 1.000  | 3.002  | 3.997  | 0.833 | 0.032   | 0.896    | (0.563, 1.429) | (1.821, 4.177) | 0.838|

Table 2. Simulation (Sub-model I).

| n   | λ₁     | λ₂     | λ₃     | ρ     | SE (MM) | 95% CI (MM) | PC  |
|-----|--------|--------|--------|-------|---------|-------------|-----|
| 20  | 0.997  | 0.779  | 0.799  | 0.765 | 0.221   | (0.564, 1.430) |    |
| 30  | 0.999  | 0.771  | 0.773  | 0.768 | 0.224   | (0.644, 1.353) |    |
| 50  | 0.999  | 0.771  | 0.773  | 0.770 | 0.224   | (0.729, 1.273) |    |
| 100 | 1.000  | 0.773  | 0.773  | 0.773 | 0.224   | (0.729, 1.273) |    |
| 500 | 1.000  | 0.774  | 0.775  | 0.774 | 0.224   | (0.729, 1.273) |    |
| 1000| 1.000  | 0.775  | 0.775  | 0.775 | 0.224   | (0.729, 1.273) |    |

Sub-Model II (i.e. in the limiting case in which λ₂ = 0): for the parameter values λ₁ = 3 and λ₃ = 4 (See Figure 3 for the mass function plot). The corresponding MM (also equal to MLE), bootstrapped SE and CI are displayed in Table 3.

We summarize Tables 1–3 with following general remarks. We note that with increase in sample size, the moment and m.l.e.’s standard error (SE) decreases and the Pearson
Figure 2. Bivariate Pseudo-Poisson (Sub-model I) mass function plot ($n = 1000$, $\lambda_1 = 1$ and $\lambda_3 = 3$).

### Table 3. Simulation (Sub-model II).

| n   | P      | MM   | SE (MM)   | 95% CI (MM)   | PC     |
|-----|--------|------|-----------|---------------|--------|
| 20  | $\lambda_1$ | 3.005 | 0.391     | (2.239, 3.771) | 0.890 |
|     | $\lambda_3$ | 3.997 | 0.262     | (3.483, 4.511) |        |
|     | $\rho$     | 0.894 | 0.006     | (0.882, 0.906) |        |
| 30  | $\lambda_1$ | 2.996 | 0.319     | (2.371, 3.621) | 0.893 |
|     | $\lambda_3$ | 3.997 | 0.213     | (3.580, 4.140) |        |
|     | $\rho$     | 0.894 | 0.005     | (0.884, 0.904) |        |
| 50  | $\lambda_1$ | 3.000 | 0.247     | (2.516, 3.484) | 0.893 |
|     | $\lambda_3$ | 4.000 | 0.165     | (3.677, 4.323) |        |
|     | $\rho$     | 0.894 | 0.004     | (0.886, 0.902) |        |
| 100 | $\lambda_1$ | 2.999 | 0.175     | (2.656, 3.342) | 0.894 |
|     | $\lambda_3$ | 4.001 | 0.116     | (3.774, 4.228) |        |
|     | $\rho$     | 0.894 | 0.003     | (0.888, 0.899) |        |
| 500 | $\lambda_1$ | 2.999 | 0.077     | (2.848, 3.149) | 0.894 |
|     | $\lambda_3$ | 4.002 | 0.051     | (3.902, 4.102) |        |
|     | $\rho$     | 0.894 | 0.001     | (0.892, 0.896) |        |
| 1000| $\lambda_1$ | 3.000 | 0.055     | (2.892, 3.108) | 0.894 |
|     | $\lambda_3$ | 4.000 | 0.037     | (3.927, 4.073) |        |
|     | $\rho$     | 0.894 | 0.001     | (0.892, 0.896) |        |

The correlation (PC) converges to the population correlation. Also, the Wald confidence interval using MLE estimators has the shortest length compared to the confidence interval constructed using moment estimators.

### 7.5.2. A particular data set I

We consider a data set which is mentioned in Islam and Chowdhury [4], the source of the data is from the tenth wave of the Health and Retirement Study (HRS) (Figure 4). The
Figure 3. Bivariate Pseudo-Poisson (Sub-model II) mass function plot ($n = 1000, \lambda_1 = 3, \text{and } \lambda_3 = 4$).

Figure 4. Health and retirement study data.

data represents the number of conditions ever had ($X_1$) as mentioned by the doctors and utilization of healthcare services (say, hospital, nursing home, doctor and home care) ($X_2$). We refer Figure 4 for the histogram plot of the data.

It has been noted that the sample Pearson correlation coefficient for the above data is 0.063. Primarily, for further analyses, the data has been tested for independence
Table 4. Health and retirement study data: full model.

| n     | Parameter | Moment | m.l.e | PC          | $-2 \log L$ |
|-------|-----------|--------|-------|-------------|-------------|
| 5567  | $\lambda_1$ | 2.643  | 2.643 | 0.063       | 32766.08    |
|       | $\lambda_2$ | 0.688  | 0.64  |             |             |
|       | $\lambda_3$ | 0.031  | 0.049 |             |             |
|       | $\rho$     | 0.057  | 0.091 |             |             |

Table 5. Health and retirement study data: Sub-model I.

| n     | Parameter | Moment | $-2 \log L$ |
|-------|-----------|--------|-------------|
| 5567  | $\lambda_1$ | 2.643  | 33077.09    |
|       | $\lambda_3$ | 0.211  |             |

Table 6. Health and retirement study data: Sub-model II.

| n     | Parameter | Moment | $-2 \log L$ |
|-------|-----------|--------|-------------|
| 5567  | $\lambda_1$ | 2.643  | 43813.17    |
|       | $\lambda_3$ | 0.291  |             |

(cf. Section 6.3.3) and the $-2 \log \Lambda_3$ value is 28.359. Consequently, the assumption of variables independence is rejected.

Further, the estimated Fisher index of $X_1$ is 0.801 (approximately equi-dispersed) and the dispersion index of $X_2$ is 1.03 (slightly over-dispersed). Moment and m.l.e’s values are displayed in Table 4. Next, we consider the sub-models and fit the same data for these models. Recall that for these sub-models the m.l.e’s and the moment estimates coincide,

- Sub-Model I: That is, when $\lambda_2 = \lambda_3$, for fitted values, c.f. Table 5.
- Sub-Model II: For $\lambda_2 = 0$, the fitted values are displaced in Table 6.

Note that using the AIC criterion, the Pseudo-Poisson Full-Model fit the data better. However, we also refer to Loukas and Kemp [11] for $\chi^2$ goodness-of-fit tests for univariate and bivariate discrete distributions.

7.5.3. A particular data set II

Here, we consider a data set which is in Leiter and Hamdani [10], the source of the data is a 50-mile stretch of Interstate 95 in Prince William, Stafford and Spotsylvania counties in Eastern Virginia. The data represents the number of accident categorized as fatal accidents, injury accidents or property damage accidents, along with the corresponding number of fatalities and injuries for the period 1 January 1969 to 31 October 1970 (c.f. Figure 5 for the histogram plot).

We consider the number of fatalities as $X_1$, since the estimated Fisher index is 1.051 and the number of injury accidents as $X_2$ (estimated Fisher index is 1.141). Moment and m.l.e’s values are displayed in Table 7. Next, we consider the sub-models and fit the same data for these models. Recall that for these sub-models the m.l.e’s and the moment estimates coincide,
Figure 5. Accidents and fatalities data.

Table 7. Accidents and fatalities data: full model.

| n  | Parameter | Moment | m.l.e  | PC   | $-2 \log L$ |
|----|-----------|--------|--------|------|-------------|
| 639| $\lambda_1$ | 0.058  | 0.058  | 0.205| 1862.076    |
|    | $\lambda_2$ | 0.812  | 0.813  |      |             |
|    | $\lambda_3$ | 0.867  | 0.843  |      |             |
|    | $\rho$      | 0.219  | 0.213  |      |             |

Table 8. Accidents and fatalities data: Sub-model I.

| n  | Parameter | Moment | $-2 \log L$ |
|----|-----------|--------|-------------|
| 639| $\lambda_1$ | 0.058  | 1862.094    |
|    | $\lambda_3$ | 0.815  |             |

Table 9. Accidents and fatalities data: mirrored sub-model II.

| n  | Parameter | Moment | $-2 \log L$ |
|----|-----------|--------|-------------|
| 639| $\lambda_1$ | 0.862  | 1847.505    |
|    | $\lambda_3$ | 0.067  |             |

- Sub-Model I: That is, when $\lambda_2 = \lambda_3$, for fitted values, see Table 8.
- Mirrored Sub-Model II (c.f. Section 7): For $\lambda_2 = 0$, the fitted values are displaced in Table 9.

Note that using the AIC criteria, the Pseudo-Poisson Mirrored Sub-Model II fits the data better (see the next section for an explanation of the term ‘mirroring’).
8. The mirrored, or permuted model

When we assume that \( X_1 \sim \mathcal{P}(\lambda_1) \) and that \( X_2 \mid X_1 = x_1 \sim \mathcal{P}(\lambda_2(x_1)) \), it is natural to think that, in some unspecified way, the variable \( X_1 \) influences or, dare we say, causes \( X_2 \). But, for many data sets, the ordering of the variables is quite arbitrary and we should also entertain the possibility that the data might be better modeled by the corresponding ‘mirrored’ model in which \( X_2 \sim \mathcal{P}(\lambda_1) \) and \( X_1 \mid X_2 = x_2 \sim \mathcal{P}(\lambda_2(x_2)) \). The original model and the mirrored model are distinct and, inevitably, one of them will fit the data better than the other (except in the less interesting case in which \( X_1 \) and \( X_2 \) are independent).

With this in mind, we return to the Health and Retirement Study data. The corresponding values of the AIC criterion for our original Pseudo-Poisson, Sub-models are displayed in Table 10, together with those for the corresponding mirrored models and the Bivariate Conway–Maxwell–Poisson (COM-Poisson) model. We refer to Sellers et al. [13] for the further discussion on the bivariate COM-Poisson model. Using the Akaike information criterion, the Bivariate COM-Poisson model appears to be the best but the computation time required for fitting this model may be a problem. Also, note that BPP MSM-II is not suitable for the Health and Retirement Study data since the Pseudo-Poisson model is only appropriate when \( X_1 = 0 \) (or mirrored \( X_2 = 0 \)) implies that \( X_2 = 0 \) (or mirrored \( X_1 = 0 \)).

Now, for the Accidents and Fatalities data, note that the considered data is not suitable for the Mirrored Full model or the Sub-Model II. We refer to Table 11 for AIC values.

| Models               | No. parameters | AIC  |
|----------------------|----------------|------|
| BPP FM               | 3              | 32772.08 |
| BPP MFM              | 3              | 32783.08 |
| BPP SM-I             | 2              | 33081.09 |
| BPP MSM-I            | 2              | 35640.46 |
| BPP SM-II            | 2              | 43817.17 |
| BPP MSM-II           | 2              | – |
| BCMP                 | 6              | 32690.18 |

Notes: AIC values for Bivariate Pseudo-Poisson Full Model (BPP FM), Bivariate Pseudo-Poisson Mirrored Full Model (BPP MFM), Bivariate Pseudo-Poisson Sub-Model I (BPP SM-I), Bivariate Pseudo-Poisson Mirrored Sub-Model I (BPP MSM-I), Bivariate Pseudo-Poisson Sub-Model II (BPP SM-II), Bivariate Pseudo-Poisson Mirrored Sub-Model II (BPP MSM-II) and Bivariate COM-Poisson (BCMP) on Health and Retirement Study data.

| Models               | No. parameters | AIC  |
|----------------------|----------------|------|
| BPP FM               | 3              | 1862.076 |
| BPP MFM              | 3              | – |
| BPP SM-I             | 2              | 1866.094 |
| BPP MSM-I            | 2              | 1865.560 |
| BPP SM-II            | 2              | – |
| BPP MSM-II           | 2              | 1847.505 |
| BCMP                 | 6              | 1854.125 |

Table 10. Health and retirement study data: AIC.

Table 11. Accidents and fatalities data: AIC.
of other models. Using the Akaike information criterion, the mirrored Bivariate Pseudo-Poisson Sub-Model II model appears to be the best. For the Accidents and Fatalities data the models BPP MFM and BPP SM-II are inappropriate. Also, note that the bivariate Pseudo-Poisson mirrored Sub-model II is exactly the same model as that considered in Leiter and Hamdan [10].

There do exist other over-dispersed models which include the bivariate COM-Poisson as a special case. However, the number of parameters to fit the data and the computation time for the analysis are less in the Pseudo-Poisson model. For example, the bivariate COM-Poisson model has 6 parameters and for the above-given data size computation is very slow because of the non-existence of closed-form expressions. Also, note that for the Pseudo-Poisson Sub-Model I, both $X_1$ and $X_2$ can take any non-negative integer values but such data sets are not plausible for the Sub-Model II or the Poisson-Poisson model in Leiter and Hamdan [10] or its mirrored models. Finally, we reiterate our recommendation that the bivariate Pseudo-Poisson model should be used when the given count data has one marginal equi-dispersed and the other over-dispersed.

**Note 8.1:** Having analyzed a bivariate data set using a Pseudo-Poisson model, the analysis of the corresponding mirrored model can be implemented by repeating the analysis with the roles of the $X_1$’s and $X_2$’s interchanged.

9. **Permutations of $k$-variate models**

In the discussion of $k$-variate pseudo-models in Sections 1 and 2, in reality it will be better to consider $k!$ related models obtained by permuting the roles of the $k$ variables in the data set to be fitted. In such a situation, it will usually be the case that one and only one of the $k!$ models will turn out to provide the best fit to the data.

10. **Conclusion**

By considering bivariate models in which one marginal distribution is assumed to be of the Poisson form while the conditional distributions of the second variable, given the first, are also assumed to be of the Poisson form, we have developed flexible models called bivariate pseudo-Poisson distributions. Distributional and inferential issues have been investigated for these models. The models discussed in this paper may be considered as viable alternatives to the various bivariate Poisson models that have been introduced in the literature. The simplicity of the structure of the pseudo models allows simple simulation and straightforward parameter estimation and model fitting. They are not a panacea but we would argue that they, and their extensions to higher dimensions and permuted variants of them, deserve a place in the modelers toolkit.

**Notes**

1. SE: standard error; PC: Pearson correlation.
2. Table value of $\chi^2$ at 95% percentile is 3.842.

**Acknowledgments**

Second author is very grateful to the UGC-SAP-DSA I Programme.
Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

B. G. Manjunath http://orcid.org/0000-0003-2687-0138

References

[1] B.C. Arnold, E. Castillo, and J.M. Sarabia, *Conditional Specification of Statistical Models*, Springer Series in Statistics, New York, 1999.
[2] J.K. Filus, L.Z. Filus, and B.C. Arnold, *Families of multivariate distributions involving 'Triangular' transformations*, Commun. Stat. - Theory Methods 39 (2009), pp. 107–116.
[3] I. Ghosh, F. Marques, and S. Chakraborty, *A new bivariate Poisson distribution via conditional specification: Properties and applications*, J. Appl. Stat. (2020), pp. 1–23. doi:10.1080/02664763.2020.1793307
[4] M.A. Islam and R.I. Chowdhury, *Analysis of Repeated Measures Data*, Springer Nature, Singapore, 2017.
[5] Y. Jelmer, Introduction to nloptr: An R interface to NLopt, preprint (2020). Available at https://cran.r-project.org/web/packages/nloptr/vignettes/nloptr.pdf.
[6] N.L. Johnson, A.W. Kemp, and S. Kotz, *Univariate Discrete Distributions*, John Wiley & Sons, New Jersey, 2005.
[7] C.C. Kokonendji and P. Puig, *Fisher dispersion index for multivariate count distributions: A review and a new proposal*, J. Multivar. Anal. 165 (2018), pp. 180–193.
[8] L. Kopylev and B. Sinha, *On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary*, Sankhyā: Ser. B 73 (2008), pp. 20–41.
[9] A. Kyriakoussis and H. Papageorgiou, *On characterization of Power Series distributions by a marginal distribution and a regression function*, Ann. Inst. Stat. Math. 41 (1989), pp. 671–676.
[10] R.E. Leiter and M.A. Hamdan, *Some bivariate probability models applicable to traffic accidents and fatalities*, Int. Stat. Rev. 41 (1973), pp. 87–100.
[11] S. Loukas and C.D. Kemp, *On the chi-square-of-fit statistic for bivariate discrete distributions*, J. R. Stat. Soc. Ser. D 35 (1986), pp. 525–529.
[12] G. Molenberghs and G. Verbeke, *Likelihood ratio, score, and Wald tests in a constrained parameter space*, Am. Stat. 61 (2007), pp. 22–27.
[13] K.F. Sellers, S.D. Morris, and N. Balakrishnan, *Bivariate Conway–Maxwell–Poisson distribution: Formulation, properties, and inference*, J. Multi. Anal. 150 (2016), pp. 152–168.
[14] V. Seshadri and G.P. Patil, *A characterization of a bivariate distribution by the marginal and the conditional distributions of the same component*, Ann. Inst. Stat. Math. 15 (1964), pp. 215–221.