Bulletin T. CLVI de l'Académie serbe des sciences et des arts − 2023
Classe des Sciences mathématiques et naturelles
Sciences mathématiques, № 48

NEW BIVARIATE PROBABILITY MODELS BASED ON PANJER-TYPE RELATIONS

MEITNER CADENA, MOHAMMAD MASJED-JAMEI,
EDWARD OMEY, REIN VESILO

(Accepted at the 8th Meeting, held on November 24, 2023)

A b s t r a c t. Large families of bivariate probability models based on Panjer-type relations are proposed. These relations are generated by differential equations of probability generating functions (p.g.fs), expressed by rational functions that incorporate parameters. In this setting, models are not explicit, but their p.g.fs may be obtained explicitly by solving such differential equations. Then, for parameter estimation, minimization of distances between p.g.fs and empirical p.g.fs is used. Numerical applications to real data sets are presented.

AMS Mathematics Subject Classification (2020): 60E05, 62E10.
Key Words: Discrete bivariate random variable, power series distribution, Panjer-type relations, probability generating function, empirical probability generating function.

1. *Introduction*

Phenomena involving bivariate random variables are frequent in fields like engineering, health, finance and insurance, among others. In fact, a well-known and useful model using a bivariate random variable (X, N) is

$$Y = \sum_{i=1}^{N} X_i, \tag{1.1}$$

assuming $Y = 0$ if $N = 0$, where N is a count random number and X_i, i = 1, 2, $\ldots$, are copies of the random variable X. This model can have different interpretations, such as the accumulation of losses due to failures of machines or the total of incomes coming from different sources. Furthermore, a bivariate random variable $\Xi = (X, Y)$ may concern only count random variables X and Y. For instance, when count random variables observed in different times are compared each other. Also, continuous random variables may be discretized because of convenience or because their observed values do not have enough decimal places.

In this paper, we focus on discrete bivariate random variables $\Xi = (X, Y)$. We start by presenting the following probability distribution (p.d.) that X and Y are supposed to follow.

In a recent paper, Masjed-Jamei et al. [23] assume that the Taylor series expansion $f(\lambda) = \sum_{n=0}^{\infty} f^{(n)}(0)\lambda^n/n!$ converges for values $\lambda \in A \subseteq (-\infty, \infty)$. If $f^{(n)}(0) \geq 0, \forall n \geq 0$ and $f(\lambda) > 0$, then the probability distribution (p.d.) $(p_n(\lambda; f), n \geq 0, \lambda \in A)$ generated by f is given by

$$P(X(\lambda; f) = n) = p_n(\lambda; f) = \frac{f^{(n)}(0)}{f(\lambda)} \frac{\lambda^n}{n!} = \Theta_n(\lambda; f)\lambda^n, \quad n \geq 0.$$

Probability distributions of this type are called power series distributions (p.s.ds). They were studied among others in [15, 28] (see also [25]). The probability generating function (p.g.f.) of $X(\lambda; f)$ is given by:

$$\phi_{X(\lambda;f)}(z) = Ez^{X(\lambda;f)} = \frac{f(z\lambda)}{f(\lambda)}, \quad \lambda \in A, \ z\lambda \in A. \tag{1.2}$$

For the bivariate random variable Ξ, we consider a generalization of discrete p.d.s that are generated by bivariate power series. The earliest references to this kind of p.d.s go back to Guldberg [12], Khatri [18], and Shoukri et al. [34]. The marginal p.ds of Ξ are p.s.ds, as shown below.

Furthermore, we are interested in p.d.s that can be computed from recursive relations in order to provide them greater flexibility. In the univariate case, well-known examples of these p.d.s are the Poisson, binomial and negative binomial distributions, whose p.d.s satisfy the Panjer recursive relation, see [29]. These p.d.s are p.s.ds. Extensions of this kind of relations in the univariate case have been given in e.g. [24] and [23].

Under a bivariate setting, we propose Panjer-type recursive relations in order to have more flexible p.d.s than those based on given models. The design of these recursive relations is based on differential equations of the p.g.f. of Ξ. These equations use parametrized rational functions. Under this configuration, the bivariate p.s.ds. are generalized, thus generating new discrete bivariate p.d.s.

The previous procedure implies that the model for Ξ is not explicit. In fact, what is explicit is its p.g.f. Hence, the maximum likelihood method cannot be applied for estimating the parameters of these new models. To circumvent this problem, we propose to use the empirical probability generating function (e.p.g.f.) for estimating such parameters. Recall that if $\xi_1 = (x_1, y_1), \ldots, \xi_\tau = (x_\tau, y_\tau)$ is a sample of Ξ, its e.p.g.f. is, for $z, t > 0$,

$$\widehat{\phi}_{\Xi,\tau}(z,t) = \frac{1}{\tau} \sum_{i=1}^{\tau} z^{x_i} t^{y_i}. \tag{1.3}$$

In a univariate setting, this is an unbiased estimator of the theoretical ϕ_Ξ and has the property, for all $x > 0$ where $\phi_\Xi(x) < \infty$, almost surely $\widehat{\phi}_{\Xi,\tau}(x) \to \phi_\Xi(x)$ as $\tau \to \infty$. This convergence is uniform on closed and bounded intervals included in $(0, 1]$, see e.g. [30] and [9].

The idea is to minimize a distance between the p.g.f. and e.p.g.f. by varying the parameters associated to Ξ, i.e. those parameters incorporated in its p.g.f. This procedure has been used as an alternative to the maximum likelihood method because the latter is sensitive to outliers, see e.g. [17, 26, 7, 35] for univariate p.d.s, and e.g. [27, 13] for multivariate p.d.s. Also, the p.g.f. and e.p.g.f. have been used to test fits for discrete distributions, see [19] and e.g. [31, 5]. To the best of our knowledge, this couple of functions has not been used before for estimating parameters in a bivariate setting when the expressions of the d.f.s are unknown.

The new models are applied to real data sets in order to present a diversity of modeling alternatives and assess their performance.

The rest of the paper is organized as follows. In the next section, our discrete bivariate p.s.ds are presented. Section 3 shows a method for estimating the parameters of these p.ds. Section 4 provides an estimation method analysis via simulation. Section 5 presents numerical results when our models and competitors are applied to two real data sets. The last section gives concluding remarks.

2. *New discrete bivariate models*

We begin this section presenting the conditions to formulate our discrete bivariate probability distributions.

2.1 A bivariate power series distribution

Let f be a bivariate continuously differentiable function. More conditions on f are given later.

The general expression for the Taylor series in two variables for $f(x, y)$ may be written as

$$
\begin{aligned}
f(a+\lambda,b+\nu) &= f(a,b)+\sum_{n=1}^{\infty}\left\{\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\frac{\partial^n f}{\partial^k x\partial^{n-k}y}(a,b)\times\lambda^k\nu^{n-k}\right\}\\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{k!(n-k)!}\frac{\partial^n f}{\partial^k x\partial^{n-k}y}(a,b)\times\lambda^k\nu^{n-k}\\
&= \sum_{k=0}^{\infty}\sum_{n=k}^{\infty}\frac{1}{k!(n-k)!}\frac{\partial^n f}{\partial^k x\partial^{n-k}y}(a,b)\times\lambda^k\nu^{n-k}\\
&= \sum_{k=0}^{\infty}\sum_{n=k}^{\infty}\frac{1}{k!m!}\frac{\partial^{m+k} f}{\partial^k x\partial^{m}y}(a,b)\times\lambda^k\nu^{m}.
\end{aligned}
$$

Hence, we have

$$f(a+\lambda,b+\nu)=\sum_{n,m\geq 0}\frac{1}{n!m!}\frac{\partial^{n+m} f}{\partial^n x\partial^{m}y}(a,b)\times\lambda^n\nu^{m}.$$

For simplicity and without loss of generality we take $a=b=0$, so that

$$f(\lambda,\nu)=\sum_{n,m\geq 0}\frac{1}{n!m!}\frac{\partial^{n+m} f}{\partial^n x\partial^{m}y}(0,0)\times\lambda^n\nu^{m}.$$

Now we assume that $f(\lambda,\nu)>0$ and

$$\frac{\partial^{n+m} f}{\partial^n x\partial^{m}y}(0,0)\geq 0.$$

Then, we can define the following bivariate discrete probability distribution, also called the bivariate power series distribution.

2.2. Definition

The p.d. generated by f and associated to the bivariate random variable (r.v.) $\Xi=(X(\lambda,\nu;f),Y(\lambda,\nu;f))$ is given by,

$$
\begin{aligned}
p_{n,m}(\lambda,\nu;f) &= P(X(\lambda,\nu;f)=n,Y(\lambda,\nu;f)=m)\\
&= \frac{1}{f(\lambda,\nu)}\frac{1}{n!m!}\frac{\partial^{n+m} f(0,0)}{\partial^n x\partial^{m}y}\lambda^n\nu^m.
\end{aligned}
$$

The joint p.g.f. of this p.d. is given by

$$E(z^{X(\lambda,\nu;f)}t^{Y(\lambda,\nu;f)})\equiv\phi_{\Xi}(z,t)=\frac{f(\lambda z,\nu t)}{f(\lambda,\nu)}.$$

For further details, the reader is referred to among others [14]. For the marginals we have

$$\begin{aligned} E(z^{X(\lambda,\nu;f)}) &= \phi_\Xi(z,1) = \frac{f(\lambda z,\nu)}{f(\lambda,\nu)} \\ E(t^{Y(\lambda,\nu;f)}) &= \phi_\Xi(1,t) = \frac{f(\lambda,\nu t)}{f(\lambda,\nu)}. \end{aligned}$$

Then, we deduce that the marginals of the p.d. $p_{n,m}(\lambda,\nu;f)$ are univariate p.s.ds.

2.3. Remarks

1) Note that

$$\ln p_{n,m}(\lambda,\nu;f) = \ln\left(\frac{\partial^{n+m} f(0,0)}{\partial^n x \partial^m y}\right) - \ln f(\lambda,\nu) + n\ln\lambda + m\ln\nu + \ln\left(\frac{1}{n!m!}\right).$$

Taking the derivative with respect to λ, we find

$$\frac{1}{p_{n,m}(\lambda,\nu;f)}\frac{\partial}{\partial\lambda}p_{n,m}(\lambda,\nu;f) = \frac{n}{\lambda} - \frac{1}{f(\lambda,\nu)}\frac{\partial}{\partial\lambda}f(\lambda,\nu).$$

2) Also note that

$$\begin{aligned} p_{n+1,m}(\lambda,\nu;f) &= \frac{1}{f(\lambda,\nu)}\frac{1}{(n+1)!m!}\frac{\partial^{n+1+m} f(0,0)}{\partial^{n+1}x\partial^m y}\lambda^{n+1}\nu^m \\ &= p_{n,m}(\lambda,\nu;f)\frac{\lambda}{(n+1)}\frac{\partial^{n+1+m} f(0,0)/\partial^{n+1}x\partial^m y}{\partial^{n+m} f(0,0)/\partial^n x\partial^m y}. \end{aligned}$$

2.4. Example

For the exponential function $f(\lambda,\nu) = \exp(\alpha\lambda + \beta\nu)$, for some α, $\beta \geq 0$, we find that

$$f(\lambda,\nu) = \sum_{n\geq 0}\sum_{m\geq 0}\frac{1}{n!m!}\alpha^n\beta^m\lambda^n\nu^m,$$

and

$$\begin{aligned} p_{n,m}(\lambda,\nu;f) &= \frac{1}{n!m!}\alpha^n\beta^m\lambda^n\nu^m e^{-\alpha\lambda-\beta\nu} \\ &= P(X(\alpha;\lambda)=n)P(X(\beta;\nu)=m), \end{aligned}$$

where $X(\alpha;\lambda)$ and $X(\beta;\nu)$ are independent Poisson random variables.

2.5 Discrete bivariate probability distributions based on recursive relations

We consider $(U,V),(U_i,V_i), i=1,2,\ldots$ i.i.d. random vectors independent from $(N,M)=(X(\lambda,\nu;f),Y(\lambda,\nu;f))$, and let $\phi_{(U,V)}(z,t)=Ez^Ut^V$. Consider the random sums $S_0=T_0=0$ and for $n,m\geq 1$,

$$(S_n,T_m)=\left(\sum_{i=1}^{n}U_i,\sum_{j=1}^{m}V_j\right).$$

Now, we study the vector of random sums (S_N,T_M). Clearly, we have

$$\begin{aligned} E\left(z^{S_N}t^{T_M}\right) &= \sum_{n,m=0}^{\infty}E(z^{S_n}t^{T_m})p_{n,m} \\ &= \sum_{n=0}^{\infty}\sum_{m=n+1}^{\infty}\phi_{(U,V)}^{n}(z,t)\phi_V^{m-n}(t)p_{n,m} \\ &\quad +\sum_{m=0}^{\infty}\sum_{n=m+1}^{\infty}\phi_{(U,V)}^{m}(z,t)\phi_U^{n-m}(z)p_{n,m}+\sum_{n=0}^{\infty}\phi_{(U,V)}^{n}(z,t)p_{n,n}. \end{aligned}$$

Hence, we have

$$\begin{aligned} E\left(z^{S_N}t^{T_M}\right) &= \sum_{r=1}^{\infty}\phi_V^{r}(t)\sum_{n=0}^{\infty}\phi_{(U,V)}^{n}(z,t)p_{n,n+r} \\ &\quad +\sum_{s=1}^{\infty}\phi_U^{s}(z)\sum_{m=0}^{\infty}\phi_{(U,V)}^{m}(z,t)p_{m+s,m}+\sum_{n=0}^{\infty}\phi_{(U,V)}^{n}(z,t)p_{n,n}. \end{aligned}$$

When the U and V are independent, we have

$$\begin{aligned} E\left(z^{S_N}t^{T_M}\right) &= \sum_{n,m=0}^{\infty}E(z^{S_n}t^{T_m})p_{n,m} \\ &= \sum_{n,m=0}^{\infty}(E(z^U))^n(E(t^V))^m p_{n,m} \\ &= \frac{f(\lambda Ez^U,\nu Et^V)}{f(\lambda,\nu)}. \qquad (2.1) \end{aligned}$$

Lazarova and Minkova [20] analyzed (2.1) assuming that U and V follow geometric distributions, i.e., $Ez^U=(1-p_U)z/(1-p_Uz)$ and $Et^V=(1-p_V)t/(1-p_Vt)$. Moreover, for various particular choices of f these authors determined detailed expressions for the p.d.f. of $\Xi=(S_N,T_M)$.

Also, we set

$$\phi_\Xi(z,t) = \sum_{i\geq 0}\sum_{j\geq 0} p_{i,j} z^i t^j. \tag{2.2}$$

If we assume that $f(x,y) = \exp(\alpha x + \beta y)$, we find that, by considering that U and V are independent,

$$E(z^{S_N} t^{T_M}) = \exp(\alpha\lambda(Ez^U - 1) + \beta\nu(Et^V - 1))$$

or

$$\phi_\Xi(z,t) = \exp(\alpha\lambda(\phi_U(z) - 1) + \beta\nu(\phi_V(t) - 1)).$$

Taking derivatives with respect to z, we find that

$$\frac{\partial\phi_\Xi}{\partial z}(z,t) = \phi_\Xi(z,t) \times \alpha\lambda\phi'_U(z), \tag{2.3}$$

and, then, with respect to t,

$$\frac{\partial^2\phi_\Xi}{\partial z\partial t}(z,t) = \phi_\Xi(z,t) \times \alpha\lambda\phi'_U(z) \times \beta\nu\phi'_V(t). \tag{2.4}$$

Now, we assume that for $i \in \{U, V\}$, we have

$$\phi_i(x) = \frac{A_i + B_i x}{C_i + D_i x + E_i x^2}.$$

Note that $\phi'_i(x)$ is of the form

$$\phi'_i(x) = \frac{\sum\limits_{k=0}^{2} a_i(k)x^k}{\sum\limits_{k=0}^{4} b_i(k)x^k}, \tag{2.5}$$

where the conditions $b_U(0) \neq 0$ and $b_V(0) \neq 0$, and $b_U(4) \neq 0$, $b_V(4) \neq 0$, $a_U(2) \neq 0$ and $a_V(2) \neq 0$, are incorporated.

Special cases of $\phi_i(x)$ are as follows:

- Minkova [24] and Momeni [25] studied geometric random variables with p.g.fs

$$\phi_U(z) = \frac{(1-p)z}{1-pz}.$$

- Masjed-Jamei et al. [23] used a modified power distribution as follows:

$$\phi_U(z) = \frac{(1-\alpha)z}{1-\alpha z}\frac{(1-\beta)z}{1-\beta z}.$$

- Gomez and Calderin-Ojeda [11] and Bakouch et al. [4] studied generating functions of the form

$$\phi_U(z) = \beta \frac{1-p}{1-pz} + (1-\beta)\frac{(1-p)^2}{(1-pz)^2}.$$

- Sankaran [32] studied p.g.fs of the similar form with $p = 1/(1+\theta)$.

2.5.1 Remark. The previous special cases show us that not all the parameters A_i, B_i, C_i, D_i and E_i in the p.g.fs associated to U and V may be unknown. For instance, if the p.d. of U is the geometric distribution, in a design as

$$\phi_U(z) = \frac{A_U + B_U x}{C_U + D_U x + E_U x^2},$$

it is convenient to fix $A_U = 0$, $C_U = 1$ and $E_U = 0$.

2.5.2 A recursion of Panjer type. The relations (2.4), (2.2) and (2.5) allow the formulation of the following recursive relations. From these equations, we have

$$\sum_{i,j\geq 0} ijp_{i,j} z^{i-1} t^{j-1} = \sum_{i,j\geq 0} p_{i,j} z^i t^j \times \alpha\lambda \frac{\sum_{k=0}^{2} a_U(k) z^k}{\sum_{k=0}^{4} b_U(k) z^k} \times \beta\nu \frac{\sum_{l=0}^{2} a_V(l) t^l}{\sum_{l=0}^{4} b_V(l) t^l}. \quad (2.6)$$

Then, it follows that

$$\sum_{k,l=0}^{4} b_V(l) b_U(k) \sum_{i,j\geq 1} ijp_{i,j} z^{i+k-1} t^{j+l-1}$$
$$= \alpha\lambda\beta\nu \sum_{k,l=0}^{2} a_U(k) a_V(l) \sum_{i\geq 0}\sum_{j\geq 0} p_{i,j} z^{i+k} t^{j+l},$$

so that

$$\sum_{k,l=0}^{4}\sum_{n\geq k}\sum_{m\geq l} b_U(k) b_V(l)(n+1-k)(m+1-l) p_{n+1-k,m+1-l} z^n t^m$$
$$= \alpha\lambda\beta\nu \sum_{k,l=0}^{2}\sum_{n\geq k}\sum_{m\geq l} a_U(k) a_V(l) p_{n-k,m-l} z^n t^m. \quad (2.7)$$

The coefficients of $z^n t^m$ for $n, m \geq 4$ are given by

$$\sum_{k=0}^{4}\sum_{l=0}^{4} b_U(k)b_V(l)(n+1-k)(m+1-l)p_{n+1-k,m+1-l} = \alpha\lambda\beta\nu \sum_{k=0}^{2}\sum_{l=0}^{2} a_U(k)a_V(l)p_{n-k,m-l}.$$

It follows that

$$\begin{aligned} &b_U(0)b_V(0)(n+1)(m+1)p_{n+1,m+1} \\ &\quad = \alpha\lambda\beta\nu \sum_{k=0}^{2}\sum_{l=0}^{2} a_U(k)a_V(l)p_{n-k,m-l} \\ &\qquad - \sum_{k,l\geq 0, \max(k,l)>0}^{4} b_U(k)b_V(l)(n+1-k)(m+1-l)p_{n+1-k,m+1-l}. \end{aligned}$$

Some straightforward calculations show that

$$\begin{aligned} &b_U(0)b_V(0)(n+1)(m+1)p_{n+1,m+1} \\ &\quad = \sum_{k,l=0}^{2} (\alpha\lambda\beta\nu a_U(k)a_V(l) - b_U(k+1)b_V(l+1)(n-k)(m-l))p_{n-k,m-l} \\ &\quad - \sum_{l=0}^{3} b_U(0)b_V(l+1)(n+1)(m-l)p_{n+1,m-l} \\ &\quad - \sum_{k=0}^{3} b_U(k+1)b_V(0)(n-k)(m+1)p_{n-k,m+1} \\ &\quad - \sum_{k=0}^{3} b_U(k+1)b_V(4)(n-k)(m-3)p_{n-k,m-3} \\ &\quad - \sum_{l=0}^{3} b_U(4)b_V(l+1)(n-3)(m-l)p_{n-3,m-l} \\ &\quad = I - II - III - IV - V. \end{aligned}$$

Now, we divide by $b_U(0)b_V(0)(n+1)(m+1)$. For the result we denote the terms on the right-hand side by $I', II', \dots, V'$. Denoting

$$\xi_{i,j,k} = b_i(j+1)\left(1 - \frac{j+1}{k+1}\right),$$

we find

$$\begin{aligned}
I' &= \frac{1}{b_U(0)b_V(0)}\sum_{k,l=0}^{2}\left(\frac{\alpha\lambda\beta\nu a_U(k)a_V(l)}{(n+1)(m+1)} - \xi_{U,k,n}\xi_{V,l,m}\right)p_{n-k,m-l}\\
II' &= \frac{1}{b_V(0)}\sum_{l=0}^{3}\xi_{V,l,m}p_{n+1,m-l}\\
III' &= \frac{1}{b_U(0)}\sum_{k=0}^{3}\xi_{U,k,n}p_{n-k,m+1}\\
IV' &= \frac{\xi_{V,3,m}}{b_U(0)b_V(0)}\sum_{k=0}^{3}\xi_{U,k,n}p_{n-k,m-3}\\
V' &= \frac{\xi_{U,3,n}}{b_U(0)b_V(0)}\sum_{l=0}^{3}\xi_{V,l,m}p_{n-3,m-l}.
\end{aligned}$$

Hence, we have

$$p_{n+1,m+1} = \begin{cases} \dfrac{\alpha\lambda\beta\nu a_U(0)a_V(0)}{b_U(0)b_V(0)}p_{0,0} & \text{if } n=0 \text{ and } m=0,\\ VI & \text{if } n=0 \text{ and } m\geq 1,\\ VII & \text{if } n\geq 1 \text{ and } m=0,\\ VIII - IX - X - XI & \text{if } 1\leq n\leq 4 \text{ or } 1\leq m\leq 4,\\ I' - II' - III' - IV' - V' & \text{if } n, m\geq 4, \end{cases} \tag{2.8}$$

where

$$\begin{aligned}
VI &= \frac{\alpha\lambda\beta\nu a_U(0)}{b_U(0)b_V(0)(m+1)}\sum_{l=0}^{\min(m,2)} a_V(l)p_{0,m-l}\\
&\quad -\frac{1}{b_V(0)}\sum_{l=0}^{\min(m,4)-1}\xi_{V,l,m}p_{1,m-l},\\
VII &= \frac{\alpha\lambda\beta\nu a_V(0)}{b_U(0)b_V(0)(n+1)}\sum_{k=0}^{\min(n,2)} a_U(l)p_{n-k,0}\\
&\quad -\frac{1}{b_U(0)}\sum_{k=0}^{\min(n,4)-1}\xi_{U,k,n}p_{n-k,1},
\end{aligned}$$

$$VIII = \frac{\alpha\lambda\beta\nu}{b_U(0)b_V(0)(n+1)(m+1)} \sum_{k=0}^{\min(n,2)} \sum_{l=0}^{\min(m,2)} a_U(k)a_V(l)p_{n-k,m-l},$$

$$IX = \frac{1}{b_U(0)b_V(0)} \sum_{k=0}^{\min(n,4)-1} \sum_{l=0}^{\min(m,4)-1} \xi_{U,k,n}\xi_{V,l,m}p_{n-k,m-l},$$

$$X = \frac{1}{b_V(0)} \sum_{l=0}^{\min(m,4)-1} \xi_{V,l,m}p_{n+1,m-l},$$

$$XI = \frac{1}{b_U(0)} \sum_{k=0}^{\min(n,4)-1} \xi_{U,k,n}p_{n-k,m+1}.$$

The Panjer-type relations given by (2.8) are those that we propose to make calculations recursively. These results are collected in the following theorem.

Theorem 2.1. *Let $\Xi = (U,V)$ be a bivariate r.v. Assume that the p.g.f. of Ξ, ϕ_Ξ, is differentiable and satisfies, for some $z, t > 0$,*

$$\frac{\partial^2 \phi_\Xi}{\partial z \partial t}(z,t) = \kappa_U \kappa_V Q_U(z) Q_V(t) \phi_\Xi(z,t), \tag{2.9}$$

where $\kappa_U = \alpha\lambda$ and $\kappa_V = \beta\nu$, where $\alpha, \lambda, \beta, \nu > 0$, and, for $\eta \in \{U,V\}$,

$$Q_\eta(s) = \frac{\sum_{k=0}^{2} a_\eta(k)s^k}{\sum_{k=0}^{4} b_\eta(k)s^k},$$

with $a_\eta(0) \neq 0$, $a_\eta(2) \neq 0$, $b_\eta(0) \neq 0$, and $b_\eta(4) \neq 0$. Then, the p.d.f. of $\Xi = (U,V)$, $p_{i,j} = P(U = i, V = j)$, satisfies (2.8), given $p_{0,0}$, $p_{i,0}$ for $i \geq 1$ and $p_{0,j}$ for $j \geq 1$.

2.5.3 Remarks.

1) The recursive relation (2.8) needs to have $p_{0,0}$ as given, and also $p_{i,0}$ for $i \geq 1$ and $p_{0,j}$ for $j \geq 1$. This issue is circumvented in the following subsection.

2) For a fixed $j \geq 0$ or $i \geq 0$, we get univariate p.d.s from (2.8), by rescaling $p_{i,j}$.

3) If the constraints $b_U(0) \neq 0$ and $b_V(0) \neq 0$, and $b_U(4) \neq 0$, $b_V(4) \neq 0$, $a_U(2) \neq 0$ and $a_V(2) \neq 0$, are not satisfied, (2.8) need to be adapted to the new conditions.

4) Our results have the following two advantages with respect to the ones provided by Vernic [37]. First, all of them are the same in some situations. For instance,

when we have $a_U(2) = a_V(2) = b_U(3) = b_U(4) = b_V(3) = b_V(4) = 0$. Second, our results generalize the ones of Vernic when the previous equalities do not hold.

5) The relations where $p_{n+1,m+1}$ is determined by previous terms of the sequence are of Panjer type. Indeed, the terms multiplying the probabilities are like $(a_1 + b_1/n)(a_2 + b_2/m)$, where we found two factors because the bivariate setting.

6) The assumption of having ϕ_i, $i \in \{U, V\}$, as rational functions leads evidently to have ϕ_i', $i \in \{U, V\}$, as rational functions too. However, (2.8) is obtained from (2.5). This fact implies that ϕ_i, $i \in \{U, V\}$, do not need to be rational functions.

2.5.4 An initialization of the previous Panjer-type relations. Following is the description of a procedure to initialize the Panjer-type relations obtained in the previous subsection. This proposal allows overcoming the remarks presented in Subsection 2.5.3. In fact, by using this procedure, we go to prove that there is needed no extra parameter.

Let us consider the case $p_{i,0}$ for $i \geq 0$. We aim to find Panjer-type relations for these probabilities, but using the parameters involved in (2.8). Using (2.3) and (2.5), we have

$$\sum_{i,j\geq 0} ip_{i,j}z^{i-1}t^j = \sum_{i,j\geq 0} p_{i,j}z^i t^j \times \alpha\lambda \frac{\sum_{k=0}^{2} a_U(k)z^k}{\sum_{k=0}^{4} b_U(k)z^k}.$$

Then, equating the coefficients for t^0 gives

$$\sum_{i\geq 0} ip_{i,0}z^{i-1} = \sum_{i\geq 0} p_{i,0}z^i \times \alpha\lambda \frac{\sum_{k=0}^{2} a_U(k)z^k}{\sum_{k=0}^{4} b_U(k)z^k}, \tag{2.10}$$

so that

$$\sum_{k=0}^{4} b_U(k)z^k \times \sum_{i\geq 0} ip_{i,0}z^{i-1} = \alpha\lambda \sum_{i\geq 0} p_{i,0}z^i \times \sum_{k=0}^{2} a_U(k)z^k.$$

It follows that

$$\sum_{k=0}^{4}\sum_{i\geq 0} b_U(k)ip_{i,0}z^{i+k-1} = \alpha\lambda \sum_{k=0}^{2}\sum_{i\geq 0} a_U(k)p_{i,0}z^{i+k},$$

i.e., we have

$$\sum_{k=0}^{4} b_U(k) \sum_{i\geq 1} i p_{i,0} z^{i+k-1} = \alpha\lambda \sum_{k=0}^{2} \sum_{i\geq 0} a_U(k) p_{i,0} z^{i+k}.$$

Equivalently, we need that

$$\sum_{k=0}^{4} \sum_{m\geq k} b_U(k)(m+1-k) p_{m+1-k,0} z^m = \alpha\lambda \sum_{k=0}^{2} \sum_{m\geq k} a_U(k) p_{m-k,0} z^m.$$

Equating the coefficients of z^m for $m \geq 4$, we find that

$$b_U(0)(m+1)p_{m+1,0} + \sum_{k=1}^{4} b_U(k)(m+1-k) p_{m+1-k,0} = \alpha\lambda \sum_{k=0}^{2} a_U(k) p_{m-k,0}.$$

From here, it follows that, by using the notation $\xi_{i,j,k}$ introduced above,

$$p_{m+1,0} = \frac{1}{b_U(0)} \sum_{k=0}^{2} \left(\alpha\lambda \frac{a_U(k)}{m+1} + \xi_{U,k,m} \right) p_{m-k,0} - \frac{\xi_{U,3,m}}{b_U(0)} p_{m-3,0}.$$

Hence, we have

$$p_{m+1,0} = \begin{cases} \frac{\alpha\lambda a_U(0)}{b_U(0)} p_{0,0} & \text{if } m = 0, \\ \frac{1}{2b_U(0)} \left(\alpha\lambda a_U(0) p_{0,0} + (\alpha\lambda a_U(0) - b_U(1)) p_{1,0} \right) & \text{if } m = 1, \\ \frac{1}{3b_U(0)} \left(\alpha\lambda \sum_{k=0}^{2} a_U(k) p_{2-k,0} - \sum_{k=1}^{2} b_U(k)(3-k) p_{3-k,0} \right) & \text{if } m = 2, \\ \frac{1}{4b_U(0)} \left(\alpha\lambda \sum_{k=0}^{2} a_U(k) p_{3-k,0} - \sum_{k=1}^{3} b_U(k)(4-k) p_{4-k,0} \right) & \text{if } m = 3, \\ \frac{1}{b_U(0)} \sum_{k=0}^{2} \left(\alpha\lambda \frac{a_U(k)}{m+1} + \xi_{U,k,m} \right) p_{m-k,0} - \frac{\xi_{U,3,m}}{b_U(0)} p_{m-3,0} & \text{if } m \geq 4. \end{cases} \tag{2.11}$$

In a similar way, we get

$$p_{0,m+1} = \begin{cases} \dfrac{\beta\nu a_V(0)}{b_V(0)} p_{0,0} & \text{if } m = 0, \\ \dfrac{1}{2b_V(0)} \left(\beta\nu a_V(0) p_{0,0} + (\beta\nu a_V(0) - b_V(1)) p_{0,1}\right) & \text{if } m = 1, \\ \dfrac{1}{3b_V(0)} \left(\beta\nu \sum\limits_{l=0}^{2} a_V(l) p_{0,2-l} - \sum\limits_{l=1}^{2} b_V(l)(3-l) p_{0,3-l}\right) & \text{if } m = 2, \\ \frac{1}{4b_V(0)} \left(\beta\nu \sum\limits_{l=0}^{2} a_V(l) p_{0,3-l} - \sum\limits_{l=1}^{3} b_V(l)(4-l) p_{0,4-l}\right) & \text{if } m = 3, \\ \dfrac{1}{b_V(0)} \sum\limits_{l=0}^{2} \left(\beta\nu \frac{a_V(l)}{m+1} + \xi_{V,l,m}\right) p_{0,m-l} - \dfrac{\xi_{V,3,m}}{b_V(0)} p_{0,m-3} & \text{if } m \geq 4. \end{cases} \tag{2.12}$$

2.5.5 Remarks.

1) The recursive relations (2.11) and (2.12) are of Panjer type.

2) The recursive relations (2.11) and (2.12) depend on only $p_{0,0}$, assumed that the parameters a_i and b_i, $i \in \{U, V\}$, are given. Next, $p_{0,0}$ is computed from the natural condition $\sum_{i,j\geq 0} p_{i,j} = 1$.

3. *Method for parameter estimation*

Let $\Xi = (U, V)$ be a discrete bivariate random vector with d.f. as defined via the recursive relations in Subsections 2.2.1 and 2.2.2. Such definitions do not provide the explicit d.f. of this r.v., but its p.g.fs involving parameters are known. In this section, we aim to estimate such parameters.

For what follows, without loss of generality we assume that U and V take non-negative integer values. Let $(u_1, v_1), \ldots, (u_\tau, v_\tau)$ be a sample of independent observations of (U, V).

Assuming the hypotheses given in the previous section, we have the equation (2.4). Denoting $Q_U(z) = \phi'_U(z)$ and $Q_V(t) = \phi'_V(t)$, all of them being rational functions, this equation can be written as, for some constant $\kappa > 0$,

$$\frac{\partial^2 \phi_\Xi}{\partial z \partial t}(z, t) = \kappa_U \kappa_V Q_U(z) Q_V(t) \phi_\Xi(z, t).$$

Consequently, a solution of this equation is

$$\phi_\Xi(z, t) = \exp\left(-\left(\kappa_U \int_z^1 Q_U(s)ds + \kappa_V \int_t^1 Q_V(s)ds\right)\right). \tag{3.1}$$

We denote by $\phi_{\Xi,\boldsymbol{\theta}}$ the function ϕ_Ξ when considering the parameters $\boldsymbol{\theta}$. We obtain $Q_{U,\boldsymbol{\theta}}$ and $Q_{V,\boldsymbol{\theta}}$ correspondingly.

In order to estimate the parameters of the p.g.f. of Ξ given the sample provided above, we use the p.g.f. and e.p.g.f. of this variable, see $\widehat{\phi}_{\Xi,\tau}$ defined in (1.3). The idea is to fit the e.p.g.f. of Ξ by using its p.g.f., $\phi_{\Xi,\boldsymbol{\theta}}$ where $\boldsymbol{\theta}$ is the vector of parameters associated to this p.g.f. To this aim, we adopt the Hellinger type distance (h.t.d.) for two dimensions

$$d_\tau(\boldsymbol{\theta}) = \int_0^1 \int_0^1 \left(\phi^\alpha_{\boldsymbol{\theta},\Xi}(z,t) - \widehat{\phi}^\alpha_{\Xi,\tau}(z,t)\right)^2 dz\, dt$$

proposed in [27]. Note that $d_\tau(\boldsymbol{\theta}) = 0$ if and only if $\phi^\alpha_{\Xi,\boldsymbol{\theta}}(z,t) = \widehat{\phi}^\alpha_{\Xi,\tau}(z,t)$, for all $\alpha > 0$ and $z, t \in [0,1]$. Jiménez-Gamero and Batsidis [13] proposed variants for this distance, by incorporating a function weight in the integral expression.

Let Θ be the set of parameters $\boldsymbol{\theta}$ associated to Ξ through its p.g.f.

The minimum rational h.t.d. (m.r.h.t.d.) estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_\tau = \min_{\boldsymbol{\theta}\in\Omega} d_\tau(\boldsymbol{\theta})$, is a strongly consistent estimator, as proved in [27]. This property of $\widehat{\boldsymbol{\theta}}_\tau$ is expressed in the following adapted result.

Proposition 3.1 (Proposition 1 given in [27])**.** *Fix $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$. Assume that the parameter space Ω is compact. Then, almost surely $\left|\phi^\alpha_{\Xi,\tau}(z,t) - \phi^\alpha_{\Xi,\boldsymbol{\theta}_0}(z,t)\right| \to 0$ implies that almost surely $\left|\widehat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}_0\right| \to 0$ as $\tau \to \infty$.*

An extra hypothesis included in Proposition 1 given in [27] is that $\phi_{\Xi,\boldsymbol{\theta}}(z,t)$ is differentiable with respect to $\boldsymbol{\theta}$ under the integral sign. However, such an assumption is obviously satisfied in our setting since $\phi_{\Xi,\boldsymbol{\theta}}$ is based on rational functions. Furthermore, we fix $\alpha = 1/2$ in order to facilitate computations. In this case, $\widehat{\boldsymbol{\theta}}_\tau$ satisfies other nice asymptotic properties because $\alpha < 1$, see e.g. [13]. Some of them are reviewed below.

In order to find nice properties for the m.r.h.t.d. estimator, we assume the following:

Assumption. *The m.r.h.t.d. has a unique minimum at $\boldsymbol{\theta}_* \in \Theta$.*

Besides this assumption, Jiménez-Gamero and Batsidis [13] assumed two more. One to guarantee continuity of the m.r.h.t.d. as a function of $\boldsymbol{\theta}$, and another to guarantee twice continuously differentiability of $\phi_{\Xi,\boldsymbol{\theta}}(z,t)$. Both of them are satisfied by the m.r.h.t.d. and $\phi_{\Xi,\boldsymbol{\theta}}(z,t)$ because of their construction based on continuously differentiable functions. Then, these authors proved that the m.r.h.t.d. estimator satisfies the following result.

Theorem 3.1 (Theorem 1 in [13])**.** *Suppose that Assumption holds and that $P(\Xi = 0) > 0$, then almost surely $\widehat{\boldsymbol{\theta}}_\tau \to \boldsymbol{\theta}_*$ as $\tau \to \infty$.*

Furthermore, the following result also proved by these authors presents the convergence rate of the previous limit.

Theorem 3.2 (Theorem 2 in [13]). *Suppose that Assumption holds, then*

$$\sqrt{\tau}\big(\widehat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}_*\big) = \frac{1}{\sqrt{\tau}}\sum_{i=1}^{\tau} \ell((u_i, v_i), \boldsymbol{\theta}_*) + o_P(1),$$

where $\ell((u,v),\boldsymbol{\theta}) = D^{-1}(\boldsymbol{\theta})h((u,v),\boldsymbol{\theta})$,

$$\begin{aligned} D(\boldsymbol{\theta}) =& \frac{1}{2}\int_0^1\int_0^1 \frac{1}{\phi_{\Xi,\boldsymbol{\theta}}(z,t)} \nabla_{\boldsymbol{\theta}}\phi_{\Xi,\boldsymbol{\theta}}(z,t)\nabla_{\boldsymbol{\theta}}\phi_{\Xi,\boldsymbol{\theta}}(z,t)'\,dtdz \\ &+ \frac{1}{2}\int_0^1\int_0^1 \frac{\phi_{\Xi}^{1/2}(z,t) - \phi_{\Xi,\boldsymbol{\theta}}^{1/2}(z,t)}{\phi_{\Xi,\boldsymbol{\theta}}^{3/2}(z,t)} \nabla_{\boldsymbol{\theta}}\phi_{\Xi,\boldsymbol{\theta}}(z,t)\nabla_{\boldsymbol{\theta}}\phi_{\Xi,\boldsymbol{\theta}}(z,t)'\,dtdz \\ &- \int_0^1\int_0^1 \frac{\phi_{\Xi}^{1/2}(z,t) - \phi_{\Xi,\boldsymbol{\theta}}^{1/2}(z,t)}{\phi_{\Xi,\boldsymbol{\theta}}^{1/2}(z,t)} \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\phi_{\Xi,\boldsymbol{\theta}}(z,t)\,dtdz, \end{aligned}$$

and

$$h((u,v),\boldsymbol{\theta}) = -\frac{1}{2}\int_0^1\int_0^1 \frac{z^u t^v - \phi_{\Xi}(z,t)}{\phi_{\Xi}^{1/2}(z,t)\phi_{\Xi,\boldsymbol{\theta}}^{1/2}(z,t)} \nabla_{\boldsymbol{\theta}}\phi_{\Xi,\boldsymbol{\theta}}(z,t)\,dtdz.$$

4. *Estimation method analysis via simulation*

In this section, we assess the performance of the method proposed for estimating parameters when the sample size varies and outliers are included in simulated data. To this aim, we consider simulations of a r.v. $\Xi = (X_1, X_2)$ following the bivariate Poisson distribution (BPD) given by $X_i = R_0 + R_i$, $i \in \{1,2\}$, where R_0, R_1, and R_2 are independent r.v.s following Poisson distributions with parameters $p_0 = 0.1$, $p_1 = 1.0$, and $p_2 = 1.0$, respectively. Such simulations are performed using the function `rbp` of the package `bzinb` available in the environment programming R. Since the model of Ξ is known, its parameters are estimated by using the maximum likelihood method implemented in the function `bp` also included in the package indicated previously.

We analyze sample sizes from 50 to 500, increasing these sizes by 50 each time. These data are also analyzed when 10 % of them are replaced by outliers. These outliers are introducing by multiplying by 3 the last 10 % of data obtained randomly.

Note that the p.g.f. associated with Ξ does not allow a representation like (2.9). In order to have a r.v. close to Ξ, but allowing a representation like (2.9), we consider $\Xi' = (R_1, R_2)$. The distribution associated to Ξ' is denoted by BM_0. For this model, the following couple of one-parameter functions Q_U and Q_V are such that:

$$Q_U(x) = \alpha_U \quad \text{and} \quad Q_V(x) = \alpha_V,$$

respectively. Furthermore, following the notations of Theorem 2.1, we find that some constants of Q_U and Q_V are fixed. More precisely, we have

$$\kappa_U = 1,\ a_U(0) = \alpha_U,\ b_U(0) = 1,\ \kappa_V = 1,\ a_V(0) = \alpha_U,\ b_V(0) = 1.$$

Also, because the introduction of outliers can lead to the fitting of models that are different from the one associated to Ξ, we consider the following two variants of our bivariate models (BM). For the first model, BM_1, its corresponding two-parameter functions Q_U and Q_V are:

$$Q_U(x) = \alpha_U \frac{1 + 0.5\beta_U x}{1 - (\beta_U x)^2} \quad \text{and} \quad Q_V(x) = \alpha_V \frac{1 + 0.5\beta_V x}{1 - (\beta_V x)^2}, \tag{4.1}$$

respectively. Note that Q_U and Q_V conform the conditions to generate the Panjer-type relations like (2.8). However, clearly, the integrals of Q_U and Q_V are not rational functions. Furthermore, following the notations of Theorem 2.1, we find that some constants of Q_U and Q_V are fixed. More precisely, we have

$$\kappa_U = 1, a_U(0) = \alpha_U, a_U(1) = 0.5\alpha_U\beta_U, b_U(0) = 1, b_U(1) = 0, b_U(2) = -\beta_U^2,$$

and

$$\kappa_V = 1, a_V(0) = \alpha_V, a_V(1) = 0.5\alpha_V\beta_V, b_V(0) = 1, b_V(1) = 0, b_V(2) = -\beta_V^2.$$

For the second model, BM_2, its corresponding two-parameter functions Q_U and Q_V are:

$$Q_U(x) = \alpha_U \frac{1 + \beta_U x + (\beta_U x)^2}{1 - 0.5(\beta_U x)^3} \quad \text{and} \quad Q_V(x) = \alpha_V \frac{1 + \beta_V x + (\beta_V x)^2}{1 - 0.5(\beta_V x)^3}, \tag{4.2}$$

respectively. In this case, Q_U and Q_V also conform the conditions to generate the Panjer-type relations like (2.8). However, the integrals of Q_U and Q_V are not rational functions. With respect to the notations of Theorem 2.1, some constants of Q_U and Q_V are fixed as follows. $\kappa_U = 1$, $a_U(0) = \alpha_U$, $a_U(1) = \alpha_U\beta_U$, $a_U(2) = \alpha_U\beta_U^2$, $b_U(0) = 1$, $b_U(1) = 0$, $b_U(2) = -\beta_U^3$, and $\kappa_V = 1$, $a_V(0) = \alpha_V$, $a_V(1) = \alpha_V\beta_V$, $a_V(2) = \alpha_V\beta_V^2$, $b_V(0) = 1$, $b_V(1) = 0$, $b_V(2) = -\beta_V^3$. For parameter estimating of our models, we use suitable adaptations of the initial values given by (2.11) and (2.12).

We study goodness-of-fits of all models of simulated data. To this aim, the Akaike criterion information (AIC), see [2], and the Bayesian criterion information (BIC), see [36], are used. If ℓ is the log-likelihood of a model given a sample size τ, and q is the number of parameters estimated, these measures are defined by $-2\ell+2q$ and $-2\ell + q \ln \tau$, respectively.

As the 3-parameter Poisson distribution is the right distribution for data without outliers, we divide the AIC and BIC values by the ones of this distribution, respectively. In this way, all those results are comparable through the diverse sample sizes taken into account. Note that, the AIC and BIC values increase when the sample size increases. Figure 1 presents those outputs. As expected, all models perform worse than the 3-parameter Poisson distribution. However, all of them except BM_1 tend to reach the AIC and BIC values of the 3-parameter Poisson distribution when the sample size increases and data do not contain outliers. Considering that the data contain outliers, all models tend to perform worse each time the sample size increases.

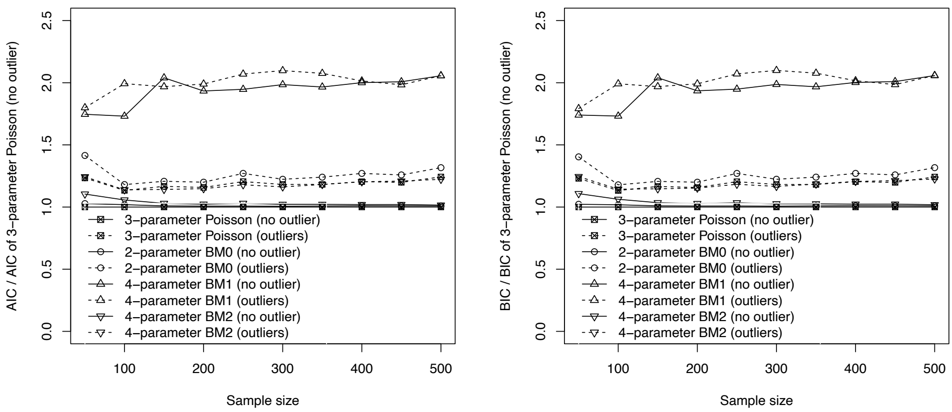

Figure 1: Goodness-of-fit of simulated data: AIC (left) and BIC (right)

In practice, where the presence of outliers is unknown, models giving better goodness-of-fits would be favorite. Since in our simulation setting the model from which simulated data are obtained is known, it would be still expected such a model will be the favorite one. Considering data with outliers, Figure 2 shows results of dividing the AIC and BIC values of the studied models by the ones of the 3-parameter Poisson distribution, respectively. As expected again, the model from which outliers come from is favorite. However, in this case, a different model may be also favorite. It corresponds to the BM_2, which shows a similar performance like that of the 3-parameter Poisson distribution.

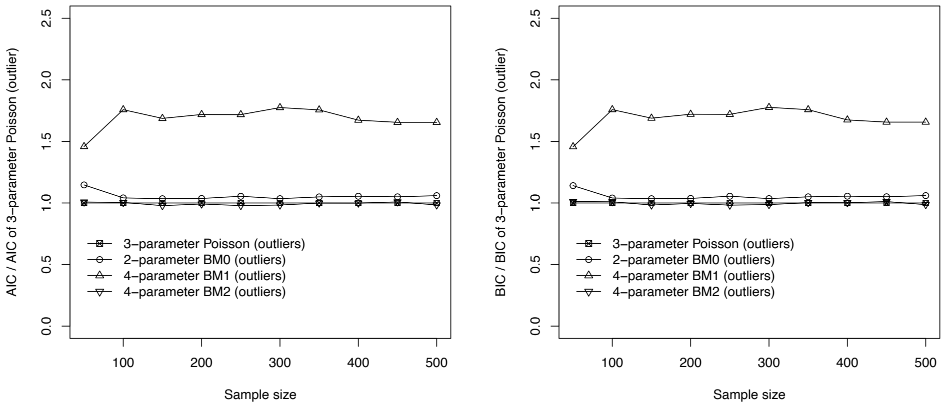

Figure 2: Goodness-of-fit of simulated data with outliers: AIC (left) and BIC (right)

5. *Numerical illustrations*

In this section, we present two applications of our models to real data sets. These data sets have already been used by scholars to assess other models. Considering those results, comparisons with competitors are shown.

As alternatives to models studied by other authors, we consider BM_1 and BM_2 introduced in the previous section. For selecting models through all the models taken into account, the AIC and BIC are used.

5.1 Number of motor vehicle accidents

Arbous and Kerrich [3] presented data on the number of accidents experienced by 122 shunters over two non-overlapping intervals of time, from 1937 to 1942 and from 1943 to 1947, see also [1]. These data have attracted the attention of scholars to assess new models. For instance, Famoye and Consul [10] proposed a six-parameter correlated bivariate generalized Poisson distribution (BGPD) and Sellers et al. [33] built the six-parameter bivariate ConwayMaxwellPoisson distribution (BCOMPD). Bivariate Poisson distributions (BPD) have been also used for modeling these data, for instance [3]. We take into account the three-parameter BPD implemented in `bivpois`, which is a package used within the R statistical language and environment. Details of this BPD are presented in e.g. [16]. Also, we consider the results for the four-parameter bivariate negative binomial distribution (BNBD) given in [33].

Estimates of the parameters of our models are presented in Table 1.

The maximum number of accidents observed were 6 during the first period and 7 during the second one. For presentation of results, we adopt the maximum of both

Model	α_U	β_U	α_V	β_V
BM_1	0.8580	0.4358	0.6456	0.4719
BM_2	0.3415	0.8345	0.3786	0.6255

Table 1: Number of motor vehicle accidents: estimates for the parameters of the models BM_1 and BM_2.

as usually done through those authors. Table 2 presents expected values obtained from the models BM_1 and BM_2 when the data about numbers of accidents are fitted.

$n \setminus m$		0	1	2	3	4	5	6	7+	Total
0	Obs.	21.00	18.00	8.00	2.00	1.00	0.00	0.00	0.00	50.00
	BM_1	19.63	16.85	9.06	4.71	2.29	1.10	0.52	0.26	54.45
	BM_2	19.70	16.44	9.66	4.89	2.27	1.00	0.42	0.15	54.56
1	Obs.	13.00	14.00	10.00	1.00	4.00	1.00	0.00	0.00	43.00
	BM_1	12.67	10.88	7.77	3.81	1.96	0.92	0.44	0.17	38.67
	BM_2	12.32	10.28	8.06	4.08	1.90	0.88	0.38	0.09	38.02
2	Obs.	4.00	5.00	4.00	2.00	1.00	0.00	1.00	0.00	17.00
	BM_1	5.58	3.60	2.58	1.44	0.73	0.36	0.17	0.10	14.60
	BM_2	6.18	3.87	3.03	1.71	0.81	0.36	0.16	0.06	16.21
3	Obs.	2.00	1.00	3.00	2.00	0.00	1.00	0.00	0.00	9.00
	BM_1	2.78	2.00	1.32	0.71	0.36	0.18	0.08	0.05	7.52
	BM_2	2.85	1.78	1.28	0.72	0.35	0.16	0.07	0.03	7.27
4	Obs.	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	2.00
	BM_1	1.28	0.83	0.54	0.30	0.15	0.07	0.03	0.03	3.27
	BM_2	1.25	0.82	0.57	0.31	0.15	0.07	0.03	0.01	3.24
5	Obs.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	BM_1	0.62	0.42	0.27	0.14	0.07	0.03	0.01	0.01	1.62
	BM_2	0.53	0.33	0.23	0.12	0.06	0.02	0.01	0.00	1.34
6	Obs.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	BM_1	0.29	0.18	0.11	0.06	0.03	0.01	0.00	0.00	0.73
	BM_2	0.22	0.14	0.09	0.05	0.02	0.01	0.00	0.00	0.55
7+	Obs.	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	BM_1	0.45	0.38	0.15	0.07	0.03	0.01	0.00	0.00	1.12
	BM_2	0.29	0.27	0.11	0.05	0.02	0.00	0.00	0.00	0.76
Total	Obs.	40.00	39.00	26.00	8.00	6.00	2.00	1.00	0.00	122.00
	BM_1	43.34	35.17	21.83	11.27	5.67	2.72	1.29	0.66	122.00
	BM_2	43.38	33.96	23.07	11.95	5.60	2.53	1.09	0.39	122.00

Table 2: Observed (Obs.) and expected, under the models BM_1 and BM_2, numbers of accidents among 122 shunters.

Now, we assess all models considered. We present in Table 3 both AIC and

BIC values of all those models. The values of ℓ are included, which were taken from [33] for the models BPD, BNBD, BGPD, and BCOMPD. Lower AIC and BIC values, better fit models. The lowest AIC and BIC values are highlighted. Therefore, the BNBD is favored when considering both the AIC and BIC. With respect to our models, BM_1 performs better than BGPD and BCOMPD when considering the BIC, and BM_2 performs better than BGPD, BCOMPD and BM_1 when considering the BIC and also than BM_1 when considering the AIC.

Statistic	BPD	BNBD	BGPD	BCOMPD	BM_1	BM_2
AIC	697.2700	**691.2200**	695.0260	695.4080	700.0423	698.8119
BIC	705.6821	**702.4361**	711.8501	712.2321	711.2584	710.0280
ℓ	-345.635	-341.610	-341.513	-341.704	-346.021	-345.406

Table 3: AIC and BIC: numbers of accidents among 122 shunters

5.2 Football scores

A number of authors have analyzed football scores by using statistical procedures, see e.g. [22, 8, 6]. We consider the data set of the football scores between ACF Fiorentina and Juventus from 1996 to 2011. These data were presented by Lee et al. [21]. It consists of 26 matches, registering the goals scored by each team. In this application, we aim to fit the scores of these matches by using our models BM_1 and BM_2. Additionally, we retake the BPD indicated in the previous application.

Estimates of the parameters of our models are presented in Table 4.

Model	α_U	β_U	α_V	β_V
BM_1	1.1806	-0.2443	1.8572	0.2545
BM_2	1.0914	0.0251	1.7540	0.0000

Table 4: Football scores: estimates for the parameters of the models BM_1 and BM_2

The maximum number of accidents observed were 6 during the first period and 7 during the second one. For presentation of results, we adopt the maximum of both maximum numbers of accidents observed as usually done through those authors. Table 5 presents expected values obtained from the models BM_1 and BM_2 when the data about numbers of accidents are fitted.

When all models considered are assessed, we find that BM_1 performs better than BPD and BM_2 as for AIC as well for BIC, Table 6.

$n \setminus m$		0	1	2	3+	Total
0	Obs.	1.00	5.00	0.00	0.00	6.00
	BM_1	1.28	2.38	2.06	0.37	6.11
	BM_2	1.34	2.36	2.07	1.21	7.00
1	Obs.	1.00	8.00	6.00	1.00	16.00
	BM_1	1.51	2.81	3.83	0.46	8.63
	BM_2	1.47	2.58	3.64	2.12	9.82
2	Obs.	0.00	0.00	0.00	0.00	0.00
	BM_1	0.80	0.94	1.29	0.38	3.43
	BM_2	0.82	0.89	1.26	1.17	4.16
3+	Obs.	1.00	0.00	1.00	2.00	4.00
	BM_1	1.86	3.41	1.91	0.62	7.81
	BM_2	0.31	0.34	0.30	0.28	1.25
Total	Obs.	3.00	13.00	7.00	3.00	26.00
	BM_1	5.46	9.57	9.10	1.85	26.00
	BM_2	3.95	6.19	7.29	4.81	26.00

Table 5: Observed (Obs.) and expected, under the models BM_1 and BM_2, numbers of goals among 26 matches

Statistic	BPD	BM_1	BM_2
AIC	135.8319	**135.5629**	136.7675
BIC	141.6856	**140.5953**	141.7998
ℓ	−64.9159	−63.7814	−64.3837

Table 6: AIC and BIC: numbers of goals among 26 matches

6. *Concluding Remarks*

We developed a large family of discrete bivariate distributions whose probability distribution functions satisfy relations of Panjer type. This family was built by modeling differential equations of the probability generating functions (p.g.f.) of their members. By incorporating parametrized rational functions in these differential equations, we facilitate the generation of a variety of bivariate distributions. When these models are used to fit data sets, we propose the minimization of distances between p.g.fs and empirical p.g.f. for estimating parameters.

Two of these models were applied for modeling a couple of real data sets. These results show that members of this new family of distributions may outperform competitors.

Acknowledgment. The work of the second author has been supported by the *Alexander von Humboldt Foundation* under the grant number: Ref. 3.4-IRN-1128637-GF-E.

REFERENCES

[1] A.M. Adelstein, *Accident Proneness: A Criticism of the Concept Based Upon an Analysis of Shunters' Accidents*, J. R. Stat. Soc. Ser. A Stat. Soc. **115** (3) (1952), 354–410.

[2] H. Akaike, *A new look at the statistical model identification*, IEEE Trans. Automat. Contr., **19** (6) (1974), 716–723.

[3] A.G. Arbous, J.E. Kerrich, *Accident Statistics and the Concept of Accident-Proneness*, Biometrics, **7** (4) (1951), 340–432.

[4] H.S. Bakouch, M. Aghababaei Jazi, S. Nadarajah, *A new discrete distribution*, Statistics **48** (1) (2014), 200–240.

[5] L. Baringhaus, N. Henze, *A goodness of fit test for the Poisson distribution based on the empirical generating function*, Stat. Probab. Lett., **13** (1992), 269–274.

[6] G. Boshnakov, T. Kharrata, I.G. McHale, *A bivariate Weibull count model for forecasting association football scores*, Int. J. Forecast., **33** (2) (2017), 458–466.

[7] L.G. Doray, S.M. Jiang, A. Luong, *Some Simple Method of Estimation for the Parameters of the Discrete Stable Distribution with the Probability Generating Function*, Commun. Stat. - Simul. Comput., **38** (9) (2009), 2004–2017.

[8] L. Egidi, F. Pauli, N. Torelli, *Combining historical data and bookmakers' odds in modelling football scores*, Stat. Modelling, **18** (2018), 436–459.

[9] M.L. Esquivel, *Some applications of probability generating function based methods to statistical estimation*, Discuss. Math. Probab. Stat., **29** (2009), 131–153.

[10] F. Famoye, P.C. Consul, *Bivariate Generalized Poisson Distribution with Some Applications*, Metrika, **42** (1995), 127–138.

[11] E. Gomez-Deniz, E. Calderin-Ojeda, *The discrete Lindley distribution: properties and applications*, J. Stat. Comput. Simul. **81** (11) (2011), 1405–1416.

[12] A. Guldberg, *On discontinuous frequency functions of two variables*, Scand. Actuar. J., **1** (1934), 89–117.

[13] M.D. Jiménez-Gamero, A. Batsidis, *Minimum distance estimators for count data based on the probability generating function with applications*, Metrika **80** (2017), 503–545.

[14] K.K. Jose, S. Jacob, *Type II bivariate generalized power series Poisson distribution and its applications in risk analysis*, Manuscript (2016).

[15] J.W. Joshi, *Some recent advances with power series distributions*, In: A Modern Course on Statistical Distributions in Scientific Work, Volume 1 – Models and Structures, D. Reidel Publishing Company, Dordrecht-Holland / Boston-USA, 1975.

[16] D. Karlis, I. Ntzoufras: *Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R*. J. Stat. Softw., **14** (2005), 2–37.

[17] C.D. Kemp, A.W. Kemp, *Rapid estimation for discrete distributions*, J. Roy. Stat. Soc. D-Sta. **37** (3) (1988), 243–255.

[18] C.G. Khatri, *On certain properties of power series distributions*, Biometrika **46** (3/4) (1959), 486–490.

[19] S. Kocherlakota, K. Kocherlakota, *Goodness of fit test for discrete distributions*, Commun. Stat. - Theory Methods, **15** (3) (1986), 815–829.

[20] M. Lazarova, L. Minkova, *A Family of bivariate inflated-parameter generalized power series distributions*, Comptes Rendus de l'Academie Bulg. des Sci. **Tome 68** (5) (2015), 577–588.

[21] H. Lee, J. H. Cha, G. Pulcini, *Modeling discrete bivariate data with applications to failure and count data*, Qual. Reliab. Eng. Int. **33** (7) (2017), 1455–1473.

[22] M.J. Maher, *Modelling association football scores*, Stat. Neerl., **36** (3) (1982), 109–118.

[23] M. Masjed-Jamei, E. Omey, M. Cadena, N. Saad, *Discrete probability models using Taylor series*, Bulletin T. CLV. de l'Académie Serbe des Sciences et des Arts, Classe des sciences mathématiques et naturelles, Sciences mathématiques, **47** (2022), 101–128.

[24] L. Minkova, *A generalization of the classical discrete distributions*, Comm. Statist. Theory Methods **31** (6) (2002), 871–888.

[25] F. Momeni, *The generalized power series distributions and their applications*, J. Math. Comp. Sc. **2** (4) (2011), 691–697.

[26] M. Nakamura, V. Pérez-Abrue, *Use of an empirical probability generating function for testing a Poisson model*, Can. J. Stat. **21** (2) (1993), 149–156.

[27] C.M. Ng, S.-H. Ong, H.M. Srivastava, *Parameter estimation by Hellinger type distance for multivariate distributions based upon probability generating functions*, Appl. Math. Model., **37** (2013), 7374–7385.

[28] A. Noack, *A class of random variables with discrete distributions*, Ann. Math. Stat. **21** (1) (1950), 127–132.

[29] H.J. Panjer, *Recursive evaluation of a family of compound distributions*, Astin Bull. **12** (1981), 22–26.

[30] B. Rémillard, R. Theodorescu, *Inference based on the empirical probability generating function for mixture of Poisson distributions*, Stat. Decis. **18** (2000), 349–366.

[31] R. Rueda, V. Pérez-Abreu, F. O'Reilly, *Goodness of fit for the Poisson distribution based on the probability generating function*, Commun. Stat. - Theory Methods, **20** (1991), 3093–3110.

[32] M. Sankaran, *The discrete Poisson-Lindley distribution*, Biometrics, **26** (1) (1970), 145–149.

[33] K.F. Sellers, D.S. Morris, N. Balakrishnan, *Bivariate Conway-Maxwell-Poisson distribution: Formulation, properties, and inference*, J. Multivar. Anal. **150** (2016), 152–168.

[34] M.M. Shoukri, P.C. Consul, *Bivariate modified power series distribution some properties, estimation and applications*, Biom. J. **24** (8) (1982), 787–799.

[35] S.Z. Sim, S.H. Ong, *Parameter estimation for discrete distributions by generalized Hellinger-type divergence based on probability generating function*, Commun. Stat. - Simul. Comput., **39** (2) (2010), 305–314.

[36] G.E. Schwarz, *Estimating the dimension of a model*, Ann. Stat. **6** (2) (1978), 461–464.

[37] R. Vernic, *Recursive evaluation of some bivariate compound distributions*, ASTIN Bull. **29** (2) (1999), 316–325.

DECE, Universidad de las Fuerzas Armadas
Sangolqui, Ecuador
e-mail: `mncadena2@espe.edu.ec`

Department of Mathematics
K.N. Toosi University of Technology, Tehran, Iran
e-mail: `mmjamei@kntu.ac.ir`

Faculty of Business and Economics
K.U. Leuven Campus Brussels
Warmoesberg 26, 1000 Brussels, Belgium
e-mail: `edward.omey@kuleuven.be`

School of Engineering
Macquarie University
NSW 2109, Australia
e-mail: `rein.vesilo@mq.edu.au`