J. of Ramanujan Society of Mathematics and Mathematical Sciences
Vol. 12, No. 2 (2025), pp. 27-40

DOI: 10.56827/JRSMMS.2025.1202.2 **ISSN (Online): 2582-5461**

ISSN (Print): 2319-1023

BIVARIATE EXTENDED POISSON-LINDLEY DISTRIBUTION BASED ON SARMANOV-LEE FAMILY

C. G. Louzayadio, R.O Malouata and M. Koukouatikissa Diafouka

Department of Mathematics,
Marien Ngouabi University,
BP 69, Brazzaville, CONGO

E-mail : gelinlouz@gmail.com, onesimero@gmail.com,
michel.koukouatikissa@umng.cg

(**Received:** *Sep. 20, 2024* **Accepted:** *May 15, 2025* **Published:** *Jun. 30, 2025)*

Abstract: In this article, we propose a new discrete bivariate distribution, called the Bivariate Extended Poisson-Lindley (BEPL) distribution, suitable for modeling overdispersed and correlated count data. The distribution is constructed within the Sarmanov-Lee family by combining two Extended Poisson-Lindley distributions via a multiplicative factor. We analyze its theoretical properties (moment-generating function, various moments) and estimate its parameters using the maximum likelihood method. The model's performance is evaluated and compared to that of existing discrete bivariate models on two real-world datasets. The proposed distribution exhibits great flexibility in capturing different types of dependencies (positive, zero, or negative correlations), providing an effective tool for modeling correlated count data.

Keywords and Phrases: Bivariate Sarmanov distribution, extended Poisson-Lindley, count data, overdispersion, maximum likelihood estimation, statistical modeling.

2020 Mathematics Subject Classification: 62H10, 62E15, 60E05.

1. Introduction

The Poisson-Lindley distribution, defined in Equation (1.1), is considered one of the main alternatives to the Poisson distribution for modeling overdispersed count data [11].

$$P(X = x) = \frac{\theta^2(\theta + x + 2)}{(\theta + 1)^{x+3}}, \quad \forall x \in \mathbb{N}, \theta > 0. \tag{1.1}$$

This preference is explained by the restrictive nature of the Poisson distribution owing to its equidispersion property (mean = variance).

To enhance its capability for modeling overdispersed count data, several two-parameter univariate Poisson-Lindley distributions have been proposed [1, 2]. Recently, [8] proposed a new generalized Poisson-Lindley distribution called the beta transformation of the Poisson-Lindley distribution. The beta transformation method, originally developed by [3], introduces an additional parameter into discrete probability distributions. In this study, we refer to this new discrete distribution as the extended Poisson-Lindley distribution.

Although univariate distributions are useful for modeling univariate count data, they do not capture relationships between two dependent variables. In many practical applications, such as analyzing the interactions between the number of treated cases and the number of deaths in an epidemic, or studying the scores of home and away teams in a football match, it is important to consider the dependence between these variables. Therefore, a bivariate approach is necessary to model simultaneously these two phenomena and to better understand their correlation [9].

Several discrete bivariate distributions have been constructed using the trivariate reduction method [6]. The bivariate Poisson distribution (BP) is a classic example; however, this approach can only model positive correlations [5]. An alternative method, based on the product of two discrete marginal distributions weighted by a multiplicative factor, has been proposed in the literature [10]. This method is known as the Sarmanov-Lee family [13], which can model positive, zero, or negative correlations [7, 9, 12, 15].

The objective of this study is to propose a bivariate version of the extended Poisson-Lindley distribution. This new model is based on the Sarmanov-Lee methodology, which combines two marginal distributions using a multiplicative factor to model their dependence. The rest of this paper is organized as follows. In Section 2, the univariate case of the extended Poisson-Lindley distribution is presented. Section 3 proposes the bivariate extended Poisson-Lindley (BEPL) distribution and several of its properties, including the moment-generating function, the moments and the correlation coefficient. Parameter estimation using the maximum likelihood method is provided in Section 4. Section 5 examines the goodness-of-fit of the BEPL distribution to real data compared with other discrete bivariate distributions.

2. Method

2.1. Univariate Extended Poisson-Lindley (EPL) Distribution

Let X be a non-negative integer-valued random variable following a Poisson-Lindley distribution with parameter $\theta > 0$ and probability mass function (PMF) given by (1.1).

Definition 2.1. *The beta transformation Y of X is a non-negative integer-valued random variable* [8], *whose probability mass function is given by*

$$p(y) = P(Y = y) = \begin{cases} \dfrac{\theta^2 + 3\theta + 1}{\beta(\theta+1)^3}, & \text{if } \; y = 0 \\ \dfrac{\theta^2(\theta + y + 2)}{\beta(\theta+1)^{y+1}}\left(\dfrac{\beta(\theta+1)(\theta+y+1)}{\theta+y+2} - 1\right), & \text{if } \; y = 1, 2, \ldots \end{cases} \tag{2.1}$$

where the parameter β is subject to the condition that

$$\beta > 1 - \frac{\theta^2(\theta+2)}{(\theta+1)^3} \; \text{and} \; \beta \geq \frac{\theta+3}{(\theta+1)(\theta+2)}. \tag{2.2}$$

In this paper, we say that Y follows the Univariate Extended Poisson-Lindley (EPL) distribution, whereas in [8] it was referred to as the beta transformation of the Poisson-Lindley distribution. Alternatively, the probability mass function (2.1) can be written as follows

$$p(y) = \left(\frac{\theta^2 + 3\theta + 1}{\beta(\theta+1)^3}\right)^{\delta_0(y)} \left[\frac{\theta^2(\theta+y+2)}{\beta(\theta+1)^{y+1}}\left(\frac{\beta(\theta+1)(\theta+y+1)}{\theta+y+2} - 1\right)\right]^{1-\delta_0(y)} \quad \forall y \in \mathbb{N}, \tag{2.3}$$

where $\delta_0(y)$ is the indicator function at 0. The moment-generating function of the random variable Y is given by;

$$M_Y(t) = \frac{1}{\beta} - \frac{(1-\beta e^t)\theta^2}{\beta(\theta+1)} \frac{(\theta+2-e^t)}{(\theta+1-e^t)^2}. \tag{2.4}$$

2.2. Methodology

Let $u_i(y_i)$ (for $i = 1, 2$) and ξ be bounded non-constant functions and a real number respectively [7, 13] satisfying the conditions

$$E[u_i(Y_i)] = 0 \; \text{and} \; 1 + \xi \prod_{i=1}^{2} u_i(y_i) \geq 0 \; \text{for} \; i = 1, 2.$$

Then, the function defined by

$$p(y_1, y_2) = p(y_1)p(y_2)\left[1 + \xi u_1(y_1)u_2(y_2)\right] \tag{2.5}$$

is a bivariate probability mass function (PMF), where $p(y_1)$ and $p(y_2)$ are marginal mass functions. Inspired by this approach, several bivariate discrete distributions have been proposed [7, 9, 12, 15].

Using the same approach, we propose a bivariate extended Poisson-Lindley (BEPL) distribution, as in the case of the bivariate Poisson distribution proposed in [7]. From (2.5), we obtain $u_i(y_i) = e^{-y_i} - E[e^{-Y_i}]$.

3. Results

3.1. Probability mass function of BEPL Distribution

Definition 3.1. *Let Y_1 and Y_2 be two Extended Poisson-Lindley random variables with parameters $(\theta_1,\ \beta_1)$ and $(\theta_2,\ \beta_2)$, respectively, and probability mass functions $p(y_1)$ and $p(y_2)$. The random pair $(Y_1,\ Y_2\ )$ has a joint distribution called the Bivariate Extended Poisson-Lindley (BEPL) distribution with parameters $(\theta_1, \theta_2, \beta_1,\ \beta_2,\ \xi)$ whose probability mass function (PMF) is given by:*

$$\begin{aligned} p(y_1, y_2) = &\left(\frac{\theta_1^2 + 3\theta_1 + 1}{\beta_1(\theta_1 + 1)^3}\right)^{\delta_0(y_1)} \\ &\times \left[\frac{\theta_1^2(\theta_1 + y_1 + 2)}{\beta_1(\theta_1 + 1)^{y_1+1}}\left(\frac{\beta_1(\theta_1 + 1)(\theta_1 + y_1 + 1)}{\theta_1 + y_1 + 2} - 1\right)\right]^{1-\delta_0(y_1)} \\ &\times \left(\frac{\theta_2^2 + 3\theta_2 + 1}{\beta_2(\theta_2 + 1)^3}\right)^{\delta_0(y_2)} \\ &\times \left[\frac{\theta_2^2(\theta_2 + y_2 + 2)}{\beta_2(\theta_2 + 1)^{y_2+1}}\left(\frac{\beta_2(\theta_2 + 1)(\theta_2 + y_2 + 1)}{\theta_2 + y_2 + 2} - 1\right)\right]^{1-\delta_0(y_2)} \\ &\times \left[1 + \xi\left(e^{-y_1} - D_1\right)\left(e^{-y_2} - D_2\right)\right],\ \ \forall (y_1, y_2) \in \mathbb{N}^2; \end{aligned} \tag{3.1}$$

where $\xi \in \mathbb{R}$, $\theta_i > 0$, and $\beta_i > 0$ satisfy the conditions (2.2) $\forall\ i = 1, 2$, and

$$D_i = E\left[e^{-Y_i}\right] = M_{Y_i}(-1) = \frac{1}{\beta_i} - \frac{(1 - \beta_i e^{-1})\theta_i^2}{\beta_i(\theta_i + 1)}\frac{(\theta_i + 2 - e^{-1})}{(\theta_i + 1 - e^{-1})^2}\ for\ i = 1,\ 2.$$

3.2. Moment-generating function

Proposition 3.1. *The moment-generating function M_{Y_1,Y_2} of the extended Poisson-Lindley random pair (Y_1, Y_2) is given by*

$$\begin{aligned} M_{Y_1,Y_2}(t_1, t_2) = \ &M_{Y_1}(t_1)M_{Y_2}(t_2) + \xi\left[M_{Y_1}(t_1 - 1) - M_{Y_1}(-1)M_{Y_1}(t_1)\right] \\ &\times\left[M_{Y_2}(t_2 - 1) - M_{Y_2}(-1)M_{Y_2}(t_2)\right]; \end{aligned} \tag{3.2}$$

where, $M_{Y_i}(t_i)$ ($i = 1, 2$) is given by (2.4)*.*
Proof. The moment-generating function is defined as

$$\begin{aligned}
M_{Y_1,Y_2}(t_1,t_2) &= E\left[e^{t_1Y_1+t_2Y_2}\right] = \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty} e^{t_1y_1+t_2y_2}p(y_1)p(y_2) \\
&\quad \times \left[1+\xi\left(e^{-y_1}-D_1\right)\left(e^{-y_2}-D_2\right)\right] \\
&= \sum_{y_1=0}^{\infty} e^{t_1y_1}p(y_1)\sum_{y_2=0}^{\infty} e^{t_2y_2}p(y_2) \\
&\quad +\xi\left(\sum_{y_1=0}^{\infty} e^{(t_1-1)y_1}p(y_1) - E\left[e^{-Y_1}\right]\sum_{y_1=0}^{\infty} e^{t_1y_1}p(y_1)\right) \\
&\quad \times\left(\sum_{y_2=0}^{\infty} e^{(t_2-1)y_2}p(y_2) - E\left[e^{-Y_2}\right]\sum_{y_2=0}^{\infty} e^{t_2y_2}p(y_2)\right) \\
&= M_{Y_1}(t_1)M_{Y_2}(t_2) + \xi\left[M_{Y_1}(t_1-1) - M_{Y_1}(-1)M_{Y_1}(t_1)\right] \\
&\quad \times\left[M_{Y_2}(t_2-1) - M_{Y_2}(-1)M_{Y_2}(t_2)\right].
\end{aligned}$$

3.3. Features

The expected value and variance of Y_i ($i = 1, 2$) are given by

$$E(Y_i) = 1 + (1-\beta_i^{-1})\frac{\theta_i+2}{\theta_i(\theta_i+1)}, \quad i = 1, 2;$$

and

$$\begin{aligned}
V(Y_i) &= \sigma_i^2 = (1-\beta_i^{-1})\frac{\theta_i^3+4\theta_i^2+6\theta_i+2}{\theta_i^2(\theta_i+1)^2} + \frac{2\beta_i^{-1}(\theta_i+2)}{\theta_i(\theta_i+1)} \\
&\quad +(1-\beta_i^{-1})\beta_i^{-1}\left(\frac{\theta_i+2}{\theta_i(\theta_i+1)}\right)^2, \quad i = 1, 2.
\end{aligned}$$

respectively.

Remark 3.1. *The formula we present for the variance of Y_i differs slightly from that given in* [8]*. Upon verification, it appears that the expression in* [8] *contains an error in the final term. The version provided here has been rigorously derived from the moment-generating function of the Extended Poisson-Lindley (EPL) distribution. This correction ensures theoretical consistency with the other properties of the marginal distribution.*

Proposition 3.2. *The covariance between random variables Y_1 and Y_2 is given by*

$$Cov(Y_1, Y_2) = \xi \left[M'_{Y_1}(-1) - M_{Y_1}(-1).E(Y_1)\right] \times \left[M'_{Y_2}(-1) - M_{Y_2}(-1).E(Y_2)\right], \tag{3.3}$$

where

$$\begin{aligned} M'_{Y_i}(-1) &= -\frac{\theta_i^2}{\beta_i(\theta_i+1)(\theta_i+1-e^{-1})^4}\left[(\theta_i+1-e^{-1})^2\left(-\beta_i e^{-1}(\theta_i+2-e^{-1})\right.\right. \\ & \left.\left. -e^{-1}+\beta_i e^{-2}\right) - (1-\beta_i e^{-1})(\theta_i+2-e^{-1})\left(-2e^{-1}(\theta_i+1-e^{-1})\right)\right] \end{aligned}$$

is the value of the derivative of the moment generating function at $t = -1$ with $i = 1, 2$.

Proof.

$$\begin{aligned} Cov(Y_1, Y_2) &= E(Y_1Y_2) - E(Y_1)E(Y_2) \\ &= \sum_{y_1,y_2\geq 0} y_1y_2p(y_1)p(y_2) \\ & \quad \left[1+\xi(e^{-y_1} - E(e^{-Y_1}))(e^{-y_2} - E(e^{-Y_2}))\right] - E(Y_1)E(Y_2) \\ &= \sum_{y_1,y_2\geq 0} y_1y_2p(y_1)p(y_2) - E(Y_1)E(Y_2) \\ & \quad +\xi \sum_{y_1,y_2\geq 0} y_1y_2p(y_1)p(y_2)(e^{-y_1} - E(e^{-Y_1}))(e^{-y_2} - E(e^{-Y_2})) \\ &= \xi \sum_{y_1,y_2\geq 0} y_1y_2p(y_1)p(y_2)(e^{-y_1} - E(e^{-Y_1}))(e^{-y_2} - E(e^{-Y_2})) \\ &= \xi \left[\sum_{y_1\geq 0} y_1e^{-y_1}p(y_1) - E(e^{-Y_1})\sum_{y_1\geq 0} y_1p(y_1)\right] \\ &\times \left[\sum_{y_2\geq 0} y_2e^{-y_2}p(y_2) - E(e^{-Y_2})\sum_{y_2\geq 0} y_2p(y_2)\right] \\ &= \xi \left[M'_{Y_1}(-1) - M_{Y_1}(-1).E(Y_1)\right] \times \left[M'_{Y_2}(-1) - M_{Y_2}(-1).E(Y_2)\right]. \end{aligned}$$

Therefore, the correlation coefficient of Y_1 and Y_2 is:

$$\rho_{Y_1Y_2} = \frac{\xi \left[M'_{Y_1}(-1) - M_{Y_1}(-1).E(Y_1)\right]\left[M'_{Y_2}(-1) - M_{Y_2}(-1).E(Y_2)\right]}{\sigma_1\sigma_2}.$$

If $\xi = 0$, the random variables Y_1 and Y_2 are independent.

4. Parameter Estimation

Consider an n-sample $(y_{1\,1}, y_{2\,1})$, $(y_{1\,2}, y_{2\,2})$,. . ., $(y_{1\,n}, y_{2\,n})$, of the pair of random variables (Y_1, Y_2) of BEPL(Ω) where Ω =$(\theta_1, \theta_2, \beta_1, \beta_2, \xi)$. The log-likelihood function for the BEPL distribution is given by

$$
\begin{aligned}
l(\theta_1, \beta_1, \theta_2, \beta_2, \xi) =& \sum_{i=1}^{n} \delta_0(y_{1i}) \log\left(\frac{\theta_1^2 + 3\theta_1 + 1}{\beta_1(\theta_1+1)^3}\right) \\
&+ \sum_{i=1}^{n} (1-\delta_0(y_{1i})) \left[\log\left(\frac{\theta_1^2(\theta_1 + y_{1i} + 2)}{\beta_1(\theta_1+1)^{y_{1i}+1}}\right)\right. \\
&\left. + \log\left(\frac{\beta_1(\theta_1+1)(\theta_1+y_{1i}+1)}{\theta_1 + y_{1i} + 2} - 1\right)\right] \\
&+ \sum_{i=1}^{n} \delta_0(y_{2i}) \log\left(\frac{\theta_2^2 + 3\theta_2 + 1}{\beta_2(\theta_2+1)^3}\right) \\
&+ \sum_{i=1}^{n} (1-\delta_0(y_{2i})) \left[\log\left(\frac{\theta_2^2(\theta_2 + y_{2i} + 2)}{\beta_2(\theta_2+1)^{y_{2i}+1}}\right)\right. \\
&\left. + \log\left(\frac{\beta_2(\theta_2+1)(\theta_2+y_{2i}+1)}{\theta_2 + y_{2i} + 2} - 1\right)\right] \\
&+ \log\left[1 + \xi\left(e^{-y_1} - D_1\right)\left(e^{-y_2} - D_2\right)\right] \\
=& \sum_{i=1}^{n} \delta_0(y_{1i}) \log\left(\frac{\theta_1^2 + 3\theta_1 + 1}{\beta_1(\theta_1+1)^3}\right) \\
&+ \sum_{i=1}^{n} (1-\delta_0(y_{1i})) \left[2\log\theta_1 - \log\beta_1 - (y_{1i}+1)\log(\theta_1+1)\right] \\
&+ \sum_{i=1}^{n} (1-\delta_0(y_{1i})) \log\left[\beta_1(\theta_1+1)(\theta_1+y_{1i}+1) - (\theta_1+y_{1i}+2)\right] \\
&+ \sum_{i=1}^{n} \delta_0(y_{2i}) \log\left(\frac{\theta_2^2 + 3\theta_2 + 1}{\beta_2(\theta_2+1)^3}\right) \\
&+ \sum_{i=1}^{n} (1-\delta_0(y_{2i})) \left[2\log\theta_2 - \log\beta_2 - (y_{2i}+1)\log(\theta_2+1)\right] \\
&+ \sum_{i=1}^{n} (1-\delta_0(y_{2i})) \log\left[\beta_2(\theta_2+1)(\theta_2+y_{2i}+1) - (\theta_2+y_{2i}+2)\right] \\
&+ \sum_{i=1}^{n} \log\left[1 + \xi\left(e^{-y_{1i}} - D_1\right)\left(e^{-y_{2i}} - D_2\right)\right]. \qquad (4.1)
\end{aligned}
$$

To estimate the unknown parameters Ω we take the partial derivatives with respect to each parameter, and then equate them to zero, that is,

$$\frac{\partial l(\Omega)}{\partial \theta_1} = 0, \frac{\partial l(\Omega)}{\partial \theta_2} = 0, \frac{\partial l(\Omega)}{\partial \beta_1} = 0, \frac{\partial l(\Omega)}{\partial \beta_2} = 0 \ and \ \frac{\partial l(\Omega)}{\partial \xi} = 0.$$

Because the above equations are not presented in closed forms, the numerical method of the six-dimensional Nelder-Mead type procedure is used to solve this system of equations. The solutions of the ML estimates of Ω using the R program with the maxLik function in the maxLik R package [4].

5. Application

In this section we examine and compare the performance of our distribution with those of the BPL [15] and BGPL [12] distributions on two real datasets.

Example 5.1. In this example, we considered a dataset concerning the number of goals scored in the Italian football championship (Serie A) for the 1991-1992 season [5]. The results of the different scores from the 306 matches in the championship are presented in Table 1. The random variables Y_1 and Y_2 represent the number of goals scored by the home team and the number of goals scored by the away team respectively.

Y_1 \ Y_2	0	1	2	3	4	8	Total
0	38	23	13	0	1	0	75
1	41	58	12	10	3	0	124
2	28	19	10	3	0	1	61
3	6	11	4	4	1	0	26
4	7	5	1	0	1	0	14
5	2	2	2	0	0	0	6
Total	122	118	42	17	6	1	306

Table 1: Distribution of match scores [5]

These data show that 128 games were won by the home team, 67 games by the away team and 111 games ended in draws, including $38 : 0-0$, $58 : 1-1$, $10 : 2-2$, $4 : 3-3$ and $1 : 4-4$. This finding suggests that playing at home provided a significant advantage for teams seeking to win matches during the 1991-1992 league season.

Table 2 presents the descriptive statistics for the random variables Y_1 and Y_2 and their correlation is included. The mean and standard deviation of the goals scored by the home team were greater than those of the away team. Moreover, the

Variance-to-Mean ratios for both the home and away teams are close to 1, suggesting that the two variables are overdispersed. The empirical correlation coefficient between Y_1 and Y_2 was 0.894. This indicated a strong positive correlation between the two variables.

Statistiques descriptives	Y_1	Y_2
Minimum	0	0
First quartile	1	0
Median	1	1
Mean	1.34	0.9314
Third quartile	2	1
Maximum	5	8
Standard deviation	1.182736	1.04577
Variance	1.398864	1.093635
Variance/Mean	1.04403	1.174219
Empirical correlation	0.894	

Table 2: Descriptive statistics for the Italian Serie A football championship (1991-1992 season)

Parameter estimates	BPL	BGPL	BEPL
$\hat{\theta}_1$	1.111090	1.540	1.777
$\hat{\theta}_2$	1.554199	1.892	2.418
$\hat{\beta}_1$		12.056	1.806
$\hat{\beta}_2$		2.702	0.884
$\hat{\xi}$	1.092292	1.475	0.815
-l	883.347625	870.43	839.82
AIC	1772.70	1750.87	1689.64
BIC	1795.31	1769.48	1708.26
CAIC	1800.31	1774.48	1713.26

Table 3: Parameter estimates and goodness-of-fit for BPL, BGPL and BEPL distributions to data [5]

In Table 3, parameter estimates for the BPL, BGPL and BEPL distributions are calculated using the r maxLik package. The dependence parameter ξ is positive for all three distributions, which means that the correlation coefficients ρ_{Y_1,Y_2} for Y_1 and Y_2 are positive. This theoretical correlation coefficient for the dataset based on the BEPL distribution is $\hat{\rho}_{Y_1,Y_2} \simeq 0.075$. This value reflects the weak positive correlation between Y_1 and Y_2 in the BEPL distribution for this dataset. However, the BEPL distribution provides a better fit to the data [5] than the other two distributions because its AIC, BIC and CAIC are small.

Example 5.2. Here, we consider accident data from 122 experienced shunters, where random variables Y_1 and Y_2 represent the number of accidents in $1937-1942$ and $1943-1947$ [14]. The results are presented in Table 4.

Y_1 \ Y_2	0	1	2	3	4	7	Total
0	21	13	4	2	0	0	40
1	18	14	5	1	0	1	39
2	8	10	4	3	1	0	26
3	2	1	2	2	1	0	8
4	1	4	1	0	0	0	6
5	0	1	0	1	0	0	2
6	0	0	1	0	0	0	1
Total	50	43	17	9	2	1	122

Table 4: Observed and expected number of accidents sustained by 122 experienced shunters over 2 successive periods of time

Table 5 shows the descriptive statistics for random variables Y_1 and Y_2.

Statistiques descriptives	Y_1	Y_2
Minimum	0	0
First quartile	0	0
Median	1	1
Mean	1.27	0.98
Third quartile	2	1
Maximum	6	7
Standard deviation	1.23	1.14
Variance	1.65	1.3
Variance/Mean	1.3	1.33
Empirical correlation	0.93	

Table 5: Descriptive statistics for the number of accidents

The results in Table 5 show a slight overdispersion *(Variance/Mean* > 1*)* for both variables, and a strong empirical correlation (0.93) is observed, indicating that accidents over the two periods are strongly linked.

Parameter estimates	BPL	BGPL	BEPL
$\hat{\theta}_1$	0.34	0.792	1.21
$\hat{\theta}_2$	0.28	0.998	1.28
$\hat{\beta}_1$	-	0.45	0.68
$\hat{\beta}_2$	-	1.99	0.82
$\hat{\xi}$	0.99	$1.58.10^{-5}$	0.999
-l	260.29	244.89	128.13
AIC	526.59	499.78	266.26
BIC	534.99	513.8	280.28
CAIC	537.99	518.8	285.28

Table 6: Parameter estimates and goodness-of-fit for BPL, BGPL and BEPL distributions to data [14]

The results in Table 6 show that the BEPL distribution provides a significantly better fit than the BPL and BGPL distributions, with lower log-likelihood, AIC, BIC and CAIC values. The estimation of the dependency parameter $\hat{\xi} = 0.999$ indicates a positive dependence between random variables Y_1 and Y_2. Therefore, the estimate of the theoretical correlation under the BEPL distribution is $\hat{\rho}_{Y_1,Y_2} \simeq 0.094 > 0$. This value is significantly lower than that of the empirical correlation (0.93), reflecting a weak but positive dependence between Y_1 and Y_2 under the BEPL distribution.

6. Conclusion

In this article, we proposed a new discrete bivariate Sarmanov distribution, called the Bivariate Extended Poisson-Lindley (BEPL) distribution, constructed within the Sarmanov-Lee family. This approach allows for flexible modeling of correlated and overdispersed count data, capturing positive, zero, and negative dependencies. The main theoretical properties of the BEPL distribution, including the moment-generating function, various marginal moments, and the correlation coefficient, were derived and analyzed. The model parameters were estimated using the maximum likelihood method, with numerical optimization based on the Nelder-Mead algorithm, demonstrating practical feasibility despite the complexity of the likelihood equations. Applications to two real-world datasets, the number of goals scored during an Italian football season and workplace accidents among shunters, showed that the BEPL model provides a better fit compared to other existing discrete bivariate distributions, particularly according to the AIC, BIC, and CAIC criteria. However, the analyses also revealed certain limitations, notably a slight underestimation of the theoretical correlation compared to the empirical correlation observed in the datasets. This suggests interesting perspectives for future research, such as by introducing additional parameters or exploring more complex dependence structures within the Sarmanov-Lee family. In conclusion, the BEPL distribution significantly contributes to the modeling of correlated count data and provides a promising foundation for developing more flexible statistical models suitable for a wide variety of practical applications.

Acknowledgement

The authors thank the editorial team and reviewers for their comments and suggestions.

References

[1] Al-Nuaami W. A. H., Heydari A. A. & Khamnei H. J., The Poisson-Lindley distribution: Some characteristics, with its application to SPC, Mathematics, 11(11) (2023), 24-28.

[2] Bhati D., Sastry D.V.S. & Qadri P. M., A new generalized Poisson-Lindley distribution: Applications and properties, Austrian Journal of Statistics, 44(4) (2015), 35-51.

[3] Dimitrov B. & Kolev N., Beta transformation. beta type self-decomposition and related characterizations, Brazilian Journal of Probability and Statistics, 14(2) (2000), 123-140.

[4] Henningsen A., & Toomet O., maxLik: A package for maximum likelihood estimation in R, Computational Statistics, 26 (2011), 443-458.

[5] Karlis D. and Ntzoufras I., Analysis of Sports Data by Using Bivariate Poisson Models, Journal of the Royal Statistical Society D (The Statistician), 52(3) (2003), 381-393.

[6] Koukouatikissa Diafouka M. & Louzayadio C. G., On a Bivariate Katz's Distribution Constructed by the Trivariate Reduction Method, Serdica Journal of Computing, 17(2) (2023), 79-93.

[7] Lakshminarayana J. Pandit S.N. & Srinivasa Rao K., On a bivariate Poisson distribution, Communications in Statistics-Theory and Methods, 28(2) (1999), 267-276.

[8] Louzayadio C. G., Koukouatikissa Diafouka M. & Malouata R. O., Beta transformation of the Poisson-Lindley distribution for overdispersed count data, Advances in Mathematics: Scientific Journal, 12(3) (2023), 457-464.

[9] Ong S. H., Gupta R. C., Ma T. & Sim S. Z., Bivariate Conway-Maxwell Poisson distributions with given marginals and correlation. Journal of Statistical Theory and Practice, 15(10) (2021), 1-19.

[10] Pandit S. N. N., Rao C. R. & Sarma Y. V. S. S., On a Bivariate Binomial Distribution, Assam Statistical Review, 2(2) (1988), 67-72.

[11] Sankaran M., 275. Note: The Discrete Poisson-Lindley Distribution, Biometrics, 26(1) (1970), 145-149.

[12] Sirinapa A. & Bodhisuwan W., On New Bivariate Poisson-Lindley distribution with Application of Correlated Bivariate Count Data Analysis, Thailand Statistician, 21(2) (2023), 228-243.

[13] Ting Lee M. L., Properties and applications of the Sarmanov family of bivariate distributions, Communications in Statistics-Theory and Methods, 25(6) (1996), 1207-1222.

[14] Zamani H., Faroughi P. & Ismail N., Bivariate Poisson-weighted exponential distribution with applications, In AIP Conference Proceedings, 1602(1) (2014), 964-968.

[15] Zamani H., Faroughi P. & Ismail N., Bivariate Poisson-Lindley distribution with application, J. Math. Stat., 11(1) (2015), 1-6.