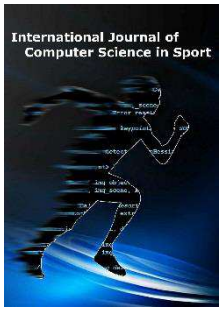
International Journal of Computer Science in Sport

Volume 23, Issue2, 2024

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2023-0009



Modified Bivariate Poisson-Lindley Model: Properties and Applications in Soccer

Allaeddine Haddari[1, 2], Halim Zeghdoudi[2] and Raman Vinoth[3]

[1]Mathematics Department, Faculty of Mathematics and Computer Science, Batna, University, Algeria, a.haddari@univ-batna2.dz

[2]Laboratory of Probability and Statistics (LaPS), BadjiMokhtar-Annaba University, P. O. Box 12, 23000 Annaba, Algeria, halim.zeghdoudi@univ-annaba.dz

[3]Quality Measurement and Evaluation, Department, Deanship of Quality and Academic Accreditation, Imam Abdulrahman Bin Faisal University, P. O. Box 1982, Dammam 31441, Saudi Arabia, vrrangan@iau.edu.sa

Abstract

This paper presents the bivariate Poisson-new XLindley distribution (BPNXLD), which may be used to represent dependent and over-dispersed countdata. Among the characteristics considered are the correlation coefficient, mean, and variance of the distribution. Acomparison with several Bivariate distributions is included. The goodness of fit of this novel model iscompared with the bivariate Poisson, bivariate negative binomial and bivariate Poisson-Lindley distributions using two data sets from a German Bundesliga season.

KEYWORDS: POISSON-NEW XLINDLEY DISTRIBUTION, BIVARIATE POISSON-LINDLEY DISTRIBUTION, ESTIMATION, SOCCER DATA SET .

Introduction

One of the most popular sports in the world is soccer. Both experts and novices love watching the games and making predictions about their results. Many studies have been conducted on numerous football modeling topics. Bradley and Terry (1952) used a variety of Bradley-Terry model extensions to simulate the odds of winning, drawing, or losing. On the other side, Reep and Benjamin (1968) model the number of passes a team successfully completes up to a change of possession, either by a shot on goal, being tackled, or an infraction, using a negative binomial distribution.

Another strategy uses a hierarchical Poisson log-linear model to directly model the number of goals scored by each team in each game. This method was inspired by the modeling frameworks of Maher (1982), Dixon and Coles (1997), and Baio and Blangiardo (2010). In addition, Karlis and Ntzoufras (2003) looked at the bivariate Poisson distribution, which is rarely used since it is difficult to apply, as a way to model scores in a few sports, namely football and water polo. Using the help of the bivariate Poisson distribution, a parameter may be used to model the relationship between two random variables, X and Y, which represent the home and away goals in a football game. In recent years, Tsokos et al. (2019) conducted acomparison research to assess how well a hierarchical Poisson log-linear model and several extensions of the Bradley-Terry model performed in predicting soccer match results (win, draw, or loss).

Many publications analyze data, forecasts, and models related to soccer using probability distribution and other mathematical techniques. Interesting works concerning soccer data and modeling can be found in Sadeghkhani and Ahmed (2020), Wheatcroft (2021), Owen (2011), Marek et al. (2014), Karlis and Ntzoufras (2009), Goddard and Asimakopoulos (2004), oddard (2005), Constantinou et al.(2012), Boshnakov et al. (2017), Shahin (2023).

By combining the Poisson and Lindley distributions in 1970, Sankaran created the Poisson-Lindley distribution, of which the probability mass function is

$$f(z;\varepsilon) = \frac{\varepsilon^2(z+\varepsilon+2)}{(1+\varepsilon)^{z+3}}, z = 0, 1, \ldots, \varepsilon > 0.$$

Many works (see Shanker (2016a, 2016b, 2017), Grine and Zeghdoudi (2017), Zeghdoudi and Nedjar (2017, 2020)) introduce new discrete distributions by compound Poisson and others introduce new continuous distributions like Poisson-Amarendra, Poisson-Sujatha, Poisson-Garima, Poisson Quasi-Lindley, Poisson Pseudo-Lindley, and Poisson-Gamma Lindley distributions. Two novel distributions, called XLindley (XLD) and new XLindley (NXLD), were recently introduced by Chouia and Zeghdoudi (2020) and Khodja et al. (2023). The suggested names for these distributions stem from the fact that they are created as a unique combination of two distributions: Lindley and Gamma.

Poisson-XLindley (PXLD) and Poisson-new XLindley (PNXLD) are two new discrete distributions that were recently created by Ahsan-ul Haq et al. (2022) and Seghier et al. (2022, 2023) by the compounding of Poisson and XLindley distributions.

To develop novel families of bivariate distribution, mixing procedures were applied. Several types of mixed distribution that follow from the univariate instances include the Bivariate Poisson-Lindley (BPLD), Bivariate Negative Binomial (BNBD), and Bivariate Poisson (BPD). For additional readings, consider the following: tests for over-dispersion and independence in BNBD were discussed in (Jung et al., 2009; Cheon et al., 2009); bivariate Poisson-Weighted Exponential was proposed in Zamani et al. (2014); bivariate Poisson-Lindley was proposed in

Zamani et al. (2015); Marshall and Olkin (1990) studied BNBD and applied it in Karlis and Ntzoufras (2003).

In this study, the BPNXLD is employed, which was constructed by merging two Poisson-new XLindley margins with a multiplicative factor parameter. We go over this new class of discrete distributions' mean, variance, correlation, and joint moment generating function. We looked at a soccer data set to show how flexible the proposed model is compared to the BPD, BPLD and BNBD. This is the first time that the classic bivariate Poisson model used by Karlis and Ntzoufras (2003) has been applied to sport in the literature using the new bivariate model.

The structure of this paper is as follows. We provide a brief overview of the univariate Poisson new XLindley distribution (PNXLD) in Section 2. The formulation of the new proposed model, which includes its mean, variance, and correlation coefficient of the BPNXLD, is explained in Section 3. In section 4, a moment estimation approach was offered for figuring out the proposed model's unidentified parameters. The BPD, BPLD, BNBD and BPNXLD are compared in Section 5.

Methods

Univariate Poisson New XLindley Distribution (PNXLD)

The NXLD has the following (p.d.f) (Khodja et al, 2023).

$$f_{NXLD}(x;\theta) = \frac{\theta}{2}(1+\theta x)exp(-\theta x), x > 0, \theta > 0. \qquad (1)$$

It can be obtained by mixture of $f_1(x) \sim Exp(\theta)$ and $f_2(x) \sim gamma(2,\theta)$, which

$$f_{NXLD}(x;\theta) = \frac{1}{2}f_1(x;\theta) + \frac{1}{2}f_2(x;\theta).$$

Where the *rth* moment of the NXLD is defined as follows

$$\mu_r' = \boldsymbol{E}(\boldsymbol{X^r}) = \frac{1}{2\theta^r}\big(\Gamma(r+1) + \Gamma(r+2)\big).$$

A random variable *Y* follows Poisson New XLindley distribution if it possesses the following stochastic representation $(Y|\lambda) \sim Poisson(\lambda)\ with\ p.m.f$

$$P(Y = y|\lambda) = \frac{e^{-\lambda}(\lambda)^y}{y!}, y = 0, 1, 2, \dots, \lambda > 0,$$

where λ represents the average rate (e.g. average number of goals)and $(\lambda|\theta) \sim NXLD(\theta)$.

For $\lambda > 0$ and $\theta > 0$. We call the unconditional distribution of *Y* the Poisson New XLindley distribution and denote it with $PNXLD(\theta)$and its probability mass function is obtained as follows

$$\begin{aligned} P(Y = y) &= \int_0^\infty P(Y = y|\lambda) f_{NXLD}(\lambda;\theta) d\lambda = \int_0^\infty \frac{e^{-\lambda}(\lambda)^y}{y!}\frac{\theta}{2}(1+\theta\lambda)e^{-\theta\lambda}d\lambda \\ &= \frac{\theta}{2\lambda!}\left[\int_0^\infty e^{-(1+\theta)\lambda}\lambda^y d\lambda + \theta\int_0^\infty e^{-(1+\theta)\lambda}\lambda^{x+1}d\lambda\right] \\ &= \frac{\theta}{2\lambda!}\left[\frac{\Gamma(x+1)}{(\theta+1)^{x+1}} + \frac{\theta\Gamma(x+2)}{(\theta+1)^{y+2}}\right], \end{aligned}$$

Which we obtained p.m.f of *PNXLD* as follows

$$P_{PNXLD}(y;\theta) = \frac{\theta(\theta y + 2\theta + 1)}{2(\theta+1)^{y+2}}, y = 0,1,2,3,\dots,\theta > 0. \quad (2)$$

With the central first four moments of the PNXLD are obtained as

$$\mu_1 = \frac{3}{2\theta}, \mu_2 = \frac{6\theta+7}{4\theta^2}, \mu_3 = \frac{6\theta^2+21\theta+15}{4\theta^3}, \mu_4 = \frac{24\theta^3+304\theta^2+612\theta+333}{16\theta^4}.$$

The probability-generating function $G_Y(t)$ and themoment-generating function $M_Y(z)$ of the PNXLD can beobtained as

$$G_Y(t) = \boldsymbol{E}(\boldsymbol{t^Y}) = \sum_{y=1}^{\infty} t^y P(x;\theta) = \frac{\theta t}{2(1+\theta)^2}\left[\frac{\theta(1+\theta)}{(\theta+1-t)^2} + \frac{(2\theta+1)}{(\theta+1-t)}\right],$$

and

$$M_Y(z) = \boldsymbol{E}(\boldsymbol{e^{zY}}) = \frac{\theta e^z}{2(1+\theta)^2}\left[\frac{\theta(1+\theta)}{(\theta+1-e^z)^2} + \frac{\theta(1+\theta)}{(\theta+1-e^z)}\right].$$

For more details on this section see Seghier et al. 2023.

Bivariate Poisson New XLindley Distribution (BPNXLD)

The joint p.m.f. of $BP(\lambda_1,\lambda_2,\alpha)$distribution, which was derived from the product of two Poisson margins witha multiplicative factor parameter, is defined as (Lakshminarayana et al.,1999):

$$P(Y_1 = y_1, Y_2 = y_2) = e^{-\lambda_1-\lambda_2}\frac{{\lambda_1}^{y_1}{\lambda_2}^{y_2}}{y_1!\,y_2!}(1 + \alpha((g_1(y_1) - \overline{g_1})(g_2(y_2) - \overline{g_2}))), \quad (4)$$

$$y_1, y_2 = 0,1,2,\dots;\ \lambda_1,\lambda_2 > 0.$$

where, $g_1(y_1)$ and $g_2(y_2)$ arebounded functions in y_1 and y_2 respectively. The value of $(1 + \alpha[(g_1(y_1) - \overline{g_1})(g_2(y_2) - \overline{g_2})])$ in *(4)* is non-negative when$g_t(y_t) = e^{-y_t}$ and $\overline{g_t} = E[g_t(Y_t)] = E(e^{-Y_t}), t = 1,2.$

In a similar manner, the joint p.m.f. of BPNXLD$(\theta_1,\theta_2,\alpha)$ is defined as:

$$P(Y_1 = y_1, Y_2 = y_2) = \frac{\theta_1(\theta_1 y_1 + 2\theta_1 + 1)}{2(\theta_1+1)^{y_1+2}}\frac{\theta_2(\theta_2 y_2 + 2\theta_2 + 1)}{2(\theta_2+1)^{y_2+2}}\big(1 + \alpha(e^{-y_1} - c_1)(e^{-y_2} - c_2)\big) \quad (5)$$

where $y_1, y_2 = 0,1,2,\dots;\ \theta_1,\theta_2 > 0$,and

$$c_t = E(e^{-Y_t}) = \frac{\theta_t e^{-1}}{2(1+\theta_t)^2}\left[\frac{\theta_t(1+\theta_t)}{(\theta_t+1-e^{-1})^2} + \frac{(2\theta_t+1)}{(\theta_t+1-e^{-1})}\right], t = 1,2. \quad (6)$$

We obtain $E(e^{-Y_t})$ in (6) by letting$z = -1$in m.g.f. (3). When$\alpha = 0$, random variables Y_1 and Y_2 areindependent, each is distributed as a marginal *PNXLD*. Therefore, $\alpha$is the parameter of independence.

The mean and variance of BPNXLD$(\theta_1,\theta_2,\alpha)$ are:

$$E(Y_t) = \mu_t = \frac{3}{2\theta_t}, \qquad t = 1,2.$$

$$Var(Y_t) = \frac{6\theta_t + 7}{4\theta_t^2}, \qquad t = 1,2,$$

where we can check that

$$Var(Y_t) > E(Y_t).$$

And

$$Cov(Y_1, Y_2) = \alpha(c_{11} - \mu_1 c_1)(c_{22} - \mu_2 c_2), \qquad (7)$$

where, $c_{tt} = E(Y_t e^{-Y_t}), t = 1,2$,differentiating m.g.f. in (3) with respect to $z$and letting$z = -1$, we

have

$$\frac{\partial}{\partial z} M_Y(z)|_{z=-1} = E(Ye^{-Y}).$$

Thus

$$c_{tt} = \frac{\theta_t e^{-1}}{2(1+\theta_t)^2}\left[\frac{\theta_t(1+\theta_t)}{(\theta_t + 1 - e^{-1})^2} + \frac{(2\theta_t + 1)}{(\theta_t + 1 - e^{-1})}\right], t = 1,2.$$

Using the variance and covariance in (7), the correlation coefficient is:

$$\rho_{12} = \frac{\rho_{12}}{\rho_1 \rho_2} = \frac{\alpha(c_{11} - \mu_1 c_1)(c_{22} - \mu_2 c_2)}{\rho_1 \rho_2} \qquad (8)$$

From (8), Y_1 and Y_2 are independent when $\alpha = 0$ and havepositive and negative correlations when $\alpha > 0$ and $\alpha < 0$respectively.

Parameter Estimation

By equating the mean and covariance in (7) with the sample moments, onemay derive the moment estimates of $\mathrm{BPNXLD}(\theta_{1,}\theta_{2,}\alpha)$. The unique moment estimate of θ_t is

$$\widehat{\theta_t} = \frac{3}{2\overline{Y_t}}, \overline{Y_t} > 0, t = 1,2.$$

The moment estimate for α can then be computed using$\hat{\alpha} = s_{12}(\hat{c}_{11} - \hat{c}_1\overline{y_1})^{-1}(\hat{c}_{22} - \hat{c}_2\overline{y_2})^{-1}$, with

$$\overline{Y_t} = \frac{\sum_{i=1}^{n} y_{it}}{n}, t = 1,2 \; and \; S_{12} = \frac{\sum_{i=1}^{n}(y_{i1} - \overline{y_1})(y_{i2} - \overline{y_2})}{n-1},$$

where $\hat{c}_1, \hat{c}_2, \hat{c}_{11}$ and $\hat{c}_{22}$ areestimated values of c_1, c_2, c_{11} and c_{22}.

The log likelihood function for $\mathrm{BPNXLD}(\theta_{1,}\theta_{2,}\alpha)$ is:

$$\begin{aligned} LogL = \sum_{i=1}^{n} \{ & log\theta_1 + \log(\theta_1 y_{i1} + 2\theta_1 + 1) - (y_{i1} + 2)\log\big(2(\theta_1 + 1)\big) \\ & + log\theta_2 + \log(\theta_2 y_{i2} + 2\theta_2 + 1) - (y_{i2} + 2)\log\big(2(\theta_2 + 1)\big) \qquad (9) \\ & + \log[1 + \alpha(e^{-y_{i1}} - c_1)(e^{-y_{i2}} - c_2)]\} \end{aligned}$$

One can derive the log likelihood estimates of $\mathrm{BPNXLD}(\theta_{1,}\theta_{2,}\alpha)$ by maximizing the log likelihood inequation (9). The negative expectation of the second derivatives of thelog likelihood can be used to obtain the Fisher Information matrix.

Results

Application and Comparison

Presentation of data

To illustrate the bivariate Poisson new XLindley model, we shall examinemany football data sets (see [1,2]).The GermanBundesliga football championship comprised 18 teams which competed in both home and away matches, resulting in a total of 306 matches. The file underinvestigation originates from the bivpois package developed by Karlis and Ntzoufras (2003) and has four variables, namely:

- the amount of goals scored by the home team (g_1);
- the amount of goals scored by the visiting team (g_2);
- *team1*: name of the home team;
- *team2*: name of the visiting team;

For the investigation's aims, let's set:

- the home team's goal total is represented by X,
- while the away team's goal total is represented by Y.

The distribution of matches by score is shown in Table 1

Table 1. Distribution of matches according to scores with the fit of BPNXLD of data set I.

X/Y	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
0	16(17.2)	12(12.3)	16(15)	2(2.1)	3(2.8)	1(1.1)	1(0.5)	51
1	23(24.4)	41(40.3)	17(16.1)	10(11)	10(8.5)	1(1.6)	2(2.1)	104
2	19(19.4)	29(28.5)	15(15.3)	8(7)	1(2.1)	2(0.9)	0(0.4)	74
3	11(9.9)	14(14.3)	12(11.4)	1(0.9)	3(3.5)	0(0.4)	1(1.4)	42
4	10(9.1)	8(8.7)	3(2.7)	0(0.01)	0(0.01)	0(0.01)	0(0.01)	21
5	4(3.1)	3(3.1)	2(2.1)	0(0.01)	0(0.01)	0(0.01)	0(0.01)	9
6	2(1.5)	1(0.8)	0(0.2)	0(0.01)	0(0.2)	0(0.01)	0(0.01)	3
7	1(1.1)	1(0.6)	0(0.01)	0(0.01)	0(0.01)	0(0.01)	0(0.01)	2
Total	86	109	65	21	17	4	4	306

Table 2. Descriptive statistics of data set I and II.

Globalstatistics	**Home goals(*X*)**		**Away goals(*Y*)**	
	SetI	SetII	SetI	SetII
Minimum	0	0	0	0
Maximum	7	8	6	6
Mean	1.76	1.66	1.35	1.55
Wins	47%	40%	29%	38%
Standard deviation	1.40	1.31	1.30	1.25
Variance	1.11	1.03	1.25	1.008
Correlation	0.11		0.08	

Table 3. General statistic of data set I and II.

	SetI	SetII		SetI	SetII
Numberofrounds	34	34	Averageofgoalspermatch	3.11	3.21
Homevictories	143	123	Bestroundingoals	14th (41goals)	10th (35goals)
Awayvictories	90	115	Totalnumberofgoals	953	982
Drawnmatches	73	68	Averagegoalsperround	28.03	28.83

Table 4. Various statistics Wins, Drawn matches and Losses of data set I.

Wins		Drawnmatches		Losses	
BayernMunchen	24	ArminiaBielefeld	13	SpVggGreutherFürth	22
BorussiaDortmund	22	EintrachtFrankfurt	12	HerthaBerlin	19
BayerLeverkusen	19	VfBStuttgart	12	ArminiaBielefeld	16
RBLeipzig	17	FCKöln	10	FCAugsburg	16
FCUnionBerlin	16	SCFreiburg	10	VfLBochum	16
SCFreiburg	15	BorussiaMönchengladb	9	VfLWolfsburg	16

Table 5. Various statistics victories,Tie games and Losses (Home case).

Home victories		Tie games at home		Losses at home	
BayernMunchen	13	ArminiaBielefeld	10	SpVggGreutherFürth	8
BorussiaDortmund	13	EintrachtFrankfurt	7	HerthaBerlin	8
RBLeipzig	11	FSVMainz05	5	VfBStuttgart	7
FSVMainz05	10	FCUnionBerlin	5	VfLWolfsburg	7
BayerLeverkusen	10	SCFreiburg	5	EintrachtFrankfurt	6
FCUnionBerlin	10	SpVggGreutherFürth	5	FCAugsburg	6

Table 6. Various statistics victoriess, Tie games and Losses of data set I (Away case).

Away victories		Tie games (Away)		Losses (Away)	
BayernMunchen	11	VfBStuttgart	8	pVggGreutherFürth	13
BorussiaDortmund	9	FCKöln	6	FSVMainz05	12
BayerLeverkusen	9	BayerLeverkusen	5	ArminiaBielefeld	11
RBLeipzig	7	BorussiaMönchengladb	5	HerthaBerlin	11
EintrachtFrankfurt	6	EintrachtFrankfurt	5	VfLBochum	11
FCUnionBerlin	6	RBLeipzig	5	FCAugsburg	10

Tables1-6 represents general statistics and descriptive statistics of data set I and II.

Crockett quick and Index of dispersion tests for bivariate Poisson new XLindley distribution

Crockett quick test

Crockett (Rayner and Best (1979)) proposed the test statistic

$$T = n\frac{\overline{Y_2}^2 Z_1^2 - 2S_{12}^2 Z_1 Z_2 - 2S_{12}^2 + \overline{Y_1}^2 Z_2^2}{2(\overline{Y_1}^2 \overline{Y_2}^2 - S_{12}^4)}$$

where $\overline{Y_1}$and$\overline{Y_2}$are the sample means;$Z_1 = S_1^2 - \overline{Y_1}, Z_2 = S_2^2 + \overline{Y_2}$; $S_1^2, S_2^2$the sample variances; and S_{12} the samplecovariance.

The statistic T asymptotically follows a χ^2distribution with twodegrees of freedom for n large and the H_0hypothesis is rejected if$T > \chi^2_{2,1-\alpha}$for a level test α.

Index of Dispersion Test

Similar to the Crockett test, Best and Rayner (1997) also suggest revisingthe statistic and recommending it.

$$I_B = n\frac{\overline{Y}_2 S_1^2 - 2S_{12}^2 + \overline{Y}_1 S_2^2}{\overline{Y_2 Y_1} - S_{12}^2}$$

The null distribution of this statistic I_B is asymptotically aχ^2 with *2n-3* degrees of freedom. Hypothesis H_0 is thereforerejected if $I_B > \chi^2_{2n-3,1-\alpha}$ for a test of level α.

In our case,we test:

- Hypothesis H_0: the data can be modeled with BPNXLD.
- Hypothesis H_1: the data can not be modeled with BPNXLD.

Table 7. Crockett quick and Index of dispersion tests values for α=5%.

	Crockett quick test	**Index of dispersion test**
Set data I (n = 8)	T= 0.66	I_B = 2.55
Set data II (n = 8)	T= 0.645	I_B = 1.88

Hypothesis H_0 is not rejected because T and $I_B < \chi^2_{\nu,\alpha}$. According table 3 to crockett quick and Index of dispersion tests,the two data sets can be modeled with BPNXLD.

Comparison between BP, BPLD, BNBD and BPNXLD

This subsection presents a comparative analysis of many bivariatedistributions, including the BPD, BPLD, BNBD with five parameters$(\theta_1, \theta_2, \alpha, \beta_1, \beta_2)$ (see Famoye (2010)) andBPNXLD.

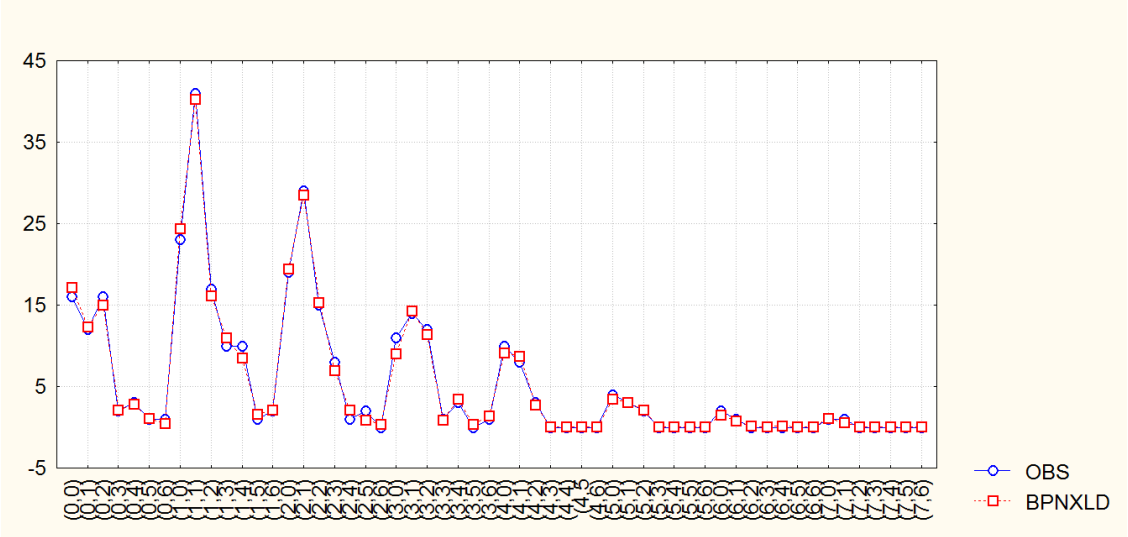


Figure 1. Estimate values using BPNXLD and real values of matches according to scores of data set I.

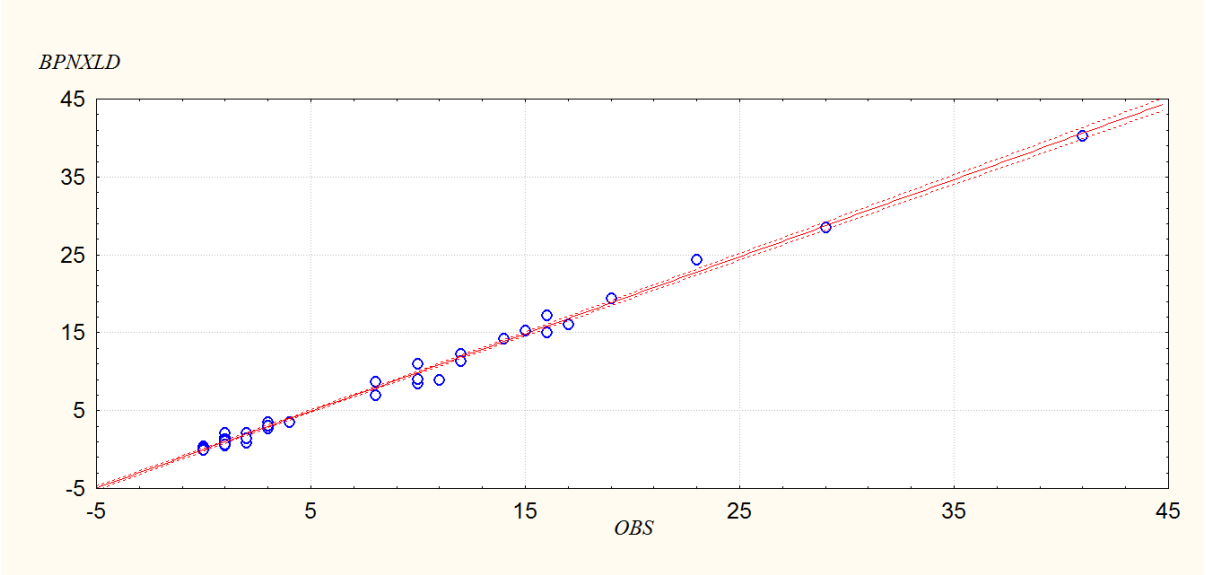


Figure 2. The P–P plot of BPNXLD according to data set I.

Table 8.

Distributions	Parameters	-LL	AIC	AICc	BIC	χ^2
BPD	$\widehat{\theta_1} = 1.185, \widehat{\theta_2} = 0.785, \hat{\alpha} = 0.086$	152.76	305.53	305.99	322.69	50.698
BPLD	$\widehat{\theta_1} = 2.014, \widehat{\theta_2} = 1.631, \hat{\alpha} = 0.84$	118.54	237.09	237.55	254.25	27.734
BNBD	$\widehat{\theta_1} = 0.40, \widehat{\theta_2} = 0.31, \hat{\alpha} = 0.94,$ $\beta_1 = 0.11, \beta_2 = 0.70$	110.97	231.94	232.14	250.56	26.692
BPNXLD	$\widehat{\theta_1} = 0.804, \widehat{\theta_2} = 1.105, \hat{\alpha} = 0.05$	89.16	179.32	179.78	195.85	6.29

Table 9.

Distributions	Parameters	-LL	AIC	AICc	BIC	χ^2
BPD	$\widehat{\theta_1} = 1.06, \widehat{\theta_2} = 0.72, \hat{\alpha} = 0.057$	188.65	383.30	383.38	394.47	100.8
BPLD	$\widehat{\theta_1} = 1.98, \widehat{\theta_2} = 1.54, \hat{\alpha} = 0.69$	154.98	315.96	316.04	327.13	77.5
BNBD	$\widehat{\theta_1} = 0.35, \widehat{\theta_2} = 0.28, \hat{\alpha} = 0.88,$ $\beta_1 = 0.10, \beta_2 = 0.67$	121.20	252.40	252.60	271.02	20.3
BPNXLD	$\widehat{\theta_1} = 0.787, \widehat{\theta_2} = 1.08, \hat{\alpha} = 0.045$	122.96	251.92	252.00	263.09	20.2

As shown in Table 8, 9 and Figures 1, 2, the BPNXLD offers the smallest χ^2, BIC, AIC and AICc values in comparison to the other bivariate distributions, and as a result, best fits the data of all the distributions taken into consideration.

Discussion

It is clear from Tables 2-6 that, on average, the home side scores more goals than the away team. This is known as the "home (field) advantage,"which is discussed here and isn't exclusive to soccer. Right now, the Poisson new XLindley distribution is a helpful idea to teach.

Additionally, Table 1 demonstrates that the variance and mean are not equal, which means that the Poisson distribution is inappropriate in this circumstance (see Karlis and Ntzoufras (2003) and Zamani et al. (2014)). However, the Poisson new XLindley distribution is appropriate because the variance is greater than the mean. Table 2 shows that the average number of goals scored at homeand away in data set II (2019-2020 season) are nearly the same because therewere no spectators present at home matches owing to quarantine measuresrelated to the Covid-19 pandemic.

Hence, using negative log-likelihood *-LL*, Akaike Information Criteria *AIC*, Akaike Information Corrected Criteria *AICc* and Bayesian Information Criteria *BIC*, the choice of the best fitted model is evaluated. Inaddition, we calculate the anticipated frequencies for each distribution, and the quality of fit of the model is evaluated using Pearson's chi-squaretest. Tables 8, 9 and Fig 1, 2 show that all distributions have positivevalues for α, which suggests a positive connection. A positive correlation is shown by a positive value for α indistributions. Onthe other hand, BNBD is the most suitable option based on *-LL* for the dataset II (2019-2020 season) because there were no spectators present owing tothe quarantine imposed by the Covid-19 pandemic. The BPNXLD offers the bestfit for the two data sets based on *AIC*, *AICc* and *BIC*. Also, we can seethat this study demonstrates the adaptability of the BPNXLD for sport datasets.

Conclusion

In this paper, the bivariate Poisson XLindley distribution was used to fit a sample of bivariate count data. The results show that, compared to the bivariate Poisson and bivariate Poisson-Lindley distributions, the bivariate Poisson new XLindley distribution fits dependent and excessively dispersed count data with a positive correlation. This implies that the distribution could be a good substitute. We believe it would be interesting to use the bivariate Poisson-new XLindley model to model actuarial or dependability data. The significance and potentiality of the proposed bivariate distribution are empirically demonstrated using several sport data sets..

Acknowledgements

The authors are grateful for the comments and suggestions by the referee and the Editor Professor Arnold Baca.Their comments and suggestions greatly improved the article.

References

Ahsan-ul-Haq, M., Al-Bossly, A., El-Morshedy, M., & Eliwa, M. S. (2022). Poisson XLindley distribution for count data: statistical and reliability properties with estimation techniques and inference. *Computational Intelligence and neuroscience*, *2022*(1), 6503670.

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, *37*(2), 253-264.

Best D. J., Rayner J. C. W. (1997). Crockett.s test of fit for the Bivariate Poisson. *Biometrical Journal*, 39(4):423.430.

Boshnakov, G., Kharrat, T., & McHale, I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2), 458-466.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete blockdesigns: I. The method of paired comparisons. *Biometrika*, 39, 324-345.

Cheon S., Song S.H., Jung B.C. (2009). Tests for independence in a bivariatenegative binomial model. *J. Korean Statist. Soc*., 38: 185-190.

Chouia, S., & Zeghdoudi, H. (2021). The XLindley distribution: Properties and application. *Journal of Statistical Theory and Applications*, *20*(2), 318-327.

Constantinou, A. C., Fenton, N. E., & Neil, M. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 36, 322-339.

Cox, D.R. and D.V. Hinkley, (1979). Theoretical Statistics. 1st Edn., CRCPress, ISBN-10: 0412161605, pp: 528.

McNeil, D. (1979). A QUICK TEST OF FIT OF A BIVARIATE DISTRIBUTION. In *Interactive Statistics: Proceedings of the Applied Statistics Conference, Sydney, February 8-9, 1979* (p. 185). North-Holland.

Dixon, M. J., Coles, S. G. (1997). Modelling association football scores andinefficiencies in the football betting market. *Applied Statistics*, 46(2), 265.

Famoye, F. and P.C. Consul, (1995). Bivariate generalized Poisson distribution with some applications. *Metrika*, 42: 127-138.

Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics*, *37*(6), 969-981.

Ghitany M. E., Atieh B., Nadarajah S. (2008). Lindley distribution and itsapplications. *Math. Comput. Simulation*, 78, pp. 493-506.

Goddard, J. (2005). Regression models for forecasting goals and matchresults in association football. *International Journal of Forecasting*, 21(2), 331-340.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of .xed-odds betting. *Journal of Forecasting*, 23(1), 51-66.

Holgate, P., (1964). Estimation for the bivariate Poisson distribution. Biometrika, 51: 241-245.

Johnson, N.L., S. Kotz and N. Balakrishnan, (1997). Discrete Multivariate Distributions. 1st Edn., Wiley, New York, ISBN-10: 0471128449, pp: 328.

Jung, B.C., M. Jhun and S.M. Han, (2009). Score test for overdispersionin the bivariate negative binomial models. *J. Statist. Comput. Simulat*., 79:11-24.

Karlis, D. and I. Ntzoufras, (2003). Analysis of sports data by using Bivariate Poisson models. *Statistician*, 52: 381-393.

Karlis, D., &Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam.s distribution for the goal difference. *IMA Journal of Management Mathematics*, 20(2), 133-145.

Kocherlakota, S., & Kocherlakota, K. (2017). *Bivariate discrete distributions*. CRC Press.

Khodja, N., Gemeay, A. M., Zeghdoudi, H., Karakaya, K., Alshangiti, A. M., Bakr, M. E., Hussam, E. (2023). Modeling voltage real data set by a new version of Lindley distribution. *IEEE Access*, *11*, 67220-67229.

Lakshminarayana, J., S.N.N. Pandit and K.S. Rao, (1999). On a bivariatePoisson distribution. *Communicat. Statist. Theory Methods*, 28: 267-276.

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 102-107.

Lord, D. and S.R. Geedipally, (2011). The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. *Accident Anal. Prevent*., 43: 1738-1742.

Loukas S. and Kemp C. D. (1986), The Index of Dispersion Test for the Bivariate Poisson Distribution. *International Biometric Society*, Vol. 42, No. 4,pp. 941-948.

Maher,M. J. (1982).Modelling association footballscores. *Statistica Neerlandica*, 36(3),109.118.

Mahmoudi E. and Zakerzadeh H. (2010), Generalized Poisson-Lindley distribution, *Communications in Statistics- Theory and Methods*, 39, 1785 - 1798.

Marek, P., .edivá, B., & µToupal, T. (2014). Modeling and prediction of icehockey match results. *Journal of quantitative analysis in sports*, 10(3), 357-365.

Mitchell, C. R., & Paulson, A. S. (1981). A new bivariate negative binomial distribution. *Naval Research Logistics Quarterly*, *28*(3), 359-374.

Owen, A. (2011). Dynamic Bayesian forecasting models of football matchoutcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*, 22(2), 99-113.

Paul, S.R. and N.I. Ho, (1989). Estimation in the bivariate poisson distribution and hypothesis testingconcerning independence. *Communicat. Statist.Theory Methods*. 18: 1123-1133.

Reep, C., & Benjamin, B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)*, *131*(4), 581-585.

Sadeghkhani, A., & Ahmed, S. E. (2020). The application of predictive distribution estimation in multiple-inflated poisson models to ice hockey data. *Model Assisted Statistics and Applications*, *15*(2), 127-137.

Sankaran, M., (1970). The discrete Poisson-Lindley distribution. *Biometrics*, 26: 145-149.

Seghier, F. Z., Zeghdoudi, H., & Raman, V. (2023). A Novel Discrete Distribution: Properties and Application Using Nipah Virus Infection Data Set. *European Journal of Statistics*, *3*, 3-3.

Shanker, R. (2016a). The discrete poisson-amarendra distribution. *Int. J. Stat. Distrib. Appl*, *2*(2), 14-21.

Shanker R. (2016b), The discrete Poisson-Sujatha distribution. *International Journal of Probability and Statistics*. 5(1):1-9.

Shanker R. (2017), The discrete poisson-garima distribution. *Biometrics & Biostatistics International Journal*, 5(2):48-53.

Seghier, F. Z., Ahsan-ul-Haq, M., Zeghdoudi, H., & Hashmi, S. (2023). A new generalization of poisson distribution for over-dispersed, count data: mathematical properties, regression model and applications. *Lobachevskii Journal of Mathematics*, *44*(9), 3850-3859.

Tsokos, A., Narayanan, S., Kosmidis, I., Baio, G., Cucuringu, M., Whitaker, G., & Király, F. (2019). Modeling outcomes of soccer matches. *Machine Learning*, *108*, 77-95.

Wheatcroft, E. (2021). Forecasting football matches by predicting match statistics. *Journal of Sports Analytics*, 7(2), 77-97.

Zamani, H., Faroughi, P., & Ismail, N. (2014, June). Bivariate Poisson-weighted exponential distribution with applications. In *AIP Conference Proceedings* (Vol. 1602, No. 1, pp. 964-968). American Institute of Physics.

Zamani, H., P. Faroughi and N. Ismail, (2015). Bivariate Poisson-Lindley Distribution with Application, *Journal of Mathematics and Statistics*, 11 (1): 1-6.

Zeghdoudi, H., & Nedjar, S. (2017). On Poisson pseudo Lindley distribution: Properties and applications. *Journal of probability and statistical science*, *15*(1), 19-28.

Shahin, S. (2023). Sports Data Analysis by using Bivariate Poisson Models in the Bayesian Framework. *Quaid-e-Awam University Research Journal of Engineering Science and Technology*, *21*(1), 7-15.

Wheatcroft, E. (2021). Forecasting football matches by predicting match statistics. *Journal of Sports Analytics*, *7*(2), 77-97.

Singh, A., Scarf, P., & Baker, R. (2023). A unified theory for bivariate scores in possessive ball-sports: the case of handball. *European Journal of Operational Research*, *304*(3), 1099-1112.

[1] Data Set I :https://www.the-sports.org/football-soccer-2021-2022-german-bundesliga-epr114545.html. 2023

[2] Data Set II :https://www.the-sports.org/football-soccer-2019-2020-german-bundesliga-epr98318.html. 2024