Thailand Statistician
April 2023; 21(2): 228-243
http://statassoc.or.th
Contributed paper

On New Bivariate Poisson - Lindley distribution with Application of Correlated Bivariate Count Data Analysis

Sirinapa Aryuyuen [a] and Winai Bodhisuwan*[b]
[a] Department of Mathematics and Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand
[b] Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, Thailand
*Corresponding author; e-mail: fsciwnb@ku.ac.th

Received: 18 May 2022
Revised: 28 July 2022
Accepted: 11 August 2022

Abstract

This paper proposes a bivariate generalized Poisson-Lindley (BGPL) distribution, which is an alternative to analyse correlated bivariate count data. The joint probability mass function, covariance, and correlation coefficient of the proposed distribution, are discussed. It has two sub-models, such as the bivariate Poisson-Lindley (BPL) and bivariate geometric (BGeo) distributions. The unknown parameters of the proposed distribution are estimated with the maximum likelihood method. In addition, some examples of correlated bivariate count data are fitted with the proposed distribution and it compares with the BPL, BGeo, and bivariate Poisson distributions. The results show that the BGPL distribution provides the lowest value of the Akaike information criterion and Bayesian information criterion than other distributions. It is indicated that the BGPL distribution can be used as an alternative flexible distribution for modeling correlated bivariate count data which are either positive or negative correlations.

Keywords: Bivariate discrete distribution, bivariate poisson Lindley, correlated bivariate count data, maximum likelihood estimation.

1. Introduction

Count data are used to describe many phenomena in various fields, from insurance and economics to biometrics and the social sciences. Statistical distributions play a vital role in modeling data in applied areas, such as engineering, economics, and sciences. However, the quality of the procedures primarily depends upon the assumed probability model of the phenomenon under consideration. Poisson distribution plays an important role in count data analysis, it can model some data with equi-dispersion.

Let X be a random variable distributed as the Poisson (Pois) distribution with a parameter λ, denoted by $X \sim \text{Pois}(\lambda)$. Its probability mass function (pmf) is

$$g(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, \text{ and } \lambda > 0. \tag{1}$$

However, in practice, it has often been found that count data presented over-dispersion (variance greater than mean). Hence, there is a demand to modify the Poisson distribution when such problems are encountered. Many researchers have developed new distributions for count data analysis. One such method has been widely used for practical modeling is a mixture of distributions (Karlis and Xekalaki, 2005; Deepesh et al., 2017).

The mixed Poisson distribution has been considered as an alternative for fitting count data with over-dispersion. For examples, the Poisson-inverse Gaussian (Holla, 1967), Poisson-Lindley (Sankaran, 1970; Ghitany et al., 2008), generalized Poisson-Lindley (Mahmoudiet and Zakerzadeh, 2010), discrete two-parameter Poisson-Lindley (Shanker et al., 2012), Poisson-weighted exponential (Zamani et al., 2014), two-parameter Poisson-Lindley (Shanker and Mishra, 2014), new generalized Poisson-Lindley (Bhati et al., 2015), and Poisson-generalised Lindley (Wongrin and Bodhisuwan, 2016) distributions.

Univariate distributions have been used to derive new bivariate version of distributions for correlated bivariate count data. Some examples of bivariate mixed Poisson distributions are bivariate negative binomial (Marshall and Olkin, 1990), bivariate Poisson (Lakshminarayana et al., 1999), bivariate Poisson Lindley (Gómez-Déniz et al., 2008; Zamani et al., 2015), and bivariate Poisson-weighted exponential (Zamani et al., 2014), where the distributions can be used for modeling correlated bivariate count data.

The rest of this paper is organized as follows. In Section 2, the new generalized Poisson-Lindley distribution (Bhati et al., 2015) is described. In Section 3, we will develop a new bivariate Poisson-Lindley distribution. Its joint pmf, covariance, and correlation coefficient of the proposed distribution, are discussed. The maximum likelihood (ML) estimation method is employed for estimating the model parameters in In Section 4. In Section 5, the simulation results are presented. Some examples of correlated bivariate count data are fitted with the proposed distribution and it compares with some exiting distributions in Section 6. Finally, conclusions are included in Section 7.

2. The New Generalized Poisson-Lindley Distribution

Firstly, we provide a new generalized Poisson-Lindley (GPL) distribution which proposed by Bhati et al. (2015). The GPL is derived by compounding the Poisson distribution and the two-parameter Lindley (TPaL) distribution. The TPaL distribution was introduced by Shanker et al. in 2013.

Let X be a random variable distributed as the GPL distribution with parameters α and β, denoted by $X \sim \text{GPL}(\alpha, \beta)$. Its pmf is

$$f(x; \alpha, \beta) = \frac{\alpha^2(1 + \alpha + \beta + \beta x)}{(\alpha + \beta)(\alpha + 1)^{x+2}}, x = 0, 1, 2, \ldots, \tag{2}$$

where $\alpha > 0$ and $\beta > 0$ (Bhati et al., 2015).

Theorem 1 *If $X \sim GPL(\alpha, \beta)$, then the moments of X is*

$$\mu'_k = \frac{\alpha^2}{\alpha + \beta}\left[\frac{\Gamma(k+1)}{\alpha^{k+1}} + \frac{\beta\Gamma(k+2)}{\alpha^{k+2}}\right], k = 1, 2, 3, \ldots, \alpha > 0 \textit{ and } \beta > 0. \tag{3}$$

Proof: Let λ be a random variable distributed as the TPaL distribution with parameters α and β, denoted by $\lambda \sim \text{TPaL}(\alpha, \beta)$. Its the probability density function (pdf) is (Shanker et al., 2013)

$$g(\lambda; \alpha, \beta) = \frac{\alpha^2}{\alpha + \beta}(1 + \beta\lambda)e^{-\alpha\lambda}, \lambda > 0, \alpha > 0 \text{ and } \beta > 0. \tag{4}$$

If $X|\lambda \sim \text{Pois}(\lambda)$ where $\lambda \sim \text{TPaL}(\alpha, \beta)$, then the kth moment about the origin of X can be obtained as

$$\mu'_k = E(X^k) = E\left[E\left(X^k|\lambda\right)\right] = \int_0^\infty \left[\sum_{x=0}^{\infty} \frac{x^k e^{-\lambda}\lambda^x}{\Gamma(x+1)}\right] \frac{\alpha^2}{\alpha + \beta}(1 + \beta\lambda)e^{-\alpha\lambda} d\lambda.$$

Since $\sum_{x=0}^{\infty} \frac{x^k e^{-\lambda}\lambda^x}{\Gamma(x+1)} = \lambda^k$, we have

$$\mu'_k = \frac{\alpha^2}{\alpha+\beta}\int_0^{\infty} \lambda^k(1+\beta\lambda)e^{-\alpha\lambda}d\lambda = \frac{\alpha^2}{\alpha+\beta}\left[\frac{\Gamma(k+1)}{\alpha^{k+1}} + \frac{\beta\Gamma(k+2)}{\alpha^{k+2}}\right].$$

When $\beta = 0$, the GPL distribution reduces to the geometric distribution.

Corollary 1 *Let $X \sim$ GPL(α, β), if $\beta = 0$ and $p = \alpha/(\alpha+1)$ $\alpha \neq -1$ then the GPL distribution reduces to the geometric distribution with the pmf:*

$$f(x;p) = p(1-p)^x; x = 0, 1, 2, \ldots \text{ and } 0 < p < 0. \tag{5}$$

For $\beta = 1$, the GPL distribution reduces to the Poisson Lindley (PL) distribution, which is introduced by Shakaran in 1970 (Sankaran, 1970). Moreover, Bhati et al. (2015) showed some properties of the GPL distribution as follows. If $X \sim \text{GPL}(\alpha, \beta)$ then its mean and variance respectively are

$$E(X) = \frac{\alpha+2\beta}{\alpha(\alpha+\beta)} \text{ and } V(X) = \frac{2\beta^2(1+\alpha)+\alpha^2(1+\alpha)+\alpha\beta(4+3\alpha)}{\alpha^2(\alpha+\beta)^2}.$$

Its probability generating function (pgf) is

$$P(x) = \frac{\alpha^2(\beta+\alpha-t+1)}{(\alpha+\beta)(\alpha-t+1)^2}, t > 0. \tag{6}$$

Additionally, its moment generating function (mgf) is

$$M_X(t) = \frac{\alpha^2(\beta+\alpha-e^t+1)}{(\alpha+\beta)(\alpha-e^t+1)^2}, t > 0. \tag{7}$$

The GPL distribution can be fitted to some real data sets. It has been found to be better and more flexible than the Poisson, negative binomial, PL (Sankaran, 1970; Ghitany et al., 2008), generalized Poisson-Lindley (Mahmoudiet and Zakerzadeh, 2010), and Poisson-weighted exponential (Zamani et al., 2014) distributions.

Some plots of the pmf of the GPL distribution are shown in Figure 1.

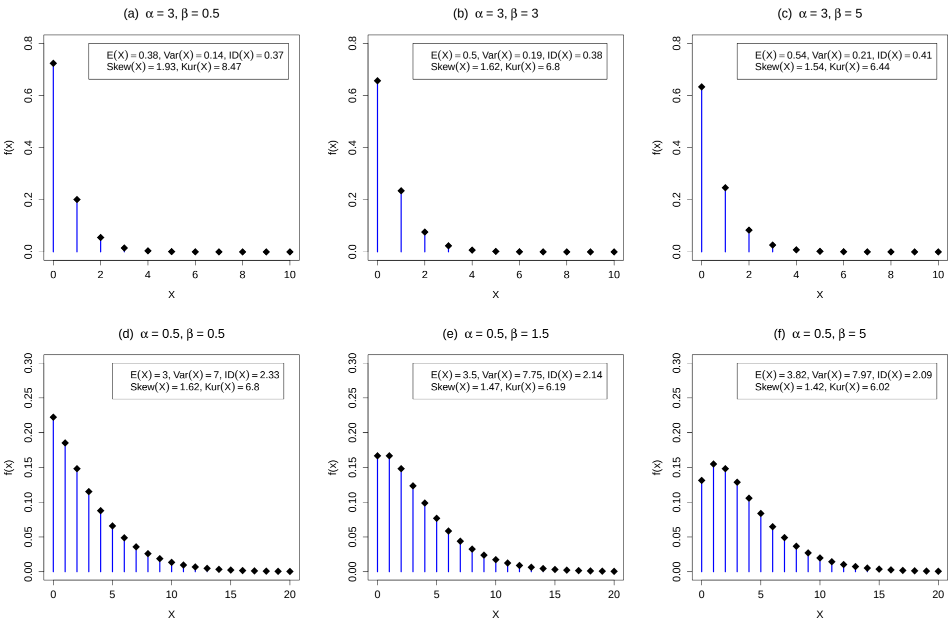

Figure 1 Some pmf plots of the GPL distribution

3. A New Bivariate Poisson-Lindley Distribution

In this section, we derive a bivariate version of the GPL distribution, namely the bivariate generalized Poisson-Lindley (BGPL) distribution, by using the concept of the bivariate Poisson (BP) distribution with parameters λ_1, λ_2 and θ, denoted by BP$(\lambda_1, \lambda_2, \theta)$. The BP distribution was introduced by Lakshminarayana et al. (1999). Its joint pmf is derived from the product of two Poisson distributions with a multiplicative factor parameter as

$$\begin{aligned} f(x_1, x_2) &= P(X_1 = x_1, X_2 = x_1) \\ &= e^{-\lambda_1-\lambda_2} \frac{\lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!} \left\{1 + \theta \left[\left(g_1(x_1) - \bar{g}_1\right)\left(g_2(x_2) - \bar{g}_2\right)\right]\right\}, \end{aligned} \tag{8}$$

where $x_1, x_2 = 0, 1, 2, \ldots,$ $g_1(x_1)$ and $g_2(x_2)$ are bounded functions in x_1 and x_2 respectively. In (8), the term $1 + \theta\left[\left(g_1(x_1) - \bar{g}_1\right)\left(g_2(x_2) - \bar{g}_2\right)\right]$ is non-negative when $g_j(x_j) = e^{x_j}$ and $\bar{g}_j = E[g_j(X_j)] = E[e^{-x_j}]$, for $j = 1, 2$.

When $\theta = 0$, random variables X_1 and X_2 are independent. The joint pmf of the BP distribution is

$$f(x_1, x_2) = \frac{\lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!} e^{-\lambda_1-\lambda_2} \left\{1 + \theta \left(e^{-x_1} - e^{-c\lambda_1}\right)\left(e^{-x_2} - e^{-c\lambda_2}\right)\right\}, \tag{9}$$

where $x_j = 0, 1, 2, \ldots, c = 1 - e^{-1}, \lambda_j > 0$ and $-\infty < \theta < \infty$ for $j = 1, 2$ (Lakshminarayana et al., 1999).

Theorem 2 *If (X_1, X_2) be random variables distributed as the BGPL distribution with parameters α_1, α_2, β_1, β_2 and θ, denoted by $(X_1, X_2) \sim BGPL(\alpha_1, \alpha_2, \beta_1, \beta_2, \theta)$. Its joint pmf is*

$$\begin{aligned} f(x_1, x_2) &= \frac{\alpha_1^2(1 + \alpha_1 + \beta_1 + \beta_1 x_1)}{(\alpha_1 + \beta_1)(\alpha_1 + 1)^{x_1+2}} \frac{\alpha_2^2(1 + \alpha_2 + \beta_2 + \beta_2 x_2)}{(\alpha_2 + \beta_2)(\alpha_2 + 1)^{x_2+2}} \\ &\quad \times [1 + \theta(e^{-x_1} - m_1)(e^{-x_2} - m_2)], x_j = 0, 1, 2, \ldots, \end{aligned} \tag{10}$$

where $\alpha_j > 0, \beta_j > 0$ *for* $j = 1, 2$, $-\infty < \theta < \infty$, *and*

$$m_j = \frac{\alpha_j^2(\beta_j + \alpha_j - e^{-1} + 1)}{(\alpha_j + \beta_j)(\alpha_j - e^{-1} + 1)^2} \textit{ for } j = 1, 2.$$

Proof: Let $(X_1, X_2) \sim \text{BP}(\lambda_1, \lambda_2, \theta)$, and $X_j|\lambda_j \sim \text{Pois}(\lambda_j)$ where $\lambda_j \sim \text{TPaL}(\alpha_j, \beta_j)$ for $j = 1, 2$. The joint pmf of (X_1, X_2) is

$$\begin{aligned} f(x_1, x_2) &= \int_0^\infty g(x_1; \lambda_1) g(\lambda_1; \alpha_1, \beta_1) d\lambda_1 \int_0^\infty g(x_2; \lambda_2) g(\lambda_2; \alpha_2, \beta_2) d\lambda_2 \\ &\quad \times \left\{1 + \theta \left[\left(e^{x_1} - \bar{g}_1\right) \left(e^{x_2} - \bar{g}_2\right) \right] \right\}, \end{aligned}$$

where $\bar{g}_1 = E[e^{-x_1}]$ and $\bar{g}_2 = E[e^{-x_2}]$. From the mgf of the TPaL distribution in (7), the joint pmf of the BGPL distribution is

$$\begin{aligned} f(x_1, x_2) &= \int_0^\infty \frac{e^{-\lambda_1} \lambda_1^{x_1}}{x_1!} \frac{\alpha_1^2}{\alpha_1 + \beta_1} (1 + \beta_1 \lambda_1) e^{-\alpha_1 \lambda_1} d\lambda_1 \int_0^\infty \frac{e^{-\lambda_2} \lambda_2^{x_2}}{x_2!} \frac{\alpha_2^2}{\alpha_2 + \beta_2} d\lambda_2 \\ &\quad \times \left[1 + \theta(e^{-x_1} - m_1)(e^{-x_2} - m_2)\right] \\ &= \frac{\alpha_1^2(1 + \alpha_1 + \beta_1 + \beta_1 x_1)}{(\alpha_1 + \beta_1)(\alpha_1 + 1)^{x_1 + 2}} \frac{\alpha_2^2(1 + \alpha_2 + \beta_2 + \beta_2 x_2)}{(\alpha_2 + \beta_2)(\alpha_2 + 1)^{x_2 + 2}} \\ &\quad \times \left[1 + \theta(e^{-x_1} - m_1)(e^{-x_2} - m_2)\right], \end{aligned}$$

where $m_j = \frac{\alpha_j^2(\beta_j + \alpha_j - e^{-1} + 1)}{(\alpha_j + \beta_j)(\alpha_j - e^{-1} + 1)^2}$ for $j = 1, 2$.

The BGPL distribution has five-parameter: β_1 and β_2 are shape parameter; α_1, α_2 and θ are scale parameter. The plots of the joint pmf of the BGPL distribution with some specified parameters are shown in Figure 2 by using the `plot3D` package in R (Soetaert , 2021).

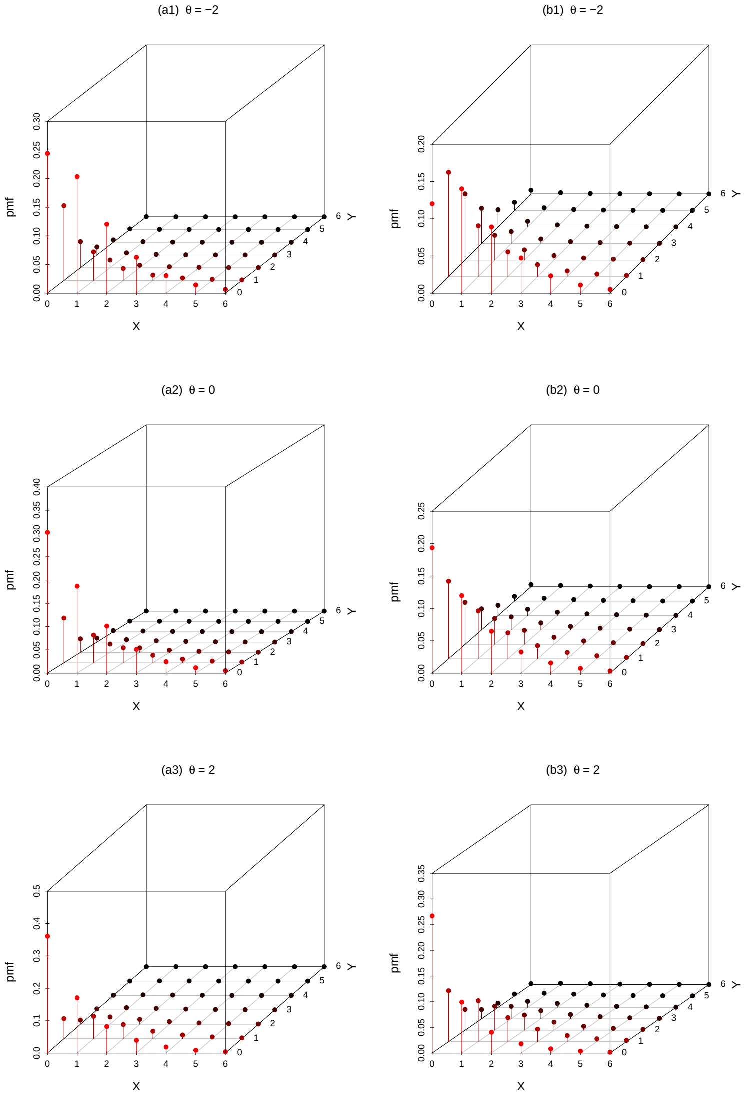

Figure 2 The joint pmf plots of (X, Y), where $(X, Y) \sim$ BGPL$(\alpha_1, \beta_1, \alpha_2, \beta_2, \theta)$: (a1)-(a3) BGPL$(1.5, 3, 3, 1.5, \theta)$; (b1)-(b3) BGPL$(1.5, 3, 1.5, 3, \theta)$

The mean, variance, and covariance of the BGPL distribution are respectivley,

$$\begin{aligned}\mathrm{E}(X_j) &= \frac{\alpha_j+2\beta_j}{\alpha_j(\beta_j+\alpha_j)}=\mu_j,\\ \mathrm{Var}(X_j) &= \frac{\alpha_j^2+4\alpha_j\beta_j+2\beta_j^2}{\alpha_j^2(\alpha_j+\beta_j)^2}=\sigma_j^2,\ \text{and}\\ \mathrm{Cov}(X_1,X_2) &= \theta\left(m_{11}-\mu_1 m_1\right)\left(m_{22}-\mu_2 m_2\right)=\sigma_{1,2},\end{aligned}$$

where $m_{jj}=E(X_je^{-X_j}), j=1,2$. By differentiating the mgf in (7) with respect to t and letting $t=-1$ that is, $\frac{\partial}{\partial t}M_X(t)\big|_{t=-1}=E(Xe^{-X})$, then

$$m_{jj}=\frac{\alpha_j^2(\alpha_j+2\beta_j-e^{-1}+1)e^{-1}}{(\alpha_j+\beta_j)(\alpha_j-e^{-1}+1)^3}, j=1,2.$$

Consequently, the correlation coefficient of X_1 and X_2 is

$$\rho_{X_1,X_2}=\frac{\sigma_{1,2}}{\sqrt{\sigma_1^2\sigma_2^2}}=\frac{\theta\left(m_{11}-\mu_1 m_1\right)\left(m_{22}-\mu_2 m_2\right)}{\sigma_1\sigma_2}. \tag{11}$$

For $\theta=0$, we have random variables X_1 and X_2 that are independent. If $\theta>0$ and $\theta<0$ then random variables X_1 and X_2 have a positive and negative correlation, respectively.

3.1. Some sub-models of the BGPL distribution

Let $(X_1,X_2)\sim$ BGPL$(\alpha_1,\beta_1,\alpha_2,\beta_2,\theta)$. Some special sub-models of the BGPL distribution are presented as the following.

- When $\beta_1=\beta_2=1$ then the BGPL distribution reduces to the BPL (Gómez-Déniz et al., 2008) with the joint pmf:

$$\begin{aligned}f(x_1,x_2) &= \frac{\alpha_1^2(2+\alpha_1+x_1)}{(\alpha_1+1)(\alpha_1+1)^{x_1+2}}\frac{\alpha_2^2(2+\alpha_2+x_2)}{(\alpha_2+1)(\alpha_2+1)^{x_2+2}}\\ &\quad\times[1+\theta(e^{-x_1}-m_1)(e^{-x_2}-m_2)], x_j=0,1,2,\ldots\end{aligned}$$

where $\alpha_j>0, -\infty<\theta<\infty$ and $m_j=\frac{\alpha_j^2(\alpha_j-e^{-1}+2)}{(\alpha_j+1)(\alpha_j-e^{-1}+1)^2}$ for $j=1,2$.

- In the case of $\beta_1=\beta_2=0$ the BGPL distribution reduces to the bivariate geometric (BGeo) distribution with the joint pmf:

$$f(x_1,x_2)=\frac{p_1p_2\left[1+\theta(e^{-x_1}-m_1)(e^{-x_2}-m_2)\right]}{(\alpha_1+1)^{x_1}(\alpha_2+1)^{x_2}}; x_j=0,1,2,\ldots,0<p_j<1, \tag{12}$$

where $\alpha_j>0, p_j=\frac{\alpha_j}{\alpha_j+1}, -\infty<\theta<\infty$ and $m_j=\frac{\alpha_j}{\alpha_j-e^{-1}+1}$ for $j=1,2$.

- If $\alpha_1=\alpha_2=\alpha$ and $\beta_1=\beta_2=\beta$ then its joint pmf is

$$\begin{aligned}f(x_1,x_2) &= \frac{\alpha^4(1+\alpha+\beta+\beta x_1)(1+\alpha+\beta+\beta x_2)}{(\alpha+\beta)^2(\alpha+1)^{x_1+2}(\alpha+1)^{x_2+2}}\\ &\quad\times[1+\theta(e^{-x_1}-m)(e^{-x_2}-m)], x_1,x_2=0,1,2,\ldots\end{aligned}$$

where $\alpha>0, \beta>0$ and $m=\frac{\alpha^2(\beta+\alpha-e^{-1}+1)}{(\alpha+\beta)(\alpha-e^{-1}+1)^2}$.

4. Parameter Estimation

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an independent and identically distributed (iid) as BGPL($\boldsymbol{\Theta}$) where $\boldsymbol{\Theta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, \theta)$. The log-likelihood function of $\boldsymbol{\Theta}$ can be written as,

$$\begin{aligned}
\ell(\boldsymbol{\Theta}) &= \log \prod_{i=1}^{n} f(x_i, y_i; \boldsymbol{\Theta}) \\
&= 2n \log \alpha_1 + 2n \log \alpha_2 - n \log(\alpha_2 + \beta_2) - (x_i + 2) \sum_{i=1}^{n} \log(\alpha_1 + 1) \\
&\quad - n \log(\alpha_1 + \beta_1) + \sum_{i=1}^{n} \log(1 + \alpha_1 + \beta_1 + \beta_1 x_i) \\
&\quad + \sum_{i=1}^{n} \log(1 + \alpha_2 + \beta_2 + \beta_2 y_i) - (y_i + 2) \sum_{i=1}^{n} \log(\alpha_2 + 1) \\
&\quad + \sum_{i=1}^{n} \log \left[1 + \theta \left(e^{-x_i} - \frac{\alpha_1^2(\beta_1 + \alpha_1 - e^{-1} + 1)}{(\alpha_1 + \beta_1)(\alpha_1 - e^{-1} + 1)^2} \right) \right. \\
&\quad \left. \times \left(e^{-y_i} - \frac{\alpha_2^2(\beta_2 + \alpha_2 - e^{-1} + 1)}{(\alpha_2 + \beta_2)(\alpha_2 - e^{-1} + 1)^2} \right) \right].
\end{aligned}$$

To estimate the unknown parameters $\boldsymbol{\Theta}$ we take the partial derivatives with respect to each parameter, and then equate them to zero, i.e.,

$$\frac{\partial \ell(\boldsymbol{\Theta})}{\partial \alpha_1} = 0, \frac{\partial \ell(\boldsymbol{\Theta})}{\partial \alpha_2} = 0, \frac{\partial \ell(\boldsymbol{\Theta})}{\partial \beta_1} = 0, \frac{\partial \ell(\boldsymbol{\Theta})}{\partial \beta_2} = 0, \frac{\partial \ell(\boldsymbol{\Theta})}{\partial \theta} = 0. \tag{13}$$

Since the above equations are not presented in closed forms, thus the numerical method of the five-dimensional Newton-Raphson type procedure is used for solving this system of equations. The solutions of the ML estimates of $\boldsymbol{\Theta}$ by using the R program with the `nlm` function in the **stats** R package (R Core Team, 2021). The R source code for estimating the parameters is provided in Appendix.

5. Simulation Study

In this section, the simulation is provided for estimating the parameters of the BGPL distribution based on ML estimation method. Some situations for the parameter estimation of the BGPL are studied. The population size is $N = 1,000$ and three sets of parameters are

- Case 1 $(X_i, Y_i) \sim$ BGPL$(1.5, 3, 1.5, 3, -2)$,
- Case 2 $(X_i, Y_i) \sim$ BGPL$(1.5, 3, 1.5, 3, 0)$,
- Case 3 $(X_i, Y_i) \sim$ BGPL$(1.5, 3, 1.5, 3, 2)$,

where $i = 1, 2, \ldots, N$.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an iid as BGPL$(\alpha_1, \beta_1, \alpha_2, \beta_2, \theta)$. Each case is considered to run with sample sizes of n as 50, 100, 150, and 200, with 1,000 replications. The ML estimate of the parameter Θ_k for $k = 1, 2, 3, 4, 5$ were obtained by including the estimated ML estimators and its root mean square error (RMSE) as:

$$\text{Estimate}(\hat{\Theta_k}) = \frac{1}{1,000} \sum_{j=1}^{1,000} \hat{\Theta}_{kj} \text{ and } \text{RMSE}(\hat{\Theta_k}) = \sqrt{\frac{1}{1,000} \sum_{j=1}^{1,000} \left(\hat{\Theta}_{kj} - \Theta_k \right)^2}.$$

Results of the simulation study are shown in Table 1, and box-plots of the parameter estimates are shown in Figures 3 - 5. Larger samples showed greater promise to give estimators closer to their parameter and the RMSE decreased. Thus, the ML method offered greater efficiency to determine the parameters as the sample size increased.

Table 1 Results of the ML estimates of $(X_i, Y_i) \sim$ BGPL$(\alpha_1, \beta_1, \alpha_2, \beta_2, \theta)$

Cases	Parameters	Statistics	Sample size (n) 50	100	150	200
1	$\alpha_1 = 1.5$	Estimate	1.707	1.643	1.631	1.544
		RMSE	0.423	0.306	0.238	1.846
	$\beta_1 = 3$	Estimate	7.419	4.548	4.326	3.549
		RMSE	7.316	3.091	2.353	1.846
	$\alpha_2 = 1.5$	Estimate	1.696	1.666	1.688	1.590
		RMSE	0.405	4.131	0.267	0.243
	$\beta_2 = 3$	Estimate	7.079	5.220	4.957	3.585
		RMSE	6.842	4.131	2.871	1.792
	θ=-2	Estimate	-2.302	-2.080	-2.164	-2.178
		RMSE	0.972	0.925	0.485	0.396
2	$\alpha_1 = 1.5$	Estimate	1.709	1.730	1.519	1.555
		RMSE	0.416	0.372	0.175	0.166
	$\beta_1 = 3$	Estimate	6.829	5.992	4.109	3.671
		RMSE	7.068	5.513	3.277	1.836
	$\alpha_2 = 1.5$	Estimate	1.721	1.709	1.519	1.502
		RMSE	0.431	0.353	0.175	0.141
	$\beta_2 = 3$	Estimate	7.154	6.554	4.697	3.711
		RMSE	7.230	5.784	3.560	1.912
	θ=0	Estimate	-0.086	-0.071	-0.093	-0.053
		RMSE	0.868	0.614	0.482	0.416
3	$\alpha_1 = 1.5$	Estimate	1.968	1.842	1.696	1.577
		RMSE	1.516	0.530	0.397	0.197
	$\beta_1 = 3$	Estimate	8.189	5.266	3.848	3.765
		RMSE	7.192	4.217	3.931	2.258
	$\alpha_2 = 1.5$	Estimate	1.837	1.729	1.651	1.498
		RMSE	0.580	0.522	0.400	0.342
	$\beta_2 = 3$	Estimate	7.506	4.999	3.895	3.558
		RMSE	6.690	3.785	3.216	3.174
	θ=2	Estimate	2.008	1.987	1.934	1.989
		RMSE	0.905	0.713	0.492	0.456

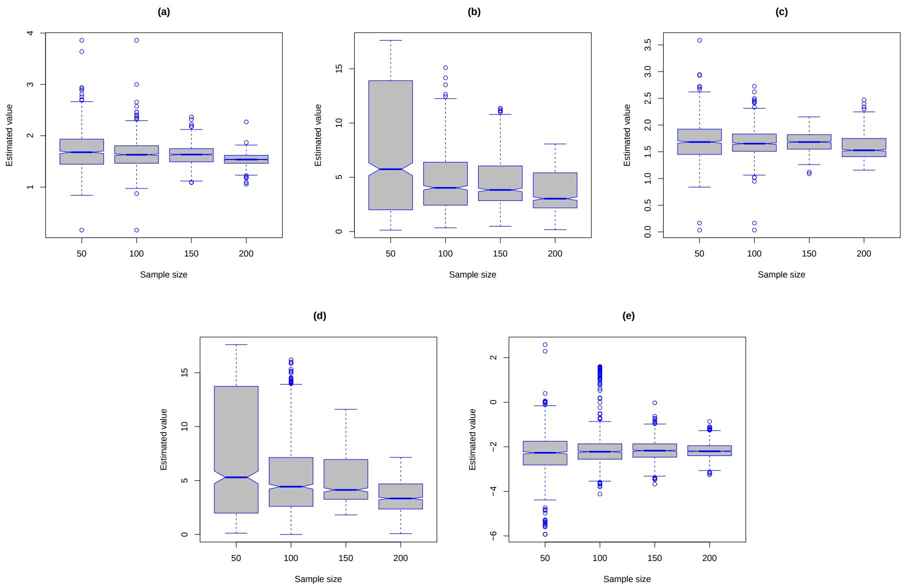

Figure 3 Box-plots of the parameter estimates for parameters (a) $\alpha_1 = 1.5$, (b) $\beta_1 = 3$, (c) $\alpha_2 = 1.5$, (d) $\beta_2 = 3$ and (e) $\theta = -2$

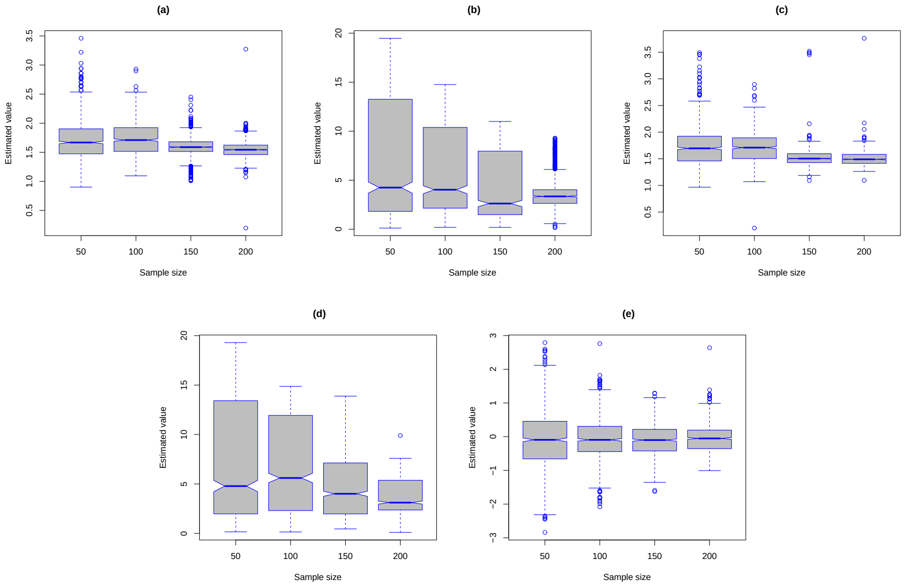

Figure 4 Box-plots of the parameter estimates for parameters (a) $\alpha_1 = 1.5$, (b) $\beta_1 = 3$, (c) $\alpha_2 = 1.5$, (d) $\beta_2 = 3$ and (e) $\theta = 0$

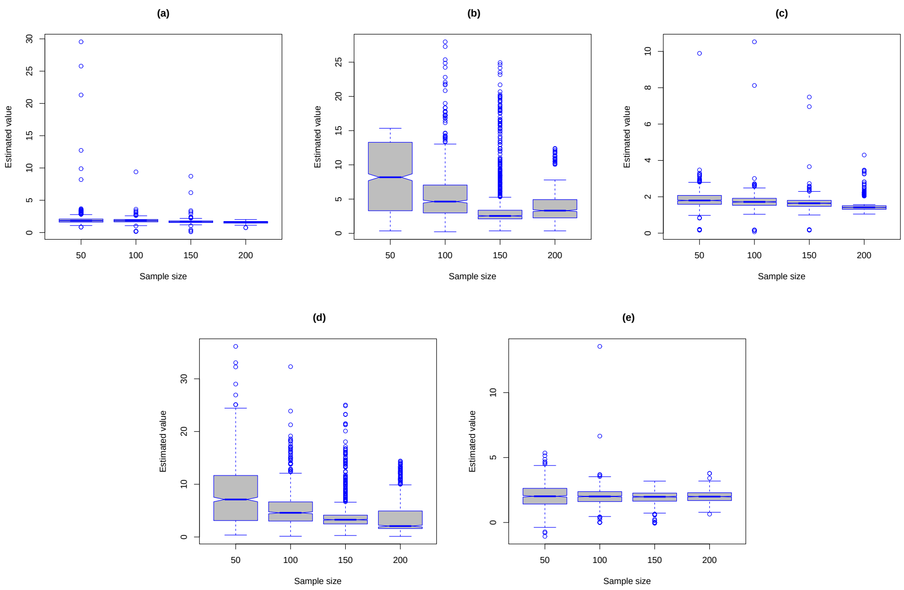

Figure 5 Box-plots of the parameter estimates for parameters (a) $\alpha_1 = 1.5$, (b) $\beta_1 = 3$, (c) $\alpha_2 = 1.5$, (d) $\beta_2 = 3$ and (e) $\theta = 2$

6. Application

In this section, the numerical illustration is provided the application of the BGPL distribution. The first data set is accident data from 122 experienced shunters, where random variables X and Y respectively represent the number of accidents in 1937-1942 and 1943-1947 (see Zamani et al. (2014)), see Table 3. In addition, the second data set the flight aborts count data from 109 aircraft, where random variables X and Y respectively represent the number of flight aborts in the first and second consecutive six months of one year (Mitchell and Paulson, 1981), is shown in Table 5.

These data are fitted with some distributions, such as the BP (Lakshminarayana et al., 1999), BGeo, BPL (Gómez-Déniz et al., 2008), and BGPL distributions. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used for a model selection criteria. If the distribution gives the smallest value of the AIC and BIC then it will be the best fitting distribution.

Table 2 shows the estimated parameters, AIC and BIC of each distribution. The observed values of the number of accidents in 1937-1942 (X) and 1943-1947 (Y), and its expected values from the fitting of the distributions are shown in Table 3. The result shows that all distributions provide a positive value of $\hat{\theta}$, which indicates that X and Y have a positive correlation. The BGPL distribution provides the lowest value of the AIC and BIC than the BP, BGeo, and BPL distributions, indicating that the proposed BGPL distribution is the best fitting distribution for this data set.

Table 2 Results of the ML estimates of fitting distributions for the number of accidents in 1937-1942 (X) and 1943-1947 (Y)

Parameter estimates	Distributions			
	BP	BGeo	BPL	BGPL
$\hat{\lambda}_1$ (s.e.)	1.2821 (0.1023)	-	-	-
$\hat{\lambda}_2$ (s.e.)	0.9811 (0.0902)	-	-	-
$\hat{\theta}$ (s.e.)	1.7259 (0.6118)	1.7763 (0.5788)	1.7238 (0.5832)	1.5579 (0.6043)
$\hat{\alpha}_1$ (s.e.)	-	0.8203 (0.0988)	1.1751 (0.1142)	1.5540 (0.1650)
$\hat{\alpha}_2$ (s.e.)	-	1.0992 (0.1410)	1.5031 (0.1581)	1.9843 (0.2425)
$\hat{\beta}_1$ (s.e.)	-	-	-	44.091 (15.7900)
$\hat{\beta}_2$ (s.e.)	-	-	-	30.548 (11.2680)
$-\log L$	354.29	351.38	348.40	343.22
AIC	714.58	708.76	702.80	696.44
BIC	722.99	717.17	711.21	710.46

Table 4 illustrates the estimated parameters, AIC and BIC for each distribution. The observed values of the number of flight aborts in the first (X) and second (Y) consecutive six months of a one-year period, and its expected values from the fitting of the distributions are shown in Table 5. The result shows that all distributions provide a negative value for $\hat{\theta}$, which indicates that X and Y have a negative correlation. The BGPL distribution provides the lowest value of the AIC and BIC than the BP, BGeo, and BPL distributions, which indicate that the proposed BGPL distribution provides the best fitting distribution for this data set.

7. Conclusion

In this paper, the BGPL distribution is presented, it is a bivariate version of the GPL distribution (Bhati et al., 2015). The BGPL distribution is derived by using the concept of the bivariate Poisson (BP) distribution (Lakshminarayana et al., 1999). Some sub-models, such as the bivariate Poisson Lindley (BPL), and bivariate geometric (BGeo) distributions are presented. The unknown parameters of the BGPL distributions are estimated by using the maximum likelihood estimation. Based on the simulation result, the ML method offered greater efficiency to determine the parameters as the sample size increased. The application of the BGPL distribution has been illustrated with correlated bivariate count data, with either positive or negative correlations. Based on the results, the BGPL distribution provides the smallest value of the AIC and BIC when it compares with the BP, BGeo,

Table 3 Observed values of the number of accidents in 1937-1942 (X) and 1943-1947 (Y), and its expected values from fitting of the distributions

		Y						
X		0	1	2	3	4	7	Total
0	Observed	21	13	4	2	0	0	40
	BP	24.34	17.81	7.98	2.68	0.71	0	53.52
	BGeo	36.91	10.88	4.01	1.70	0.78	0.08	54.36
	BPL	34.02	10.80	4.08	1.72	0.76	0.07	51.45
	BGPL	27.64	11.23	4.38	1.75	0.70	0.04	45.74
1	Observed	18	14	5	1	0	1	39
	BP	11.71	16.99	10.17	3.84	1.07	0.01	43.79
	BGeo	13.80	8.24	4.22	2.06	0.99	0.11	29.42
	BPL	14.01	8.77	4.57	2.21	1.02	0.09	30.67
	BGPL	14.53	10.76	5.59	2.52	1.06	0.06	34.52
2	Observed	8	10	4	3	1	0	26
	BP	3.56	7.33	4.72	1.82	0.51	0	17.94
	BGeo	6.27	4.98	2.72	1.36	0.66	0.07	16.06
	BPL	6.61	5.47	3.04	1.50	0.70	0.06	17.38
	BGPL	7.06	6.74	3.74	1.72	0.73	0.04	20.03
3	Observed	2	1	2	2	1	0	8
	BP	0.85	2.04	1.35	0.52	0.15	0	4.91
	BGeo	3.18	2.83	1.58	0.79	0.38	0.04	8.8
	BPL	3.35	3.10	1.76	0.87	0.41	0.04	9.53
	BGPL	3.40	2.89	2.03	0.94	0.4	0.02	9.68
4	Observed	1	4	1	0	0	0	6
	BP	0.16	0.42	0.28	0.11	0.03	0	1.00
	BGeo	1.69	1.57	0.88	0.44	0.22	0.02	4.82
	BPL	1.74	1.67	0.96	0.48	0.22	0.02	5.09
	BGPL	1.61	1.77	1.01	0.47	0.2	0.01	5.07
5	Observed	0	1	0	1	0	0	2
	BP	0.03	0.07	0.05	0.02	0.01	0	0.18
	BGeo	0.92	0.87	0.49	0.25	0.12	0.01	2.66
	BPL	0.90	0.88	0.50	0.25	0.12	0.01	2.66
	BGPL	0.75	0.83	0.48	0.22	0.09	0.01	2.38
6	Observed	0	0	1	0	0	0	1
	BP	0	0.01	0.01	0	0	0	0.02
	BGeo	0.50	0.48	0.27	0.14	0.07	0.01	1.47
	BPL	0.46	0.45	0.26	0.13	0.06	0.01	1.37
	BGPL	0.34	0.38	0.22	0.1	0.04	0	1.08

Table 4 Results of the ML estimates of fitting distributions for the number of flight aborts in the first (X) and second (Y) consecutive six months of a one-year period

Parameter	Distributions			
estimates	BP	BGeo	BPL	BGPL
$\hat{\lambda}_1$ (s.e.)	0.5977 (0.0736)	-	-	-
$\hat{\lambda}_2$ (s.e.)	0.6634 (0.0772)	-	-	-
$\hat{\theta}$ (s.e.)	-1.8269 (0.7097)	-1.9211 (0.6466)	-1.9247 (0.6522)	-1.9253 (0.6503)
$\hat{\alpha}_1$ (s.e.)	-	1.6102 (0.2477)	2.1372 (0.2760)	1.6172 (2.0420)
$\hat{\alpha}_2$ (s.e.)	-	1.4571 (0.2172)	1.9572 (0.2433)	1.9389 (1.0739)
$\hat{\beta}_1$ (s.e.)	-	-	-	-0.0006 (2.0232)
$\hat{\beta}_2$ (s.e.)	-	-	-	0.9565 (3.5950)
$-\log L$	251.31	242.64	242.67	237.59
AIC	508.62	491.28	491.34	485.18
BIC	516.69	499.35	499.41	498.64

Table 5 Observed values of the number of flight aborts in the first (X) and second (Y) consecutive six months of a one-year period, and its expected values from fitting of the distributions

X		Y 0	1	2	3	4	Total
0	Observed	34	20	4	6	4	68
	BP	24.80	23.90	8.84	2.03	0.34	59.91
	BGeo	33.34	19.12	8.62	3.63	1.50	66.21
	BPL	32.54	19.52	8.79	3.60	1.41	65.86
	BGPL	32.89	19.66	8.86	3.64	1.43	66.48
1	Observed	17	7	0	0	0	24
	BP	22.13	10.19	2.83	0.58	0.09	35.82
	BGeo	18.39	4.84	1.57	0.58	0.23	25.61
	BPL	18.65	5.13	1.67	0.60	0.22	26.27
	BGPL	18.20	4.98	1.62	0.58	0.22	25.60
2	Observed	6	4	1	0	0	11
	BP	7.42	2.59	0.58	0.10	0.02	10.71
	BGeo	7.84	1.50	0.36	0.11	0.04	9.85
	BPL	7.91	1.59	0.38	0.11	0.04	10.03
	BGPL	7.74	1.55	0.37	0.11	0.04	9.81
3	Observed	0	4	0	0	0	4
	BP	1.54	0.48	0.10	0.02	0	2.14
	BGeo	3.11	0.53	0.10	0.03	0.01	3.78
	BPL	3.04	0.54	0.11	0.03	0.01	3.73
	BGPL	3.07	0.54	0.11	0.03	0.01	3.76
5	Observed	2	0	0	0	0	2
	BP	0.03	0.01	0	0	0	0.04
	BGeo	0.47	0.07	0.01	0	0	0.55
	BPL	0.40	0.07	0.01	0	0	0.48
	BGPL	0.46	0.08	0.01	0	0	0.55

and BPL distributions. It is indicated that the BGPL distribution is the best fitting distribution for these data sets. Therefore, the BGPL distribution is an alternative for fitting correlated and over-dispersed bivariate count data, with either positive or negative correlations.

References

Bhati D, Sastry DVS, Qadri PM. A new generalized Poisson-Lindley distribution: Applications and properties. Austrian J Stat. 2015; 44(4): 35-51.

Deepesh B, Kumawat P, Gómez-Déniz E. Mixed Poisson distributions. Commun Stat Theory. 2017; 46(22): 11060-11076.

Ghitany ME, Atieh B, Nadarajah S. Lindley distribution and its application. Math Comput Simul. 2008; 78(4): 493-506

Gómez-Déniz E, Sarabia JM, Calderín-Ojeda E. Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. Insur.: Math Econ. 2012; 42(1): 39-49.

Holla MS. On a Poisson-inverse Gaussian distribution, Metrika. 1967; 11(1): 115-121.

Karlis D, Xekalaki E. Mixed Poisson distributions. Internat Statist Rev. 2005; 73(1): 35-58.

Lakshminarayana J, Pandit SNN, Rao KS. On a bivariate Poisson distribution. Commun Stat Theory. 1999; 28(2): 267-276.

Mahmoudi E, Zakerzadeh H. Generalized Poisson-Lindley Distribution. Commun Stat Theory. 2010; 39(10): 1785-17.

Marshall A, Olkin I. Multivariate distributions generated from mixtures of convolution and product families, Lect. Notes Monograph Ser. 1990; 16: 371-393.

Mitchell CR, Paulson AS. A new bivariate negative binomial distribution. Nav Res Logist Q. 1981; 28(3): 359-374.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/(2021).

Sankaran M. The discrete Poisson-Lindley distribution. Biometrics. 1970; 26(1): 145-149.

Shanker R, Sharma S, Shanker R. A discrete two-parameter Poisson Lindley distribution. J Ethiop Stat Assoc. 2012; 21: 15-22.

Shanker R, Mishra A. A two-parameter Poisson-Lindley distribution, Int J Stat Syst. 2014; 9(1):79-85.

Shanker R, Sharma S, Shanker R. A two-parameter Lindley distribution for modeling waiting and survival times data. Appl Math. 2013; 4(2): 363-368.

Soetaert K. plot3D: Plotting Multi-Dimensional Data. R package version 1.4. Available from: https://CRAN.R-project.org/package=plot3D(2021).

Wongrin W, Bodhisuwan W. The Poisson-generalised Lindley distribution and its applications. Songklanakarin J Sci Technol. 2016; 38(6): 645-656.

Zamani H, Ismail N, Faroughi P. Poisson-weighted exponential univariate version and regression model with applications, J Math Stat. 2014; 10(2): 148-154.

Zamani H, Ismail N, Faroughi P. Bivariate Poisson-weighted exponential distribution with applications. AIP Conf Proc. 2014; 1602(1): 964-968.

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

Appendix

I. R code for the pmf of the BGPL distribution

```
dGPLx<-function(x,alpha1,beta1){
   k<-numeric(length(x));
   GPL<-function(x,alpha1,beta1){
    p1<-alpha1^2*(1+alpha1+beta1+beta1*x);
    p2<-(alpha1+beta1)*(alpha1+1)^(x+2);
    p<-p1/p2; p
   }
   for(i in 1:length(x)){
    k[i]<-GPL(x[i],alpha1,beta1);
   }
  k
  }
dGPLy<-function(y,alpha2,beta2){
   k<-numeric(length(y));
   GPL<-function(y,alpha2,beta2){
    p1<-alpha2^2*(1+alpha2+beta2+beta2*y);
    p2<-(alpha2+beta2)*(alpha2+1)^(y+2);
    p<-p1/p2; p
   }
   for(i in 1:length(y)){
    k[i]<-GPL(y[i],alpha2,beta2)
    }
  k
}
dBGPL<-function(x,y,alpha1,beta1,alpha2,beta2,theta){
 B<-matrix(data=NA,nrow=length(x),ncol=1);
 for(i in 1:length(x)){
    m1<-(alpha1^2*(beta1+alpha1-exp(-1)+1))/((alpha1+beta1)*
        (alpha1-exp(-1)+1)^2);
    m2<-(alpha2^2*(beta2+alpha2-exp(-1)+1))/((alpha2+beta2)*
        (alpha2-exp(-1)+1)^2);
    mx<-dGPLx(x[i],alpha1,beta1);
    my<-dGPLy(y[i],alpha2,beta2);
    B1<-mx*my;
    B2<-1+theta*(exp(-x[i])-m1)*(exp(-y[i])-m2)
    B[i]<-B1*B2; B
  }
  print(B)
}
```

II. R code for the ML estimation of the BGPL distribution

```
logBGPL<-function(x,y,t){
    alpha1<-t[1]; beta1<-t[2];
    alpha2<-t[3]; beta2<-t[4]; theta<-t[5];
    prob<-dBGPL(x,y,alpha1,beta1,alpha2,beta2,theta)
    lpdf1<-log(prob)
    loglike<- -sum(lpdf1)
    return(loglike)
}
t.start<-c(alpha10,beta10,alpha20,beta20,theta0)
nlm(logBGPL,x=x,y=y,t.start,hessian=T)
```