REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000–000
https://doi.org/00.00000/revstat.v00i0.000

On a Poisson Generalized Lindley Distribution: Related Models, Bivariate Extensions and Insurance Applications

Authors: Maria Vardaki ✉
– School of Health Sciences,
National and Kapodistrian University of Athens,
75, Mikras Asias str.,
Athens, 11527, Greece
mvardaki@uoa.gr

H. Papageorgiou
– Department of Mathematics,
National and Kapodistrian University of Athens,
Panepistemiopolis
Athens 15784, Greece
hpapageo@math.uoa.gr

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- In this paper we demonstrate that univariate and bivariate Poisson generalized Lindley (PGL) distributions can be derived by the different procedures of mixing, generalizing and addition of random variables. We also introduce and study a univariate PGL distribution which is over-dispersed and appears as marginal distribution in three bivariate PGL models, indicating potential use of the corresponding bivariate models to describe correlated count data. Furthermore, we derive two bivariate PGL distributions and examine a variety of their properties including various recurrences, conditional distributions and regression functions. Applications to different bivariate sets of automobile insurance claims data are also included.

Keywords:

- *Poisson-Lindley distributions; addition of random variables; insurance data.*

AMS Subject Classification:

- 60E05, 62E15, 62P05.

✉ Corresponding author

1. INTRODUCTION

Count data are collected in many scientific fields, not only for one variable of interest, but also for two (or more) dependent variables. For example, some of these areas are accident and insurance studies (Gómez-Dèniz, 2016; Gómez-Dèniz et al., 2012; Papageorgiou, 1986), ecology (Holgate, 1966), biology, sociology and demography (Rao et al., 1973; Papageorgiou and David, 1994). Very often in univariate or bivariate count data samples, the variance or the marginal variances are greater than the mean or the marginal means respectively, as in the bivariate data sets recorded in the above papers. Therefore, for modeling them we should employ univariate distributions with the property of over-dispersion or bivariate models with over-dispersed marginals.

A recently developed class of distributions with this property based on the seminal paper by Sankaran (1970) and applications in a variety of areas is the class of Poisson-Lindley distributions. Among its members, we refer to the univariate models derived by Chesneau et al. (2023); Atikankul (2023); Irshad et al. (2023); Bhati et al. (2015); Mahmoudi and Zakerzadeh (2010), and to the bivariate or multivariate models by Irshad et al. (2024); Vardaki and Papageorgiou (2024); Papageorgiou and Vardaki (2022); Gómez-Dèniz et al. (2012).

The purpose of this paper is threefold. (i) To introduce and study an alternative univariate Poisson generalized Lindley (APGL) distribution, which appears as marginal distribution in all bivariate models presented in this paper. In addition, utilizing the general methodology developed by Papageorgiou and Vardaki (2022) we introduce two alternative bivariate PGL distributions and examine their characteristics in detail. (ii) To demonstrate that the univariate new Poisson generalized Lindley (PGL) distribution introduced by Atikankul (2023) and Irshad et al. (2023) and the bivariate PGL distribution introduced by Irshad et al. (2024), as well as various versions of them, can be derived not only by mixing but also by employing the techniques of generalization and addition of random variables (r.v.'s). Finally, (iii) To indicate possible applicability of our bivariate models to automobile insurance problems by fitting them to various sets of claims data.

The rest of the paper is organized as follows. In Section 2, we point out that the new PGL distribution introduced by Atikankul (2023) is the same distribution introduced by Irshad et al. (2023) under the name two-parameter Poisson generalized Lindley (TPPGL) distribution and it is also related to the models of Bakouch et al. (2022) and Mahmoudi and Zakerzadeh (2010). A derivation of the PGL distribution as convolution of Sankaran's Poisson-Lindley with a negative binomial is also given. In Section 3, we introduce and study in detail a univariate APGL distribution, demonstrating that this model can be derived by the techniques of mixing, generalizing and addition of r.v.'s. In Subsection 4.1 the bivariate PGL distribution introduced by Irshad et al. (2024) as a mixture of two independent Poissons with a PGL distribution is briefly mentioned. However, in Subsection 4.2 we derive another bivariate PGL distribution by using the PGL as the mixing distribution of a bivariate correlated Poisson model.

In Section 5, we obtain a bivariate PGL distribution by generalizing a bivariate binomial distribution assuming that its exponent follows a PGL model and we examine in detail several of its properties. In Section 6, we derive the three bivariate PGL distributions described in Subsections 4.1, 4.2 and Section 5 by adding bivariate distributions previously studied in the literature. In Section 7, we fit the bivariate models which we introduced in Subsection 4.2 and Section 5 to different sets of automobile insurance claims data. In Subsections 8.1 we fit our models described in Subsection 7.1 to two additional sets of automobile insurance claims data. In addition in Subsections 8.1 and 8.2 comparisons of our models with other models fitted to the same data sets are given. Finally, Section 9 concludes.

2. THE NEW POISSON GENERALIZED LINDLEY (PGL) DISTRIBUTION

The new PGL distribution was introduced by Atikankul (2023) and also by Irshad et al. (2023) under the name two-parameter Poisson generalized Lindley (TPPGL) distribution, as a Poisson mixture. In particular, they assumed that the Poisson parameter is a continuous random variable (r.v.) Λ, following the new generalized Lindley distribution studied by Abouammoh et al. (2015) et al. (2015), with probability density function (p.d.f.)

$$f(\lambda;\theta,\alpha)=\frac{\theta^{\alpha}}{(\theta+1)}\,\frac{\lambda^{\alpha-2}(\lambda+\alpha-1)}{\Gamma(\alpha)}e^{-\theta\lambda},\qquad \lambda>0,\alpha>1,\theta>0, \tag{2.1}$$

and moment generating function (m.g.f.)

$$M_{\Lambda}(s)=\frac{\theta^{\alpha}}{\theta+1}\,\frac{\theta-s+1}{(\theta-s)^{\alpha}}. \tag{2.2}$$

It is well known that a discrete r.v. Y derived as a Poisson mixture, has probability generating function (p.g.f.) given by the relation

$$G_Y(s)=M_{\Lambda}(s-1). \tag{2.3}$$

Consequently, if the m.g.f. of the r.v. Λ is given by equation (2.2), we obtain

$$G_Y(s)=\frac{\theta^{\alpha}}{\theta+1}\,\frac{\theta-s+2}{(\theta-s+1)^{\alpha}}. \tag{2.4}$$

Calculating the y-th derivative of $G_Y(s)$ with respect to s, we have

$$\frac{\partial G^{y}(s)}{\partial s^{y}}=\frac{\theta^{\alpha}}{\theta+1}\,\frac{\Gamma(\alpha+y-1)}{\Gamma(\alpha)}\,\frac{(\alpha-1)(\theta-s+2)+y}{(\theta-s+1)^{\alpha+y}}. \tag{2.5}$$

From equation (2.5) we immediately obtain expressions for the probabilities

$$P(Y=y;\theta,\alpha)=\frac{1}{y!}\,\frac{\theta^{\alpha}}{\theta+1}\,\frac{\Gamma(\alpha+y-1)}{\Gamma(\alpha)}\,\frac{(\alpha-1)(\theta+2)+y}{(\theta+1)^{\alpha+y}}, \tag{2.6}$$

$y=0,1,\ldots$, $\alpha>1$, $\theta>0$ and the factorial moments

$$\mu_{[r]:Y}=\frac{\Gamma(\alpha+r-1)}{\Gamma(\alpha)}\,\frac{(\alpha-1)(\theta+1)+r}{\theta^{r}(\theta+1)},\qquad r=1,2,\ldots \tag{2.7}$$

where

$$\mu_{[r]:Y}=E(Y^{(r)})\quad\text{ and }\quad Y^{(r)}=Y(Y-1)\ldots(Y-r+1).$$

From equation (2.7)

$$E(Y) = \frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)}. \tag{2.8}$$

A simple characterization of the distributions of the r.v.'s Λ and Y is given by the following lemma, derived by a theorem proved by Papageorgiou and Vardaki (2022) for Poisson mixtures.

Lemma 2.1. *For Poisson mixtures of the form (2.3) with $\Lambda > 0$ a continuous r.v., the conditional expectation*

$$E[\Lambda \mid Y = y] = \frac{\alpha+y-1}{\theta+1}\,\frac{(\alpha-1)(\theta+2)+y+1}{(\alpha-1)(\theta+2)+y} \tag{2.9}$$

determines uniquely both the distributions of Λ and Y.

2.1. Another derivation and relation with other distributions

Proposition 2.1. *Consider two independent r.v.'s Y_1 and Y_2. If Y_1 follows a Poisson-Lindley distribution (Sankaran, 1970) with p.g.f. given by the equation*

$$G_{Y_1}(s) = \frac{\theta^2}{\theta+1}\,\frac{\theta-s+2}{(\theta-s+1)^2} \tag{2.10}$$

and Y_2 is distributed as a negative binomial with parameters $\alpha > 2$, $\frac{\theta}{1+\theta}$ and p.g.f. given by the equation

$$G_{Y_2}(s) = \left(\frac{\theta}{1+\theta-s}\right)^{\alpha-2}. \tag{2.11}$$

Then, the r.v. $Y_1 + Y_2$ follows a distribution with p.g.f.

$$\begin{aligned} G_{Y_1+Y_2}(s) &= G_{Y_1}(s)G_{Y_2}(s) \\ &= \frac{\theta^\alpha}{\theta+1}\,\frac{\theta-s+2}{(\theta-s+1)^\alpha}, \end{aligned}$$

which is the p.g.f. of a PGL distribution with parameters, $\alpha > 2$, $\theta > 0$.

Corollary 2.1. *If the negative binomial r.v. Y_2 with p.g.f. given by equation (2.11) is replaced by a geometric r.v. Y_3 with p.g.f.*

$$G_{Y_3}(s) = \frac{\theta}{1+\theta-s}, \tag{2.12}$$

then the distribution of the sum $Y_1 + Y_3$ is a PGL distribution with parameters $\alpha = 3$, $\theta > 0$. Consequently, its p.g.f. and probability function (p.f.) are given by the equations

$$G_Y(s) = \frac{\theta^3}{\theta+1}\,\frac{\theta-s+2}{(\theta-s+1)^3} \tag{2.13}$$

and from equation (2.6)

$$P(Y=y;\theta)=\frac{\theta^3}{\theta+1}\,\frac{(y+1)[2(\theta+2)+y]}{2(\theta+1)^{y+3}} \tag{2.14}$$

respectively.

It is of interest to note that the above procedure was used by Bakouch et al. (2022) to derive and extensively study a distribution named by them SPLG, with p.g.f. given by equation (2.13) and p.f. given, after some re-arranging, by equation (2.14) respectively.

Proposition 2.2. *Consider two independent r.v.'s Y and Y_3. The r.v. Y is a PGL with p.g.f. given by equation (2.4) and the r.v. Y_3 is geometric with p.g.f. given by equation (2.12). Then the distribution of the sum*

$$Y_4 = Y + Y_3$$

follows a generalized Poisson-Lindley distribution with parameters $\alpha > 1$, $\theta > 0$ and p.g.f.

$$G_{Y_4}(s)=\frac{\theta^{\alpha+1}}{\theta+1}\,\frac{\theta-s+2}{(\theta-s+1)^{\alpha+1}}$$

introduced by Mahmoudi and Zakerzadeh (2010).

Remark 2.1. When $\alpha = 2$ the PGL distribution with p.g.f. defined by equation (2.4), becomes the Poisson-Lindley model with p.g.f. given by equation (2.10).

Remark 2.2. As expected the addition of two Poisson-Lindley r.v.'s with the same parameter θ and p.g.f. given by equation (2.10) results to a distribution with p.g.f.

$$G_{X_1+X_1}=\frac{\theta^4}{(\theta+1)^2}\,\frac{(\theta-s+2)^2}{(\theta-s+1)^4}$$

which corresponds to the distribution obtained and studied by Chesneau et al. (2023).

3. AN ALTERNATIVE POISSON GENERALIZED LINDLEY (APGL) DISTRIBUTION

The main motivation of introducing and studying this distribution is that it appears as marginal distribution in bivariate versions of PGL models derived in Sections 4 and 5. This distribution is also of interest since it has simple forms for its p.g.f., p.f. and moments. Furthermore, it has the properties of overdispersion, monotonicity and increasing failure rate (IFR). In addition, it is derived by the different procedures of mixing, generalizing and addition of r.v.'s.

3.1. Derivation and properties

Consider a mixture of a Poisson distribution with parameter $\phi\lambda$ where $\phi > 0$ and λ is a r.v. following a new generalized Lindley distribution with p.d.f. and m.g.f. given by equations (2.1) and (2.2) respectively. Then the p.g.f. of the mixed r.v. Z is

$$\begin{aligned} G_Z(s) &= M_\Lambda(\theta(s-1)) \\ &= \frac{\theta^\alpha}{\theta+1}\,\frac{\theta+\phi-\phi s+1}{(\theta+\phi-\phi s)^\alpha}. \end{aligned} \tag{3.1}$$

Since

$$\frac{\partial^z G(s)}{\partial s^z} = \phi^z \frac{\theta^\alpha}{\theta+1}\,\frac{\Gamma(\alpha+z-1)}{\Gamma(\alpha)}\,\frac{(\alpha-1)(\theta+\phi-\phi s+1)+z}{(\theta+\phi-\phi s)^{\alpha+z}}, \tag{3.2}$$

we obtain

$$\begin{aligned} P(Z=z;\phi,\theta,\alpha) &= \frac{\phi^z}{z!}\,\frac{\theta^\alpha}{\theta+1}\,\frac{\Gamma(\alpha+z-1)}{\Gamma(\alpha)}\,\frac{(\alpha-1)(\theta+\phi+1)+z}{(\theta+\phi)^{\alpha+z}}, \\ &\qquad \alpha>1,\ \theta,\phi>0 \end{aligned} \tag{3.3}$$

and

$$\mu_{[r]:Z} = \phi^r \mu_{[r]:Y}. \tag{3.4}$$

Consequently,

$$E(Z) = \phi\frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)} \tag{3.5}$$

and

$$Var(Z) = \phi\frac{\phi((\alpha-1)\theta^2+2\alpha\theta+\alpha)+\theta(\theta+1)((\alpha-1)\theta+\alpha)}{\theta^2(\theta+1)^2}.$$

The dispersion index DI_Z is greater than one. This is expected, since from relation (3.4)

$$\begin{aligned} Var(Z) &= \phi^2[Var(Y)-E(Y)]+\phi E(Y) \\ &= c(Y)+E(Z) \quad \text{where} \quad c(Y)>0. \end{aligned}$$

Since Atikankul (2023) proved that $DI_Y > 1$, also $DI_Z > 1$. In fact,

$$DI_Z = \frac{\phi((\alpha-1)\theta^2+2\alpha\theta+\alpha)}{\theta(\theta+1)((\alpha-1)\theta+\alpha)}+1.$$

The probabilities can be calculated recursively by using the relation

$$P(Z=z+1) = \frac{\phi}{\theta+\phi}\,\frac{\alpha+z-1}{z+1}\,\frac{(\alpha-1)(\theta+\phi+1)+z+1}{(\alpha-1)(\theta+\phi+1)+z}P(Z=z)$$

with

$$P(Z=0) = G_Z(0) = \frac{\theta^\alpha}{\theta+1}\,\frac{\theta+\phi+1}{(\theta+\phi)^\alpha}.$$

The APGL distribution is unimodal and IFR. This can be proved by following an approach suggested by Johnson et al. (2005, p. 209) and used among others by Bhati et al. (2015) and Atikankul (2023).

The expression

$$\frac{P(Z=z+1)}{P(Z=z)}=\frac{\phi}{\theta+\phi}\,\frac{\alpha+z-1}{z+1}\left(1+\frac{1}{(\alpha-1)(\theta+\phi+1)+z}\right)$$

is clearly a decreasing function of z implying unimodality.

Furthermore, for $\alpha>2$ the ratio

$$\frac{P(Z=z+2)P(Z=z)}{P(Z=z+1)P(Z=z+1)}=\frac{z+1}{z+2}\,\frac{\alpha+z}{\alpha+z-1}\left[1-\frac{1}{[(\alpha-1)(\theta+\phi+1)+z+1]^2}\right]<1.$$

As such, $P(Z=z)$ is log-concave and hence the APGL distribution has an increased failure rate (IFR).

Remark 3.1. When $\alpha=2$, equation (3.1) becomes

$$G_Z(s)=\frac{\theta^2}{\theta+1}\,\frac{\theta+\phi-\phi s+1}{(\theta+\phi-\phi s)^2}, \tag{3.6}$$

which is the p.g.f. of a two-parameter Poisson-Lindley distribution discussed by Gómez-Dèniz et al. (2012). This distribution also appeared as marginal distribution of the bivariate version of the one-parameter Poisson-Lindley distribution introduced and studied by the above authors.

Remark 3.2. When $\phi=1$, as expected, the APGL distribution becomes the PGL distribution.

3.2. Additional derivations

Proposition 3.1. *Consider a r.v. Z_1 with p.g.f.*

$$E(s^{Z_1}\mid N=n)=(q+ps)^n,\qquad 0<p<1,\qquad q=1-p$$

and N is a non-negative inter-valued r.v. with p.f. $P(N=n)$ and p.g.f.

$$E(s^N)=h_N(s). \tag{3.7}$$

Then the p.g.f. of the r.v. Z_1 is

$$G_{Z_1}(s)=h_N(q+ps).$$

If N is distributed as a PGL distribution with p.g.f. given by equation (2.4) then

$$G_{Z_1}(s)=\frac{\theta^\alpha}{\theta+1}\,\frac{\theta+p-ps+1}{(\theta+p-ps)^\alpha} \tag{3.8}$$

which is the p.g.f. of an APGL distribution with parameters $\alpha>1$, $0<p<1$, $\theta>0$.

Remark 3.3. When $\alpha = 2$ equation (3.8) becomes

$$G_{Z_1}(s) = \frac{\theta^2}{\theta+1} \frac{\theta + p - ps + 1}{(\theta + p - ps)^2}. \tag{3.9}$$

This is the p.g.f. of a Binomial-discrete Poisson-Lindley distribution introduced and studied by Chesneau et al. (2022). However, it should be noted that the above distribution has also appeared as a marginal distribution in two of the bivariate Poisson-Lindley models developed by Papageorgiou and Vardaki (2022).

Proposition 3.2. *Consider two independent r.v.'s Z and Z_2. Let the r.v. Z follow a two-parameter Poisson-Lindley distribution with p.g.f. given by equation (3.6) and Z_2 be distributed as a negative binomial with parameters $\alpha > 2$, $\frac{\theta}{\theta+\phi}$ and p.g.f.*

$$G_{Z_2}(s) = \left(\frac{\theta}{\theta + \phi - \phi s}\right)^{\alpha-2}.$$

Then the r.v. $Z + Z_2$ follows a distribution with p.g.f.

$$G_{Z+Z_2}(s) = \frac{\theta^\alpha}{\theta+1} \frac{\theta + \phi - \phi s + 1}{(\theta + \phi - \phi s)^\alpha},$$

which is the p.g.f. of an APGL distribution with parameters, $\alpha > 2$, $\phi, \theta > 0$.

4. BIVARIATE PGL DISTRIBUTIONS DERIVED BY MIXING

A general family of mixed (compounded) bivariate Poisson distributions with p.g.f.

$$G_{X_1,X_2}(s_1, s_2) = \exp\{\lambda[\phi_1(s_1 - 1) + \phi_2(s_2 - 1) + \phi_{12}(s_1 s_2 - 1)]\}, \tag{4.1}$$

where $\phi_1, \phi_2, \phi_{12}$ are positive constants and λ is a r.v. with m.g.f. $M_\Lambda(\cdot)$ was examined by Kocherlakota and Kocherlakota (1992, Chapter 8). They proved that

$$G_{X_1,X_2}(s_1, s_2) = M_\Lambda[\phi_1(s_1 - 1) + \phi_2(s_2 - 1) + \phi_{12}(s_1 s_2 - 1)]$$

and this representation enabled them to derive several general properties.

Two simpler forms of the equation (4.1) are

$$G_{X_1,X_2}(s_1, s_2) = M_\Lambda[\phi_1(s_1 - 1) + \phi_2(s_2 - 1)] \tag{4.2}$$

and

$$G_{X_1,X_2}(s_1, s_2) = M_\Lambda[q(s_1 - 1) + p(s_1 s_2 - 1)], \tag{4.3}$$

where $0 < p < 1$ and $q = 1 - p$.

Equation (4.2) was utilized by Gómez-Dèniz et al. (2012) to derive a bivariate Poisson-Lindley distribution and by Irshad et al. (2024) to introduce a bivariate PGL model. Alternatively, Papageorgiou and Vardaki (2022) derived some general properties of the class of distributions with p.g.f. given by equation (4.3) and examined a number of illustrative examples.

4.1. A bivariate PGL distribution based on the structure (4.2)

This model with p.g.f.

$$G_{X_1,X_2}(s_1,s_2) = \frac{\theta^\alpha}{\theta+1}\,\frac{\theta+\phi_1+\phi_2-\phi_1 s_1-\phi_2 s_2+1}{(\phi+\phi_1+\phi_2-\phi_1 s_1-\phi_2 s_2)^\alpha} \tag{4.4}$$

was introduced and studied by Irshad et al. (2024).

The marginals X_1 and X_2 are AGPL distributions with parameters (ϕ_1,θ,α) and (ϕ_2,θ,α) respectively. Consequently, all characteristics of these distributions can be obtained from the corresponding characteristics of the APGL distribution examined in Section 3.

4.2. A bivariate PGL distribution based on the structure (4.3)

Assume that the m.g.f. of a r.v. Λ is given by equation (2.2). Then equation (4.3) becomes

$$G_{X_1,X_2}(s_1,s_2) = \frac{\theta^\alpha}{\theta+1}\,\frac{\theta-qs_1-ps_1s_2+2}{(\theta-qs_1-ps_1s_2+1)^\alpha}, \tag{4.5}$$

which is the p.g.f. of a bivariate PGL distribution with parameters $\alpha>1$, $0<p<1$, $\theta>0$.

The marginals X_1 and X_2 have p.g.f.'s

$$G_{X_1}(s_1) = \frac{\theta^\alpha}{\theta+1}\,\frac{\theta-s_1+2}{(\theta-s_1+1)^\alpha} \tag{4.6}$$

and

$$G_{X_2}(s_2) = \frac{\theta^\alpha}{\theta+1}\,\frac{\theta+p-ps_2+1}{(\theta+p-ps_2)^\alpha} \tag{4.7}$$

respectively. Consequently, the r.v. X_1, follows a PGL distribution documented in Section 2, but X_2 follows an APGL distribution extensively studied in Section 3. For a bivariate discrete r.v. (X_1,X_2) with p.g.f. $G_{X_1,X_2}(s_1,s_2)$ given by equation (4.5), the conditional p.g.f. $G_{X_2|X_1=x_1}(s)$ of X_2 on X_1 is

$$G_{X_2|X_1=x_1}(s) = (q+ps)^{x_1}.$$

Also, the conditional p.g.f. $G_{X_1|X_2=x_2}(s)$ of X_1 on X_2 is

$$G_{X_1|X_2=x_2}(s) = s^{x_2}\,\frac{(\alpha-1)(\theta-qs+1)+\alpha+x_2-1}{(\theta-qs+1)^{\alpha+x_2}}\,\frac{(\theta+p)^{\alpha+x_2}}{(\alpha-1)(\theta+p)+\alpha+x_2-1}. \tag{4.8}$$

Their derivation is based on the following result due to Subrahmaniam (1966).

For a bivariate discrete r.v. (X_1,X_2) with p.g.f. $G_{X_1,X_2}(s_1,s_2)$, the conditional p.g.f. of X_2 on X_1 is

$$G_{X_2|X_1=x_1}(s) = \frac{G^{(x_1,0)}(0,s)}{G^{x_1,0)}(0,1)} \tag{4.9}$$

where

$$G^{(x_1,x_2)}(u,v) = \frac{\partial^{x_1+x_2} G(s_1,s_2)}{\partial s_1^{x_1} \partial s_2^{x_2}}\Big|_{s_2=v}^{s_1=u}.$$

Since

$$P(X_2 \mid X_1 = x_1) = \binom{x_1}{x_2} p^{x_2} q^{x_1-x_2} \tag{4.10}$$

and the p.f. of the marginal distribution of X_1 is given by equation (2.6), we have

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2) &= \binom{x_1}{x_2} p^{x_2} q^{x_1-x_2} P(X_1 = x_1) \\ &= \frac{\theta^\alpha}{(\theta+1)} \frac{p^{x_2} q^{x_1-x_2}}{x_2!(x_1-x_2)!} \frac{\Gamma(\alpha+x_1-1)}{\Gamma(\alpha)} \frac{(\alpha-1)(\theta+2)+x_1}{(\theta+1)^{\alpha+x_1}}. \end{aligned}$$

The probabilities can be calculated using the recurrences

$$\begin{aligned} P(X_1 = x_1+1, X_2 = x_2) =& \frac{q}{\theta+1} \frac{\alpha+x_1-1}{x_1+1-x_2} \frac{(\alpha-1)(\theta+2)+x_1+1}{(\alpha-1)(\theta+2)+x_1} \\ & P(X_1 = x_1, X_2 = x_2), \quad x_1 = 0,1,\ldots, \quad x_2 = 0,1,\ldots,x_1, \end{aligned} \tag{4.11}$$

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2+1) = \frac{p}{q} \frac{x_1-x_2}{x_2+1} P(X_1 = x_2, X_2 = x_2), \\ x_1 = 1,2,\ldots, \quad x_2 = 0,1,\ldots,x_1-1. \end{aligned} \tag{4.12}$$

It is worth noting that this recurrence does not depend on the parameters α or θ. In addition,

$$P(X_1 = 0, X_2 = 0) = G(0,0) = \frac{\theta^\alpha}{\theta+1} \frac{\theta+2}{(\theta+1)^\alpha}. \tag{4.13}$$

Moments of the marginal distributions can be oftained from Sections 2 and 3. In particular, the marginal mean of the r.v. X_1 is given by equation (2.8) and

$$E(X_2) = p\frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)}. \tag{4.14}$$

Differentiating equation (4.5) we obtain

$$E(X_1 X_2) = p\frac{(\alpha-1)\theta^2+\alpha^2\theta+\alpha^2+\alpha}{\theta^2(\theta+1)}.$$

Hence,

$$Cov(X_1,X_2) = p\frac{(\alpha-1)\theta^3+(3\alpha-2)\theta^2+3\alpha\theta+\alpha}{\theta^2(\theta+1)}$$

which is positive.

The conditional expectation $E[X_1 \mid X_2 = x_2]$ is given by

$$E[X_1 \mid X_2 = x_2] = x_2 + \frac{q}{\theta+p} \frac{\alpha+x_2-1}{\alpha-1} \frac{(\alpha-1)(\theta+p)+\alpha+x_2}{(\alpha-1)(\theta+p)+\alpha+x_2-1}.$$

5. BIVARIATE PGL DISTRIBUTION DERIVED BY GENERALIZING

Bivariate Poisson-Lindley distributions derived by generalizing a bivariate binomial model with respect to its exponent were examined by Papageorgiou and Vardaki (2022). In this section we introduce and study a bivariate Poisson generalized Lindley distribution obtained by this technique.

5.1. Probabilities and Moments

Consider a bivariate binomial distribution with p.g.f.

$$E(s_1^{X_1} s_2^{X_2} \mid N = n) = (qs_1 + ps_2)^n, \qquad 0 < p < 1,$$

$q = 1 - p$, where N is a non-negative integer valued r.v. with p.g.f. $h_N^{(s)}$.

Consequently, the joint distribution of X_1 and X_2 is given by the p.g.f.

$$G_{X_1,X_2}(s_1, s_2) = h_N(qs_1 + ps_2). \tag{5.1}$$

If N follows a PGL distribution with p.g.f. given by equation (2.4) then

$$G_{X_1,X_2}(s_1, s_2) = \frac{\theta^\alpha}{\theta + 1} \frac{\theta - qs_1 - ps_2 + 2}{(\theta - qs_1 - ps_2 + 1)^\alpha}. \tag{5.2}$$

The p.g.f.'s of the marginals are

$$G_{X_1}(s_1) = \frac{\theta^\alpha}{\theta + 1} \frac{\theta + q - qs_1 + 1}{(\theta + q - qs_1)^\alpha}$$

and

$$G_{X_2}(s_2) = \frac{\theta^\alpha}{\theta + 1} \frac{\theta + p - ps_2 + 1}{(\theta + p - ps_2)^\alpha}.$$

Both marginals are APGL distributions with parameters $\alpha > 1$, $\theta > 0$, $0 < q < 1$ and $\alpha > 1$, $\theta > 0$, $0 < p < 1$ respectively. However, the p.g.f. of the distribution of the sum $X_1 + X_2$ given by

$$G_{X_1+X_2}(s) = \frac{\theta^\alpha}{\theta + 1} \frac{\theta - s + 2}{(\theta - s + 1)^\alpha}$$

is a PGL distribution with parameters $\alpha > 1$, $\theta > 0$.

From equation (5.1) we obtain

$$\frac{\partial^{x_1+x_2} G(s_1, s_2)}{\partial s_1^{x_1} \partial s_2^{x_2}} = q^{x_1} p^{x_2} h_N^{(x_1+x_2)}(qs_1 + ps_2). \tag{5.3}$$

Consequently,

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2) &= \frac{q^{x_1} p^{x_2}}{x_1! x_2!} h_N^{(x_1+x_2)}(0) \\ &= \binom{x_1 + x_2}{x_1} q^{x_1} p^{x_2} P(N = x_1 + x_2). \end{aligned}$$

Since N follows a PGL distribution, from equation (2.6)

$$P(X_1=x_1,X_2=x_2)=\frac{\theta^{\alpha}}{\theta+1}\,\frac{q^{x_1}p^{x_2}}{x_1!x_2!}\,\frac{\Gamma(\alpha+x_1+x_2-1)}{\Gamma(\alpha)}\,\frac{(\alpha-1)(\theta+2)+x_1+x_2}{(\theta+1)^{\alpha+x_1+x_2}},$$
$$x_i=0,1,\ldots,\quad i=1,2. \tag{5.4}$$

Furthermore from equation (5.3)

$$\mu_{[r,k]}=q^r p^k \mu_{[r+k]:N}$$

where

$$\mu_{[r,k]}=E(X_1^{(r)}X_2^{(k)}).$$

In addition, from equation (2.7)

$$\mu_{[r,k]}=q^r p^k\frac{\Gamma(\alpha+r+k-1)}{\Gamma(\alpha)}\,\frac{(\alpha-1)(\theta+1)+r+k}{\theta^{r+k}(\theta+1)},\qquad r,k=0,1,2,\ldots$$

with

$$E(X_1)=q\frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)}, \tag{5.5}$$

$$E(X_2)=p\frac{(\alpha-1)\theta+\alpha}{\theta(\theta+1)}, \tag{5.6}$$

$$E(X_1X_2)=\alpha pq\frac{(\alpha-1)(\theta+1)+2}{\theta^2(\theta+1)}$$

and

$$Cov(X_1,X_2)=pq\frac{(\alpha-1)\theta^2+2\alpha\theta+\alpha}{\theta^2(\theta+1)^2}$$

which is positive.

From equation (5.4) the following simple recurrences for the probabilities are obtained

$$P(X_1=x_1+1,X_2=x_2)=\frac{q}{\theta+1}\,\frac{\alpha+x_1+x_2-1}{x_1+1}\,\frac{(\alpha-1)(\theta+2)+x_1+x_2+1}{(\alpha-1)(\theta+2)+x_1+x_2}$$
$$P(X_1=x_1,X_2=x_2),\qquad x_i=0,1,\ldots,\qquad i=1,2 \tag{5.7}$$

and

$$P(X_1=x_1,X_2=x_2+1)=\frac{p}{\theta+1}\,\frac{\alpha+x_1+x_2-1}{x_2+1}\,\frac{(\alpha-1)(\theta+2)+x_1+x_2+1}{(\alpha-1)(\theta+2)+x_1+x_2}$$
$$P(X_1=x_1,X_2=x_2),\qquad x_i=0,1,\ldots,\qquad i=1,2 \tag{5.8}$$

with $P(X_1=0,X_2=0)$ given by equation (4.13).

5.2. Conditional probabilities and moments

From equations (5.1) and (4.9) we can obtain that

$$G_{X_2|X_1=x_1}(s)=\frac{h_N^{(x_1)}(ps)}{h_N^{(x_1)}(p)}. \tag{5.9}$$

Utilizing equation (2.5)

$$G_{X_2|X_1=x_1}(s) = \frac{(\alpha-1)(\theta-ps+2)+x_1}{(\theta-ps+1)^{\alpha+x_1}}\;\frac{(\theta+q)^{\alpha+x_1}}{(\alpha-1)(\theta+q+1)+x_1}.$$

Also, since

$$\frac{\partial^{x_2}}{\partial s^{x_2}} G^{(s)}_{X_2|X_1=x_1} = p^{x_2}\frac{h_N^{(x_1+x_2)}(ps)}{h_N^{(x_1)}(p)} \tag{5.10}$$

$$P(X_2=x_2 \mid X_1=x_1) = \frac{p^{x_2}}{x_2!}\;\frac{\Gamma(\alpha+x_1+x_2-1)}{\Gamma(\alpha+x_1-1)}\;\frac{(\alpha-1)(\theta+2)+x_1+x_2}{(\alpha-1)(\theta+q+1)+x_1}\;\frac{(\theta+q)^{\alpha+x_1}}{(\theta+1)^{\alpha+x_1+x_2}}.$$

These probabilities can be calculated recursively from the relation

$$P(X_2=x_2+1 \mid X_1=x_1) = \frac{p^{x_2}}{(x_1+1)}\;\frac{(\alpha+x_1+x_2-1)}{(\theta+1)}\;\frac{(\alpha-1)(\theta+2)+x_1+x_2+1}{(\alpha-1)(\theta+2)+x_1+x_2}$$
$$P(X_2=x_2 \mid X_1=x_1), \qquad x_i=0,1,\dots, \qquad i=1.2$$

with

$$P(X_2=0 \mid X_1=x_1) = \frac{(\alpha-1)(\theta+2)+x_1}{(\alpha-1)(\theta+q+1)+x_1}\;\frac{(\theta+q)^{\alpha+x_1}}{(\theta+1)^{\alpha+x_1}}.$$

Similar relations can be derived for $P(X_1 \mid X_2=x_2)$.

In addition from equation (5.10)

$$\mu_{[r|X_1=x_1]} = p^r\frac{h_N^{(x_1+r)}(p)}{h_N^{(x_1)}(p)}$$

and from equation (2.5)

$$\mu_{[r|X_1=x_1]} = \frac{p^r}{(\theta+q)^r}\;\frac{\Gamma(\alpha+x_1+r-1)}{\Gamma(\alpha+x_1-1)}\;\frac{(\alpha-1)(\theta+q+1)+x_1+r}{(\alpha-1)(\theta+q+1)+x_1}.$$

Consequently,

$$E[X_2 \mid X_1=x_1] = p\frac{\alpha+x_1-1}{\theta+q}\;\frac{(\alpha-1)(\theta+q+1)+x_1+1}{(\alpha-1)(\theta+q+1)+x_1}.$$

6. Bivariate PGL distributions derived by addition of r.v.'s

Since the univariate PGL distribution can be obtained as the convolution of a Poisson-Lindley model with a negative binomial distribution, bivariate PGL distributions can also be derived by addition of r.v.'s.

The technique of constructing bivariate models by adding r.v.'s is well known in the literature, see for example Holgate (1966); Papageorgiou (1986); Kocherlakota (1989); Kocherlakota and Kocherlakota (1992).

In this section we obtain bivariate PGL distributions generated by the general structure

$$R_1 = U_1 + V_1$$

$$R_2 = U_2 + V_2,$$

where (U_1, U_2) and (V_1, V_2) are independently distributed bivariate discrete r.v.'s. Then the p.g.f. of (R_1, R_2) is given by

$$G_{R_1,R_2}(s_1, s_2) = G_{U_1,U_2}(s_1, s_2)G_{V_1,V_2}(s_1, s_2).$$

This structure was initially suggested by Papageorgiou and Piperigou (1997) and also reported by Johnson et al. (1997, p. 138).

Proposition 6.1. *Let us assume that the bivariate r.v. (U_1, U_2) follows a bivariate Poisson-Lindley distribution with parameters $\theta > 0$, $\phi_1 > 0$, $\phi_2 > 0$ introduced by Gómez-Dèniz et al. (2012) and has p.g.f.*

$$G_{U_1,U_2}(s_1, s_2) = \frac{\theta^2}{\theta+1} \frac{\theta + \phi_1 + \phi_2 - \phi_1 s_1 - \phi_2 s_2 + 1}{(\theta + \phi_1 + \phi_2 - \phi_1 s_1 - \phi_2 s_2)^2}.$$

Also let the r.v. (V_1, V_2) follow a negative trinomial distribution with parameters $\alpha > 2$, $\frac{\phi_1}{\theta+\phi_1+\phi_2}$, $\frac{\phi_2}{\theta+\phi_1+\phi_2}$ and p.g.f. given by the equation

$$G_{V_1,V_2}(s_1, s_2) = \left(\frac{\theta}{\theta + \phi_1 + \phi_2 - \phi_1 s_1 - \phi_2 s_2}\right)^{\alpha-2}.$$

Then the p.g.f. of the r.v. (R_1, R_2) is

$$G_{R_1,R_2}(s_1, s_2) = \frac{\theta^\alpha}{\theta+1} \frac{\theta + \phi_1 + \phi_2 - \phi_1 s_1 - \phi_2 s_2 + 1}{(\theta + \phi_1 + \phi_2 - \phi_1 s_1 - \phi_2 s_2)^\alpha},$$

which is a bivariate PGL distribution with parameters $\alpha > 2$, $\theta, \phi_i > 0$, $i = 1, 2$ mentioned in Subsection 4.1.

Proposition 6.2. *Consider a bivariate Poisson-Lindley distribution with parameters $0 < p < 1$, $\theta > 0$ introduced by Papageorgiou and Vardaki (2022) and p.g.f. given by the equation*

$$G_{U_1,U_2}(s_1, s_2) = \frac{\theta^2}{\theta+1} \frac{\theta - qs_1 - ps_1s_2 + 2}{(\theta - qs_1 - ps_1s_2)^2}.$$

In addition, assume that (V_1, V_2) follows a negative binomial-Bernoulli model with parameters $\alpha > 2$, $\frac{q}{\theta+1}$, $\frac{p}{\theta+1}$ and p.g.f. given by the equation

$$G_{(V_1,V_2)}(s_1, s_2) = \left(\frac{\theta}{\theta - qs_1 - ps_1s_2 + 1}\right)^{\alpha-2}.$$

Models of this form were introduced by Cacoullos and Papageorgiou (1982). Then the p.g.f. of the r.v., (R_1, R_2) is

$$G_{R_1,R_2}(s_1, s_2) = \frac{\theta^\alpha}{\theta+1} \frac{\theta - qs_1 - ps_1s_2 + 2}{(\theta - qs_1 - ps_1s_2 + 1)^\alpha},$$

which is a bivariate PGL distribution with parameters $\alpha > 2$, $0 < p < 1$, $\theta > 0$ discussed in Subsection 4.2.

Proposition 6.3. *Assume that the r.v. (U_1, U_2) follows a distribution with p.g.f. given by equation (4.5) and (V_1, V_2) is distributed as a geometric-Bernoulli with p.g.f.*

$$G_{V_1,V_2}(s_1, s_2) = \frac{\theta}{\theta - qs_1 - ps_1 s_2}.$$

Then the p.g.f. of the r.v. (R_1, R_2) is given by the equation

$$G_{R_1,R_2}(s_1, s_2) = \frac{\theta^{\alpha+1}}{\theta + 1} \frac{\theta - qs_1 - ps_1 s_2 + 2}{(\theta - qs_1 - ps_1 s_2 + 1)^{\alpha+1}}$$

with parameters $\alpha > 1$, $0 < p < 1$, $\theta > 0$.

This distribution was introduced by Papageorgiou and Vardaki (2022) and it is a bivariate version of the generalized Poisson-Lindley distribution introduced and studied by Mahmoudi and Zakerzadeh (2010).

Proposition 6.4. *Suppose that (U_1, U_2) follows bivariate Poisson-Lindley distribution with parameters $0 < p < 1$, $\theta > 0$ introduced by Papageorgiou and Vardaki (2022) and p.g.f. given by the equation*

$$G_{U_1,U_2}(s_1, s_2) = \frac{\theta^2}{\theta + 1} \frac{\theta - qs_1 - ps_2 + 2}{(\theta - qs_1 - ps_2 + 1)^2}$$

and (V_1, V_2) follows a negative trinomial distribution with parameters $\alpha > 2$, $\frac{q}{\theta+1}$, $\frac{p}{\theta+1}$ and p.g.f.

$$G_{V_1,V_2}(s_1, s_2) = \left(\frac{\theta}{\theta - qs_1 - ps_2 + 1}\right)^{\alpha-2}.$$

Then the r.v. (R_1, R_2) has the p.g.f.

$$G_{R_1,R_2}(s_1, s_2) = \frac{\theta^{\alpha}}{\theta + 1} \frac{\theta - qs_1 - ps_2 + 2}{(\theta - qs_1 - ps_2 + 1)^{\alpha}},$$

which is a bivariate PGL distribution with parameters $\alpha > 2$, $0 < p < 1$, $\theta > 0$ studied in Section 5.

Proposition 6.5. *Assume that (U_1, U_2) follows a distribution with p.g.f. given by equation (5.2) and (V_1, V_2) is distributed as a bivariate geometric with p.g.f.*

$$G_{V_1,V_2}(s_1, s_2) = \frac{\theta}{\theta - qs_1 - ps_2 + 1}.$$

Then, the p.g.f. of the r.v. (R_1, R_2) is given by the equation

$$G_{R_1,R_2}(s_1, s_2) = \frac{\theta^{\alpha+1}}{\theta + 1} \frac{\theta - qs_1 - ps_2 + 2}{(\theta - qs_1 - ps_2 + 1)^{\alpha+1}}$$

with parameters $\alpha > 1$, $0 < p < 1$, $\theta > 0$.

This distribution was introduced and studied by Papageorgiou and Vardaki (2022) and it is another bivariate version of the generalized Poisson-Lindley distribution introduced by Mahmoudi and Zakerzadeh (2010).

7. EXAMPLES FROM AUTOMOBILE INSURANCE CLAIMS DATA

In this section we indicate possible applications of the bivariate PGL distributions introduced in Subsection 4.2 and Section 5 to real data sets previously reported in the literature. For this reason, we assumed that $\alpha = 3$ in agreement with the univariate PGL model developed by Bakouch et al. (2022). Furthermore, for $\alpha = 2$ the resulting model becomes a bivariate Poisson-Lindley distribution introduced by Papageorgiou and Vardaki (2022). Since there are structural differences between the models considered in Subsection 4.2 and the one in Section 5, we fitted them to different sets of data.

Both sets of data have positive correlation and originate from portfolios of automobile insurance claims. The first set reported by Gómez-Dèniz (2016) refers to the number of claims (variable X_1) and the size of claims (variable X_2) in an Australian portfolio containing 67856 one year insurance policies taken out in 2004 or 2005. The second set previously used, among others, by Partrat (1994) and Gómez-Dèniz et al. (2012) refers to claims corresponding to a large automobile insurance portfolio in France, including 181038 liability policies issued during the year 1989. The yearly claim frequencies have been divided into material damage and bodily injury, corresponding to variables X_1 and X_2.

7.1. Fitting bivariate PGL distributions derived in Subsection 4.2

For $\alpha = 3$ and θ replaced by θ_1 from equations (4.5), (2.8) and (4.14) the p.g.f. and the marginal means of the resulting distribution are

$$G_{X_1,X_2}(s_1, s_2) = \frac{\theta_1^3}{\theta_1 + 1} \frac{\theta_1 - qs_1 - ps_1s_2 + 2}{(\theta_1 - qs_1 - ps_1s_2 + 1)^3}, \tag{7.1}$$

$$E(X_1) = \frac{2\theta_1 + 3}{\theta_1(\theta_1 + 1)}. \tag{7.2}$$

and

$$E(X_2) = p\frac{2\theta_1 + 3}{\theta_1(\theta_1 + 1)}. \tag{7.3}$$

In addition, from equation (4.11)

$$P(X_1 = x_1 + 1, X_2 = x_2) = \frac{q}{\theta_1 + 1} \frac{x_1 + 2}{x_1 + 1 - x_2} \frac{2\theta_1 + x_1 + 5}{2\theta_1 + x_1 + 4} P(X_1 = x_1, X_2 = x_2), \quad x_1 = 0, 1, \ldots, \quad x_2 = 0, 1, \ldots, x_1. \tag{7.4}$$

The recurrence for $P(X_1 = x_1, X_2 = x_2 + 1)$, as expected, is given by equation (4.12). Also from equation (4.13) or from (7.1)

$$P(X_1 = 0, X_2 = 0) = \frac{\theta_1^3(\theta_1 + 2)}{(\theta_1 + 1)^4}. \tag{7.5}$$

For $\alpha = 2$ and θ replaced by θ_2, the corresponding characteristics of this bivariate model are

$$G_{X_1,X_2}(s_1, s_2) = \frac{\theta_2^2}{\theta_2 + 1} \frac{\theta_2 - qs_1 - ps_1s_2 + 2}{(\theta_2 - qs_1 - ps_1s_2 + 1)^2}, \tag{7.6}$$

$$E(X_1) = \frac{\theta_2 + 2}{\theta_2(\theta_2 + 1)}, \tag{7.7}$$

$$E(X_2) = p\frac{\theta_2 + 2}{\theta_2(\theta_2 + 1)}, \tag{7.8}$$

$$\begin{aligned} P(X_1 = x_1 + 1, X_2 = x_2) = \frac{q}{\theta_2 + 1} \frac{x_1 + 1}{x_1 + 1 - x_2} \frac{\theta_2 + x_1 + 3}{\theta_2 + x_1 + 2} P(X_1 = x_1, X_2 = x_2), \\ x_1 = 0, 1, \ldots, \quad x_2 = 0, 1, \ldots, x_1. \end{aligned} \tag{7.9}$$

$P(X_1 = x_1, X_2 = x_2 + 1)$ is given by equation (4.12) and

$$P(X_1 = 0, X_2 = 0) = \frac{\theta^2(\theta_2 + 2)}{(\theta_2 + 1)^3}. \tag{7.10}$$

Moment estimators are easily derived by using the marginal means. For both models

$$\tilde{p} = \frac{\bar{X}_2}{\bar{X}_1}.$$

However, estimators for the parameters θ_1 or θ_2 result as the solution of a quadratic equation. In particular, from equation (7.2)

$$\tilde{\theta}_1 = \frac{-(\bar{X}_1 - 2) + \sqrt{(\bar{X}_1 - 2)^2 + 12\bar{X}_1}}{2\bar{X}_1}.$$

Similarly, from equation (7.7)

$$\tilde{\theta}_2 = \frac{-(\bar{X}_1 - 1) + \sqrt{(\bar{X}_1 - 1)^2 + 8\bar{X}_1}}{2\bar{X}_1}.$$

Table 1 presents number of claims (X_1) and total number of claims with claim size larger than a threshold monetary value (X_2) $\psi = 500$ from a portfolio of 67856 automobile insurance policies; see Gómez-Dèniz (2016). The first line represents the observed frequencies and the second and third lines the expected frequencies for the models given by equation (7.1), $(\alpha = 3)$ and equation (7.6), $(\alpha = 2)$ respectively. Calculated parameter estimates and χ^2 values are also given.

Size of claims / Number of claims	0	1	2	3	4	Total
0	63232 63175.84 63252.78					63232 63175.84 63252.78
1	1840 1842.57 1783.02	2493 2592.71 2508.92				4333 4435.28 4291.94
2	37 40.29 50.10	117 113.40 140.99	117 79.78 99.20			271 233.47 290.29
3	1 0.78 1.40	5 3.31 5.93	5 4.66 8.34	7 2.18 3.91		18 10.93 19.58
4	0 0.01 0.04	0 0.08 0.22	1 0.17 0.47	0 0.16 0.44	1 0.06 0.15	2 0.48 1.32
Total	65110 65059.49 65087.34	2615 2709.50 2656.06	123 84.61 108.01	7 2.34 4.35	1 0.06 0.15	67856 67856 67855.91
For $\alpha = 3$ $\tilde{p} = 0.584566$ $\tilde{\theta}_1 = 27.963303$ $\chi^2(3) = 3.245$ For $\alpha = 2$ $\tilde{p} = 0.584566$ $\tilde{\theta}_2 = 14.624072$ $\chi^2(3) = 10.295$						

Table 1: Observed and expected frequencies for 65856 observations with claim size larger than 500.

To compute the bivariate probabilities for model (7.1), ($\alpha = 3$) we used the recurrences, see also equations (7.4) and (4.12)

$$P(X_1 = x_1 + 1, X_2 = 0) = \frac{q}{\theta_1 + 1} \frac{x_1 + 2}{x_1 + 1} \frac{2\theta_1 + x_1 + 5}{2\theta_1 + x_1 + 4} P(X_1 = x_1, X_2 = 0), \quad x_1 = 0, 1, 2, 3$$

and

$$P(X_1 = x_1, X_2 = x_2 + 1) = \frac{p}{q}\frac{x_1 - x_2}{x_2 + 1}P(X_1 = x_1, X_2 = x_2),$$
$$x_1 = 1, 2, 3, 4, \quad x_2 = 0, 1, 2, 3, \tag{7.11}$$

with $P(X_1 = 0, X_2 = 0)$ given by equation (7.5).

For model (7.6), ($\alpha = 2$), from equation (7.9),

$$P(X_1 = x_1 + 1, X_2 = 0) = \frac{q}{\theta_2 + 1}\frac{\theta_2 + x_1 + 3}{\theta_2 + x_1 + 2}P(X_1 = x_1, X_2 = 0),$$
$$x_1 = 0, 1, 2, 3.$$

The required relations $P(X_1 = x_1, X_2 = x_2 + 1)$ and $P(X_1 = 0, X_2 = 0)$ are given by equations (7.11) and (7.10) respectively.

Since $\bar{X}_1 = 0.072757$ and $\bar{X}_2 = 0.042531$ parameter estimates and χ^2 values are also reported in Table 1. To compute the chi squared test statistic, for this set of data, we follow an approach suggested by Gómez-Dèniz (2016). We grouped classes to produce a theoretical class of 5 or larger. The degrees of freedom of the relative χ^2 statistic obtained were $n_0 - k - 2$, where n_0 is the number of classes considered and k the number of parameters.

Seven categories were considered by grouping the classes $\{(1, 1), (2, 2), (3, 3)\}$, $\{(3, 0), (3, 1)\}$, $\{(3, 2), (4, 0), (4, 1), (4, 2), (4, 3), (4, 4)\}$. The calculated χ^2 values were $\chi^2 = 3.245$ (for $\alpha = 3$) and $\chi^2 = 10.295$ (for $\alpha = 2$) with 3 degrees of freedom. Since the 5% critical value of χ^2 with the same degrees of freedom is 7.815 we can assume that the model with p.g.f. given by equation (7.1), ($\alpha = 3$) provides a satisfactory fit to the set of insurance claims data under consideration.

7.2. Fitting the bivariate PGL distribution derived in Section 5

In the previous subsection we considered two versions of the bivariate PGL distribution with p.g.f. given by equation (4.5) for $\alpha = 3$ and $\alpha = 2$. In this subsection we consider the bivariate PGL distribution with p.g.f. given by equation (5.2) and we examine only the case for $\alpha = 3$. The corresponding version of this model for $\alpha = 2$, becomes a bivariate Poisson-Lindley distribution introduced by Papageorgiou and Vardaki (2022) and fitted to the same set of 181038 observations by Vardaki and Papageorgiou (2024).

For $\alpha = 3$ from equations (5.2), (5.5), (5.6), (5.7) and (5.8) the p.g.f., marginal means and recurrences for probabilities become

$$G_{X_1,X_2}(s_1, s_2) = \frac{\theta^3}{\theta + 1}\frac{\theta - qs_1 - ps_2 + 2}{(\theta - qs_1 - ps_2 + 1)^3} \tag{7.12}$$

$$E(X_1) = q\frac{2\theta + 3}{\theta(\theta + 1)} \tag{7.13}$$

$$E(X_2) = p\frac{2\theta + 3}{\theta(\theta + 1)} \tag{7.14}$$

$$P(X_1 = x_1 + 1, X_2 = x_2) = \frac{q}{\theta+1} \frac{x_1 + x_2 + 2}{x_1 + 1} \frac{2\theta + x_1 + x_2 + 5}{2\theta + x_1 + x_2 + 4}$$
$$P(X_1 = x_1, X_2 = x_2), \quad x_i = 0, 1, \ldots, \quad i = 1, 2, \tag{7.15}$$

$$P(X_1 = x_1, X_2 = x_2 + 1) = \frac{p}{\theta+1} \frac{x_1 + x_2 + 2}{x_2 + 1} \frac{2\theta + x_1 + x_2 + 5}{2\theta + x_1 + x_2 + 4}$$
$$P(X_1 = x_1, X_2 = x_2), \quad x_i = 0, 1, \ldots, \quad i = 1, 2. \tag{7.16}$$

We observe that

$$E(X_1 + X_2) = \frac{2\theta + 3}{\theta(\theta + 1)}. \tag{7.17}$$

Consequently, moment estimators are derived by using the marginal means. In particular from equation (7.17)

$$\tilde{\theta} = \frac{-(\bar{X}_1 + \bar{X}_2 - 2) + \sqrt{(\bar{X}_1 + \bar{X}_2 - 2)^2 + 12(\bar{X}_1 + \bar{X}_2)}}{2(\bar{X}_1 + \bar{X}_2)}$$

and from equations (7.13) and (7.14)

$$\tilde{p} = \frac{\bar{X}_2}{\bar{X}_1 + \bar{X}_2}.$$

Table 2 presents material damage (X_1) and bodily injury (X_2) claims from a portfolio of 181038 liability policies, see Partrat (1994); Gómez-Déniz (2012). The first line represents observed frequencies and the second one the expected frequencies for the model given by equation (7.12), ($\alpha = 3$). Calculated parameter estimates and χ^2 value are also included.

Bodily injury Material damage	0	1	≥ 2	Total
0	171345 171221.11	918 920.71	2 3.71	172265 172145.53
1	8273 8493.32	73 68.50	0 0.37	8346 8562.19
2	389 315.93	5 3.40	0 0.02	394 319.35
3	31 10.44	1 0.14	0 0.00	32 10.58
≥ 4	1 0.32	0 0.01	0 0.00	1 0.33
Total	180039 181041.12	997 992.76	2 4.10	181038 181037.98
$\alpha = 3 \quad \tilde{p} = 0.097802 \quad \tilde{\theta} = 35.856181 \quad \chi^2(3) = 5.265$				

Table 2: Material damage (X_1) and bodily injury (X_2) claims from a portfolio of 181038 liability policies.

To compute the bivariate probabilities we used the recurrence, see also equation (7.15)

$$P(X_1 = x_1 + 1, X_2 = 0) = \frac{q}{\theta + 1} \frac{x_1 + 2}{x_1 + 1} \frac{2\theta + x_1 + 5}{2\theta + x_1 + 4} P(X_1 = x_1, X_2 = 0), \quad x_1 = 0, 1, 2, 3.$$

In addition we use the recurrence (7.16) for $x_1 = 0, 1, \ldots, 4$ and $x_2 = 0, 1$. The probability $P(X_1 = 0, X_2 = 0)$ is given by equation (7.5).

Since $\bar{X}_1 = 0.051006$ and $\bar{X}_2 = 0.005529$ parameter estimates and the corresponding χ^2 value are given in Table 2.

Seven categories were considered for computing the χ^2 goodness of fit test by grouping the classes $\{(1,0),(2,0),(3,0)\}$ and $\{(1,2),(2,2),(3,1),(3,2),(4,0),(4,1),(4,2)\}$. The calculated χ^2 value was $\chi^2 = 5.265$ with 3 degrees of freedom (which is below 7.815, the 5% critical value of χ^2 with three degrees of freedom). Therefore, it appears that the model fits the data satisfactory.

8. ON DISTRIBUTIONS FITTED TO DATA SETS FROM AUTOMOBILE INSURANCE PORTFOLIOS CONSIDERED IN THE PREVIOUS SECTION

Our initial intention was simply to demonstrate the applicability of our models defined by equations (7.1), (7.6) and (7.12) by using real data sets. However, since the data sets selected, as well as alternative arrangements of them, were previously considered by a number of other researchers a comparison of our models with relative models can be of interest.

8.1. Models fitted to accident data reporting claim frequencies and size of claims

In automobile insurance, bonus-malus systems (BMS) are used to determine that fair premiums are paid by all policyholders.

Gómez-Dèniz (2016) initially and other authors subsequently (Gómez-Dèniz and Calderin-Ojeda, 2018, 2020; Moumeesri and Pongsart, 2022; Hernawati et al., 2017) proposed models derived by Bayesian methodoloty to determine premiums based on a BMS system that distinguished two types of claims. Those less than a threshold or critical monetary value denoted by ψ and those that exceed this value. They used a portfolio of 67856 one year automobile insurance policies taken out in 2004 or 2005. The data set is available at the website of the Faculty of Business and Economics, Macquaire University (Sydney, Australia), see also, De-Jong and Heller (2008). Out of the 67856 policies in that portfolio, 4624 claims were made. There were 4333 policyholders who made claims once, 271 twice, 18 three times and 2 four times.

Table 1 records these data when the threshold value is $\psi = 500$. Gómez-Dèniz (2016) fitted these data by using a four parameter Poisson gamma-binomial beta distribution. He reported a satisfactory fit after grouping cells to satisfy the rule of five. The same set of data was fitted by Moumeesri and Pongsart (2022) utilizing two bivariate models, each one with three parameters, a Poisson exponential-binomial beta and a Poisson-Lindley-binomial beta. In their calculation of the χ^2 goodness of fit tests they did not report any grouping of the expected frequencies. They concluded that the Poisson-Lindley-binomial beta distribution gave a better fit to the data.

For threshold values $\psi = 1000$ and $\psi = 3000$ Gómez-Dèniz and Calderin-Ojeda (2020) fitted both the resulting data set tables by two bivariate distributions. A two parameter bivariate Poisson type model and from this distribution by using Bayesian methodology they obtained a four parameter bivariate Poisson gamma-binomial beta distribution similar to the model used by Gómez-Dèniz (2016). Combining cells to comply with the rule of five, they reported that although their Poisson gamma-binomial beta distribution fitted both sets of data adequately the corresponding fits of the bivariate Poisson models can not be considered as satisfactory. For comparison purposes we have fitted our model considered in Subsection 7.1 to both sets of data given by Gómez-Dèniz and Calderin-Ojeda (2020).

Tables 3 and 4 present the number of claim (X_1) from a portfolio of 67856 automobile insurance policies (Gómez-Dèniz and Calderin-Ojeda, 2020) when the claim value is larger than a threshold monetary value (X_2) $\psi = 1000$ and $\psi = 3000$ respectively. The first line represents the observed frequencies and the second and third lines the expected frequencies for the models given by equation (7.1), ($\alpha = 3$) and equation (7.6), ($\alpha = 2$) respectively. Calculated parameter estimates and χ^2 values are also given.

To comply with the rule of five for computing the χ^2 goodness of fit test, eight categories were considered by grouping the classes $\{(1,1),(2,1)\}$ and $\{(3,2),(3,3),(4,0),(4,1),(4,2),(4,3),(4,4)\}$. Table 4 presents the corresponding results when $\psi = 3000$. In this case we have used seven categories by grouping the classes $\{(1,1),(2,1)\}$ and $\{(3,1),(3,2),(3,3),(4,0),(4,1),(4,2),(4,3),(4,4)\}$. For the bivariate Poisson-Lindley distribution defined by model (7.6), ($\alpha = 2$) the calculated χ^2 values were for $\psi = 1000$ $\chi^2 = 1.542$ with three degrees of freedom and for $\psi = 3000$ $\chi^2 = 2.443$ with four degrees of freedom (which are below 9.488 the 5% critical value of the χ^2 with four degrees of freedom). Therefore, it appears that the bivariate Poisson-Lindley distribution fitted both sets of data satisfactory. For $\psi = 274$ and $\psi = 3755$ tables of the derived data sets were given by Hernawati et al. (2017). They fitted both sets of data using a three parameter bivariate Poisson exponential-binomial beta distribution. They did not report any groupings in the calculation of the χ^2 goodness of fit tests. It appears that their model gave a better fit for the 274 observations.

Results on trivariate extensions and multiple threshold values were given by Gómez-Dèniz and Calderin-Ojeda (2018).

8.2. Models fitted to accident data reporting claim frequencies on material damage and bodily injury

The set of data recorded in Table 2, consists of claims experience of a large automobile portfolio in France including 181038 liability policies observed during the year 1989. The corresponding yearly claim frequencies have been divided into material damage and bodily injury claims. This set of data was used by several authors to illustrate the applicability of various bivariate models.

Size of claims / Number of claims	0	1	2	3	4	Total
0	63232 63175.84 63252.78					63232 63175.84 63252.78
1	2551 2624.15 2539.35	1782 1811.12 1752.60				4333 4435.27 4291.95
2	109 81.73 101.63	114 112.81 140.28	48 38.93 48.41			271 233.47 290.32
3	5 2.26 4.06	6 4.69 8.40	6 3.23 5.80	1 0.74 1.29		18 10.92 19.55
4	1 0.06 0.16	0 0.16 0.45	0 0.69 0.46	1 0.08 0.21	0 0.01 0.04	2 1.00 1.32
Total	65898 65884.04 65897.98	1902 1928.78 1901.73	54 42.85 54.67	2 0.82 1.50	0 0.01 0.04	67856 67856.50 67855.92
For $\alpha = 3$ $\tilde{p} = 0.408345$ $\tilde{\theta}_1 = 27.963303$ $\chi^2(4) = 20.662$ For $\alpha = 2$ $\tilde{p} = 0.408345$ $\tilde{\theta}_2 = 14.624072$ $\chi^2(4) = 1.542$						

Table 3: Observed and expected frequencies for 67856 observations with claim size larger than 1000.

Size of claims / Number of claims	0	1	2	3	4	Total
0	63232 63175.84 63252.78					63232 63175.84 63252.78
1	3576 3688.72 3569.52	757 746.37 722.42				4333 4435.09 4291.94
2	216 161.49 200.79	44 65.35 54.75	11 6.61 8.22			271 233.45 290.28
3	12 6.28 11.26	4 3.81 6.84	2 0.77 1.38	0 0.05 0.09		18 10.91 19.57
4	2 0.23 0.63	0 0.19 0.51	0 0.06 0.16	0 0.01 0.02	0 0.00 0.00	2 0.49 1.32
Total	67038 67032.56 67034.98	805 815.72 811.04	13 7.44 9.76	0 0.06 0.11	0 0.00 0.00	67856 67855.78 67855.89

For $\alpha = 3$ $\tilde{p} = 0.168321$ $\tilde{\theta}_1 = 27.963303$ $\chi^2(3) = 31.778$
For $\alpha = 2$ $\tilde{p} = 0.168321$ $\tilde{\theta}_2 = 14.624072$ $\chi^2(3) = 2.443$

Table 4: Observed and expected frequencies for 67856 observations with claim size larger than 3000.

Partrat (1994) initially fitted a standard bivariate Poisson, a bivariate Poisson-Gamma and a bivariate Poisson-inverse Gaussian distribution, each with three parameters, to this set of data. He concluded that the standard bivariate Poisson model should be rejected but the other two models gave a satisfactory fit, with the bivariate Poisson-inverse Gaussian giving a better fit. Gómez-Dèniz et al. (2012) fitted a three parameter bivariate Poisson-Lindley distribution and a three parameter bivariate Sarmanov model. They suggested that both models performed better than the bivariate Poisson-Gamma and the bivariate Poisson-inverse

Gaussian considered by Partrat (1994).

Vernic (1997) fitted the same set of data with a six parameter bivariate generalized Poisson model and Diafouka et al. (2022) with a five parameter bivariate Katz's distribution. Finally, Vardaki and Papageorgiou (2024) used a two parameter bivariate Poisson-Lindley distribution which gave a satisfactory fit.

It is worth noting that only Vardaki and Papageorgiou (2024) used the same grouping categories suggested by Gómez-Dèniz and Calderin-Ojeda (2020). All other authors used different grouping categories.

9. Conclusions

In this paper using p.g.f.'s, five univariate and bivariate PGL distributions, with two of them already appeared in the literature, were derived by addition of r.v.'s. Similar models like the univariate generalized Poisson-Lindley distribution introduced by Mahmoudi and Zakerzadeh (2010), as well as its bivariate versions can be also obtained by adding r.v.'s. We also introduced a univariate APGL distribution which appears as marginal distribution in all bivariate models examined and it also has the attractive properties of over-dispersion, monotonicity and IFR. In addition, we considered two bivariate PGL distributions derived one by mixing and the other by generalization; a variety of properties including various recurrences and conditional expectations were obtained. Finally, applications of our bivariate distributions to various sets of automobile insurance claims data were given. It appears that our models are appropriate for fitting bivariate data when there is a large number of observations in the $(0, 0)$ cell. Furthermore, it is possible worth examining, if alternative bivariate models introduced by Papageorgiou and Vardaki (2022) can be useful in interpreting these types of data.

ACKNOWLEDGMENTS

We thank both referees for their comments and suggestions. We are particularly grateful to the referee who suggested the line of research of Section 8.

REFERENCES

Abouammoh, A. M., Alshangiti, A. M., and Ragab, I. (2015). A new generalized Lindley distribution. *Journal of Statistical Computation and Simulation*, 85(18):3662–3678.

Atikankul, Y. (2023). A new Poisson generalized Lindley regression model. *Austrian Journal of Statistics*, 52(1):39–50.

Bakouch, H. S., Karakaya, K., Chesneau, C., and Adkŏgan, Y. (2022). A non-negative integer-valued model: estimation, count regression and practical examples. *Applicable Analysis and Discrete Mathematics*, 16(2):467–484.

Bhati, D., Sastry, D. V. S., and Quadri, P. Z. M. (2015). A new generalized Poisson-Lindley distribution: applications and properties. *Austrian Journal of Statistics*, 44(1):35–51.

Cacoullos, T. and Papageorgiou, H. (1982). Bivariate negative binomial-poisson and negative binomial-bernoulli models with an application to accident data. In Kallianpur, G., Krishnaiah, P. R., and Ghosh, J. K., editors, *Statistics and Probability: Essays in Honor of CR Rao*, pages 155–168. North-Holland Publishing Company.

Chesneau, C., Bakouch, H. S., Akdoğan, Y., and Karakaya, K. (2022). The binomial-discrete Poisson-Lindley model: modelling and applications to count regression. *Communications In Mathematical Research*, 38(1):28–51.

Chesneau, C., D'cruz, V., Maya, R., and Irshad, M. R. (2023). A novel discrete distribution based on the mixture of Poisson and sum of two Lindley random variable: Accepted - september 2023. *REVSTAT - Statistical Journal*.

DeJong, P. and Heller, G. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press.

Diafouka, M. K., Louzayadio, C. G., Malouata, R. O., Ngabassaka, N. R., and Bidounga, R. (2022). On a bivariate Katz's distribution. *Advances In Mathematics: Scientific Journal*, 11(10):955–968.

Gómez-Dèniz, E. (2016). Bivariate credibility bonus-malus premiums distinguishing between two types of claims. *Insurance: Mathematics and Economics*, 70:117–124.

Gómez-Dèniz, E. and Calderin-Ojeda, E. (2018). Multivariate credibility in bonus-malus systems distinguishing between different types of claims. *Risks*, 6(2):34.

Gómez-Dèniz, E. and Calderin-Ojeda, E. (2020). A survey of the individual claim size and other risk factors using credibility bonus-malus premiums. *Risks*, 8(1):20.

Gómez-Dèniz, E., Sarabia, J. M., and Balakrishnan, N. (2012). A multivariate discrete Poisson-Lindley distribution: extensions and actuarial applications. *Astin Bulletin*, 42(2):655–678.

Hernawati, L., Purnaba, I. G. P., and Mangku, I. W. (2017). Comparing bonus malus premiums of two types claim which is assessed using Bayesian method. *Applied Mathematical Sciences*, 11(34):1687–1694.

Holgate, P. (1966). Bivariate generalizations of Neyman's type A distribution. *Biometrika*, 53(1-2):241–244.

Irshad, M. R., D'cruz, V., Maya, R., and Khan, N. M. (2023). Inferential properties with a novel two parameter Poisson generalized Lindley distribution with regression and application to INAR (1) process. *Journal of Biopharmaceutical Statistics*, 70(2):117–124.

Irshad, M. R., D'cruz, V., Maya, R., Soobhug, A. D., and Khan, N. M. (2024). Bivariate Poisson generalized Lindley distributions and the associated BINAR (1) processes. *Austrian Journal of Statistics*, 53(1):1–28.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. John Wiley and Sons, New York.

Johnson, N. L., Kotz, S., and Kemp, A. W. (2005). *Univariate Discrete Distributions, 2nd Edition*. John Wiley and Sons, New York.

Kocherlakota, S. (1989). A note on the bivariate binomial distribution. *Statistics and Probability Letters*, 7(2):21–24.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Marcel Dekker Inc., New York.

Mahmoudi, E. and Zakerzadeh, H. (2010). Generalized Poisson-Lindley distribution. *Communications in Statistics-Theory and Methods*, 39(10):1785–1798.

Moumeesri, A. and Pongsart, T. (2022). Bonus-malus premiums based on claim frequency and size of claims. *Risks*, 20(9):181.

Papageorgiou, H. (1986). Bivariate "short" distributions. *Communications in Statistics-Theory and Methods*, 15(2):893–905.

Papageorgiou, H. and David, K. M. (1994). On countable mixtures of bivariate binomial distributions. *Biometrical Journal*, 36(5):581–601.

Papageorgiou, H. and Piperigou, V. E. (1997). On bivariate "short" and related distributions. In Johnson, N. L. and Balakrishnan, N., editors, *Advances in the Theory and Practice of Statistics: A Volume in Honor of Samule Kotz*, pages 397–413. John Wiley and Sons, New York.

Papageorgiou, H. and Vardaki, M. (2022). Bivariate Poisson-Lindley distributions. *Journal of Statistical Theory and Practice*, 16:30.

Partrat, C. (1994). Compound model for two dependent kinds of claim. *Insurance: Mathematics and Economics*, 15:219–231.

Rao, B. R., Mazumdar, S., Waller, J. H., and Li, C. C. (1973). Correlation between the number of two types of children in a family. *Biometrics*, 29:271–279.

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

Subrahmaniam, K. (1966). A test of "intrinsic correlation" in the theory of accident proneness. *Journal of the Royal Statistical Society, Series B*, 28:180–189.

Vardaki, M. and Papageorgiou, H. (2024). Multivariate discrete Poisson-Lindley distributions. *Journal of Statistical Theory and Practice*, 18:33.

Vernic, R. (1997). On the bivariate generalized Poisson distribution. *Astin Bulletin*, 27(1):23–32.