THE PARALOGISTIC-NEGATIVE BINOMIAL DISTRIBUTION AND ITS APPLICATIONS

Sirithip Wasinrat – Boonyarit Choopradit

Abstract

The objectives of this research were fourfold: 1) to develop a model that combined the paralogistic distribution with the negative binomial distribution, 2) to obtain the proposed model's statistical properties including survival function, hazard function, quantile function, and likelihood function, 3) to use the maximum likelihood method to estimate model parameters, and 4) to apply the model to two insurance datasets. Results demonstrated the effectiveness of the new four-parameter model, named the paralogistic-negative binomial distribution, which showed more flexibility when modeling lifetime data. The proposed model's statistical properties including the survival function, hazard function, quantile function, and likelihood function were derived. Model parameters were estimated using the maximum likelihood method, and the observed information matrix was obtained. A simulation study was carried out to investigate the accuracy of model parameter estimation. Two insurance datasets were used to compare the proposed model's flexibility to other traditional lifetime models and the new distribution proved to be more accurate.

Keywords: mixture distribution, quantile function, maximum likelihood method

JEL Code: C19, C46

Introduction

The paralogistic (PL) continuous distribution function is well known in insurance and economic applications (Kleiber & Kotz, 2003), with probability density function (PDF) of a random variable X

$$f_{PL}(x) = \frac{\alpha^2 \left(x/\theta\right)^{\alpha}}{x\left[\left(x/\theta\right)^{\alpha}+1\right]^{\alpha+1}}, \quad x>0, \theta>0, \alpha>0. \tag{1}$$

The corresponding cumulative distribution function (CDF) can be represented as**:**

$$F_{PL}(x) = 1-\left(\frac{1}{\left(x/\theta\right)^{\alpha}+1}\right)^{\alpha}, \quad x>0, \theta>0, \alpha>0. \tag{2}$$

Where $\theta>0$ is a scale parameter and $\alpha>0$ is a shape parameter.

Previous researchers expanded the PL distribution. Idemudia & Ekhosuehi (2019) proposed a new PL distribution with three parameters by adding the location parameter, while Bhati et al. (2019) extended the PL distribution, called PL-MPLG distribution, with four parameters. Continuous development of new distribution models has evolved to match the increasing diversity of data and changing circumstances. A method for combined distribution as the finite mixture distribution was presented by Hall & Zhou (2003) and Balakrishnan et al. (2009), while the infinite mixture distribution was suggested by Gómez-Déniz et al. (2008) and Withers & Nadarajah (2011). The negative binomial (NB) distribution is also used in mixture distribution. Rodrigues et al. (2011) proposed the Weibull NB distribution, while Ortega et al. (2012) proposed the NB-beta Weibull distribution, Yusuf et al. (2016) proposed the inverse Burr NB distribution, and Zubair et al. (2018) proposed the power-Cauchy NB distribution.

This paper proposed a new model that combined the PL and NB distributions and presented the survival function, hazard function, and quantile function mathematical properties. Model parameters were estimated using the maximum likelihood method and a simulation study was performed. This new distribution model demonstrated potential as flexibility when dealing with two real datasets.

1 Mixture distributions

1.1 The G-negative binomial distribution

Percontini et al. (2013) proposed a general family of continuous distributions called the G-negative binomial (G-NB) family. Let Z be a random variable from zero truncated NB (ZTNB) probability mass function (PMF) with parameters $s>0$ and $\beta\in(0,1)$ given by

$$P(z;s,\beta)=\beta^{z}\binom{z+s-1}{z}\left[(1-\beta)^{-s}-1\right]^{-1},\mathrm{z}\in\mathrm{N}$$

Let $W_1,W_2,...,W_Z$ be a random sample from any density function $g(x)$ where Z and W are independent random variables. Let $X=\min(W_1,W_2,...,W_Z)$, then the conditional CDF of X given Z is $F(x\mid z)=1-P(X\geq x\mid z)=1-\left[1-P(W_1\leq x)\right]^{z}=1-\left[1-G(x)\right]^{z}$.

The unconditional CDF of X becomes

$$F(x)=\sum_{z=0}^{\infty}\beta^{z}\binom{s+z-1}{z}\left\{(1-\beta)^{-s}-1\right\}^{-1}\left\{1-\left[1-G(x)\right]^{z}\right\},$$

for $x>0$, s and β are shape parameters. Then, the CDF of X reduces to

$$F(x)=\left[(1-\beta)^{-s}-\left\{1-\beta[1-G(x)]\right\}^{-s}\right]\left[(1-\beta)^{-s}-1\right]^{-1} \tag{3}$$

The PDF corresponding to (3) is given by

$$f(x)=s\beta\cdot g(x)\left\{1-\beta\left[1-G(x)\right]\right\}^{-s-1}\left[(1-\beta)^{-s}-1\right]^{-1} \tag{4}$$

The survival function and hazard rate function of X are given by

$$S(x)=\left[\left\{1-\beta\left[1-G(x)\right]\right\}^{-s}-1\right]\left[(1-\beta)^{-s}-1\right]^{-1} \tag{5}$$

and
$$h(x)=\left[s\beta\cdot g(x)\left\{1-\beta\left[1-G(x)\right]\right\}^{-s-1}\right]\left[\left\{1-\beta\left[1-G(x)\right]\right\}^{-s}-1\right]^{-1} \tag{6}$$

The random variable X following the family (3) - (6) follows a G-NB distribution. The G distribution is a sub-model when $s=1$ and $\beta\to 0$. This generalization is obtained by increasing the number of parameters compared to the G distribution or $g(x)$, thereby adding more flexibility to the generated distribution (Percontini et al., 2013).

1.2 The paralogistic-negative binomial distribution

A new mixture distribution of the PL distribution and NB distribution using the G-NB distribution was proposed in Section 1.1

Definition 1. Let $X>0$ be a random variable of paralogistic-negative binomial (PLNB) distribution with parameters $s, \beta, \theta,$ and α denoted as $X \sim PLNB(s,\beta,\theta,\alpha)$ with $s>0$, $0<\beta<1$, $\theta>0$ and $\alpha>0$.

Theorem 1. Let $X \sim PLNB(s,\beta,\theta,\alpha)$, then the PLNB's CDF is

$$F(x)=\left[(1-\beta)^{-s}-\left\{1-\beta\left[\left(\frac{1}{(x/\theta)^{\alpha}+1}\right)^{\alpha}\right]\right\}^{-s}\right]\left[(1-\beta)^{-s}-1\right]^{-1} \tag{7}$$

Proof: In Equation (3) substituting $G(x)$ with Equation (2) completes the proof of the theorem.

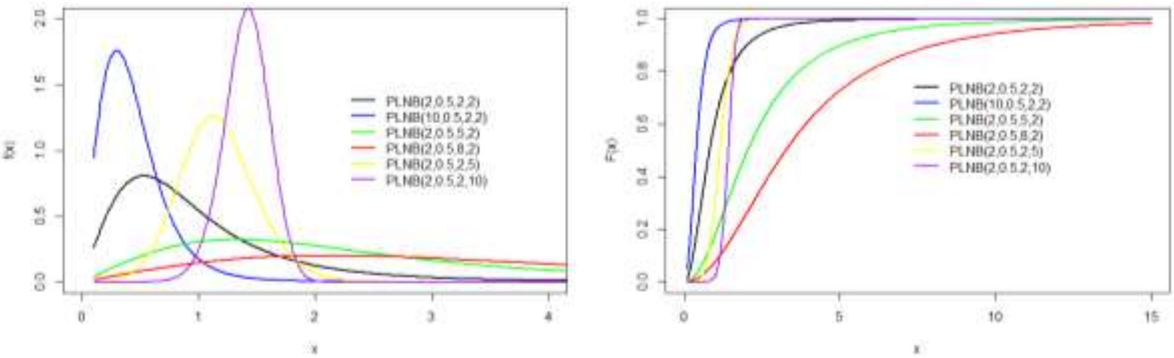
Theorem 2. Let $X \sim PLNB(s,\beta,\theta,\alpha)$, then the PLNB's PDF is

$$f(x)=\frac{s\beta\alpha^{2}(x/\theta)^{\alpha}}{x\left[(1-\beta)^{-s}-1\right]\left[(x/\theta)^{\alpha}+1\right]^{\alpha+1}}\left\{1-\beta\left[\left(\frac{1}{(x/\theta)^{\alpha}+1}\right)^{\alpha}\right]\right\}^{-s-1} \tag{8}$$

Proof: In Equation (4) substituting $g(x)$ and $G(x)$ with Equations (1) and (2), respectively completes the proof of the theorem.

Fig. 1 displays the PDF and CDF curves for the PLNB distribution with selected values for the parameters.

Fig. 1: The PLNB's PDF and CDF curves with some specified parameter values



Source: Own research

2 Some properties of the PLNB distribution

2.1 Survival function and hazard function

Theorem 3. Let $X \sim PLNB(s,\beta,\theta,\alpha)$, then the PLNB's survival function is

$$S(x)=\left[\left\{1-\beta\left[\left(\frac{1}{(x/\theta)^{\alpha}+1}\right)^{\alpha}\right]\right\}^{-s}-1\right]\left[(1-\beta)^{-s}-1\right]^{-1}$$

Proof: In Equation (5) substituting $G(x)$ with Equation (2) completes the proof of the theorem.

Theorem 4. Let $X \sim PLNB(s,\beta,\theta,\alpha)$, then the PLNB' s hazard function is

$$h(x)=s\beta\alpha^{2}(x/\theta)^{\alpha}\left\{1-\beta\left[\left(\frac{1}{(x/\theta)^{\alpha}+1}\right)\right]\right\}^{-s-1}x^{-1}\left[(x/\theta)^{\alpha}+1\right]^{-(\alpha+1)}\left[\left\{1-\beta\left[\left(\frac{1}{(x/\theta)^{\alpha}+1}\right)\right]\right\}^{-s}-1\right]^{-1}$$

Proof: In Equation (6) substituting $g(x)$ and $G(x)$ with Equations (1) and (2), respectively completes the proof of the theorem.

2.2 Quantile function

Let $X \sim PLNB(s,\beta,\theta,\alpha)$, then the quantile function (QF) is denoted by $Q(p)$ and $Q(p)=F^{-1}(p)$, where $p\in(0,1)$.

Theorem 5. Let $X \sim PLNB(s,\beta,\theta,\alpha)$, then the QF of X is given as:

$$Q(p)=\theta\left[\left[\frac{1-\left\{(1-\beta)^{-s}-p\left[(1-\beta)^{-s}-1\right]\right\}^{-\frac{1}{s}}}{\beta}\right]^{-\frac{1}{\alpha}}-1\right]^{\frac{1}{\alpha}}$$

Proof. Since $Q(p)=F^{-1}(p)$, $p\in(0,1)$. This implies that $F(Q(p))=p$. As $X \sim PLNB(s,\beta,\theta,\alpha)$, or the CDF of X as noted in Equation (7), we can solve for $Q(p)$.

3 Parameter estimation of the PLNB distribution

The parameters of the PLNB were estimated using the maximum likelihood estimation (MLE) function as follows:

$$L(s,\beta,\theta,\alpha)=\prod_{i=1}^{n}\frac{s\beta\alpha^{2}(x_i/\theta)^{\alpha}}{x_i\left[(1-\beta)^{-s}-1\right]\left[(x_i/\theta)^{\alpha}+1\right]^{\alpha+1}}\left\{1-\beta\left[\left(\frac{1}{(x_i/\theta)^{\alpha}+1}\right)^{\alpha}\right]\right\}^{-s-1}$$

The log-likelihood function of the above expression is given by

$$\begin{aligned}\ln L(s,\beta,\theta,\alpha)&=n\ln(s)+n\ln(\beta)+2n\ln(\alpha)+\alpha\sum_{i=1}^{n}\ln(x_i/\theta)\\&+(-s-1)\sum_{i=1}^{n}\ln\left\{1-\beta\left[\left(\frac{1}{(x_i/\theta)^{\alpha}+1}\right)^{\alpha}\right]\right\}-\sum_{i=1}^{n}\ln(x_i)\\&-n\ln\left[(1-\beta)^{-s}-1\right]-(\alpha+1)\sum_{i=1}^{n}\ln\left[(x_i/\theta)^{\alpha}+1\right]\end{aligned}$$

The MLE solutions for s, β, θ and α can be obtained by simultaneously solving the resulting equations as follow:

$$\frac{\partial\ln L(s,\beta,\theta,\alpha)}{\partial s}=0,\quad \frac{\partial\ln L(s,\beta,\theta,\alpha)}{\partial \beta}=0,\quad \frac{\partial\ln L(s,\beta,\theta,\alpha)}{\partial \theta}=0 \text{ and } \frac{\partial\ln L(s,\beta,\theta,\alpha)}{\partial \alpha}=0.$$

A numerical procedure is used, such as the Newton-Raphson method. In this study, the MLE estimates of $\hat{s}$, $\hat{\beta}$, $\hat{\theta}$ and $\hat{\alpha}$ were obtained using the "mledist" function in the R software "fitdistrplus" package suite (Delignette-Muller & Dutang, 2015).

4 Simulation study

A simulation study was conducted to assess the effectiveness of the MLE of the parameters s, β, θ and α in the previous section. The estimates of s,β,θ and α were obtained using the "mledist" function in the R software "fitdistrplus" package suite (Delignette-Muller & Dutang, 2015).

The study was based on 2,000 simulated samples from the PLNB with different sample sizes: n = 25, 50, and 200. Random variables were generated from the PLNB using the inverse of the distribution function. Consider the identity $F(X) = U \Rightarrow X = F^{-1}(U)$ where U is the standard uniform distribution, or the uniform (0, 1). Let $X \sim PLNB(s,\beta,\theta,\alpha)$, then the random variable can be generated from

$$X = \theta\left[\left[\frac{1-\left\{(1-\beta)^{-s} - U\left[(1-\beta)^{-s} - 1\right]\right\}^{-\frac{1}{s}}}{\beta}\right]^{-\frac{1}{\alpha}} - 1\right]^{\frac{1}{\alpha}}.$$

Tab. 1: Mean estimates, standard deviation and root mean squared errors of s, β, θ and α

Distribution	n	Parameter	Mean estimate	SD	Bias	RMSE
PLNB (2,0.5,2,2)	25	s	0.9257	1.7671	-1.0742	2.0673
		β	0.6791	0.3405	0.1791	0.3846
		θ	2.4735	1.0658	0.4735	1.1658
		α	2.2431	0.3930	0.2431	0.4620
	50	s	1.0343	1.7407	-0.9656	1.9899
		β	0.6772	0.3333	0.1772	0.3774
		θ	2.5184	1.0410	0.5184	1.1625
		α	2.1518	0.2547	0.1518	0.2964
	200	s	1.3725	1.9954	-0.6274	2.0907
		β	0.6783	0.3183	0.1783	0.3647
		θ	2.6623	1.13290	0.6623	1.3118
		α	2.0826	0.1523	0.0826	0.1732
PLNB (5,0.8,8,10)	25	s	5.0663	9.0458	0.0663	9.0415
		β	0.7298	0.3748	-0.0701	0.3811
		θ	8.1346	3.5840	0.1346	3.5847
		α	14.0292	5.8646	4.0292	7.1130
	50	s	5.3943	8.7087	0.3943	8.7133
		β	0.7146	0.3694	-0.0853	0.3790
		θ	8.3753	3.3998	0.3753	3.4187
		α	11.9584	3.5199	1.9584	4.0265
	200	s	7.10969	8.1914	2.1096	8.4547
		β	0.6933	0.3429	-0.1066	0.3590
		θ	8.7770	2.8628	0.7770	2.9650
		α	10.3842	1.2167	0.3842	1.2753

Source: Own research

Tab. 1 presents mean values of the parameter estimates as well as the standard deviation (SD), bias and root mean squared errors (RMSEs) of the parameter estimates for different sample sizes. The estimates of $s,\ \beta,\ \theta$ and α were close to the true values, while the RMSE values for the estimates of $s,\ \beta,\ \theta$ and α decreased when the sample size *n* increased.

5 Applications

The proposed model's efficiency was evaluated by applying it to two real datasets for automobile insurance claims and automobile bodily injury claims which are applied in insurance. These datasets were fitted to the PLNB as PL, inverse PL (IPL) and log logistic (LOL) PDF distributions as follows:

$$f_{PL}(x) = \alpha^2 \left(x/\theta\right)^{\alpha} x^{-1}\left[\left(x/\theta\right)^{\alpha} + 1\right]^{-(\alpha+1)},\ x>0,\ \theta>0,\ \alpha>0\,,$$

$$f_{IPL}(x) = \nu^2 \left(x/\mu\right)^{\nu^2} x^{-1}\left[1+\left(x/\mu\right)^{\nu}\right]^{-(\nu+1)},\ x>0,\ \mu>0,\ \nu>0\,,$$

and $$f_{LOL}(x) = \gamma\left(x/\kappa\right)^{\gamma} x^{-1}\left[1+\left(x/\kappa\right)^{\gamma}\right]^{-2},\ x>0,\ \kappa>0,\ \gamma>0\,,$$ respectively.

The MLE was then used to estimate parameters in the PLNB and other comparative models.

Model comparison was conducted using Akaike's information criterion (AIC) and Bayesian information criterion (BIC), given by

$$AIC = -2LL(\hat{\underset{\sim}{\delta}}) + 2k \text{ and } BIC = -2LL(\hat{\underset{\sim}{\delta}}) + k\log(n),$$

where $LL(\hat{\underset{\sim}{\delta}})$ denotes the log-likelihood function with a vector estimated parameter $\hat{\underset{\sim}{\delta}}$, *k* is the number of estimated parameters, and *n* is the sample size. The model with the smallest value for these criteria was used as the preferred model to describe each dataset.

5.1 Automobile insurance claims

The automobile insurance claims dataset included 6,773 observations of the amount paid on a closed claim in US dollars, using claims experience from a large midwestern US property and casualty insurer for private passenger automobile insurance (Frees, 2010). Some descriptive statistics are listed in Tab. 2. Tab. 3 presents the estimated parameters and comparison criteria AIC and BIC for the PLNB, PL, IPL and LOL distributions. The PLNB provided the best data fit among all the models considered.

Tab. 2: Descriptive statistics of the amount paid on closed claims in US dollars

n	Minimum	Maximum	Median	Mean	Standard deviation
6,773	9.5	60,000	1001.70	1853.00	2,646.91

Source: Own research

Tab. 3: Parameter estimates, AIC, and BIC from automobile insurance claims dataset

Distribution	Estimate	AIC	BIC
PLNB	$\hat{s}$ =0.4197, $\hat{\beta}$ =0.8437, $\hat{\theta}$=3,389.25, $\hat{\alpha}$ =1.5936	114356.5	114363.8
PL	$\hat{\alpha}$ =1.4531, $\hat{\theta}$=1,507.01	114412.7	114426.4
IPL	$\hat{v}$=1.4622, $\hat{\mu}$=723.28	114387.0	114400.6
LOL	$\hat{\gamma}$=2.7322, $\hat{\kappa}$=1.9768	114360.3	114373.9

Source: Own research

5.2 Automobile bodily injury claims

The automobile bodily injury claims dataset included 1,340 observations of the claimant's total economic loss (in thousands). The dataset used was from the Insurance Research Council (IRC), a division of the American Institute for Chartered Property Casualty Underwriters and the Insurance Institute of America (Frees, 2010). Some descriptive statistics are listed in Tab. 4. Tab. 5 presents the estimated parameters and comparison criteria AIC and BIC for the PLNB, PL, IPL and LOL distributions. The PLNB provided the best data fit among all the models considered.

Tab. 4: Descriptive statistics of the claimant's total economic loss (in thousands)

n	Minimum	Maximum	Median	Mean	Standard deviation
1,340	0.005	1,067.697	2.331	5.954	33.1362

Source: Own research

Tab. 5: Parameter estimates, AIC and BIC from the automobile insurance claims dataset

Distribution	Estimate	AIC	BIC
PLNB	$\hat{s}$ =19.0570, $\hat{\beta}$ =0.2232, $\hat{\theta}$=12.7534, $\hat{\alpha}$ =1.0466	6275.7	6296.5
PL	$\hat{\alpha}$ =1.1844, $\hat{\theta}$=2.2931	6300.1	6310.5
IPL	$\hat{v}$=1.1349, $\hat{\mu}$=1.5573	6331.0	6341.4
LOL	$\hat{\gamma}$=1.2255, $\hat{\kappa}$=1.8678	6314.6	6325.0

Source: Own research

Conclusions

This paper defined a new distribution, the so-called PLNB distribution that combined the PL and NB distributions using the G-NB distribution. Plots of PDF and CDF were presented. Various standard properties were derived. Model parameters were estimated using the maximum likelihood method. A simulation study was conducted to determine the MLE and the results were stable and approached the true values as the sample sizes increased. Finally, the PLNB model was fitted to two real datasets to show the potential of the new distribution.

Acknowledgment

This research was funded by the Institute of Research and Development, Chandrakasem Rajabhat University.

References

Balakrishnan, N., Leiva, V., Sanhueza, A., & Cabrera, E. (2009). Mixture inverse Gaussian distributions and its transformations, moments and applications. *Statistics*, *43*(1), 91–104. https://doi.org/10.1080/02331880701829948

Bhati, D., Calderín–Ojeda, E., & Meenakshi, M. (2019). A new heavy tailed class of distributions which includes the Pareto. *Risks*, *7*(4), 99. https://doi.org/10.3390/risks7040099

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, *64*(4), 1–34. https://doi.org/10.18637/jss.v064.i04

Frees, E. W. (2010). *Regression modeling with actuarial and financial applications*. Cambridge University Press.

Gómez-Déniz, E., Sarabia, J. M., & Calderín-Ojeda, E. (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics*, *42*(1), 39–49. https://doi.org/10.1016/j.insmatheco.2006.12.001

Hall, P., & Zhou, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *The Annals of Statistics, 31*(1), 201–224. https://doi.org/10.1214/aos/1046294462

Idemudia, R., & Ekhosuehi, N. (2019). A new three-parameter paralogistic distribution: Its properties and application. *Journal of Data Science*, *17*(2), 239–258. https://doi.org/10.6339/jds.201904_17(2).0001

Kleiber, C., & Kotz, S. (2003). *Statistical Size Distributions in economics and Actuarial Sciences*. John Wiley and Sons.

Ortega, E. M., Cordeiro, G. M., & Kattan, M. W. (2012). The negative binomial–beta weibull regression model to predict the cure of prostate cancer. *Journal of Applied Statistics, 39*(6), 1191–1210. https://doi.org/10.1080/02664763.2011.644525

Percontini, A., Cordeiro, G. M., & Bourguignon, M. (2013). The G-Negative Binomial Family: General Properties and Applications. *Advances and Applications in Statistics*, *35*(2), 127–160.

Rodrigues, C., Cordeiro, G. M., Demétrio, C. G. B., & Ortega, E. M. M. (2011). The Weibull negative binomial distribution. *Advances and Applications in Statistics*, *22*(1), 25–55.

Withers, C. S., & Nadarajah, S. (2011). On the compound Poisson-gamma distribution. *Kybernetika*, *47*, 15–37.

Yusuf, A., Mikail, B. B., Aliyu, A. I., & Sulaiman, A. L. (2016). The inverse Burr negative binomial distribution with application to Real Data. *Journal of Statistics Applications & Probability*, *5*(1), 53–65. https://doi.org/10.18576/jsap/050105

Zubair, M., Tahir, M. H., Cordeiro, G. M., Alzaatreh, A., & Ortega, E. M. (2018). The power-cauchy negative-binomial: Properties and regression. *Journal of Statistical Distributions and Applications*, *5*(1). https://doi.org/10.1186/s40488-017-0082-3

Contact

Sirithip Wasinrat

Faculty of Science, Chandrakasem Rajabhat University, 39/1 Ratchadaphisek Road, Chandrakasem, Chatuchak, Bangkok 10900, Thailand

Mail: sirithip.w@chandra.ac.th

Boonyarit Choopradit

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, 99 Moo 18 Paholyothin Road, Klong Luang, Pathum Thani 12120, Thailand

Mail: boonyarit@mathstat.sci.tu.ac.th