

A Poisson-Generalized Mixture Exponential Distribution

Chadarat Tapan^{1*} and Phatcharawadee Kamboon²

Received: 20 March 2025; Revised: 4 August 2025; Accepted: 30 October 2025

Abstract

In this study, a new Poisson-generalized exponential distribution is proposed. Its statistical properties, including the cumulative distribution function, probability generating function, moment generating function, survival function, and index of dispersion, are derived. The proposed distribution has three parameters and exhibits a right-skewed, unimodal shape. Parameter estimation is performed using the maximum likelihood method. The distribution is then applied to two real overdispersed count datasets and compared with the Poisson and negative binomial distributions. The results indicate that the Poisson-generalized exponential distribution provides a better fit to both datasets based on the Akaike information criterion, Bayesian information criterion, negative log-likelihood, and the Kolmogorov–Smirnov goodness-of-fit test. These findings suggest that it is a viable alternative for modeling overdispersed count data. Future research may explore its extension to regression frameworks and other applied contexts.

Keywords: count data, maximum likelihood, overdispersion, index of dispersion, mixed Poisson

*E-mailAddress(CorrespondingAuthor): chadarat@psru.ac.th

^{1,2}Department of Mathematics, Faculty of Science and Technology, Pibulsongkram Rajabhat University

1. Introduction

Count data refers to non-negative integer values representing the frequency of events within a given context. Typical examples include the number of customer visits to a store, production defects, or daily attendance at an event. Such data are frequently encountered in diverse fields such as business, economics, public health, and scientific research. Depending on the relationship between the mean and variance, count data can be categorized into three types: equidispersed (variance equals the mean), overdispersed (variance exceeds the mean), and underdispersed (variance is less than the mean) (Cameron & Trivedi, 2013).

The Poisson distribution is widely used for modeling count data due to its simplicity and well-defined properties. However, its applicability is limited by a key assumption equidispersion which rarely holds in real-world data (Agresti, 2007). Many empirical datasets display overdispersion or underdispersion, rendering the Poisson model inadequate. A review of several research studies indicates that the mixed Poisson distribution model is more appropriate for data exhibiting overdispersion, such as the negative binomial distribution or the inverse Gaussian Poisson distribution (Hilbe, 2014). This growing recognition of the limitations of the Poisson distribution has led to the development of various mixed Poisson distributions aimed at addressing issues such as overdispersion in real-world count data.

Recent years have seen increased interest in Poisson-compounded models. For instance, Das et al. (2018) introduced the three-parameter Lindley Poisson distribution for modeling overdispersed data and demonstrated its usefulness with real datasets. Shanker and Shukla (2019) proposed a weighted Lindley Poisson distribution tailored for highly skewed and high-variance data. Ramesh and Vijay (2021) explored the generalized Rayleigh Poisson distribution, which combines the Poisson and Rayleigh distributions, primarily for lifetime data. In 2022, Ahsan-ul-Hap et al. examined the XLindley distribution, which also effectively captures overdispersion. An examination of these models reveals that most compound the Poisson distribution with a lifetime distribution. For example, the exponential distribution has frequently been used in this context, leading to several models such as the bivariate Poisson-weighted exponential distribution (Zamani et al., 2014), the Poisson-binomial exponential distribution (Habibi & Asgharzadeh, 2018), the Poisson-weighted exponential distribution (Hamedani, 2020), and the two-parameter Poisson-weighted exponential distribution (Chesneau et al., 2023).

Building on these developments, this paper introduces a new mixed distribution termed the Poisson-generalized mixture exponential (PGME) distribution for modeling count data exhibiting overdispersion. The proposed model is derived by compounding the Poisson distribution with the generalized mixture exponential distribution. We investigate its statistical properties, including the probability mass function, cumulative distribution function, and index of dispersion. Parameters are estimated via the maximum likelihood estimation method. The model's applicability is demonstrated through real data analyses, and its performance is compared with existing count data models.

2. The proposed distribution

This section provides an overview of the generalized mixture exponential distribution and the Poisson-generalized mixture exponential distribution, which is derived from a combination of the Poisson distribution and the generalized mixture exponential (GME) distribution.

2.1 The generalized mixture exponential distribution

The generalized mixture exponential distribution is a distribution suitable for lifetime data. This distribution is derived through a mixing of the negative binomial distribution and the Erlang-Lindley distribution (Yaoting Yang et al., 2021). The probability density function (pdf) is defined as follows:

Definition 2.1 If random variable X follows the GME distribution, then the pdf of the GME distribution is

$$f(x; \theta, \alpha, \beta) = \frac{(\alpha+1)\theta}{\alpha+1+\alpha\beta} e^{-\theta x} \left[1 + \beta(1 - e^{-\alpha\theta x}) \right], \quad x > 0, \quad (1)$$

where, $\alpha > 0$ is the scale parameter, $\theta > 0$ and $\beta \geq -1$ are the shape parameter.

The mean of the GME distribution is given by

$$E(X) = \frac{(1+\alpha)^2(1+\beta) - \beta}{\theta(\alpha+1+\alpha\beta)(1+\alpha)} \quad (2)$$

The variance of the GME distribution is given by

$$Var(X) = \frac{(2(1+\alpha)^3(1+\beta) - 2\beta)(\alpha+1+\alpha\beta) - (1+2\alpha+2\alpha\beta+\alpha^2+\alpha^2\beta)^2}{\theta^2(\alpha+1+\alpha\beta)^2(1+\alpha)^2} \quad (3)$$

2.2 The Poisson-generalized mixture exponential distribution

This section presents the derivation of the Poisson-generalized mixture exponential distribution through a mixing of the Poisson and GME distributions and discusses its probability function properties.

Theorem 2.2 If random variable X follows the PGME distribution, with parameters α , θ and β , denote by $X \sim \text{PGME}(\alpha, \theta, \beta)$. The pmf of the PGME distribution is

$$f(x) = \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{(1+\theta)^{x+1}} - \frac{\beta}{(1+\theta+\alpha\theta)^{x+1}} \right), \quad x = 0, 1, 2, \dots, \quad (4)$$

where, $\alpha > 0$ is the scale parameter, $\theta > 0$ and $\beta \geq -1$ are the shape parameter.

Proof. A mixed Poisson distribution has the following pmf:

$$f(x) = \int_{\lambda=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \cdot g(\lambda) d\lambda, \quad \lambda > 0, x = 0, 1, 2, \dots, \quad (5)$$

when the mixing distribution is continuous with a pdf $g(\lambda)$ (Johnson et al., 2005). Thus, the PGME distribution can be obtained as

$$\begin{aligned} f(x) &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \cdot \frac{(\alpha+1)\theta}{\alpha+1+\alpha\beta} e^{-\theta\lambda} [1 + \beta(1 - e^{-\alpha\theta\lambda})] d\lambda \\ &= \frac{(\alpha+1)\theta}{x!(\alpha+1+\alpha\beta)} \int_0^{\infty} e^{-\lambda-\theta\lambda} \lambda^x + \beta e^{-\lambda-\theta\lambda} \lambda^x - \beta e^{-\lambda-\theta\lambda-\alpha\theta\lambda} \lambda^x d\lambda \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1}{(1+\theta)^{x+1}} + \frac{\beta}{(1+\theta)^{x+1}} - \frac{\beta}{(1+\theta+\alpha\theta)^{x+1}} \right) \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{(1+\theta)^{x+1}} - \frac{\beta}{(1+\theta+\alpha\theta)^{x+1}} \right) \end{aligned}$$

This proof is completed. □

The properties of probability function are $f(x) \geq 0$ for all x and $\sum_{x=0}^{\infty} f(x) = 1$.

$$\begin{aligned} \sum_{x=0}^{\infty} f(x) &= \sum_{x=0}^{\infty} \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{(1+\theta)^{x+1}} - \frac{\beta}{(1+\theta+\alpha\theta)^{x+1}} \right) \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{\theta} - \frac{\beta}{\theta+\alpha\theta} \right) \\ &= \frac{1}{(\alpha+1+\alpha\beta)} (1 + \alpha + \beta\alpha) \\ &= 1 \end{aligned}$$

Figure 1 presents the pmf plots of the PGME distribution for different values of α , θ and β . The PGME distribution is skewed to the right. Obviously and unimodal because the GME distribution is unimodal (Yaoting Yang et al., 2021).

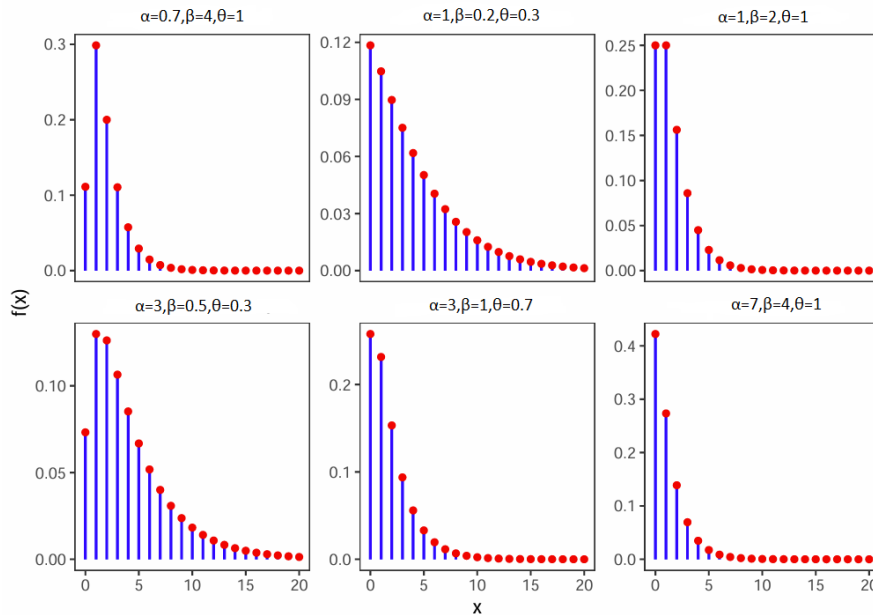


Figure1 The pmf plots of the PGME distribution with different parameter values

Theorem 2.3 If random variable X follows the PGME distribution, with parameters α , θ and β . The cumulative distribution function (cdf) of the PGME distribution is

$$F(x) = \frac{(\alpha + \alpha\beta + 1 + \beta)((1+\theta)^x - 1)(1 + \theta + \alpha\theta)^x - (\beta(1 + \theta + \alpha\theta)^x - \beta)(1 + \theta)^x}{(\alpha + 1 + \alpha\beta)(1 + \theta)^x(1 + \theta + \alpha\theta)^x} \quad (6)$$

where, $x = 0, 1, 2, \dots$, $\alpha > 0$, $\theta > 0$ and $\beta \geq -1$.

Proof. The cdf of the PGME distribution is the following:

$$\begin{aligned} F(x) &= \sum_{t=0}^x f(t) \\ &= \sum_{t=0}^x \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{(1+\theta)^{t+1}} - \frac{\beta}{(1+\theta+\alpha\theta)^{t+1}} \right) \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{(1+\beta)((1+\theta)^x - 1)}{(1+\theta)^x \theta} - \frac{\beta((1+\theta+\alpha\theta)^x - 1)}{(1+\theta+\alpha\theta)^x (\theta + \alpha\theta)} \right) \\ &= \frac{(\alpha + \alpha\beta + 1 + \beta)((1+\theta)^x - 1)}{(\alpha + 1 + \alpha\beta)(1+\theta)^x} - \frac{\beta(1+\theta+\alpha\theta)^x - \beta}{(\alpha + 1 + \alpha\beta)(1+\theta+\alpha\theta)^x} \\ &= \frac{(\alpha + \alpha\beta + 1 + \beta)((1+\theta)^x - 1)(1 + \theta + \alpha\theta)^x - (\beta(1 + \theta + \alpha\theta)^x - \beta)(1 + \theta)^x}{(\alpha + 1 + \alpha\beta)(1 + \theta)^x(1 + \theta + \alpha\theta)^x} \end{aligned}$$

This proof is completed. $\square$

Figure 2 presents the cdf of the PGME distribution for various combinations of the parameters α , β , and θ . As expected, the cdf are non-decreasing functions, reflecting the fundamental property of any valid

probability distribution. However, the plots also reveal how the shape of the cdf changes in response to different parameter values, sharper increases in the cdf curves indicate that a higher probability mass is concentrated in the lower range of the distribution, whereas flatter slopes suggest more dispersion or heavier tails. These behaviors highlight the flexibility of the PGME distribution in modeling a wide range of data patterns, especially when skewness or overdispersion is present.

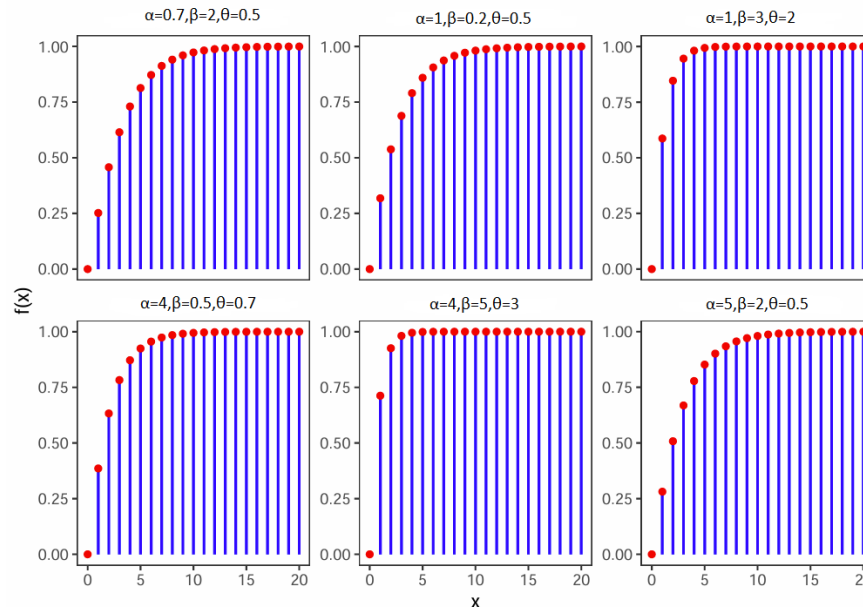


Figure2 The cdf plots of the PGME distribution with different parameter values

3. Statistics properties

This section showed the probability generating function (pgf), expectation, variance, and moment generating function (mgf), survival function, and index of dispersion (ID).

3.1 Probability generating function

Theorem 3.1 Let X be a random variable with the PGME distribution, then the pgf of the PGME distribution is

$$G(z) = \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{1+\theta-z} - \frac{\beta}{1+\theta+\alpha\theta-z} \right), \quad (7)$$

where $\alpha > 0$, $\theta > 0$, $\beta \geq -1$ and z is a real number.

Proof. The pgf of the PGME distribution can be shown as follows:

$$\begin{aligned} G(z) &= \sum_{i=0}^{\infty} f(x) z^x \\ &= \sum_{x=0}^{\infty} \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{(1+\theta)^{x+1}} - \frac{\beta}{(1+\theta+\alpha\theta)^{x+1}} \right) \cdot z^x \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\sum_{x=0}^{\infty} \frac{z^x}{(1+\theta)^{x+1}} + \sum_{x=0}^{\infty} \frac{\beta z^x}{(1+\theta)^{x+1}} - \sum_{x=0}^{\infty} \frac{\beta z^x}{(1+\theta+\alpha\theta)^{x+1}} \right) \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1}{1+\theta-z} + \frac{\beta}{1+\theta-z} - \frac{\beta}{1+\theta+\alpha\theta-z} \right) \end{aligned}$$

$$= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{1+\theta-z} - \frac{\beta}{1+\theta+\alpha\theta-z} \right)$$

This proof is completed. $\square$

3.2 Expectation, Variance and Moment Generating Function

3.2.1 Expectation

From Theorem 3.1, the expectation of X is $E(X) = G'(1)$. The expectation of the PGME distribution can be derived as follows:

$$\begin{aligned} G'(z) &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left((1+\beta)(-1)(1+\theta-z)^{-2}(-1) + \beta(1+\theta+\alpha\theta-z)^{-2}(-1) \right) \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{(1+\theta-z)^2} - \frac{\beta}{(1+\theta+\alpha\theta-z)^2} \right) \\ G'(1) &= \frac{\alpha+1}{\alpha+1+\alpha\beta} \left(\frac{1+\beta}{\theta} - \frac{\beta}{\theta(1+\alpha)^2} \right) \end{aligned}$$

Thus, the expectation of X is

$$E(X) = \frac{\alpha+1}{\alpha+1+\alpha\beta} \left(\frac{1+\beta}{\theta} - \frac{\beta}{\theta(1+\alpha)^2} \right) \quad (8)$$

3.2.2 Variance

From Theorem 3.1, the variance of X is $Var(X) = G''(1) + G'(1) - (G'(1))^2$. The variance of the PGME distribution can be derived as follows:

$$\begin{aligned} G''(z) &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left((1+\beta)(-2)(1+\theta-z)^{-3}(-2) - \beta(-2)(1+\theta+\alpha\theta-z)^{-3}(-1) \right) \\ &= \frac{2(\alpha+1)\theta}{\alpha+1+\alpha\beta} \left(\frac{\beta}{(1+\theta+\alpha\theta-z)^3} - \frac{1+\beta}{(1+\theta-z)^3} \right) \\ G''(1) &= \frac{2(\alpha+1)}{\alpha+1+\alpha\beta} \left(\frac{\beta}{\theta^2(1+\alpha)^3} - \frac{1+\beta}{\theta^2} \right) \end{aligned}$$

Thus, the variance of X is

$$Var(X) = \frac{1}{\theta(1-\alpha)(\alpha+1+\alpha\beta)} \left[\frac{2\beta+2(1+\beta)(1+\theta)^3}{\theta(1+\alpha)} - (1+\beta)(1+\alpha)^2 - \beta - \frac{((1+\beta)(1+\alpha)^2 - \beta)^2}{\theta(1+\alpha)(1+\alpha+\alpha\beta)} \right] \quad (9)$$

3.2.3 Moment generating function

From Theorem 3.1, set $z = e^t$ and we will obtain the mgf of the PGME distribution.

$$\begin{aligned} M_X(t) &= G(e^t) \\ &= \frac{(\alpha+1)\theta}{(\alpha+1+\alpha\beta)} \left(\frac{1+\beta}{1+\theta-e^t} - \frac{\beta}{1+\theta+\alpha\theta-e^t} \right) \end{aligned} \quad (10)$$

The mgf of the PGME distribution not only facilitates the derivation of its moments but also enhances the understanding of the distribution's characteristics. Additionally, the mgf serves as a fundamental tool in further statistical analyses, such as proving convergence and assessing model properties.

3.3 Survival function

The survival function is a concept used in survival analysis, commonly found in statistics, data science, and other related fields. This function is useful in various disciplines, such as in medicine, where it is used to study the survival rate of patients after treatment.

Let $F(x)$ is the cdf of the PGME distribution, the survival function of the PGME distribution is given

$$S(x) = 1 - F(x) = 1 - \frac{(\alpha + \alpha\beta + 1 + \beta)((1 + \theta)^x - 1)(1 + \theta + \alpha\theta)^x - (\beta(1 + \theta + \alpha\theta)^x - \beta)(1 + \theta)^x}{(\alpha + 1 + \alpha\beta)(1 + \theta)^x(1 + \theta + \alpha\theta)^x}, \quad (11)$$

Where $x = 0, 1, 2, \dots$, $\alpha > 0$, $\theta > 0$, and $\beta \geq -1$.

The PGME distribution could be extended to survival analysis due to its flexibility in accommodating overdispersion and underdispersion data. By modeling survival times using the PGME distributions, one could derive corresponding survival and hazard functions that are well-suited for complex time-to-event data. Incorporating the PGME distribution into survival models may improve model fit and enhance interpretability in biomedical and reliability research.

3.4 Index of dispersion and Coefficient of variation

3.4.1 Index of dispersion

The ID is a measure of dispersion; it can be derived from the variance-to-mean ratio, which is the most basic definition. The ID of the PGME distribution is given as

$$\begin{aligned} \text{ID} &= \frac{\text{Var}(X)}{E(X)} \\ &= \frac{1}{\theta(1 + \alpha)(\alpha + 1 + \alpha\beta)} \left[\frac{2\beta + 2(1 + \beta)(1 + \theta)^3}{\theta(1 + \alpha)} - (1 + \beta)(1 + \alpha)^2 - \beta - \frac{((1 + \beta)(1 + \alpha)^2 - \beta)^2}{\theta(1 + \alpha)(1 + \alpha + \alpha\beta)} \right] \\ &= \frac{\frac{\alpha + 1}{\alpha + 1 + \alpha\beta} \left(\frac{1 + \beta}{\theta} - \frac{\beta}{\theta(1 + \alpha)^2} \right)}{\frac{1 + \theta}{(\alpha + 1)((1 + \beta)(1 + \alpha)^2 - \beta)}} \left[\frac{2\beta + 2(1 + \beta)(1 + \theta)^3}{\theta(1 + \alpha)} - (1 + \beta)(1 + \alpha)^2 - \beta - \frac{((1 + \beta)(1 + \alpha)^2 - \beta)^2}{\theta(1 + \alpha)(1 + \alpha + \alpha\beta)} \right] \quad (12) \end{aligned}$$

Figure 3 presents two surface plots illustrating how the ID of the PGME distribution varies with different values of the parameters α and β . Figure 3(a) shows that for low values of α and high values of β , the distribution exhibits overdispersion ($\text{ID} \geq 1$). Conversely, as α increases or β decreases, the ID declines, indicating underdispersion. Figure 3(b), which includes negative values of β , demonstrates that the ID is less than one when $-1 < \beta < 0$, suggesting that the PGME distribution exhibits underdispersion in this range. Such behavior of the ID has important implications for selecting the PGME distribution in distribution real-world count data exhibiting either overdispersion or underdispersion. It also highlights the need for clearly defined parameter constraints to avoid invalid or non-interpretable outcomes, particularly when β takes negative values. Overall, the dispersion properties illustrated in Figure 3 contribute to a deeper understanding of the distribution's behavior and support its use in flexible count data distribution.

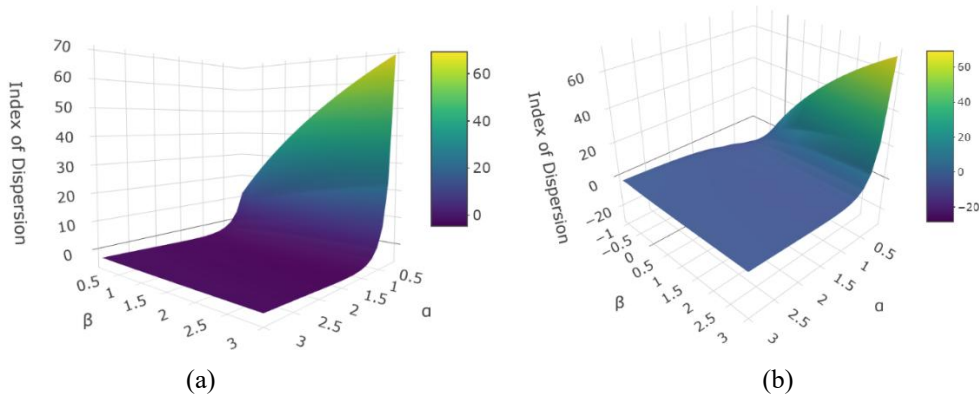


Figure 3 Index of dispersion of the PGME distribution

4. Parameter estimation

Maximum likelihood estimation (MLE) is one of the most commonly employed methods for estimating the unknown parameters of a specified probability distribution by finding the parameter values that maximize the likelihood of the observed data. The process involves defining a likelihood function, which expresses the probability of the observed data as a function of the parameters. Then, the parameter values that maximize this likelihood function are found by taking derivatives and solving for where the function reaches its maximum. This study discuss parameter estimation based on the MLE method. Let $X_1, X_2, \dots, X_n$ denote random samples n drawn from the PGME distribution with parameters θ, α, β , and $x_1, x_2, \dots, x_n$ be the observed value. The log-likelihood function $l(\theta, \alpha, \beta; \underline{x})$ of the PGME distribution is given by:

$$l(\theta, \alpha, \beta; \underline{x}) = \sum_{i=1}^n \left[\ln(\alpha + 1)\theta - \ln(\alpha + \alpha\beta + 1) + \ln \left(\frac{1 + \beta}{(1 + \theta)^{x_i + 1}} - \frac{\beta}{(1 + \theta + \beta)^{x_i + 1}} \right) \right] \quad (13)$$

To estimate the parameters θ, α , and β of the PGME distribution, we utilized the *optim* function from the R package (Steenbergen, 2006). Since Equation (13) is not an explicit function, a closed-form analytical solution is not feasible.

5. Applications

In this section, the PGME distribution is applied to real dataset and compared with the Poisson and negative binomial (NB) distributions. The analysis in this study is performed using the R programming language, with parameter estimation carried out using the *optim* function from the R stats package (R Core Team, 2024). The distributions are compared based on the negative log-likelihood (-LL), Akaike information criterion (AIC), Bayesian Information Criterion (BIC), and Kolmogorov-Smirnov (KS) test (Arnold & Emerson, 2011). The first data set represents the number of doctor visits from the Ecdat package in R language with 485 observations. The mean, variance, and ID of the number of doctor visits are 1.6103, 11.2011, and 6.9559, respectively. The ID indicates a case of overdispersion count data. The behavior of data set is reverse J-shape. The second data set comprises 189 observations representing the number of physician visits in 1st trimester (Hosmer and Lemeshow, 2000). The mean, variance, and ID for the number of complaints are 0.7937, 1.1221, and 1.4138, respectively. The ID value indicates the presence of overdispersion count data. The distribution of data set exhibits a reverse J-shaped pattern.

Tables 1 and 2 present the expected and observed frequencies, negative -LL, AIC, BIC, KS test statistics, and corresponding p-values for the PGME, Poisson, and NB distributions based on the first and second data sets, respectively. The KS test results and associated p-values indicate that both the PGME and NB distributions provide a better fit to the data compared to the Poisson distribution. Furthermore, the PGME distribution consistently yields lower -LL, AIC, and BIC values than the NB distribution, suggesting that it is the most appropriate model for both datasets. This conclusion is further corroborated by Figures 4 and 5, which illustrate the observed and expected frequencies for the two data sets, respectively.

Table 1 Observed, expected frequencies, estimated parameters, and p-value of KS test for the first data set.

Number of doctor visits	Observed frequencies	Expected frequencies		
		PGME	Poisson	NB
0	241	367	97	162
1	96	90	156	126
2	55	22	126	82
3	30	5	67	49
4	14	1	27	29
5	8	0	9	16
6	12	0	2	9
7	7	0	1	5
8	5	0	0	3
9	8	0	0	2
10	3	0	0	1
11	1	0	0	0
12	1	0	0	0
13	0	0	0	0
⋮	⋮	⋮	⋮	⋮
48	1	0	0	0
Total	485	485	485	485
Parameter estimates		$\theta=3.099$ $\alpha=0.001$ $\beta=2.501$	$\lambda=1.610$	$r=1.509$ $p=0.937$
-LL		818.180	1214.477	877.1255
AIC		1642.361	2430.954	1758.251
BIC		1654.913	2435.138	1766.619
KS		0.2857	0.4424	0.2995
p-value		0.0703	0.0035	0.1034

Table 2 Observed, expected frequencies, estimated parameters, and p-value of KS test for the second data set.

Number of physician visits in 1st trimester	Observed frequencies	Expected frequencies		
		PGME	Poisson	NB
0	100	157	86	109
1	47	27	68	44
2	30	5	27	20
3	7	1	7	9
4	4	0	1	4
5	0	0	0	2
6	1	0	0	1
Total	189	189	189	189
Parameter estimates		$\theta=4.899$ $\alpha=0.0007$ $\beta=2.4013$	$\lambda=0.794$	$r=0.815$ $p=1.0263$
-LL		171.486	237.295	234.072
AIC		348.973	476.590	472.144
BIC		358.698	479.832	478.627
KS		0.476	0.4823	0.419
p-value		0.0560	0.0472	0.120

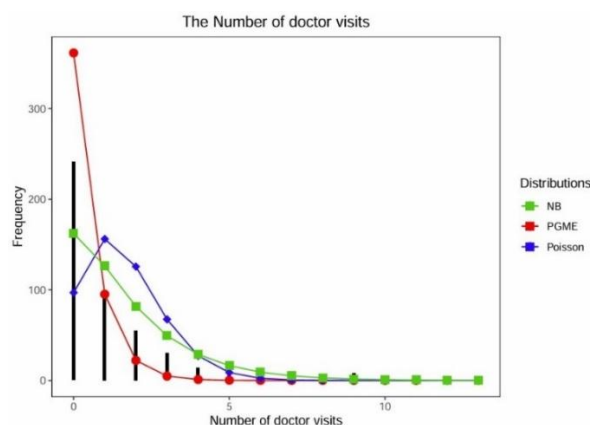


Figure 4 Plot of the observed and fitted distribution of the first data set

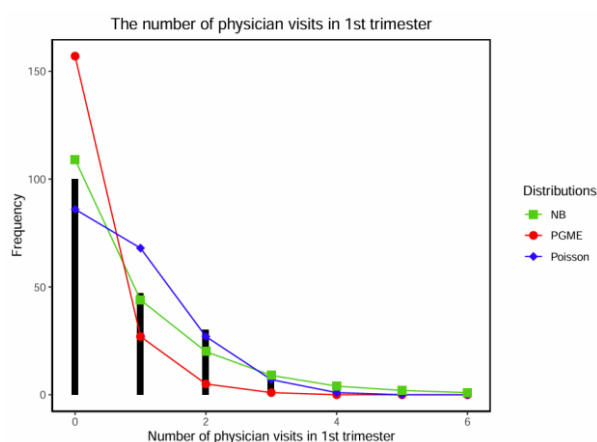


Figure 5 Plot of the observed and fitted distribution of the second data set

6. Conclusion

This paper introduces the PGME distribution as a flexible discrete probability distribution for both overdispersed and underdispersed count data. The pmf and cdf were derived, along with several key statistical properties, including the pgf, mgf, survival function, and ID. The proposed distribution has three parameters and exhibits a right-skewed, unimodal shape. Parameter estimation is performed using the method of MLE, and the model's goodness-of-fit is assessed by comparing it to the classical Poisson and negative binomial (NB) distributions using two overdispersed count datasets. Notably, the PGME distribution consistently yielded lower AIC, BIC, and negative log-likelihood values, indicating superior performance based on standard model selection criteria. Therefore, the PGME distribution may serve as a useful alternative for modeling overdispersed count data. Future work could extend the PGME framework to regression modeling and other applied contexts.

7. References

- Arnold, T.B. & Emerson, J.W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. **The R Journal**, 3(2), 34-39.
- Agresti, A. (2007). **An introduction to categorical data analysis**. New Jersey: John Wiley & Sons.
- Ahsan-ul-Haq, M., Al-Bossly, A., El-Morshedy, M., & Eliwa, M.S. (2022). Poisson XLindley distribution for count data: Statistical and Reliability Properties with Estimation Techniques and Inference. **Computational Intelligence and Neuroscience**, 2022, <https://doi.org/10.1155/2022/6503670>.
- Cameron, A.C., & Trivedi, P.K. (2013). **Regression Analysis of Count Data (2nd Edition)**. Cambridge University Press.
- Chesneau, C., Tomy, L., Jose, M. (2023). On a New Two-parameter Weighted Exponential Distribution and Its Application to the COVID-19 and Censored Data. **Thailand Statistician**, 22(1), 17-30.

- Das, K.K., Ahmed, I., & Bhattacharjee, S. (2018). A new three-parameter Poisson-Lindley distribution for modelling over-dispersed count data. **International Journal of Applied Engineering Research**, 13(23), 16468–16477.
- Hamedani, G.G. (2020). Characterizations of the new Poisson-weighted exponential and the exponentiated weibull-geometric discrete distributions. **Pakistan Journal of Statistics and Operation Research**, 16(4), 851–854.
- Hilbe, J.M. (2014). **Modeling Count Data**. Cambridge University Press.
- Shanker, R. & Shukla, K.K. (2019). On Poisson-Weighted Lindley Distribution and Its Applications. **Journal of Scientific Research**, 11(1), 1–13.
- Hosmer, D and Lemeshow, S. (2000), **Applied Logistic Regression**, Wiley.
- Ramesh, K. J.& Vijay, K. (2021). Poisson Generalized Rayleigh Distribution: Properties and Applications. **International Journal of Statistics and Applied Mathematics**, 6(1), 90-99.
- R Core Team. 2024. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing. Vienna: Austria.
- Steenbergen, M.R. (2006). **Maximum Likelihood Programming in R**. Department of Political Science, University of North Carolina, Chapel Hill.
- Yang, Y., Tian, W. & Tong, T. (2021). Generalized mixtures of exponential distribution and associated inference. **Mathematics**, 9(1371), <https://doi.org/10.3390/math9121371>.
- Zamani, H., Faroughi, P. & Ismail, N. (2014). **Bivariate Poisson-weighted exponential distribution with applications**. AIP Conference Proceedings, 964-968, doi:10.1063/1.4882600.