



OPEN

## Discrete Poisson Haq distribution with mathematical properties and count data modeling

Abdullah M. Alomair<sup>1</sup>✉, Faisal Ayyaz<sup>2</sup>, Saadia Tariq<sup>2</sup> & Muhammad Ahsan-ul-Haq<sup>3</sup>

A new one-parameter discrete distribution, namely the *Poisson Haq (PH) distribution*, is proposed by a mixture of the Poisson variable and an independently distributed Haq random variable. This model effectively analyzes over-dispersed count datasets by extending Poisson distribution. Various useful statistical properties of the PH distribution are derived and discussed. The failure rate of the proposed distribution is “increasing” and “upside bathtub” shaped. The model parameter estimation is performed using renowned estimation approaches, method of moments, and method of maximum likelihood estimation. A parametric regression model tailored for count datasets is also developed using the proposed distribution. A simulation study is conducted to demonstrate the performance and behavior of the proposed estimators. The present study validates that the new count model adequately explains the medical datasets, which are the number of infected patients with the Nipah virus, the number of mammalian cytogenetic dosimetry lesions, and the Length of Hospital Stay. Additionally, we also estimate the model parameter using the Bayesian approach with gamma prior. Compared to widely used alternatives such as the Poisson (AIC = 145.16, BIC = 147.19), Poisson moment exponential (AIC = 137.53, BIC = 139.56), Poisson-XLindley (AIC = 135.86, BIC = 137.88) distributions and others, our model demonstrates improved fitting accuracy, as evidenced by lower AIC (135.78) and BIC (137.81) values for first data and similarly for second data applications. Finally, to validate the fit of the PH regression model, it is applied to the Length of Hospital Stay dataset.

**Keywords** Discrete distribution, Haq distribution, Mixed Poisson, Estimation, Medical data

Data modeling complexity has risen significantly because of excessive data collection across fields like engineering, medicine, ecology, epidemiology, and renewable energy<sup>1–7</sup>. Poisson distribution stands as the principal statistical tool for examining data sets with count values. A critical aspect of the Poisson distribution will be referred to as equal dispersion (variance is equal to mean). Practically there are situations where Poisson distribution is unsuitable for modeling data with a high degree of under and overdispersion. Overdispersion is a phenomenon that is frequently seen in count data and presents serious difficulties for statistical modeling.

The mixed Poisson models are adaptable tools for the analysis of count data exhibiting heterogeneity and overdispersion. Over the years, various researchers have been introduced by mixing Poisson distribution with various continuous distributions. Some examples of Poisson mixture models are; The generalized Poisson-Lindley distribution<sup>8</sup> and applied in reliability and biological studies. Poisson Amarendra distribution<sup>9</sup> arises by compounding the Poisson and Amarendra distributions. It is utilized to model the ecological and insurance claims datasets. Poisson Garima distribution<sup>10</sup> effectively captures overdispersed count data and has been applied in health and social sciences. Poisson Shanker distribution proposed by<sup>11</sup> has been shown to perform well in actuarial and demographic studies. Poisson pseudo-Lindley distribution introduced by<sup>12</sup> explored its several statistical properties, including descriptive measures, quantile function, and Lorenz curve. The parameters were estimated using maximum likelihood estimation (MLE) and the applicability of the proposed model was demonstrated using two real-world datasets. Bernoulli Poisson moment exponential distribution<sup>13</sup> is studied along with its various generating functions and mathematical characteristics. Parameter estimation was conducted using the MLE approach and its applicability was established using three datasets. Poisson Quasi-Lindley distribution was put forward by<sup>14</sup>. They comprehensively studied its different reliability properties. Additionally, an extensive simulation study was conducted and parameter estimation was performed using the MLE method. Ref<sup>15</sup>. Introduced one parameter Poisson Agu Eghwerido and derived its various properties, including factorial moments, generating function, and entropies. The parameter of the said model was estimated

<sup>1</sup>Department of Quantitative Methods, School of Business, King Faisal University, Al-Ahsa 31982, Saudi Arabia.

<sup>2</sup>School of Statistics, Minhaj University Lahore, Lahore, Pakistan. <sup>3</sup>College of Statistical Sciences, University of the Punjab, Lahore, Pakistan. ✉email: ama.alomair@kfu.edu.sa

using both classical and Bayesian approaches. Ref<sup>16</sup> proposed discrete Poisson Aradhana distribution to study overdispersed data sets. To explore the behavior of the proposed model in depth, various statistical measures were derived. Parameter estimation was performed using the maximum likelihood method, along with a comprehensive simulation analysis. Ref<sup>17</sup> proposed Poisson moment exponential distribution and applied it to four datasets to demonstrate its practicality. The parameter estimation was performed using seven different estimation approaches. Ref<sup>18</sup> investigated the Poisson XLindley distribution to study right skewed, leptokurtic, and overdispersed datasets. Numerous reliability characteristics of the derived model were explored. The parameter of the new model was estimated by utilizing six different estimation techniques. Poisson Mirra, its regression model, and its first-order integer-valued autoregressive process (INAR-1) were introduced by<sup>19</sup>. The INAR-1 model parameters were estimated using Yule-Walker, conditional maximum likelihood, and conditional least squares methods. Poisson's new XLindley was instigated<sup>20</sup> and its fundamental mathematical and statistical properties were explicitly studied. Ref<sup>21</sup> developed a two-parameter discrete Poisson mixing distribution and derived its key properties. A new count regression model was proposed, and the model's applicability was demonstrated using asymmetric datasets. Ref<sup>22</sup> introduced Poisson entropy-based weighted exponential distribution for modeling right-skewed data with heavy tails. Parameters were estimated using MLE and Bayesian methods. Its applicability was demonstrated with three real-world datasets. Ref<sup>23</sup> developed Poisson quasi-Shanker distribution, derived key properties, and estimated parameters, and endorsed its applicability through simulation and real datasets. Ref<sup>24</sup> proposed Poisson Quasi XLindley distribution, a two-parameter discrete model. Its several key statistical properties were analyzed explicitly. The model was applied to two real datasets and integrated into a count regression framework as well. Some more examples of discrete models are; discrete Poisson-Lindley distribution<sup>25</sup> new geometric distribution<sup>26</sup> discrete extended odd Weibull exponential distribution<sup>27</sup> discrete Half-Logistic distribution<sup>28</sup> discrete exponentiated moment exponential distribution<sup>29</sup> Poisson Komal distribution<sup>30</sup> and Poisson Xrama distribution<sup>31</sup>.

Each of these compound models offers specific benefits and draws limitations in their usage. The acceleration of technological development produces a tremendous amount of complex high-dimensional data that spreads across healthcare domains with finance and social science and engineering operations. Numerous contemporary data models find difficulty in accurately reflecting intricate statistical patterns which include extreme overdispersion, zero inflation, and non-normal distributional forms. The rapid generation of emerging datasets demands swiftly developing interpretative statistical models that remain practical and computationally efficient to perform accurate analyses of complex datasets. Therefore, this study attempts to introduce a novel Poisson compound model based on the Haq distribution, which is designed to better capture the variations present in complex datasets. Haq distribution is a one-parameter powerful and flexible probability model designed to handle complex datasets with overdispersion, skewness, and reliability characteristics. Its derivation from a mixture of exponential and Xgamma distributions allows it to capture variations in real-world data more effectively than many existing models. The Haq distribution offers both a heavier right tail and higher flexibility for describing over-dispersion patterns better than standard distributions including the Lindley and moment-exponential distributions. Application of the Haq distribution in the Poisson framework gives users better control over tail effects and extreme distributions in addition to skewness properties that align with real-world count data patterns. The mathematical structure of the Haq distribution enables smooth integration-based operations for generating the pmf and calculating moments as well as additional properties while maintaining behavioral richness in its modeling framework.

The Haq distribution was originally presented by<sup>32</sup>. It was obtained by mixing Xgamma ( $f_1(x)$ ) and exponential ( $f_2(x)$ ) distributions with scale parameter  $\theta$  for both and mixing proportions  $p_1 = \frac{1}{1+\theta}$  and  $p_2 = \frac{\theta}{1+\theta}$ . The probability density function of the Haq distribution is given as

$$f(x) = \left( \frac{\theta}{1+\theta} \right)^2 \left( 2 + \theta + \frac{\theta x^2}{2} \right) e^{-\theta x}, \quad x > 0, \quad \theta > 0. \quad (1)$$

As stated earlier, the current work introduces the Poisson-Haq distribution, a novel mixed Poisson distribution formed by combining the Poisson and Haq distributions. The study's particular aims are:

- To develop the new Poisson-Haq distribution and derive its key mathematical properties, including probability mass function, moments, and other essential characteristics.
- To estimate the parameters of the proposed distribution using the maximum likelihood and method of moments techniques. Additionally, conduct a comprehensive simulation study to evaluate the performance and reliability of these estimators.
- To further estimate the distribution's parameters using Bayesian estimation methods, providing an alternative approach to inference.
- To demonstrate the practical applicability and adequacy of the Poisson-Haq distribution by fitting it into two medical datasets, thereby illustrating its utility in real-world scenarios.

The structure of the study is as follows: Sect. "Poisson Haq distribution" presents the derivation of the Poisson Haq distribution, along with a visual depiction of its probability mass function and hazard rate function. Section "Mathematical characteristics of PH distribution" focuses on the calculation of the key theoretical properties. Section "Parameter estimation" details the statistical inference of the new distribution parameters using both the method of moments and maximum likelihood estimation. A new count regression model is presented in Sect. "Poisson Haq regression model". Section "Data applications" represents the application of the new model to

medical field datasets. Section “Bayesian analysis” covers an analysis based on the Bayesian approach, while Sect. “Conclusion” contains the concluding remarks.

### Poisson Haq distribution

A random variable  $X$  is said to follow a PH distribution if it satisfies the following representation.

$$(X|\lambda) \sim \text{Poisson}(\lambda) \text{ and } (\lambda|\theta) \sim \text{Haq}(\theta).$$

Using Eq. (2) and the probability mass function of the Poisson distribution, the PH distribution is obtained as follows:

$$\begin{aligned} P(X=x;\theta) &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \left(\frac{\theta}{1+\theta}\right)^2 \left(2+\theta + \frac{\theta \lambda^2}{2}\right) e^{-\theta \lambda} d\lambda, \\ &= \left(\frac{\theta}{1+\theta}\right)^2 \frac{1}{x!} \int_0^{\infty} \left((2+\theta) \lambda^x e^{-\lambda-\lambda\theta} + \frac{\theta \lambda^{2+x}}{2} e^{-\lambda-\lambda\theta}\right) d\lambda, \\ &= \left(\frac{\theta}{1+\theta}\right)^2 \frac{1}{x!} \left[ (2+\theta) \int_0^{\infty} \lambda^x e^{-\lambda(1+\theta)} d\lambda + \frac{\theta}{2} \int_0^{\infty} \lambda^{2+x} e^{-\lambda(1+\theta)} d\lambda \right], \\ &= \left(\frac{\theta}{1+\theta}\right)^2 \frac{1}{x!} \left[ (2+\theta) \int_0^{\infty} \lambda^{(x+1)-1} e^{-\lambda(1+\theta)} d\lambda + \frac{\theta}{2} \int_0^{\infty} \lambda^{(x+3)-1} e^{-\lambda(1+\theta)} d\lambda \right], \end{aligned}$$

Applying the standard gamma function formula:  $\int_0^{\infty} x^{n-1} e^{-\alpha x} = \frac{\Gamma(n)}{(\alpha)^n}$

$$\begin{aligned} &= \left(\frac{\theta}{1+\theta}\right)^2 \frac{1}{x!} \left[ \frac{(2+\theta) \Gamma(x+1)}{(1+\theta)^{x+1}} + \frac{\theta \Gamma(x+3)}{2(1+\theta)^{x+3}} \right], \\ &= \left(\frac{\theta}{1+\theta}\right)^2 \left( \frac{1}{(1+\theta)^{x+1}} \right) \left[ (2+\theta) + \frac{\theta(2+x)(1+x)}{2(1+\theta)^2} \right], \\ P(X=x;\theta) &= \frac{\theta^2}{(1+\theta)^{x+3}} \left[ \frac{2(2+\theta)(1+\theta)^2 + \theta(2+x)(1+x)}{2(1+\theta)^2} \right], \quad x > 0, \theta > 0. \end{aligned} \quad (2)$$

The behavior of the pmf at the lower and upper limits is described by

$$\lim_{x \rightarrow 0} P(x) = \lim_{x \rightarrow 0} \left[ \frac{\theta^2}{(1+\theta)^{x+3}} \left( \frac{2(2+\theta)(1+\theta)^2 + \theta(2+x)(1+x)}{2(1+\theta)^2} \right) \right] = \frac{\theta^2(2+\theta)}{(1+\theta)^3},$$

and

$$\lim_{x \rightarrow \infty} P(x) = 0.$$

Figure 1 depicts a graphical depiction of the PH distribution for various parameter values. It demonstrates that the distribution is unimodal and skewed to the right.

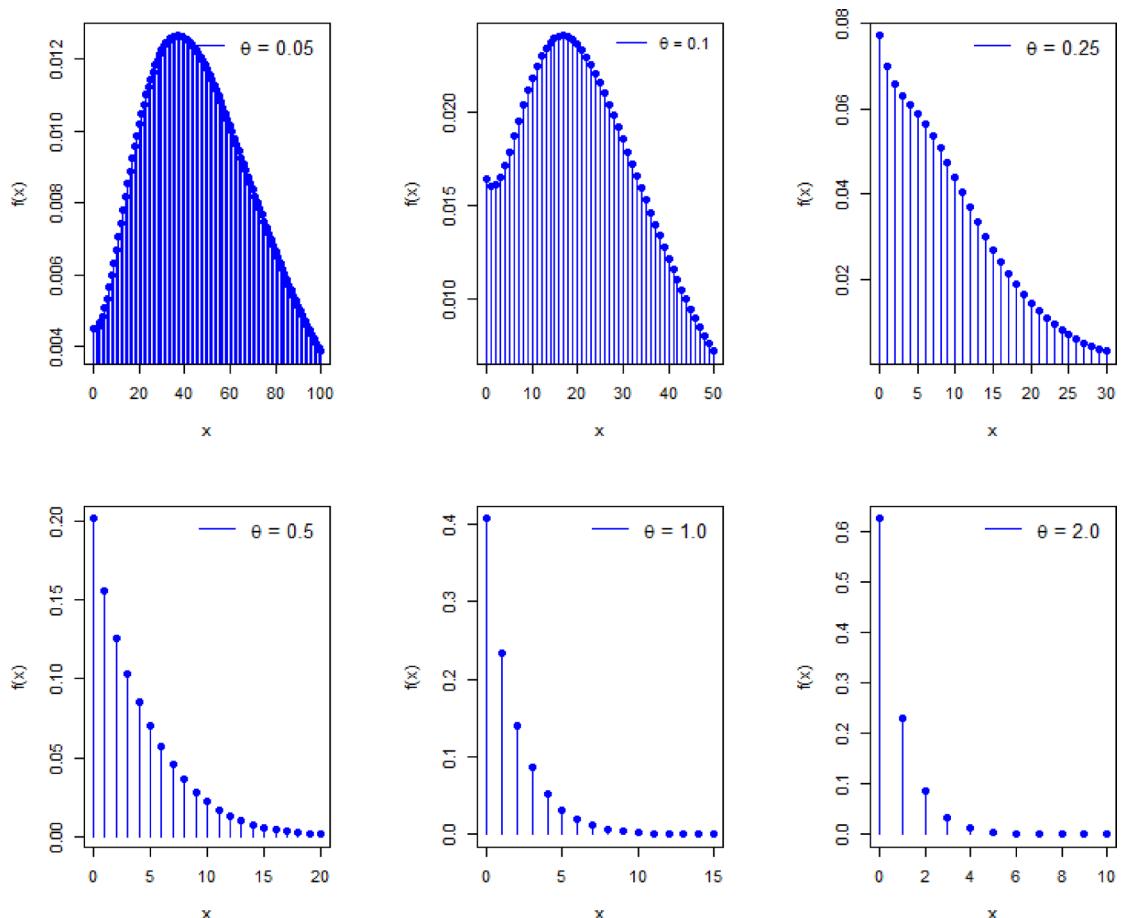
Equation (3) provides the expression for the cumulative distribution function (cdf) of the PH distribution.

$$F(x) = 1 - \frac{(2+\theta)(x^2\theta + x(2+5\theta) + 2(1+\theta)(5+\theta(3+\theta)))}{2(1+\theta)^{5+x}}. \quad (3)$$

Using the Eq. (3), the survival function is gained and given as

$$S(x) = \frac{2+\theta(x^2\theta + x(2+5\theta) + 2(1+\theta)(5+\theta(3+\theta)))}{2(1+\theta)^{5+x}}. \quad (4)$$

The hazard function also known as the failure rate is defined by taking the ratio of pmf to the survival function. The failure rate shows how the risk of an event changes over time. An increasing failure rate makes events happen more frequently over time like machines breaking down. A decreasing failure rate indicates events occur earlier



**Fig. 1.** Density plots for different choices of parameters.

in the time when products tend to break down soon after production. The hazard rate of PH distribution is given by

$$h(x) = \frac{\theta^2 (4 + \theta (x(3+x) + 2(6 + \theta (4 + \theta))))}{2 + \theta (x^2\theta + x(2 + 5\theta) + 2(1 + \theta)(5 + \theta (3 + \theta)))}. \quad (5)$$

The reverse hazard rate (also called the past failure rate) is a reliability metric that measures the probability of failure in a small interval before time  $t$ . It tells us how likely it is that an event has already occurred by a certain time, which is useful for early detection in areas like equipment monitoring or healthcare. Mathematically, it is defined as the ratio of pmf to the distribution function and is given as

$$rh(x) = \frac{\theta^2 (4 + \theta (x(3+x) + 2(6 + \theta (4 + \theta))))}{2(1 + \theta)^{5+x} - 2 + \theta (x^2\theta + x(2 + 5\theta) + 2(1 + \theta)(5 + \theta (3 + \theta)))} \quad (6)$$

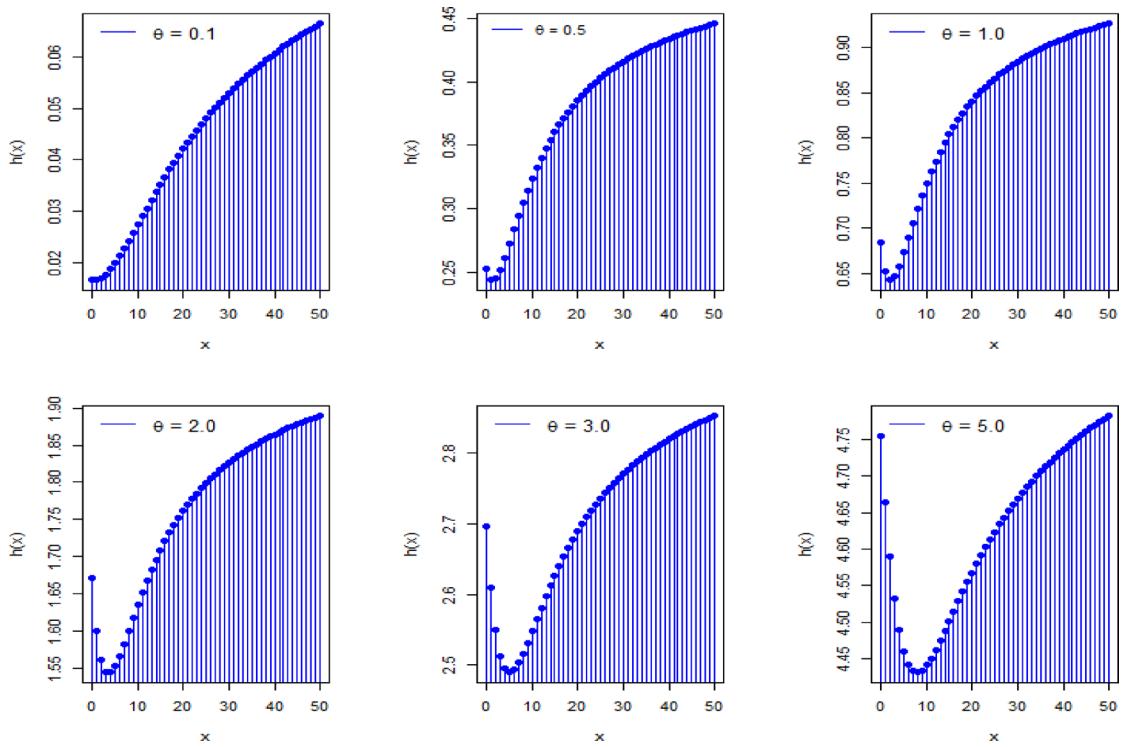
Figure 2 presents visual representations of the hazard function of PH distribution using different parameter values. The distribution curve for  $h(x)$  undergoes distinct shape modifications as the parameter  $\theta$  value increases. The hazard rate exhibits an upward trend in its pattern when  $\theta$  maintains small values. The hazard rate shows a heightened steepness when parameter values fall within moderate ranges based on the curve changes. For larger  $\theta$ , the shape shifts to a bathtub form: hazard rate first dips and then rises again.

### Mathematical characteristics of PH distribution

In this section, we derive many essential mathematical aspects of the Poisson Haq distribution and conduct a thorough investigation of their behavior.

#### Moments

The  $r^{\text{th}}$  factorial moments of random variable  $X$  can be obtained as



**Fig. 2.** Failure rate visuals for various parameter selections.

$$\begin{aligned}
 u'_{(r)} &= E [E (X^{(r)} | \lambda)] \\
 &= \left( \frac{\theta}{1+\theta} \right)^2 \int_0^\infty \left( \sum_{x=0}^\infty x^r \frac{e^{-\lambda} \lambda^x}{x!} \right) \left( 2 + \theta + \frac{\theta \lambda^2}{2} \right) e^{-\theta \lambda} d\lambda, \\
 \text{where } X^{(r)} &= X(X-1)(X-2)\dots(X-r+1) \\
 u'_{(r)} &= \frac{\theta^2}{(1+\theta)^2} \int_0^\infty \lambda^r \left( \sum_{x=r}^\infty \frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!} \right) \left( 2 + \theta + \frac{\theta \lambda^2}{2} \right) e^{-\theta \lambda} d\lambda, \\
 \text{Let } y &= x-r, \text{ we get the following.} \\
 u'_{(r)} &= \frac{\theta^2}{(1+\theta)^2} \int_0^\infty \lambda^r \left( \sum_{y=0}^\infty \frac{e^{-\lambda} \lambda^y}{y!} \right) \left( 2 + \theta + \frac{\theta \lambda^2}{2} \right) e^{-\theta \lambda} d\lambda, \\
 u'_{(r)} &= \frac{\theta^2}{(1+\theta)^2} \left[ (2+\theta) \int_0^\infty \lambda^r e^{-\theta \lambda} d\lambda + \frac{\theta}{2} \int_0^\infty \lambda^{r+2} e^{-\theta \lambda} d\lambda \right], \\
 u'_{(r)} &= \frac{\theta^2}{(1+\theta)^2} \left[ (2+\theta) \frac{r!}{\theta^{r+1}} + \frac{\theta}{2} \frac{(r+2)!}{\theta^{r+3}} \right], \\
 u'_{(r)} &= \left[ r! \left( \frac{2\theta (2+\theta) + (r+2)(r+1)}{2\theta^r (1+\theta)^2} \right) \right]; \quad \theta > 0
 \end{aligned} \tag{7}$$

### Moment generating function

The moment-generating function can be derived as

$$M_x(t) = \sum_{x=0}^\infty e^{tx} P(x),$$

$$\begin{aligned}
M_x(t) &= \sum_{x=0}^{\infty} e^{tx} \left[ \frac{\theta^2(2+\theta)}{(1+\theta)^{x+3}} + \frac{\theta^3(2+x)(1+x)}{2(1+\theta)^{x+5}} \right], \\
M_x(t) &= \frac{\theta^2(2+\theta)}{(1+\theta)^3} \left[ \sum_{x=0}^{\infty} \frac{e^{tx}}{(1+\theta)^x} \right] + \frac{\theta^3}{2(1+\theta)^5} \left[ \sum_{x=0}^{\infty} \frac{e^{tx}(2+x)(1+x)}{(1+\theta)^x} \right], \\
M_x(t) &= \frac{\theta^2(2+\theta)}{(1+\theta)^3} \left[ \sum_{x=0}^{\infty} \frac{e^{tx}}{(1+\theta)^x} \right] + \frac{\theta^3}{2(1+\theta)^5} \left[ 2 \sum_{x=0}^{\infty} \frac{e^{tx}}{(1+\theta)^x} + 3 \sum_{x=0}^{\infty} \frac{xe^{tx}}{(1+\theta)^x} + \sum_{x=0}^{\infty} \frac{x^2e^{tx}}{(1+\theta)^x} \right],
\end{aligned}$$

Using the geometric series formula  $\sum_{x=0}^{\infty} ar^x = \frac{a}{1-r}$  and after simplification, we obtain the following result.

$$\sum_{x=0}^{\infty} \frac{e^{tx}}{(1+\theta)^x} = \frac{1+\theta}{1-e^t+\theta}; \quad \sum_{x=0}^{\infty} \frac{3xe^{tx}}{(1+\theta)^x} = \frac{3e^t(1+\theta)}{(1-e^t+\theta)^2}; \quad \sum_{x=0}^{\infty} \frac{x^2e^{tx}}{(1+\theta)^x} = \frac{e^t(1+\theta)(1+e^t+\theta)}{(1-e^t+\theta)^3}$$

After simplification,

$$M_x(t) = \frac{\theta^2 (2e^t(1+\theta)(2+\theta) - 2 - e^{2t}(2+\theta) - \theta(6+\theta(4+\theta)))}{(e^t - \theta - 1)^3(1+\theta)^2}. \quad t \neq \ln(1+\theta) \quad (8)$$

Similarly, the characteristic function (CF) and the probability-generating function of the PH distribution are derived using the same approach as the moment-generating function and are given below, respectively.

$$\varphi_x(it) = \frac{\theta^2 (2e^{it}(1+\theta)(2+\theta) - 2 - e^{2it}(2+\theta) - \theta(6+\theta(4+\theta)))}{(e^{it} - \theta - 1)^3(1+\theta)^2}. \quad it \neq \ln(1+\theta) \quad (9)$$

and

$$P_x(t) = \frac{\theta^2 (2t(1+\theta)(2+\theta) - 2 - t(2+\theta) - \theta(6+\theta(4+\theta)))}{(t - \theta - 1)^3(1+\theta)^2}. \quad t \neq (1+\theta) \quad (10)$$

The first four moments about the origin are

$$\begin{aligned}
E(X) &= \frac{\theta^2 + 2\theta + 3}{\theta(1+\theta)^2}, \\
E(X^2) &= \frac{12 + 7\theta + 4\theta^2 + \theta^3}{\theta^2(1+\theta)^2}, \\
E(X^3) &= \frac{60 + 48\theta + 21\theta^2 + 8\theta^3 + \theta^4}{\theta^3(1+\theta)^2}, \\
E(X^4) &= \frac{360 + 408\theta + 180\theta^2 + 67\theta^3 + 16\theta^4 + \theta^5}{\theta^4(1+\theta)^2},
\end{aligned}$$

Variance (*Var*) and Index of dispersion (*ID*) of PH distribution are given by

$$\begin{aligned}
Var(X) &= E(X^2) - (E(X))^2, \\
Var(X) &= \frac{3 + 19\theta + 20\theta^2 + 12\theta^3 + 5\theta^4 + \theta^5}{\theta^2(1+\theta)^4},
\end{aligned}$$

and

$$ID(X) = \frac{Var(X)}{Mean(X)} = \frac{3 + 19\theta + 20\theta^2 + 12\theta^3 + 5\theta^4 + \theta^5}{\theta(1+\theta)^2(3 + 2\theta + \theta^2)}.$$

The coefficient of skewness (CS) and the coefficient of kurtosis (CK) can be determined using the following formulas.

$$CS(X) = \frac{E(X^3) - 3E(X^2)E(X) + 2(E(X))^3}{(E(X^2))^{\frac{3}{2}}},$$

and

$$CK(X) = \frac{E(X^4) - 4E(X^3)E(X) + 6E(X^2)(E(X))^2 - 3(E(X))^4}{(E(X^2))^2}.$$

- Table 1 shows that as  $\theta$  increases, the mean decreases rapidly, while the variance decreases more slowly, resulting in a consistently high Index of Dispersion (ID). This constant overdispersion reflects the typical behavior of real-world count data involving rare but extreme events, such as insurance claims, equipment failures, or hospital admissions.
- At the same time, both the coefficient of skewness (CS) and coefficient of kurtosis (CK) increase significantly with  $\theta$ . The increase in skewness indicates a growing asymmetry, with a longer right tail, meaning that large counts become increasingly probable relative to the mean. The sharp increase in kurtosis signifies heavier tails and a higher peak, characteristic of datasets with a preponderance of small counts punctuated by occasional large outliers.

## Parameter Estimation

This section is based on parameter estimation using two widely recognized estimation approaches: the method of moments and maximum likelihood estimation. Both approaches are explored in detail, providing insight into their usefulness in estimating the parameters of the PH distribution. Furthermore, a detailed simulation study is carried out to assess the behavior accuracy of these estimates under different scenarios. This comprehensive analysis aids in understanding the practical performance of estimation techniques in real-world applications.

### Maximum likelihood estimator (MLE)

The log-likelihood function for the PH distribution can be written as

$$l(\theta) = 2n\log(\theta) - \log(2) - \log(1+\theta) \sum_{i=1}^n (x_i + 5) + \sum_{i=1}^n \log(2(2+\theta)(1+\theta)^2 + \theta(2+x_i)(1+x_i)). \quad (11)$$

To optimize the equation above, we calculate the partial derivative concerning  $\theta$  and obtain

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{2n}{\theta} - \frac{1}{(1+\theta)} \sum_{i=1}^n (x_i + 5) + \sum_{i=1}^n \frac{(1+x_i)(2+x_i) + 2(1+\theta)^2 + 4(1+\theta)(2+\theta)}{(1+x_i)(2+x_i)\theta + 2(1+\theta)^2(2+\theta)}. \quad (12)$$

$\theta$	Mean	Var	CS	CK	CV	ID
0.05	56.2812	1316.83	1.10002	4.83147	0.64476	23.3973
0.50	3.77778	15.2894	1.57280	6.27145	1.03486	4.04575
1.00	1.50000	3.75000	1.96231	8.30667	1.29099	2.50000
1.50	0.88000	1.70560	2.16994	9.60034	1.48406	1.93818
2.00	0.61111	1.01543	2.30112	10.3951	1.64894	1.66162
2.50	0.46531	0.69941	2.40410	10.9577	1.79733	1.50312
3.00	0.37500	0.52604	2.49702	11.4299	1.93410	1.40278
3.50	0.31393	0.41896	2.58610	11.8742	2.06181	1.33454
4.00	0.27000	0.34710	2.67325	12.3152	2.18204	1.28556
4.50	0.23691	0.29588	2.75896	12.7615	2.29595	1.24887
5.00	0.21111	0.25765	2.84326	13.2156	2.40441	1.22048
6.00	0.17347	0.20460	3.00738	14.1455	2.60755	1.17947
7.00	0.14732	0.16962	3.16527	15.0977	2.79561	1.15138
8.00	0.12809	0.14486	3.31700	16.0650	2.97146	1.13095
9.00	0.11333	0.12641	3.46290	17.0421	3.13719	1.11542
10.0	0.10165	0.11215	3.60340	18.0260	3.29437	1.10323
11.0	0.09217	0.10078	3.73894	19.0143	3.44419	1.09338
15.0	0.06719	0.07174	4.23937	22.9911	3.98639	1.06770
20.0	0.05023	0.05276	4.79293	27.9828	4.57320	1.05045
30.0	0.03340	0.03452	5.74253	37.9813	5.56235	1.03347

**Table 1.** Mean, var, CS, CK, CV, and ID of the PH distribution.

The ML estimator may be produced by numerically solving Eq. (10), which cannot be represented simply in terms of parameters.

### Method of moments estimator (MME)

The MME can be obtained by setting the population mean equal to the sample mean. Thus, the MME of  $\theta$ , denoted as  $\hat{\theta}$ , is derived by solving the following equation.

$$\bar{x} = \frac{\hat{\theta}^2 + 2\hat{\theta} + 3}{\hat{\theta} (1 + \hat{\theta})^2}. \quad (13)$$

### Simulation

We present here a comprehensive simulation study to assess the performance of both estimation methods discussed in the previous subsection. To investigate how the estimators behave under different sample sizes, both small and large sample sizes are considered. Specifically, we considered the following sample sizes of  $n = 25, 50, 100, 200$ , and  $300$  to generate a sample. The simulation process is repeated  $N = 10,000$  times to ensure robust results. For each generated sample, we calculate the four key performance metrics to evaluate the behavior of estimators: The average estimate (AE), absolute bias (AB), mean relative error (MRE), and mean square error (MSE).

$$AB = \frac{1}{N} \sum_{i=1}^N |\hat{\theta} - \theta|, \quad MRE = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{\theta} - \theta|}{\theta}, \quad \text{and} \quad MSE = \frac{1}{N} \sum_{i=1}^N (\hat{\theta} - \theta)^2.$$

The calculated values for each of these performance measures are summarized in Table 2. Additionally, the heatmap based on AB, MRE, and MSE measures are presented in Fig. 3.

Table 2; Fig. 3 present simulation findings for the performance of MLE and MME estimation methods for different samples and various choices of parameter  $\theta$ . It is observed that.

- In the majority of cases, the maximum likelihood estimation approach provides smaller values of AB, and MRE compared to MME, indicating that MLE tends to be more accurate, especially for higher parameter values and sample sizes.
- As the sample size increases both estimators tend towards the true value of the parameter. This is a characteristic behavior for estimators that tend to converge to the true parameter.
- As the parameter  $\theta$  increases, the difference between MME and MLE becomes more noticeable, with MLE consistently showing smaller errors.
- The MLE and MME estimators show a reduction in AB and MSE as the sample size increases. However, MME tends to have slightly larger AB and MSE values than MLE, particularly for higher values of parameter.

### Possion Haq regression model

A new count regression model based on Poisson Haq distribution is proposed in this section. The PMF is defined in terms of parameter  $\theta$ , which is a transformation of the mean parameter  $\mu > 0$ . The transformation  $\theta = \theta(\mu)$  is presented by:

$$\theta(\mu) = -\frac{2\mu - 1}{3\mu} + \frac{\mu^2 + 2\mu + 1}{3\mu(A)^{1/3}} + \frac{(A)^{1/3}}{3\mu}$$

where  $A = 1 + 3\mu + 30\mu^2 + \mu^3 + 3\sqrt{3}\sqrt{2\mu^2 + 6\mu^3 + 33\mu^4 + 2\mu^5}$ . (Proof is given in Appendix).

The PMF of the PHaq distribution for count variable  $Y$  is articulated as:

$$P(Y = y | \mu) = 0.5 \left( \frac{\theta}{1 + \theta} \right)^2 \frac{2(2 + \theta)(1 + \theta)^2 + \theta(2 + y)(1 + y)}{2(1 + \theta)^{y+3}}, \quad y = 0, 1, 2, \dots$$

We denote this probability model  $Y(\theta, \mu)$ , where  $\mu$  serves as the mean parameter.

Assume the count response variable  $Y_i$  for the  $i$ -th observation follows the  $PH(\theta, \mu)$  model. Let  $E(Y) = \mu_i$  and to relate the mean  $\mu_i$  to the random variable, we utilize the log-link function:

$$\mu_i = \exp(\beta x_i^T), \quad i = 0, 1, 2, \dots$$

where  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  is the vector of regression coefficients, and  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$  is the vector of explanatory variables for the  $i$ -th observation. Substituting the log-link function  $\mu_i = \exp(\beta x_i^T)$  and re-parameterization  $\theta(\mu)$ , we obtain the regression model that can be fitted to real-world count data. The likelihood function of the new count regression model is given by

Parameter	n	MLE				MME			
		$E(\hat{\theta})$	$AB(\hat{\theta})$	$MRE(\hat{\theta})$	$MSE(\hat{\theta})$	$E(\hat{\theta})$	$AB(\hat{\theta})$	$MRE(\hat{\theta})$	$MSE(\hat{\theta})$
$\theta = 0.5$	25	0.5333	0.0333	0.0666	0.0099	0.5317	0.0758	0.1517	0.0098
	50	0.5240	0.0240	0.0481	0.0045	0.5248	0.0508	0.1016	0.0046
	100	0.5190	0.0190	0.0380	0.0022	0.5190	0.0373	0.0747	0.0023
	200	0.5174	0.0174	0.0347	0.0012	0.5175	0.0276	0.0553	0.0013
	300	0.5162	0.0162	0.0324	0.0009	0.5162	0.0237	0.0474	0.0009
$\theta = 1.0$	25	1.1328	0.1328	0.1328	0.0909	1.1182	0.2075	0.2075	0.0903
	50	1.0987	0.0987	0.0987	0.0402	1.1030	0.1531	0.1531	0.0410
	100	1.0810	0.0810	0.0810	0.0205	1.0775	0.1084	0.1084	0.0198
	200	1.0751	0.0751	0.0751	0.0123	1.0715	0.0847	0.0847	0.0113
	300	1.0732	0.0732	0.0732	0.0097	1.0723	0.0802	0.0802	0.0097
$\theta = 1.5$	25	1.7959	0.2959	0.1972	0.4404	1.7848	0.4223	0.2815	0.3944
	50	1.7073	0.2073	0.1382	0.1567	1.7229	0.2963	0.1975	0.1748
	100	1.6759	0.1759	0.1173	0.0826	1.6643	0.2087	0.1391	0.0751
	200	1.6574	0.1574	0.1049	0.0476	1.6601	0.1772	0.1181	0.0497
	300	1.6518	0.1518	0.1012	0.0383	1.6511	0.1605	0.1070	0.0383
$\theta = 2.0$	25	2.5386	0.5386	0.2693	1.3454	2.4957	0.6827	0.3413	1.1850
	50	2.3448	0.3448	0.1724	0.4592	2.3648	0.4699	0.2350	0.4627
	100	2.2895	0.2895	0.1447	0.2138	2.2886	0.3476	0.1738	0.2091
	200	2.2448	0.2448	0.1224	0.1196	2.2503	0.2793	0.1397	0.1194
	300	2.2361	0.2361	0.1180	0.0946	2.2492	0.2595	0.1297	0.1004
$\theta = 2.5$	25	3.2719	0.7719	0.3088	2.5904	3.3809	1.1391	0.4556	4.7998
	50	2.9931	0.4931	0.1972	0.9343	3.0060	0.6672	0.2669	0.9549
	100	2.8883	0.3883	0.1553	0.4216	2.8965	0.4909	0.1964	0.4399
	200	2.8426	0.3426	0.1370	0.2384	2.8571	0.3937	0.1575	0.2629
	300	2.8285	0.3285	0.1314	0.1869	2.8241	0.3473	0.1389	0.1862
$\theta = 3.0$	25	3.9815	0.9815	0.3272	4.2620	4.1697	1.5121	0.5040	7.7775
	50	3.6994	0.6994	0.2331	1.8749	3.6636	0.8976	0.2992	1.8394
	100	3.4898	0.4898	0.1633	0.7736	3.5343	0.6504	0.2168	0.8134
	200	3.4288	0.4288	0.1429	0.4025	3.4305	0.4939	0.1646	0.4129
	300	3.4056	0.4056	0.1352	0.3032	3.4186	0.4496	0.1499	0.3110

**Table 2.** Parameter estimates, bias, mres, and MSEs of PH distribution for parameters.

$$\log L(\beta) = \sum_{i=1}^n \log \left\{ 0.5 \left( \frac{\theta_i}{1+\theta_i} \right)^2 \frac{2(2+\theta_i)(1+\theta_i)^2 + \theta_i(2+y_i)(1+y_i)}{2(1+\theta_i)^{y_i+3}} \right\}.$$

where  $\theta_i = \theta(\mu_i)$  and  $\mu_i = \exp(\beta x_i^T)$ .

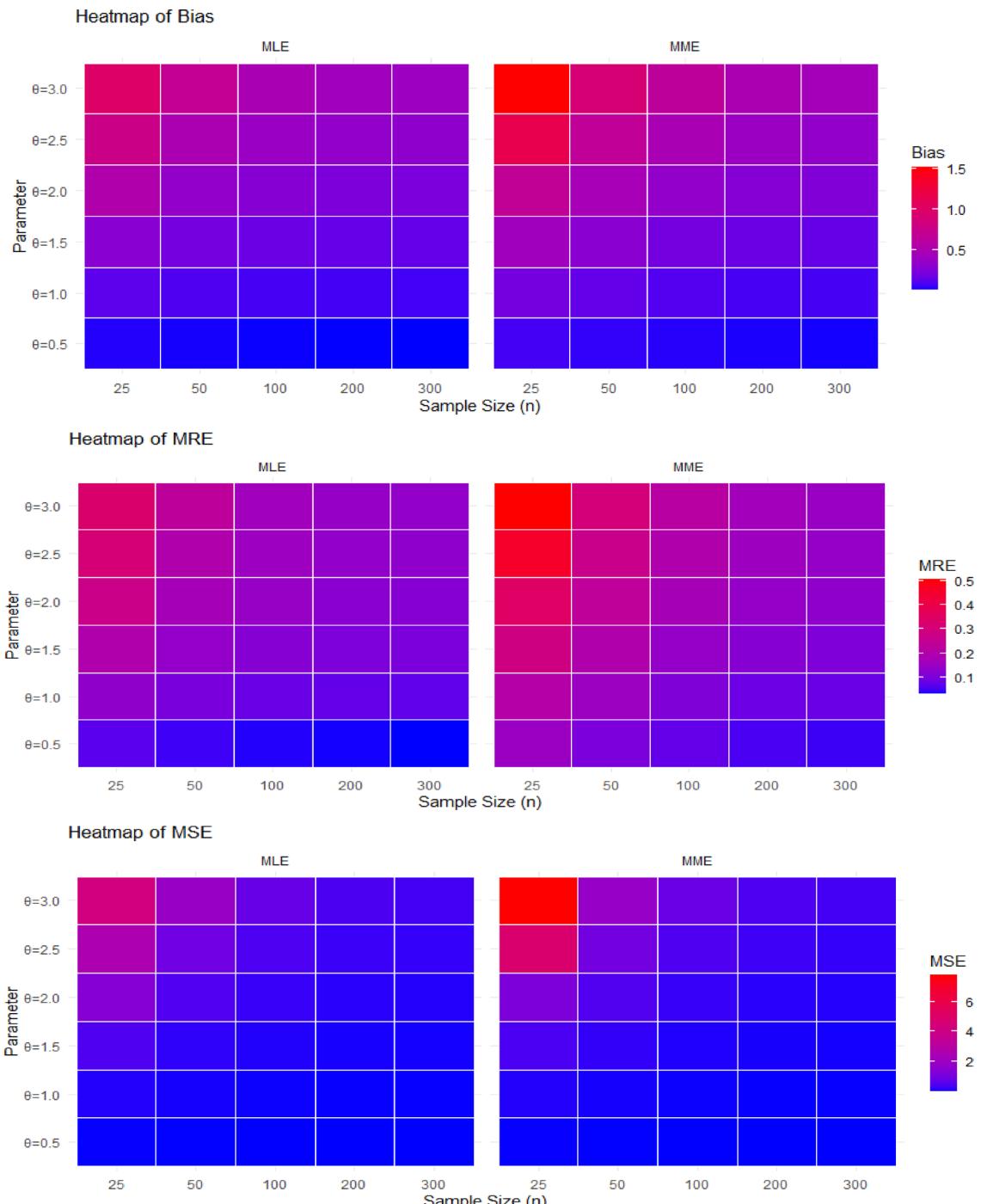
Now the estimates of the proposed regression parameters by maximizing the above log-likelihood function with respect to  $\beta$  using numerical optimization methods using R software.

### Data applications

In this section, we compare the new count distribution with widely recognized discrete probability distributions to assess its applicability and adequacy. To facilitate the analysis, we utilized three datasets from different domains: one is the number of mammalian cytogenetic dosimetry lesions, and the other examines remission times (in months) of Nipah virus infection. To perform the comparison, we consider different count distributions, each model offering unique features suited to different types of data. The competitive distributions include Poisson moment exponential (PME)<sup>17</sup> Poisson XLindley (PXL)<sup>18</sup> Poisson Ramos-Louzada (PRL)<sup>33</sup> Poisson entropy-based weighted exponential (PEWE)<sup>22</sup>. Additionally, we examine the standard Poisson distribution, widely used for modeling rare events. The density functions of these models are given below.

$$P(X=x; \theta) = \frac{\left(1 + \frac{1}{\theta}\right)^{-x} (x-1+\theta(\theta-1))}{(\theta-1)(1+\theta)^2}; x=0,1,2,3,\dots, \theta \geq 2,$$

$$P(X=x; \theta) = \frac{\theta^2 (x+\theta^2+3(1+\theta))}{(1+\theta)^{x+4}}; x=0,1,2,3,\dots, \theta > 0,$$



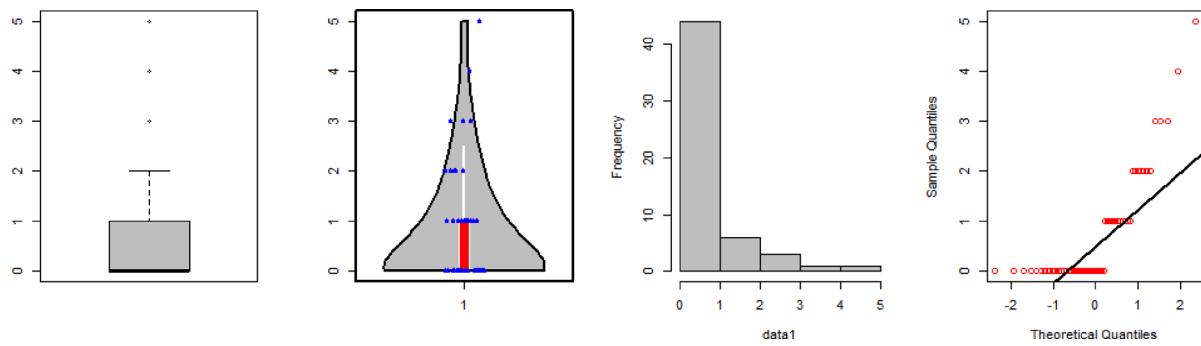
**Fig. 3.** Heatmaps for Bias, MRE, and MSE based on different choices of parameter and sample size.

$$P(X = x; \theta) = \frac{\theta^x (1+x)}{(1+\theta)^{x+2}}; x = 0, 1, 2, 3, \dots, \theta > 0,$$

$$P(X = x; \theta) = \frac{\theta ((1+\theta) \ln(\theta) - (1+x)\theta)}{(1+\theta)^{x+2} (\ln(\theta) - 1)}; x = 0, 1, 2, 3, \dots, \theta > 0,$$

and

$$P(X = x; \theta) = \frac{e^{-\theta} \theta^x}{x!}; x = 0, 1, 2, 3, \dots, \theta > 0.$$



**Fig. 4.** Descriptive plots for the Nipah Virus dataset.

X	Obs. Fr.	Exp. Fr.					
		PH	PRL	PXL	PME	PEWE	Poisson
0	33	32.319	2.0391	31.886	29.616	13.823	26.453
1	12	13.401	8.2212	13.807	16.157	14.908	19.839
2	6	5.7576	9.5403	5.9290	6.6110	10.957	7.4398
3	3	2.5247	8.7520	2.5284	2.4045	7.0041	1.8599
4	1	1.1168	7.2374	1.0718	0.8199	4.1627	0.3487
5	1	0.8810	20.210	0.7777	0.3921	5.1457	0.0596
Total	56	56	56	56	56	56	56
MLE	$\hat{\alpha}$	1.6979	2.7573	1.5406	0.3749	1.1287	0.7499
GOF Measures	$-l$	66.892	108.84	66.928	67.767	83.838	71.582
	AIC	135.78	219.67	135.86	137.53	169.68	145.16
	BIC	137.81	221.70	137.88	139.56	171.70	147.19
	$\chi^2$	0.2112	500.56	0.3219	1.5147	37.443	4.8903
	df	1.0	4.0	1.0	1.0	3.0	1.0
	p-value	0.6458	0.0000	0.5704	0.2184	0.0000	0.0270

**Table 3.** The mles, frequencies, and goodness-of-fit values for the Nipah virus dataset.

The parameters of all considered distributions are estimated using the MLE method, which ensures that the selected parameters maximize the likelihood function based on the observed data. For the determination of best-fitted probability distribution, a detailed assessment is conducted using various information criteria and goodness-of-fit measures. These measures include: the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Chi-Square goodness-of-fit. The determination of the best-fit distribution is based on the minimum values of AIC, BIC, and Chi-Square test statistic and higher the values log-likelihood and Chi-Square p-values.

### Infected patients of Nipah virus data

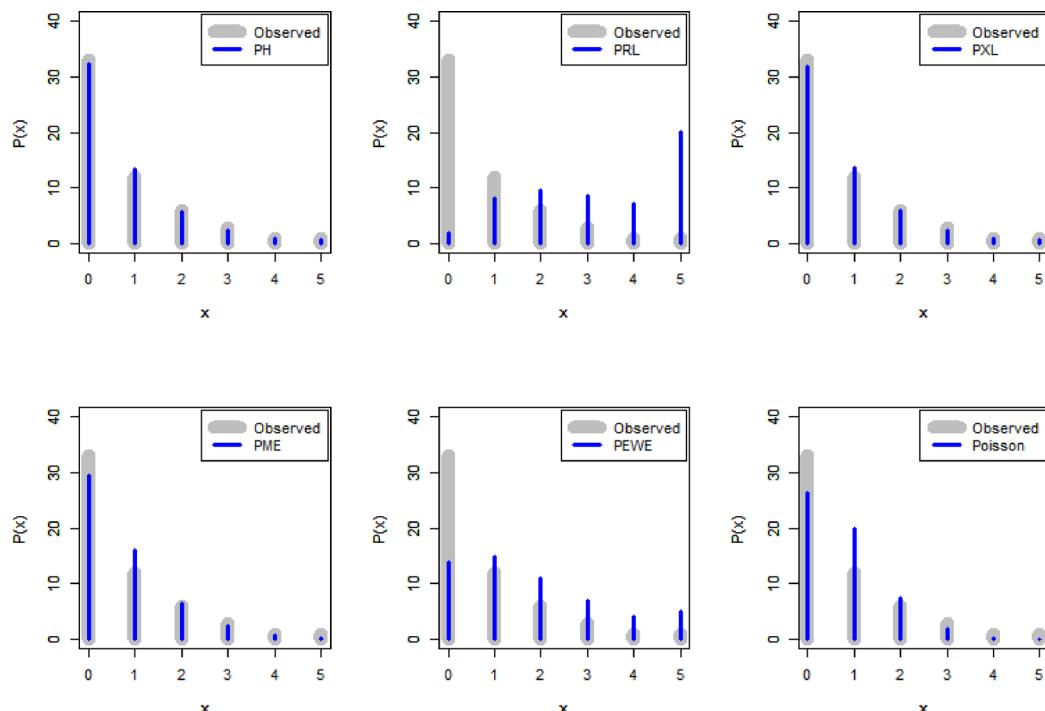
The first dataset is about the survival time (in months) of Kerela, the Indian state, resident who was infected by the Nipah virus in 2017<sup>34</sup>. Some descriptive statistics for this dataset are mean = 0.7636, variance = 0.8311, skewness = 1.8575 and kurtosis = 6.1792. Figure 4 presents various descriptive plots to visually explore the dataset, including a Violin plot to combine distributional information with summary statistics, a Boxplot to highlight the distribution and potential outliers, a Q-Q plot to assess the normality of the data by comparing it to a theoretical distribution, and a Histogram to show the frequency distribution of the data.

The parameter estimates observed frequencies (Obs. Fr.) and expected frequencies (Exp. Fr.) of all fitted distributions along with goodness-of-fit measures for the Nipah Virus dataset are given in Table 3.

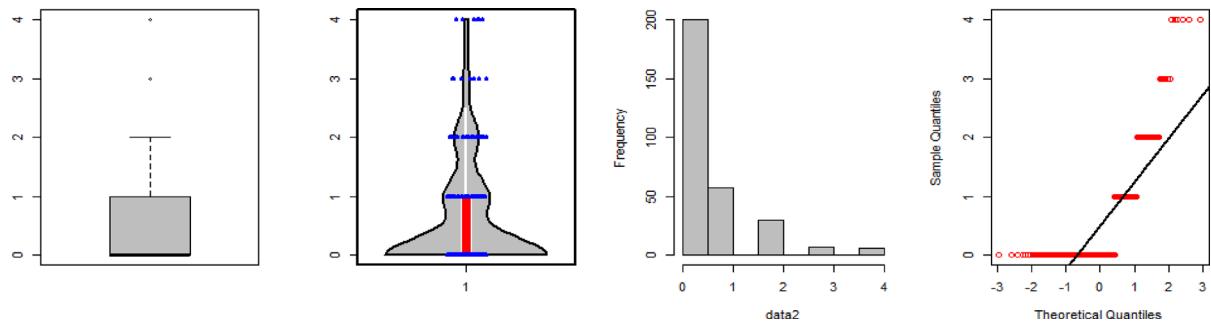
Table 3 reveals that the PH distribution achieves the smallest AIC, BIC, and Chi-Square test statistic values compared to other distributions, indicating that it offers the best fit to the observed data among all the competitive distributions. Furthermore, Fig. 5 presents a comparative visualization of the Obs. and Exp. Frequencies for the considered count models. This graphical representation allows for a clearer understanding of how well each probability distribution aligns with real-life data.

### Mammalian cytogenetic dosimetry lesions data

The second dataset represents the number of mammalian cytogenetic dosimetry lesions induced by exposure to streptogramin (NSC-45383) in rabbit lymphoblasts at a dosage of 70 3bc g/kg<sup>35</sup>. We first computed some



**Fig. 5.** Obs. and Exp. Frequencies of all fitted distributions for the Nipah Virus dataset.



**Fig. 6.** Descriptive plots for the cytogenetic dosimetry lesions dataset.

descriptive metrics of these datasets to illustrate their behavior such as the mean lesion count is 0.5400 with a variance of 1.3319, the skewness is 1.7109, and the coefficient of kurtosis is 5.6532. These descriptive measures indicate a positively skewed with heavy tails, suggesting a relatively low frequency of extreme lesion counts.

Additionally, to better understand these patterns some descriptive plots including Boxplot, Histogram, Violin, and Q-Q plots are generated and are listed in Fig. 6. These descriptive plots complement the numerical summaries by highlighting the overall pattern, central tendency, shape, and outliers of the data.

The MLEs, Obs. and Exp. Frequencies and goodness-of-fit metrics for cytogenetic dosimetry lesions data are presented in Table 4. Further, the Obs. and Exp. Frequencies are also generated for all fitted distributions and listed in Fig. 7.

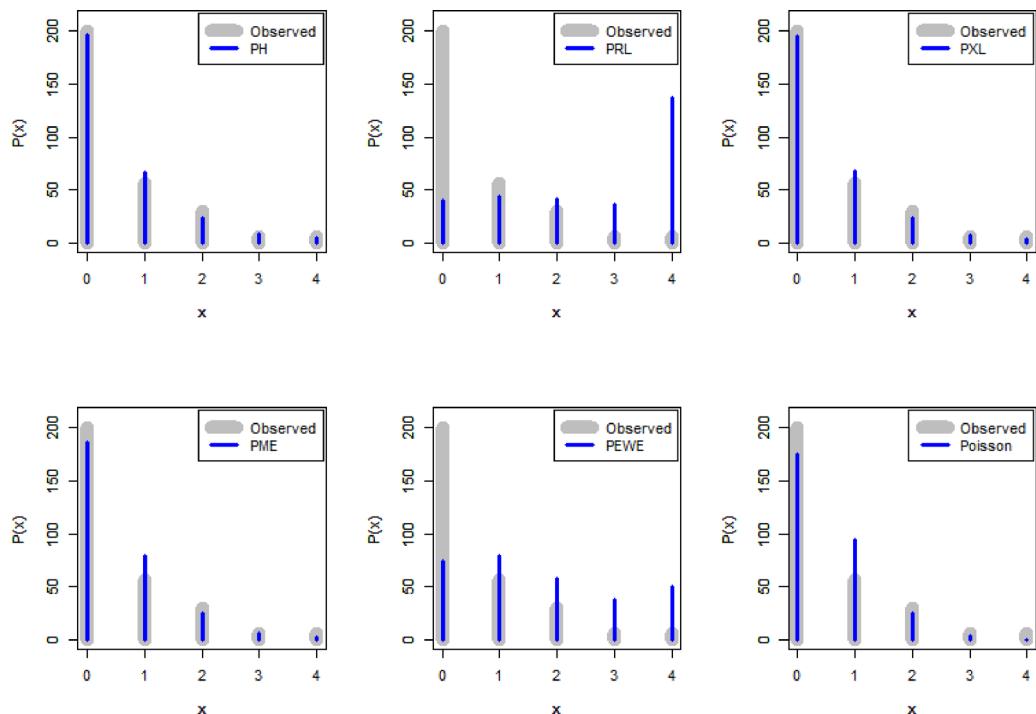
It is observed that the Poisson-Haq distribution efficiently analyzes this dataset as compared to other distributions.

### Application of the PH regression model

Now we take into account the dataset that was reported in<sup>36</sup>. The information shows the length of cardiovascular patients' stay at the hospital. A total of 3589 observations were taken into account using the COUNT package in R software. The considered dataset is known as AZPRO data. The purpose of this study is to examine how the response variable "length of patients' stays at the hospital (los)", is impacted by the variables such as cardiovascular procedure (procedure), gender (sex), Admission type (admit), and age. The barplot of the quantity of doctor visits is shown in Fig. 8.

Now by utilizing the following log-link function to associate the variables with the response mean.

X	Obs. Fr.	Exp. Fr.					
		PH	PRL	PXL	PME	PEWE	Poisson
0	200	195.9764	40.267	194.5782	186.0004	74.266	174.8245
1	57	66.9847	44.300	68.5440	79.0868	79.276	94.4052
2	30	23.6390	41.910	24.0289	25.2206	58.397	25.4894
3	7	8.5153	36.708	8.3879	7.1491	37.515	4.5881
4+	6	4.8846	136.81	4.4611	2.5431	50.547	0.6928
Total	300	300	300	300	300	300	300
MLE	$\hat{\alpha}$	2.2094	2.8186	2.0512	0.2700	1.1133	0.5400
GOF Measures	$-l$	299.21	575.24	299.31	302.41	434.29	314.23
	AIC	600.42	1152.5	600.63	606.83	870.59	630.45
	BIC	604.13	1156.2	604.33	610.53	874.29	634.16
	$\chi^2$	3.2945	789.78	3.5809	9.2566	297.02	30.527
	df	2.0	3.0	2.0	2.0	3.0	2.0
P-value		0.1925	0.000	0.1668	0.0097	0.0000	0.0000

**Table 4.** The mles, frequencies, and goodness-of-fit values for the cytogenetic dosimetry lesions dataset.**Fig. 7.** Obs. and Exp. Frequencies of all fitted distributions for the cytogenetic dosimetry lesions dataset.

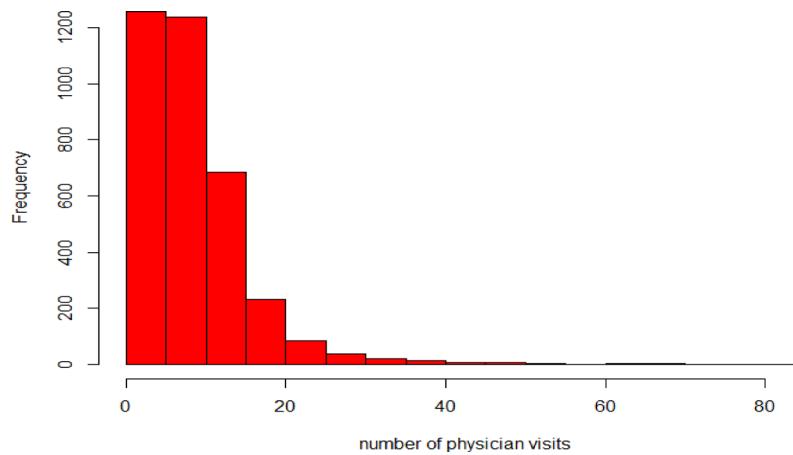
$$\log(\mu_i) = \beta_0 + \beta_1 * \text{procedure} + \beta_2 * \text{sex} + \beta_3 * \text{admit} + \beta_4 * \text{age}$$

The data is fitted with Poisson regression, Poisson-transmuted record type exponential (PTRTE)<sup>37</sup> and the proposed PH regression model. The parameter estimates and model selection measures are presented in Table 5.

Table 5 displays all findings of the proposed regression model along with considered competitive regression models. It is conclusive that all the competitive regression models are significant, as the p-value for each model is less than 5%. Furthermore, the table also confirms that the PH regression model fits the data best with the highest log-likelihood and the lowest AIC and BIC scores as compared to others.

### Bayesian analysis

In this section, the Bayesian estimation approach is employed to estimate the parameters of the new count distribution. Bayesian technique offers a powerful framework for parameter estimation by incorporating prior information about the parameter(s) in the system of prior distribution(s). For this purpose, we assume gamma prior to the parameter  $\theta$ , reflecting our preliminary identification or convention about its possible values. The

**Fig. 8.** The bar plot of the number of doctor visits.

Para.	PH		PTRTE		Poisson	
	Esti. (SE)	p-value	Esti. (SE)	p-value	Esti. (SE)	p-value
$\beta_0$	1.4060(0.0381)	< 0.0001	2.9942(0.0429)	< 0.0001	1.4560(0.0158)	< 0.0001
$\beta_1$	0.9580(0.0294)	< 0.0001	0.9655(0.0362)	< 0.0001	0.9603(0.0122)	< 0.0001
$\beta_2$	-0.1256(0.0306)	< 0.0001	-0.1255(0.0364)	< 0.0001	-0.1239(0.0118)	< 0.0001
$\beta_3$	0.3683(0.0305)	< 0.0001	0.3940(0.0366)	< 0.0001	0.3266(0.0121)	< 0.0001
$\beta_4$	0.1190(0.0324)	< 0.0001	-0.0107(0.0043)	< 0.0001	0.1222(0.0124)	< 0.0001
$\alpha$	-	-	4.7501(0.1664)	-	-	-
$-l$	10,822		11,164		11,190	
AIC	21,653		22,338		22,390	
BIC	21,684		22,369		22,421	

**Table 5.** The findings of the PH regression model.

posterior density combines both observed data and prior information and is obtained using Bayes' theorem. This density involves the multiplication of the likelihood function and prior information. The resultant posterior distribution delivers an updated estimate of the parameter  $\theta$  that combines both prior information and the evidence from the data.

$$p(\theta | x) \propto L(\theta | x) p(\theta)$$

The gamma distribution is widely considered prior distribution. It is suitable due to its conjugation with many likelihood functions commonly used in the Bayesian paradigm.

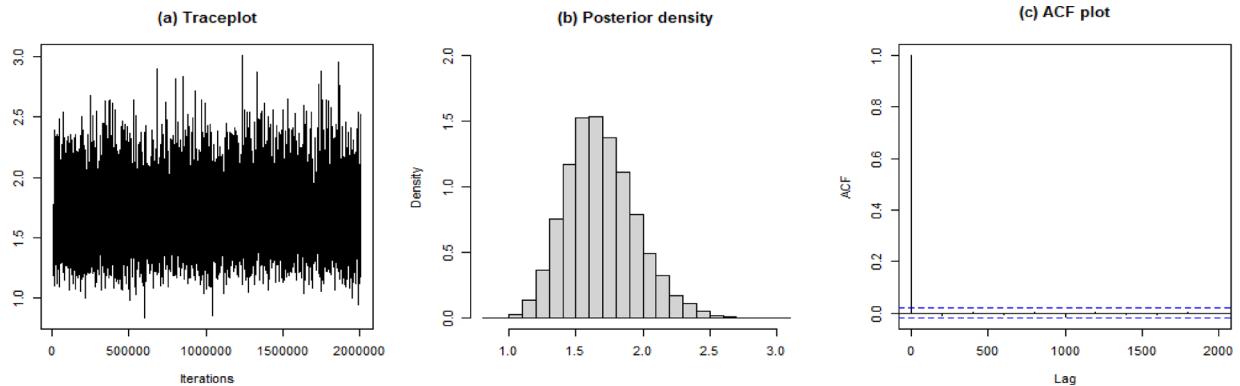
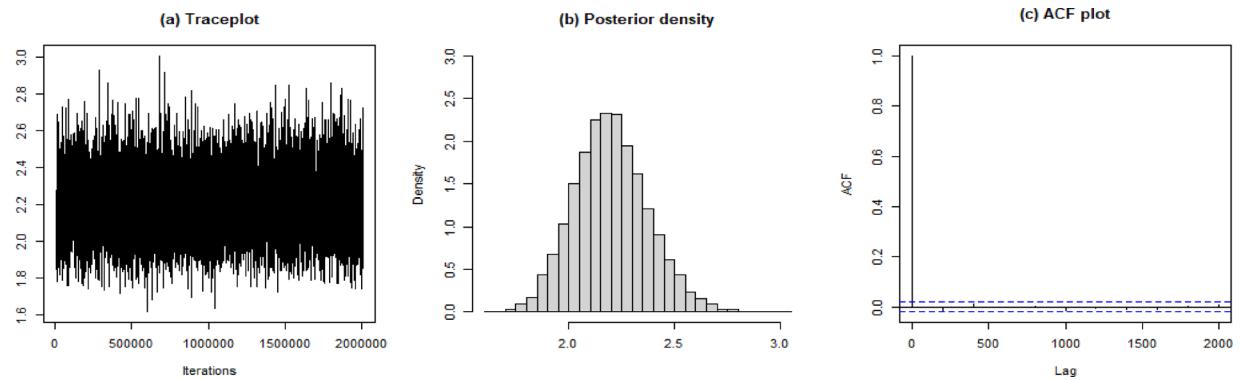
The posterior density for parameter  $\theta$  is given by

$$p(\theta | x) = \frac{b^a \theta^{a-1} e^{-\theta b}}{\Gamma(a)} \prod_{i=1}^n \frac{\theta^2}{(1+\theta)^{x_i+3}} \left[ \frac{2(2+\theta)(1+\theta)^2 + \theta(2+x_i)(1+x_i)}{2(1+\theta)^2} \right]. \quad (14)$$

It is easily seen from Eq. (11) that the posterior density is not available in closed form, necessitating the use of computational techniques for parameter estimation. To acquire posterior summaries of interest, we utilized the Markov Chain Monte Carlo (MCMC) technique which is well-suited for sampling from complex posterior density functions.

The simulation program generated 1,007,000 samples from the joint posterior distribution. The first 7,000 samples are discarded as a burn-in period to mitigate the impact of seed values on the final parameter estimates. A thinning interval of 200 is utilized to reduce autocorrelation among successive samples, ensuring the resulting set of samples is approximately independent. The mean of these samples is utilized to compute Bayes estimates. To ensure the accuracy and validity of results, coverage diagnostics are performed. Trace plots of the sample values are examined to assess the stability of the Markov chain. Additionally, we also use the Geweke diagnostic, which uses a z-score to compare the means of two non-overlapping segments of the chain, which are normalized by the asymptotic standard error of the difference. The convergence deemed satisfactory of the absolute value of the z-score is less than 1.96. All calculations and analyses are performed using the **MCMCpack** package with the R software.

Dataset	Bayes Estimate	S.D.	S.E.	HPD	Geweke z-score
Nipah Virus	1.6873	0.2646	0.0026	(1.2109, 2.2324)	-0.8895
Lesions	2.1985	0.1740	0.0017	(1.8703, 2.5490)	0.2059

**Table 6.** The posterior summaries of the PH distribution for both datasets.**Fig. 9.** Traceplot, histogram, and ACF plot for the Nipah Virus dataset.**Fig. 10.** Traceplot, histogram, and ACF plot for the cytogenetic dosimetry lesions dataset.

The posterior summaries, including the posterior mean, standard deviation, Geweke diagnostic scores, and highest posterior density interval (HPD) intervals, are listed in Table 6. These findings deliver a comprehensive overview of the Bayesian estimates and their associated uncertainties. The posterior samples for the PH distribution parameter for both datasets are shown in Figs. 9 and 10.

## Conclusion

This study presents and examines a novel one-parameter probability distribution for count data. This novel count probability model is derived using the mixed Poisson compounding technique and named as Poisson Haq (PH) distribution. The PH distribution is capable of addressing the need for modeling count datasets exhibiting overdispersion. The failure rate of PH distribution is showing an increasing pattern.

The essential statistical and reliability characteristics of PH distribution were mathematically and numerically derived, including cumulative distribution function, survival and hazard functions, moment, and associated measures. It has been observed that the mean and variance of PH distribution decreases with an increase in parameter values. This implies that the new distribution tends more concentrated around lower values with higher values of the parameter. The coefficients of skewness and kurtosis also show an increasing pattern for higher parameter values. The distribution becomes more skewed to the right, more pronounced peak and heavier tails. The dispersion index gradually decreases for larger values of the parameter.

The distribution parameter has been estimated using the methods of moments and maximum likelihood estimation. It has been observed that as the sample size increases both estimators tend towards the true value of the parameter. The MLE and MME estimators show a reduction in Bias and MSE as the sample size increases. However, MME tends to have slightly larger Bias and MSE values than the MLE, particularly for higher values

of the parameter. Further, it has been seen that as the parameter  $\theta$  increases, the difference between MME and MLE becomes more noticeable, with MLE consistently showing smaller errors.

To demonstrate the versatility of the new distribution, three datasets related to medical science are considered. The first data is associated with the number of infected patients with the Nipah virus, the second one is about the number of mammalian cytogenetic dosimetry lesions, and the third data is related to the length of hospital stay. Comparative analyses show that the new distribution analyzed these datasets adequately as compared to considered competitive distributions. Additionally, the Bayesian estimation approach is also employed to estimate the parameter of PH distribution.

Future research aims to investigate additional modifications and applications of the PH distribution. Potential directions but not limited to, exploring actuarial measures, examining reliability characteristics like mean residual life function and entropy, and forms of truncated, zero-inflated, and neutrosophic models. Additionally, this distribution can be applied to population size estimation. These advancements are anticipated to enhance the versatility and applicability of the proposed distribution, solidifying its position as a robust and competitive model in the realm of statistical literature.

## Data availability

Data availability: All data generated or analysed during this study are included in this published article.

## Appendix

The solution of re-parameterization for  $\theta$  in terms of  $\mu$ .  
We are given the equation:

$$\frac{(\theta^2 + 2\theta + 3)}{\theta(1 + \theta)^2} = \mu.$$

Multiply both sides by the denominator:

$$\theta(1 + \theta)^2 \mu = \theta^2 + 2\theta + 3$$

Expand the left-hand side:

$$(\theta^3 + 2\theta^2 + \theta) \mu = \theta^2 + 2\theta + 3$$

Move all terms to one side:

$$\begin{aligned} \mu\theta^3 + 2\mu\theta^2 + \mu\theta - \theta^2 - 2\theta - 3 &= 0 \\ \mu\theta^3 + (2\mu - 1)\theta^2 + (\mu - 2)\theta - 3 &= 0 \end{aligned}$$

Divide through by  $\mu$ :

$$\theta^3 + \left(\frac{2\mu - 1}{\mu}\right)\theta^2 + \left(\frac{\mu - 2}{\mu}\right)\theta - \frac{3}{\mu} = 0$$

This is a cubic equation in  $\theta$ . Using the cubic formula, we get:

$$\theta(\mu) = -\frac{2\mu - 1}{3\mu} + \frac{\mu^2 + 2\mu + 1}{3\mu(A)^{1/3}} + \frac{(A)^{1/3}}{3\mu}$$

where  $A = 1 + 3\mu + 30\mu^2 + \mu^3 + 3\sqrt{3}\sqrt{2\mu^2 + 6\mu^3 + 33\mu^4 + 2\mu^5}$ .

Received: 15 January 2025; Accepted: 13 June 2025

Published online: 02 July 2025

## References

1. Saha, P., Biswas, S. K., Biswas, M. H. A. & Ghosh, U. An SEQAIHR model to study COVID-19 transmission and optimal control strategies in Hong kong. 2022. *Nonlinear Dyn.* **111**, 6873–6893 (2023).
2. Saha, P., Mondal, B. & Ghosh, U. Dynamical behaviors of an epidemic model with partial immunity having nonlinear incidence and saturated treatment in deterministic and stochastic environments. *Chaos Solitons Fractals*. **174**, 113775 (2023).
3. Saha, P., Sikdar, G. C. & Ghosh, U. Transmission dynamics and control strategy of single-strain dengue disease. *Int. J. Dyn. Control*. **11**, 1396–1414 (2023).
4. Saha, P., Pal, K. K., Ghosh, U. & Tiwari, P. K. Effects of vaccination and saturated treatment on COVID-19 transmission in india: deterministic and stochastic approaches. *J Biol. Syst.*, 1–47 (2024).
5. Saha, P., Pal, K., Ghosh, K. & Kumar Tiwari, P. U. Dynamic analysis of deterministic and stochastic SEIR models incorporating the Ornstein–Uhlenbeck process. *Chaos Interdiscip J. Nonlinear Sci* **35**, (2025).
6. Tian, R., Zhang, F., Du, H. & Wang, P. Optimization design method for multi-stress accelerated degradation test based on Tweedie exponential dispersion process. *Appl. Math. Model.* **135**, 684–707 (2024).
7. Tian, R., Zhang, F., Du, H. & Wang, P. Exponential dispersion accelerated degradation modelling and reliability assessment considering initial value and processes heterogeneity. *Maint Reliab. I Nizawodn* **26**, (2024).

8. Bhati, D. & Sastry, D. V. S. Maha qadri, P. Z. A new generalized poisson-lindley distribution: applications and properties. *Austrian J. Stat.* **44**, 35–51 (2015).
9. Shanker, R. The discrete poisson-amarendra distribution. *Int. J. Stat. Distrib. Appl.* **2**, 14–21 (2016).
10. Shanker, R. The discrete poisson-garima distribution. *Biometrics Biostat Int. J.* **5**, 1–7 (2017).
11. Shanker, R. R., Fesshaye, H., Shanker, R. R., Leonida, T. A. & Sium, S. On discrete Poisson-Shanker distribution and its applications. *Biometrics Biostat Int. J.* **5**, 121 (2017).
12. Zeghdoudi, H. & Nedjar, S. On Poisson pseudo Lindley distribution: properties and applications. *J. Probab. Stat. Sci.* **15**, 19–28 (2017).
13. Alrumeay, A. Bernoulli Poisson Moment Exponential Distribution : Mathematical Properties, Regression Model, and Applications. *Int. J. Math. Math. Sci.* (2024). (2024).
14. Grine, R. & Zeghdoudi, H. On Poisson quasi-lindley distribution and its applications. *J. Mod. Appl. Stat. Methods.* **16**, 403–417 (2017).
15. Alamri, O. A. Classical and bayesian Estimation of discrete Poisson Agu-Eghwerido distribution with applications. *Alexandria Eng. J.* **109**, 768–777 (2024).
16. Shanker, R. & Shukla, K. K. A quasi Poisson-Aradhana distribution. *Hung. Stat. Rev.* **3**, 3–17 (2020).
17. Ahsan-ul-Haq, M. On Poisson moment exponential distribution with applications. *Ann. Data Sci.* <https://doi.org/10.1007/s40745-022-00400-0> (2022).
18. Ahsan-ul-Haq, M., Al-Bossly, A., El-Morshedy, M. & Eliwa, M. S. Poisson XLindley Distribution for Count Data: Statistical and Reliability Properties with Estimation Techniques and Inference. *Comput. Intell. Neurosci.* (2022). (2022).
19. Maya, R., Irshad, M. R., Chesneau, C., Nitin, S. L. & Shibu, D. S. On discrete Poisson–Mirra distribution: regression, INAR (1) process and applications. *Axioms* **11**, 193 (2022).
20. Seghier, F. Z., Ahsan-ul-Haq, M., Zeghdoudi, H. & Hashmi, S. A. New generalization of Poisson distribution for Over-dispersed, count data: mathematical properties, regression model and applications. *Lobachevskii J. Math.* **44**, 3850–3859 (2023).
21. Alrumeay, A. & Khogeer, H. A. A new Two-Parameter discrete distribution for overdispersed and asymmetric data: its properties, estimation, regression model, and applications. *Symmetry (Basel)* **15**, 1289 (2023).
22. Alomair, A. & Ahsan-ul-Haq, M. A new extension of Poisson distribution for asymmetric count data: theory, classical and bayesian Estimation with application to lifetime data. *PeerJ Comput. Sci.* **9**, e1748 (2023).
23. Zaagan, A. A. & Mahnashi, A. M. Analysis of leukemia and forest fires data using new Poisson Quasi-Shanker distribution. *Alexandria Eng. J.* **104**, 701–709 (2024).
24. Alghamdi, F. M. et al. Discrete Poisson Quasi-XLindley distribution with mathematical properties, regression model, and data analysis. *J. Radiat. Res. Appl. Sci.* **17**, 100874 (2024).
25. Al-Babtain, A. A., Gemeay, A. M. & Afify, A. Z. Estimation methods for the discrete Poisson-Lindley and discrete Lindley distributions with actuarial measures and applications in medicine. *J. King Saud Univ.* **33**, 101224 (2021).
26. Alosey, A. R., El & Gemeay, A. M. A novel version of geometric distribution: method and application. *Comput. J. Math. Stat. Sci.* **4**, 1–16 (2025).
27. Nagy, M. et al. The New Novel Discrete Distribution with Application on COVID-19 Mortality Numbers in Kingdom of Saudi Arabia and Latvia. *Complexity* (2021). (2021).
28. Teamah, A. E. A. M., Elbanna, A. A. & Gemeay, A. M. Discrete Half-Logistic distribution: statistical properties, estimation, and application. *J. Stat. Appl. Probab.* **13**, 273–284 (2024).
29. Ahmad, K. et al. Statistical inference on the exponentiated moment exponential distribution and its discretization. *J. Radiat. Res. Appl. Sci.* **17**, 101116 (2024).
30. Alomair, A. M. & Ahsan-ul-Haq, M. A new mixed Poisson Komal distribution with application on radiation, agricultural and medical sciences data. *J. Radiat. Res. Appl. Sci.* **18**, 101500 (2025).
31. Alomair, A. M. & Ahsan-ul-Haq, M. Analysis of radiation and corn borer data using discrete Poisson Xrama distribution. *J. Radiat. Res. Appl. Sci.* **18**, 101388 (2025).
32. Ahsan-ul-Haq, M. Statistical analysis of Haq distribution: Estimation and applications. *Pakistan J. Stat.* **38**, 473–490 (2022).
33. Alkhairy, I. Classical and bayesian inference for the discrete Poisson Ramos-Louzada distribution with application to COVID-19 data. *Math. Biosci. Eng.* **20**, 14061–14080 (2023).
34. Seghier, E. Z., Zeghdoudi, H. & Benchaabane, A. A size-biased Poisson-gamma Lindley distribution with application. *Eur. J. Stat.* **1**, 132–147 (2021).
35. Catcheside, D. G., Lea, D. E. & Thoday, J. M. Types of chromosome structural change induced by the irradiation of Tradescantia microspores. *J. Genet.* **47**, 113–136 (1946).
36. Hosmer, D. W. & Lemeshow, S. Applied Logistic Regression. (2000).
37. Erbayram, T. & Akdoğan, Y. A new discrete model generated from mixed Poisson transmuted record type exponential distribution. *Ric Di Mat.* <https://doi.org/10.1007/s11587-022-00755-9> (2023).

## Acknowledgements

“This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU251035].”

## Author contributions

Author contributions: Abdullah M. Alomair: conceptualization and methodology; Faisal Ayyaz: writing manuscript; Saadia Tariq: supervision and editing; Muhammad Ahsan-ul-Haq: conceptualization, methodology, writing. All authors have read and agreed to the published version of the manuscript.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

Correspondence and requests for materials should be addressed to A.M.A.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025