Poisson- exponential and gamma distribution: properties and applications

Siriporn Samutwachirawong[1*]

Received: 13 March 2021; Revised: 23 June 2021; Accepted: 2 July 2021

Abstract

The objective of research is a new mixed Poisson distribution for count data based on the exponential and gamma distribution, namely the Poisson–exponential and gamma (Poisson-EG) distribution. Its probability mass function, moments, mean, variance and index of dispersion have been obtained. The method of parameter estimation for the proposed distribution is maximum likelihood estimation. Additionally, the Poisson–exponential and gamma distribution was applied to fit some real data sets using the maximum likelihood estimation. The results, based on p-value of the discrete Anderson–Daring test and the log–likelihood values show that the Poisson-EG distribution is the most appropriate among the considered distributions for the real data sets.

Keywords: Poisson distribution, Exponential and gamma distribution, Count data

[*] siriporn_sm@mju.ac.th
[1] Section of Statistics, Faculty of Science, Maejo University

1. Introduction

The count data are non-negative integers, $x = 0, 1, 2, \ldots$, some examples include the number of patients receiving services in the hospital, the number of the car during rush hours every weekday. The most popular method to model count data is the Poisson distribution. The Poisson distribution should have been named after Bortkiewicz in 1898, he showed that those numbers follow the Poisson distribution, the mean is equal to variance (Haight, 1967). Since the concept of the Poisson distribution had gained a lot of attention. The Poisson-gamma distribution or negative binomial (NB) distribution, it has become a popular alternative distribution to the Poisson distribution. Much research gives importance to the overdispersion or underdispersion issue which can be addressed by the use of mixed Poisson distributions (Panjer, 2006). The mixed Poisson distributions are determined where the mean of Poisson is raised as a random variable with some lifetime distribution. The distribution of the Poisson mean is the so-called mixed distribution (Everitt and Hand, 1981). In the literature, several developments of the Mixed Poisson distribution have been suggested such as Poisson-gamma distribution (Greenwood and Yule, 1920), Poisson-Exponential Beta distribution (Pielou, 1962), Poisson-Lindley distribution (Sankaran,1970), Poisson-Pareto distribution (Willmot, 1993), Poisson-Exponential distribution (Devroye, 1986), Poisson-Xgamma distribution (Bilal *et al*, 2020). It has been found that the general characteristics of mixed Poisson distributions follow some characteristics of its mixing distribution.

This paper offers one parameter Poisson-exponential and gamma distribution, as a model for count data. The probability density function (pdf) and the cumulative distribution function (cdf) of a two-component mixture of exponential and gamma distribution for modeling lifetime data (Shukla, 2018) given by

$$g(x,\theta) = \frac{\theta^6}{\theta^5 + 120}\left(1 + x^5\right)e^{-\theta x}, \tag{1}$$

$$G(x,\theta) = 1 - \left[1 + \frac{\theta x\left(\theta^4 x^4 + 5\theta^3 x^3 + 20\theta^2 x^2 + 60\theta x + 120\right)}{\left(\theta^5 + 120\right)}\right]e^{-\theta x}, \tag{2}$$

where $x > 0, \theta > 0.$ The exponential and gamma distribution is a mixture of exponential distribution with scale parameter θ and gamma distribution with parameter 6 and scale parameter θ.

Some of the important one parameter lifetime distributions are the Lindley distribution (the mixture of an exponential distribution with scale parameter θ and gamma distribution with parameter 4 and scale parameter θ) (Lindley, 1958), the mixture of an exponential distribution with scale parameter θ and gamma distribution with parameter 2 and scale parameter θ (Shanker, 2015), the mixture of an exponential distribution with scale parameter θ and gamma distribution with parameter 4 and scale parameter θ (Shukla, 2019).

In this paper, new mixed Poisson distribution for count data based on the exponential and gamma distribution is introduced. Its probability mass function (pmf), moments, mean, variance and index of dispersion have been obtained. We consider the application of the Poisson-EG distribution to the data sets and showed that the Poisson-EG distribution provide a satisfactory better fit in data sets.

2. Methods

The methods of this research are:

1. The probability mass function using the mixed distribution technique, the mixing distribution can be continuous, discrete or a distribution with probability at a finite number of points (Raghavachari *et al*., 1997) and some mathematical properties of the Poisson-EG distribution have been investigated

2. The appropriate parameter estimation method for the Poisson-EG distribution using maximum likelihood estimation.

3. Application of the Poisson-EG distribution to real data sets have been studied by comparing to the Poisson, Poisson-lindley, Poisson-Xgamma and Poisson-exponential distributions using the discrete Anderson-Darling (AD) test and log-likelihood (LL).

3. Results and discussion

The proposed distribution is a mixture of the Poisson distribution and the exponential and gamma distribution. Moreover; the pmf, moments, mean, variance, index of dispersion, the steps for random variate generation and applications of proposed distribution are shown in this section.

The probability mass function of Poisson- exponential and gamma distribution

Let $X \mid \lambda$ be a random variable following a Poisson distribution with parameters λ, $X \mid \lambda \sim Poisson(\lambda)$. If λ is distribution as a two-component mixture of exponential distribution and gamma in Eq. (1), denoted $\lambda \sim EG(\theta)$, then X is called a Poisson-EG random variable.

Theorem 1. The pmf of a Poisson- exponential and gamma distribution i.e., Poisson-EG $(X;\theta)$ is given by

$$f(x)=\frac{\theta^{6}(\theta+1)^{-(x+6)}\left((\theta+1)^{5}+\frac{\Gamma(x+6)}{\Gamma(x+1)}\right)}{\theta^{5}+120}\ ; x=0,1,2,\ldots \text{ and } \theta>0 \tag{3}$$

Proof A random variable X is said to be a Poisson random variable with parameters λ, its pmf is

$$f(x;\lambda)=\frac{\exp(-\lambda)\lambda^{x}}{x!}, \text{ for } x=0,1,2,\ldots \text{ and } \lambda>0.$$

If λ is distributed as a two-component mixture of exponential and gamma distribution with pdf in Eq.(1), then X is called the Poisson- exponential and gamma distribution random variable, the pmf of X is

$$f(x)=\int_0^\infty \frac{\exp(-\lambda)\lambda^{x}}{x!}g(\lambda)d\lambda,$$

$$=\int_0^\infty \frac{\exp(-\lambda)\lambda^{x}}{x!}\frac{\theta^{6}}{\theta^{5}+120}(1+\lambda^{5})e^{-\theta\lambda}d\lambda,$$

$$=\frac{\theta^{6}}{(\theta^{5}+120)x!}\int_0^\infty \exp(-(\theta+1)\lambda)(\lambda^{x}+\lambda^{x+5})d\lambda,$$

$$=\frac{\theta^{6}(\theta+1)^{-(x+6)}\left((\theta+1)^{5}+\frac{\Gamma(x+6)}{\Gamma(x+1)}\right)}{\theta^{5}+120}. \qquad \square$$

Some pmf plots of the Poisson-EG distribution with different values of θ has been presented in figure 1.

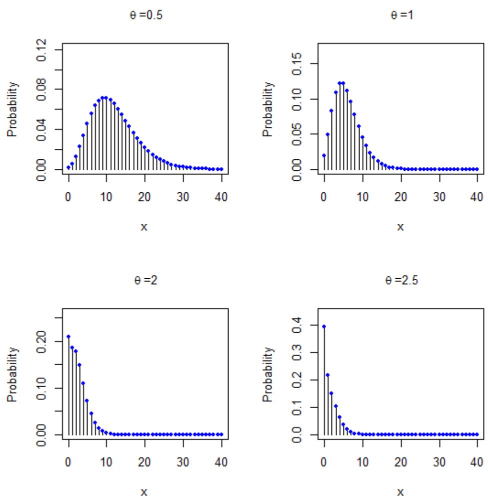

Figure 1 Some pmf plots of the Poisson-EG distribution with different values of θ

The r^{th} factorial moment of the Poisson-EG distribution is given by

$$\mu'_{(r)} = E\left[(X)_r\right] = \int_0^\infty \lambda^r g(\lambda) d\lambda.$$

$$\mu'_{(r)} = \int_0^\infty \lambda^r \frac{\theta^6}{\theta^5 + 120}(1+\lambda^5)e^{-\theta\lambda} d\lambda,$$

$$= \frac{\theta^6}{(\theta^5+120)} \int_0^\infty \lambda^r \left(1+\lambda^5\right) e^{-\theta\lambda} d\lambda,$$

$$= \frac{\theta^6}{(\theta^5+120)} \left(\frac{\Gamma(r+1)}{\theta^{r+1}} + \frac{\Gamma(r+6)}{\theta^{r+6}} \right). \tag{4}$$

Substituting $r=1,2,3$ and 4 in Eq.(4), the first four factorial moments about origin can be obtained and using the relationship between factorial moments about origin and moments about origin, the first four moment about origin of Poisson-EG distribution are obtained as

$$\mu_1' = \frac{\theta^6}{(\theta^5+120)} \left(\frac{720}{\theta^7} + \frac{1}{\theta^2} \right),$$

$$\mu_2' = \frac{\theta^6}{(\theta^5+120)} \left(\frac{5,040}{\theta^8} + \frac{2}{\theta^3} \right),$$

$$\mu_3' = \frac{\theta^6}{(\theta^5+120)} \left(\frac{40,320}{\theta^9} + \frac{6}{\theta^4} \right),$$

$$\mu_4' = \frac{\theta^6}{(\theta^5+120)} \left(\frac{326,880}{\theta^{10}} + \frac{24}{\theta^5} \right),$$

Thus, mean, variance and ID of $X \sim$ Poisson-EG (λ) are given by

$$\text{Mean} = \frac{\theta^6}{(\theta^5+120)} \left(\frac{720}{\theta^7} + \frac{1}{\theta^2} \right),$$

$$\text{Variance} = \left(\mu_2'(X) + E(X)\right) - \left(E(X)\right)^2 = \frac{\theta^{11} + \theta^{10} + 840\theta^6 + 3480\theta^5 + 86400\theta + 86400}{\theta^2(\theta^5+120)^2},$$

$$\text{ID} = \frac{\theta^{11} + \theta^{10} + 840\theta^6 + 3480\theta^5 + 86400\theta + 86400}{\left(\frac{720}{\theta^7} + \frac{1}{\theta^2} \right)\theta^8(\theta^5+120)^3}.$$

The ID plots of Poisson-EG distribution are shown in Figure 2 which illustrates that the different values of the parameter θ.

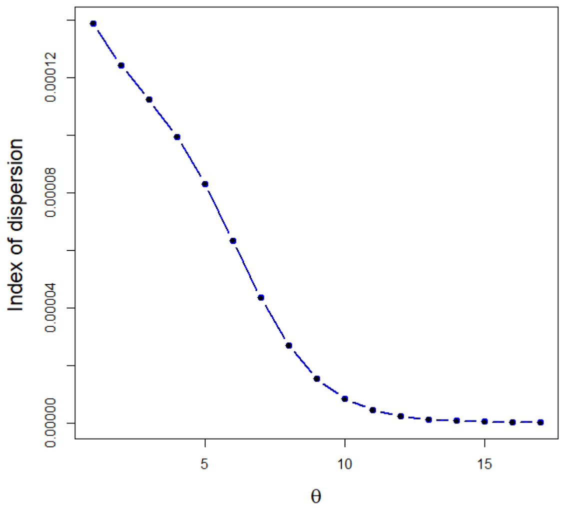

Figure 2 Some ID plots of the Poisson-EG distribution with different values of θ

Maximum Likelihood Estimate (MLE)

Suppose that the data $x_1, x_2, ..., x_n$ are a random sample of size n from the Poisson-EG distribution.

Let $\Theta = (\theta)^T$ be the vector of the parameters. Its likelihood function L of Poisson-EG distribution is given by

$$L(\Theta) = \prod_{i=1}^{n} \frac{\theta^6 (\theta+1)^{-(x_i+6)} \left((\theta+1)^5 + \frac{\Gamma(x_i+6)}{\Gamma(x_i+1)} \right)}{\theta^5 + 120}.$$

The log likelihood function is thus obtained as

$$\log L = n \log\left(\frac{\theta^6}{\theta^5 + 120} \right) - \sum_{x=1}^{k} (x_i + 6) \log(\theta+1) + \sum_{x=1}^{k} \log\left[(\theta+1)^5 + \frac{\Gamma(x_i+6)}{\Gamma(x_i+1)} \right]$$

The first derivative of the log likelihood function is given by

$$\frac{d \log L}{d\theta} = \frac{n(\theta^5 + 720)}{\theta\left(\theta^5 + 120\right)} - \frac{n\left(\bar{x} + 6\right)}{\theta + 1} + \sum_{x=1}^{n} \frac{5(\theta+1)^4 \Gamma(x+1)}{\left[(\theta+1)^5 \Gamma(x+1) + \Gamma(x+6) \right]}.$$

Although these equations are non-linear equation can be solved by any numerical iteration methods were obtained by Newton Raphson method in R language to obtain the parameter estimates.

Applications

The Poisson-EG distribution has been fitted to the data sets to test its goodness of fit along with Poisson, Poisson-lindley, Poisson-exponential and Poisson-Xgamma distributions. In each of these distributions, the parameters are estimated by using the maximum likelihood method. We use the estimated LL and the AD test for discrete distributions to compare the expected and observed values of each data set. The AD statistic can be obtained by using dgof package (Arnold and Emerson, 2011) in R language. The Poisson, Poisson-lindley, Poisson-exponential, Poisson-Xgamma and Poisson-EG distributions are given in Table 1.

Table 1 pmf of the fitted distributions

Distribution	pmf
Poisson	$\dfrac{e^{-\lambda}\lambda^{x}}{x!},\ \lambda>0$ (Poisson, 1837)
Poisson-lindley	$\dfrac{\lambda^{2}(x+\lambda+2)}{(\lambda+1)^{x+3}},\ \lambda>0$ (Sankaran, 1970)
Poisson-exponential	$(1-\lambda)^{x}\lambda,\ \lambda>0$ (Devroye, 1986)
Poisson-Xgamma	$\dfrac{\lambda^{2}}{2(1+\lambda)^{x+4}}(2(1+\lambda)^{2}+\lambda(x+1)(x+2),\ \lambda>0$ (Bilal *et al*, 2020)
Poisson-EG	$\dfrac{\theta^{6}\left(\theta+1\right)^{-(x+6)}\left(\left(\theta+1\right)^{5}+\dfrac{\Gamma(x+6)}{\Gamma(x+1)}\right)}{\theta^{5}+120},\ \theta>0$

The first data set is the number of epileptic seizures (Chakraborty,2010), this data set is overdispersed with a mean of 1.544, variance of 2.883 and ID of 1.867. The second data set is the number of patients who have come to hospital service, the research to fine the optimal queueing model which suitable for hospital policy (Ornsuda, 2001), this data set is overdispersed with a mean of 5.042, variance of 10.891 and ID of 2.160. The third data set is the number of mistakes in copying groups of random digits (Kemp, 1965), this data set is overdispersed with a mean of 0.783, variance of 1.257 and ID of 1.605. This shows that all data sets are over dispersed. We use the estimated LL and the AD test for discrete distributions to compute the expected frequencies for fitting Poisson, Poisson-lindley, Poisson-exponential, Poisson-Xgamma and Poisson-EG distributions are given in Table 2-4. Based on *p*-value, from AD statistic, we observe that Poisson-EG distribution provides a satisfactorily, better fit for the data sets compared to other distributions.

Table 2 Distribution of number of epileptic seizures

Number of epileptic seizures	Observed count	Expected frequencies				
		Poisson	Poisson-lindley	Poisson-exponential	Poisson-Xgamma	Poisson-EG
0	126	75.250	129.234	138.510	134.117	132.076
1	80	115.885	87.689	84.491	83.994	76.015
2	59	89.232	55.724	51.539	53.709	54.844
3	42	45.806	33.978	31.439	33.678	38.087
4	24	17.635	20.137	19.178	20.479	23.904
5	8	5.432	11.688	11.698	12.077	13.640
6	5	1.394	6.678	7.136	6.929	7.196
7	4	0.307	3.768	4.353	3.883	3.563
8	3	0.059	2.104	2.655	2.133	1.675
Parameter estimates		$\hat{\lambda}=1.54$	$\hat{\lambda}=0.97$	$\hat{\lambda}=0.39$	$\hat{\lambda}=1.23$	$\hat{\theta}=2.45$
-LL		636.046	595.181	598.396	595.344	593.947
AD statistic		0.145	0.027	0.041	0.030	0.021
p-value		<0.001	0.965	0.560	0.9053	0.998

Table 3 Distribution of number of patients who have come to hospital service

Number of patients	Observed count	Expected frequencies				
		Poisson	Poisson-lindley	Poisson-exponential	Poisson-Xgamma	Poisson-EG
0	6	0.313	5.980	9.137	6.556	1.829
1	2	1.575	6.370	7.583	6.053	3.740
2	4	3.970	6.177	6.294	5.744	5.665
3	3	6.669	5.672	5.224	5.382	6.781
4	6	8.403	5.025	4.336	4.915	6.906
5	8	8.470	4.342	3.599	4.368	6.267
6	4	7.115	3.682	2.987	3.784	5.219
7	4	5.123	3.077	2.479	3.206	4.065
8	1	3.227	2.542	2.058	2.664	3.002
9	4	1.807	2.081	1.708	2.178	2.122
10	3	0.911	1.690	1.418	1.755	1.447
>10	3	0.417	1.363	1.177	1.397	0.957
Parameter estimates		$\hat{\lambda}=5.04$	$\hat{\lambda}=0.34$	$\hat{\lambda}=0.17$	$\hat{\lambda}=0.47$	$\hat{\theta}=1.20$
-LL		136.074	126.310	130.125	125.240	124.864
AD statistic		0.138	0.142	0.213	0.138	0.088
p-value		0.324	0.285	0.026	0.319	0.850

Table 4 Distribution of number of mistakes in copying groups of random digits

Number of mistakes	Observed count	Expected frequencies				
		Poisson	Poisson-lindley	Poisson-exponential	Poisson-Xgamma	Poisson-EG
0	35	27.539	33.495	34.057	34.064	35.770
1	11	21.481	15.493	15.753	14.929	12.974
2	8	8.377	6.847	6.962	6.723	6.273
3	4	2.178	2.934	2.984	2.989	3.306
4	2	0.425	1.230	1.250	1.295	1.677
Parameter estimates		$\hat{\lambda}=0.78$	$\hat{\lambda}=1.74$	$\hat{\lambda}=0.56$	$\hat{\lambda}=2.10$	$\hat{\theta}=3.08$
-LL		77.546	73.351	73.375	73.206	73.130
AD statistic		0.125	0.038	0.039	0.038	0.028
p-value		0.306	1	1	1	1

4. Conclusions

The new distributions for count data have been developed by mixed Poisson distribution for count data based on the exponential and gamma distribution, namely the Poisson–exponential and gamma distribution. Its probability mass function, moments, mean, variance, index of dispersion and the steps for random variate generation have been obtained. The methods of parameter estimation for the proposed distribution which is maximum likelihood estimation. Additionally, the Poisson-exponential and gamma distribution was applied to fit some real data sets using the maximum likelihood estimation. The results, based on *p*-value, from AD statistic, we observe that Poisson-EG distribution provides a satisfactorily better fit for the data sets compared to other distributions. Also, the mixed Poisson distribution is the developed Poisson distribution for count data where the

Poisson is not equidispersion. As a result, the propose distribution is an alternative discrete distribution to analyze some kinds of count data.

Acknowledgement

I would like to express a special thanks Faculty of Science, Maejo University for their financial support during my study.

References

Arnold, T. A., & Emerson W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. **The R Journal**, 3, 34-39.

Paraa, B. A., Jan, T. R., & Bakouchc, H. S. (2020). Poisson Xgamma distribution: A discrete model for count data analysis. **Model Assisted Statistics and Applications**, 15, 139-151.

Chakraborty, S. (2010). On some distributional properties of the family of weighted generalized Poisson distribution. **Communications in Statistics–Theory and Methods**, 39, 2767-2688.

Devroye, L. (1986). **Non-Uniform Random Variate Generation**. New York: Springer-Verlag.

Everitt B., & Hand D. (1981). **Finite Mixture Distributions**. Netherlands: Chapman and Hall.

Greenwood, M., & Yule, G. (1920). An Enquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. **Journal of Royal Statistical Society**, 83, 255-279.

Haight, F. (1967). **Handbook of the Poisson distribution**. New York: John Wiley and Sons.

Kemp, C. D., & Kemp, A. W. (1965). Some properties of the Hermite distribution, Biometrika, 52, 381-394.

Lindley, D. V. (1958). Fiducial distributions and Bayes theorem. **Journal of the Royal Statistical Society**, 20, 102- 107.

Narkthewan, O. (2011). **Queueing system analysis of Maharatnakhonsrithammarat hospital services**. Master's thesis of science in statistics department of statistics Chulalongkorn university.

Panjer, H. H. (2006). **Mixed Poisson Distributions**. New York: John Wiley and Sons.

Pielou, E. (1962). Run of One Species with Respect to Another in Transects Through Plant Population. **Journal of Biometrics**, 18, 579-593.

Raghavachari, M., Srinivasan A., & Sullo P. (1997). Poisson mixture yield models for integrated circuits. A critical review. **Microelectronics Reliability**, 37, 565-580.

R Core Team. (2020). **R: A Language and Environment for Statistical Computing**. cited 2021 March 18 Retrieved from https://www.r-project.org

Sankaran, M. (1970). The Discrete Poisson–Lindley Distribution. **Journal of Biometrics**, 26, 145-149.

Shanker, R. (2015). Akash distribution and Its Applications. **International Journal of Probability and Statistics**, 4, 65-75.

Shukla, K.K. (2018). Pranav distribution and its properties and applications, **Biometrics and Biostatistics International Journal**, 7, 244-254.

Shukla, K. K., & Shanker, R. (2019). A Discrete Poisson Ishita distribution and its Application. **International Journal of Statistics and Economics**, 4, 50-59.

Willmot, G. E. (1993). On Recursive Evaluation of Mixed Poisson Probabilities and Related Quantities. **Journal of Scandinavian Actuarial**, 18, 114-133.