Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120111

Negative Binomial Quasi Akash Distribution and its Applications

Showkat Ahmad Dar[1,*], *Anwar Hassan*[1] *and Peer Bilal Ahmad*[2]

[1]Department of Statistics, University of Kashmir, Srinagar (J&K), India
[2]Department of Mathematical Sciences, Islamic University of Science & Technology, Awantipora (J&K), India

Received: 9 Mar. 2020, Revised: 20 May 2020, Accepted: 27 May 2020
Published online: 1 Jan. 2023

Abstract: In this paper, we obtained a new model for count data by compounding of negative binomial with quasi Akash distribution. Important mathematical and statistical properties of the distribution have been derived and discussed. Then, parameter estimation is discussed using maximum likelihood method of estimation. Finally, the potential of the proposed model has been tested by chi-square goodness of fit test by modeling the real count data set.

Keywords: Negative binomial distribution, quasi Akash distribution, compound distribution, count data.

1 Introduction

The count data is generated from various fields of real life like atmospheric science, health, actuarial science, financial science, linguistics etc. and for modeling of count data we need discrete probability models. Now the particular discrete probability model to be fitted to count data will depend on value of mean, variance, index of dispersion among other factors. Researchers have generalized the classical discrete models for achieving more flexibility and for extracting more variation. One of the technique for generalizing a discrete probability model is compounding. Researchers have been proved that mixed Poisson and mixed negative distributions provides better fit to count data as compared to old existing models. Altun (2019) introduced Poisson cousi Lindly regression model for analyzing over dispersed count data. Zamani and Ismail (2010) introduced a new count data model by compounding negative Binomial distribution with Lindley distribution and find its application in insurance claims. Hassan et al. (2020, 2021) obtained a new several count data models by using compounding technique and obtained various properties of that model. Adamidis and Loukas (1998) introduced a life time distribution by compounding Exponential and Geometric distribution. Tahmasbi and Rezaei (2008) proposed Exponential Logarithmic distributions and having decreasing failure rate. Christensen et al. (2003) studied a hierarchical version and finds its application to model environmental monitoring. Hassan, Dar and Ahmad (2019) introduced a new compounding probability model for count data by compounding Poisson distribution with Ishita distribution that finds its applications in epileptic seizure counts. Chesneau et al (2020) introduced Cosine geometric distribution for count data modeling. Atikankul et al. (2020) derived a new compound distribution by mixing Poisson distribution with weighted Lindley distribution finds its application in insurance claims.

In this paper, we propose a new compounding distribution by compounding Negative Binomial distribution with quasi Akash distribution, as there is a need to find more flexible model for analysing statistical data.

2 Materials and Methods

A random variable Z said to have a negative binomial distribution (NBD) if its pmf is given by

$$P(Z=z)=\binom{r+z-1}{z}p^r q^z\,; z=0,1,2....,\ r>0, \text{ and } 0<p<1 \tag{1}$$

The factorial moments, mean and of variance of NBD are given as

[*]Corresponding author e-mail: darshowkat2429@gmail.com

$$\mu_{[k]}=\frac{\Gamma r+k}{\Gamma r}\left(\frac{q}{p}\right)^{k},\quad k=1,2....$$

(2)

$$E(Z)=\frac{rq}{p}\text{ and }V(Z)=\frac{rq}{p^{2}}$$

A random variable Z said to have Quasi Akash distribution (QAD) if its pdf is given by

$$f(Z;\alpha,\sigma)=\frac{\sigma^{2}}{\alpha\sigma+2}(\alpha+\sigma z^{2})e^{-\sigma z},\ \alpha,\sigma>0,z>0 \tag{3}$$

MGF of QAD is given as

$$M_{z}(t)=\frac{\sigma^{2}}{\alpha\sigma+2}\left(\frac{\alpha}{(\sigma-t)}+\frac{2\sigma}{(\sigma-t)^{3}}\right) \tag{4}$$

Quasi Akash distribution is a newly proposed lifetime model formulated by Rama Shanker (2016) for engineering application and calculated its various properties including moments, mean deviation, median, stochastic ordering, order statistics, Renyi entropy, mean residual life, and ML estimation. The Quasi Akash distribution has better flexibility in handling lifetime data as compared to exponential and Lindley, Akash, Lognormal, Gamma, Weibull and weighted Lindley distribution.

Usually the parameters r and p in NBD are fixed constants but here we have considered a problem in which the probability parameter p is itself a random variable following QAD with pdf given in equation (3)

3 Definition of Proposed Model (Negative Binomial Quasi Akash Distribution)

If $Z|\lambda\sim$ NB$\left(r,p=e^{-\lambda}\right)$, where λ is itself a random variable following quasi Akash distribution with parameter σ,α, then determining the distribution that results from marginalizing over λ will be known as a compound of negative Binomial distribution with that of quasi Akash distribution, which is denoted by $NBQAD\left(r,\alpha,\sigma\right)$. It may be noted that proposed model will be a discrete since the parent distribution Negative Binomial distribution is discrete.

Theorem 1:

The probability mass function of a NBQAD is given by

$$P\left(z\right)=\binom{z+r-1}{z}\sum_{j=0}^{z}\binom{z}{j}(-1)^{j}\frac{\sigma^{2}}{(\alpha\sigma+2)}\left(\frac{\alpha}{(r+j+\sigma)}+\frac{2\sigma}{(r+j+\sigma)^{3}}\right)$$

Proof: The pmf of NBD is given as

$$g(z\,|\,\lambda)=\binom{r+z-1}{z}e^{-\lambda r}(1-e^{-\lambda})^{z}$$

When its parameter λ follows QAD with pdf given as

$$h(\lambda;\alpha,\sigma)=\frac{\sigma^{2}}{\alpha\sigma+2}(\alpha+\sigma\lambda^{2})e^{-\sigma\lambda}$$

The pmf of a compound NBD and QAD can be obtained as

$$P(Z;\sigma,\alpha,r)=\int_{0}^{\infty}g(z/\lambda)\,h(\lambda;\sigma,\alpha)d\lambda$$

$$P(z)\ =\int_{0}^{\infty}\binom{z+r-1}{z}e^{-\lambda r}(1-e^{-\lambda})^{z}\frac{\sigma^{2}}{\alpha\sigma+2}(\alpha+\sigma\lambda^{2})e^{-\sigma\lambda}d\lambda$$

$$P(z)=\frac{\sigma^2}{\alpha\theta\sigma+2}\binom{z+r-1}{z}\sum_{j=0}^{z}\binom{z}{j}(-1)^j\int_0^\infty e^{-\lambda(r+j+\sigma)}(\alpha+\sigma\lambda^2)d\lambda$$

$$P(z)=\frac{\sigma^2}{\alpha\sigma+2}\binom{z+r-1}{z}\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{\alpha}{(r+j+\sigma)}+\frac{2\sigma}{(r+j+\sigma)^3}\right),$$

$$z=01,2,...,\sigma,\alpha,r>0 \quad (5)$$

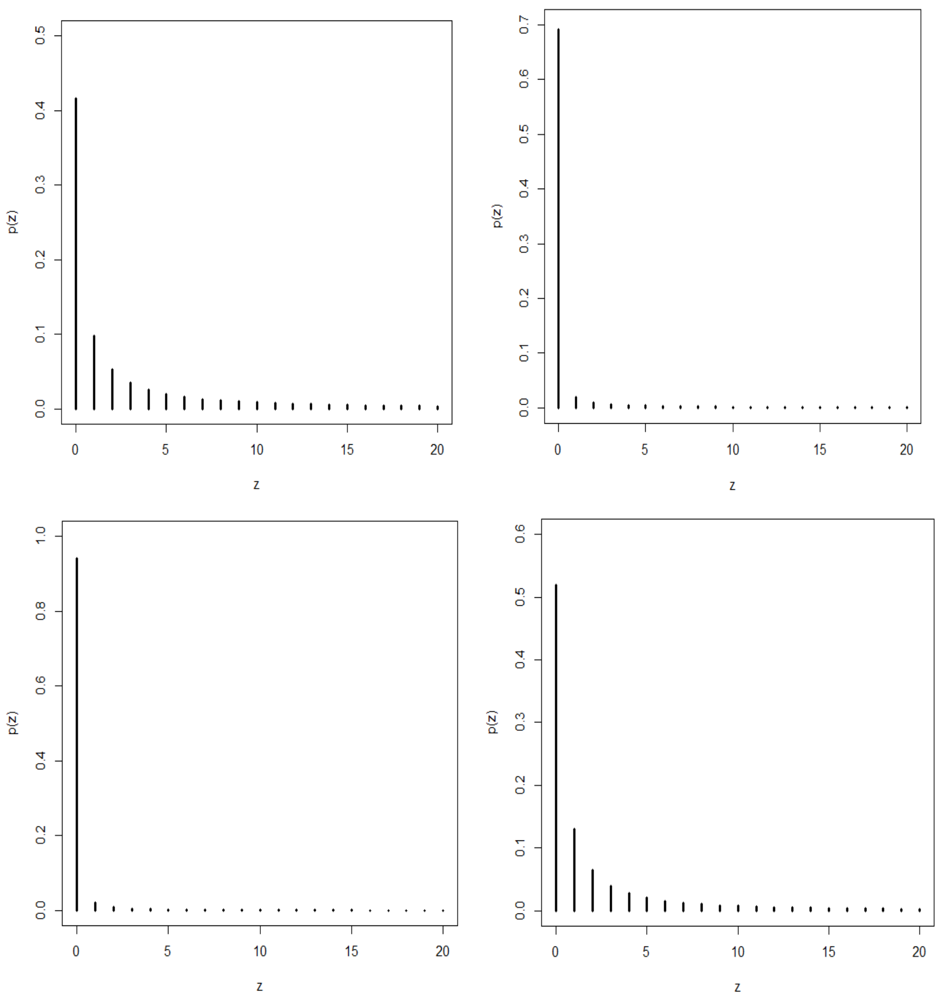

Figure 1: The pmf Plot for Different Values of Parameters

The above figures shows the pmf plot for different values of parameters.

4 Nested Distributions

In this particular section we show that the proposed model can be nested to different models under specific parameter setting

Case 1:

If we put $r = 1$ in equation (5) the NBQAD reduces to Geometric Quasi Akash Distribution with pmf given as

$$P_1(Z=z)=\frac{\sigma^2}{\alpha\sigma+2}\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{\alpha}{(1+j+\sigma)}+\frac{2\sigma}{(1+j+\sigma)^3}\right),$$

$$z=0,1,2,..,\alpha,\sigma>0$$

Case 2:

If we put $r = 1$ and $\alpha = \sigma$ in equation (5) the NBQAD reduces to Geometric Akash Distribution with pmf given as

$$P_2(Z=z) = \frac{\sigma^3}{\sigma^2+2}\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{1}{(1+j+\sigma)}+\frac{2}{(1+j+\sigma)^3}\right),$$
$$z = 0,1,2,..,\sigma > 0$$

Case 3:

If we put $\alpha = \sigma$ in equation (5) the NBQAD reduces to Negative Binomial Akash Distribution with pmf given as

$$P_3(Z=z) = \frac{\sigma^3}{\sigma^2+2}\binom{z+r-1}{z}\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{1}{(r+j+\sigma)}+\frac{2}{(r+j+\sigma)^3}\right),$$
$$z = 0,1,2,..,r > 0,\sigma > 0$$

Case 4:

If we put $\alpha = 0$ in equation (5) the NBQAD reduces to Negative Binomial Gamma Distribution with pmf as below

$$P_4(Z=z) = \binom{z+r-1}{z}\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{1}{(r+j+\sigma)^3}\right),$$
$$z = 0,1,2....,r > 0,\sigma > 0$$

5 Mean and Variance

The mean and variance of NBQAD can be obtained by using the property of conditional mean and variance

i) Since $Z|\lambda \sim NBD(r, p = e^{-\lambda})$ where λ is itself a random variable following $QAD \sim (\alpha,\sigma)$, therefore we have

$$E(Z^k) = E_\lambda\left(Z^k|\lambda\right)$$

$$E(Z^k) = r\sum_{j=0}^{k}\binom{k}{j}(-1)^j\int_0^\infty e^{\lambda(k-j)}\frac{\sigma^2}{\alpha\sigma+2}(\alpha+\sigma\lambda^2)e^{-\sigma\lambda}$$

$$E(Z^k) = r\sum_{j=0}^{k}\binom{k}{j}(-1)^j\frac{\sigma^2}{\alpha\sigma+2}\left(\frac{\alpha(\sigma-(k-j))^2+3\sigma}{(\sigma-(k-j))^3}\right)$$

Put $k = 1$ we get mean of NBQAD

$$E(Z) = r\frac{\sigma^2}{\alpha\sigma+2}\left(\frac{\alpha(\sigma-1)^2+2\sigma}{(\sigma-1)^3}-\frac{\alpha\sigma+2}{\sigma^2}\right)$$

ii) Conditional expectation identity $V(Z) = E_\lambda(V(Z|\lambda))+V_\lambda(E(Z|\lambda))$

$$r\frac{\sigma^2}{\alpha\sigma+2}\left(\left((r+1)\frac{\alpha(\sigma-2)^2+2\sigma}{(\sigma-2)^3}-2\frac{\alpha(\sigma-1)^2+2\sigma}{(\sigma-1)^3}+\frac{\alpha\sigma+2}{\sigma^2}\right)+\left(\frac{\alpha(\sigma-1)^2+2\sigma}{(\sigma-1)^3}-\frac{\alpha\sigma+2}{\sigma^2}\right)\right)$$
$$-\left(r\frac{\sigma^2}{\alpha\sigma+2}\left(\frac{\alpha(\sigma-1)^2+2\sigma}{(\sigma-1)^3}-\frac{\alpha\sigma+2}{\sigma^2}\right)\right)^2$$

6 Factorial Moment Of The Proposed Model

Theorem 2:

The factorial moments of order K of the proposed model is given by

$$\mu_{[k]}(Z)=\frac{\Gamma r+k}{\Gamma r}\sum_{j=0}^{k}\binom{k}{j}(-1)^j\frac{\sigma^2}{\alpha\sigma+2}\left(\frac{\alpha(\sigma-(k-j))^2+2\sigma}{(\sigma-(k-j))^3}\right)$$

Proof: The factorial moments of order K of NBD is

$$m_k(Z\mid\lambda)=\frac{\Gamma r+k}{\Gamma r}\left(\frac{1-e^{-\lambda}}{e^{-\lambda}}\right)^k$$

λ is itself a random variable following QAD

$$\mu_{[k]}(Z)=E_\lambda(m_k(Z\mid\lambda))$$

$$\mu_{[k]}(Z)=\frac{\Gamma r+k}{\Gamma r}\sum_{j=0}^{k}\binom{k}{j}(-1)^j E(e^{\lambda(k-j)})$$

$$\mu_{[k]}(Z)=\frac{\Gamma r+k}{\Gamma r}\sum_{j=0}^{k}\binom{k}{j}(-1)^j\frac{\sigma^2}{\alpha\sigma+2}\left(\frac{\alpha(\sigma-(k-j))^2+2\sigma}{(\sigma-(k-j))^3}\right). \qquad (6)$$

7 Maximum Likelihood Estimation

The log likelihood function of NBQAD is given by

$$LogL=\log\sum_{i=1}^{n}\binom{z+r-1}{z}+n\log\sigma^2-n\log(\alpha\sigma+2)$$
$$+\sum_{i=1}^{n}\left(\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{\alpha(r+j+\sigma)^2+2\sigma}{(r+j+\sigma)^3}\right)\right)$$

$$\frac{\partial}{\partial r}\log L=\sum_{i=1}^{n}\psi(x_i+r)-n\psi(r)$$
$$+\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{2\alpha(r+j+\sigma)^4-3(r+j+\sigma)^2\{\alpha(r+j+\sigma)^2+2\sigma)\}}{(r+j+\sigma)^6}\right)}{\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{\alpha(r+j+\sigma)^2+2\sigma}{(r+j+\sigma)^3}\right)}\right)=0$$

$$\frac{\partial}{\partial\sigma}\log L=\frac{n}{\sigma^2}-\frac{n}{(\alpha\sigma+2)}$$
$$+\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{((2\alpha)(r+j+\sigma)+2)-\{(\alpha(r+j+\sigma)^2+2\sigma)3(r+j+\sigma)^2\}}{(r+j+\sigma)^6}\right)}{\sum_{j=0}^{z}\binom{z}{j}(-1)^j\left(\frac{\alpha(r+j+\sigma)^2+2\sigma}{(r+j+\sigma)^3}\right)}\right)=0$$

$$\frac{\partial}{\partial\alpha}=\frac{-n\alpha}{(\alpha\sigma+2)}+\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{z}\binom{z}{j}(-1)^{j}\left(\frac{1}{(r+j+\sigma)}\right)}{\sum_{j=0}^{z}\binom{z}{j}(-1)^{j}\left(\frac{\alpha}{(r+j+\sigma)}+\frac{2\sigma}{(r+j+\sigma)^{3}}\right)}\right)=0$$

The above equations can be solved numerically by using R software (3.5.3).

8 Applications of Negative Binomial Quasi Akash distribution

In order to demonstrate the flexibility and applicability of proposed model we fit a data set which is already in statistical literature, for illustrating the claim that NBQAD fits statistically well when compared to ZINBD, NBD, ZIPD and PD. The data set is given in Table 2.

Table 1: Data Set has Number of Hospital Stays by United States Residents Aged 66 and over (see Flynn (2009))

Number of Hospital Days	0	1	2	3	4	5	6	7	8
Observed Frequencies	3541	599	176	48	20	12	5	1	4

We calculate the expected frequencies for fitting negative binomial quasi Akash distribution (NBQAD), zero inflated Poisson distribution (ZIPD), negative binomial distribution (NBD), zero inflated negative binomial distribution (ZINBD) and Poisson distribution (PD). In order to check the goodness of fit of the model and estimation of parameters of the model, Pearson's chi-square test and R studio statistical software has been used. The results are given in the table 2 for data set given in table 1. It is clear from the expected frequencies and the corresponding value of chi-square that negative binomial quasi Akash distribution provides a satisfactorily better fit for the data set given in table 1.

Table 2: Fitted NBQAD and other Competing Models to a Data Set 1

No. of Hospital Days	Observed Frequencies	PD	ZIPD	NBD	ZINBD	NBQAD
0	3541	3296.85	1816.5	3531.65	3541.1	3542.3
1	599	956	1609.5	588	533.4	609.7
2	176	138	712.9	181.25	217.7	161.2
3	48	13.4	210.6	64.4	68.7	53.9
4	20	0.95	46.7	24.4	18.9	21.2
5	12	0.5	8.4	9.6	4.4	9.4
6	5	0.1	1.3	3.9	1.3	4.6
7	1	0.1	0.1	1.6	0.5	2.4
8	4	0.1	0	0.65	0	1.3
Parameter estimates		$\theta=0.29$	$\theta=0.66$ $\lambda=0.88$	$r=0.37$ $p=0.55$	$\theta=0.6$ $\lambda=3.96$ $p=0.84$	$r=1.43$ $\theta=5.58$ $\beta=288.1$
Chi-square		531	2831.9	7.27	58.31	3.32
Degree of freedom		2	3	4	2	3
p-value		0	0.00	0.12	0.00	0.34

AIC (Akaike information criterion) and BIC (Bayesian information criterion) criterions has been used for comparing our proposed model with other competing model. The lower values of AIC and BIC corresponds to better fitting of the model.

It is clear from the Table 3, that the negative binomial quasi Akash distribution has lesser values of AIC and BIC as compared to other competing models. Hence we can concluded that the negative binomial quasi Akash distribution leads to a better fit than the other competing model for analysing the data set given in table 1. It is also clear from figure 1 the value of expected frequencies that negative binomial quasi Akash distribution provides a closer fit than that provided by other competing models.

Table 3: AIC and BIC for fitted PD, NBD, ZIPD, ZINBD and NBQAD for Data Set 1

Criterion	PD	NBD	ZIPD	ZINBD	NBQAD
AIC	6611	6023.2	6122	6078	6021.9
BIC	6611.2	6023.6	6135	6097	6021.49

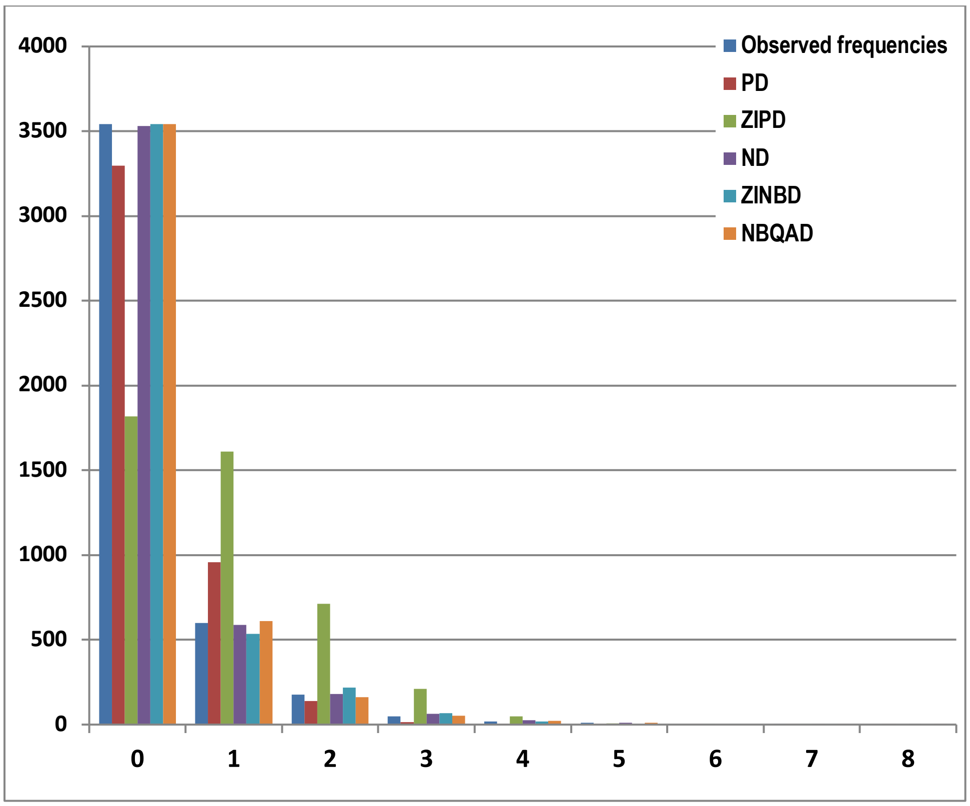

Figure 2: Graphical Overview of Fitted Models to a Data Set Given in Table 1

10 Conclusions

A new probability distribution is introduced using compounding technique. Statistical properties of the proposed model are studied and application in handling count data set representing number of hospital stays by United States residents aged 66 and over.

References

[1] Adamidis, K. and Loukas, S. A life time distribution with decreasing failure rate. Statistics and Probability Letters, **39**, 35-42(1998)

[2] Tahmasbi, R. and Rezaei, S. A two parameter life time distribution with decreasing failure rate. Computational Statistics and Data Analysis, **52(8),** 3889-3901 (2008)

[3]Christensen, A., Melgaard, H., Iwersen, J. and Thyregod, P. Environmental monitoring based on a hierarchical Poisson-

Gamma model. Journal of Quality Technology, **35(3),** 275-285 (2003)
[4]Zamani, H. and Ismail, N. Negative binomial-Lindley distribution and its application. Journal of Mathematics and Statistics, **6(1),** 4-9(2010)
[5]Shanker, R. A Quasi Akash distribution and its applications. Assam statistical Review, 30(1), 135-160(2016).
[6]Chesneau, C., Bakouch, H.S., Hussain, T. and Para, B.A. The cosine geometric distribution with count data modeling. Journal of Applied Statistics**, 48(1),** 124-137. (2021)
[7]Atikankul, Y., Thongteeraparp, A. and Bodhisuwan, W. The new Poisson mixed weighted Lindley distribution with applications to insurance claims data. Songklanakarin Journal of Science & Technology, **42(1),** 152-162 (2020)
[8]Hassan, A., Dar, S.A., Wani, S.A. and Akhtar, N. Inverse Rayleigh Minimax Distribution: Properties and Applications. Pakistan Journal of Statistics, **36(4),** 341-353(2020)
[9]Hassan, A., Dar, S.A., Peer, B.A., Para, B.A. On Size Biased Poisson Ailamujia Distribution and Its Applications. Pakistan Journal of Statistics, **37(1),** 19-38(2021)
[10]Hassan, H., Dar, S.A. and Ahmad, P.B. Poisson Ishita distribution: A new compounding probability model. IOSR Journal of Engineering (IOSRJEN), **9(2),** 38-46 (2019)
[11]Altun, E. A new model for over-dispersed count data: Poisson quasi-Lindley regression model. Mathematical Sciences, **13(3),** 241-247(2019)
[12]Flynn, M., & Francis, L. A. More flexible GLMs zero inflated models and hybrid models. Casualty Actuarial Society, 148-224(2009)
[13] R Development Core Team (2019). R version 3.5.3: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/