International Journal of Advance Research in Science and Engineering
Volume No.06, Special Issue No.(03), December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

A NEW DISCRETE DISTRIBUTION FOR VEHICLE BUNCH SIZE DATA: MODEL, PROPERTIES, COMPARISON AND APPLICATION

Adil Rashid[1], Zahoor Ahmad[2] and T. R. Jan[3]

[1,2,3] P.G Department of Statistics, University of Kashmir, Srinagar(India)

ABSTRACT

In the present paper we shall construct a new probability distribution by compounding Borel distribution with beta distribution and it will be named as Borel-Beta Distribution (BBD). The newly proposed distribution finds its application in modeling traffic data sets where the concentration of observed frequency corresponding to the very first observation is maximum. BBD reduces to Borel-uniform distribution on specific parameter setting. Furthermore, we have also obtained compound version of Borel-Tanner distributions. Some mathematical properties such as mean, variance of some compound distributions have also been discussed. The estimation of parameters of the proposed distribution has been obtained via maximum likelihood estimation method. Finally the potentiality of proposed distribution is justified by using it to model the real life data set..

Keywords –Borel distribution, Borel-Tanner distribution, beta distribution, Consul distribution, Compound distribution

I. INTRODUCTION

Probability distributions are used to model the real world phenomenon so that scientists may predict the random outcome with somewhat certainty but to many complex systems in nature one cannot expect the behavior of random variables to be identical everytime. Most of the times researchers encounter the scientific data that lacks the sequential pattern and hence and to deal with such a randomness one must be well equipped knowledge of probability. The aim of researchers lies in predicting the random behavior of data under consideration efficiently and this can be done by using a suitable probability distributions but the natural question that arises here, is that limited account of probability distributions can't be used to model vast and diverse forms of data sets. Keeping this in mind researchers have developed some new probability distributions using different techniques. The most popular and innovative technique that is employed to obtain new distributions is Compounding of probability distributions. Compound distribution arises when all or some parameters of a distribution known as parent distribution vary according to some probability distribution called the compounding distribution for instance negative binomial distribution can be obtained from Poisson distribution when its parameter λ follows gamma distribution. If the parent distribution is discrete then resultant compound distribution will also be discrete and if the parent distribution is continuous then resultant compound distribution will also be continuous i,e. the support of the original (parent) distribution determines the support of compound distributionsBased on the same compounding mechanism the discrete mixtures of NBD were obtained by various authors. For instance, Gomez, Sarabia and Ojeda (2008) proposed a new compound negative binomial distribution by ascribing an inverse Gaussian distribution to the probability parameter in NBD. The authors also discussed some basic properties of this distribution. Zamaniet al. (2010) reparameterized the probability parameter by using the transformation $p=e^{-\lambda}$ and treated λ as a random variable following Lindley distribution the resulting compound distribution was named as negative binomial-Lindley distribution. Aryuyuen and Bodhisuwan (2013) obtained a compound of NBD with that of generalized exponential distribution by using the transformation $p=e^{-\lambda}$ and used a generalized exponential distribution as compounding distribution for parameter λ. Adil and Jan (2015a, 2016a) proposed two new competitive count data models, one is obtained by compounding Consul with Kumaraswamy distribution and other is obtained by compounding negative binomial with Kumaraswamy distribution.

In several research papers it has been found that compound distributions are very flexible and can be used efficiently to model different types of data sets. With this in mind many compound probability distributions have been constructed. Sankaran (1970) obtained a compound of Poisson distribution with that of Lindley distribution, Zamani and Ismail (2010) constructed a new compound distribution by compounding negative binomial with one parameter Lindley distribution that provides good fit for count data where the probability at zero has a large value. The researchers like Adil Rashid and Jan (2013,2014,2015,2016,2017) obtained several compound distributions with diverse application in engineering, medical and social sciences.

II. INGREDIENTS AND MECHANISM

A random variable X is said to have a Borel distribution (BD), if its p.m.f is given by

$$f_1(x,\lambda)=\begin{cases}\dfrac{(x\lambda)^{x-1}e^{-\lambda x}}{x!}, & x=1,2,3,\ldots\\ 0 & otherwise\end{cases} \tag{1}$$

where $0<\lambda<1$

Suppose a queue is initiated with one member and has traffic intensity under Poisson arrivals and and contant service time. Haight and Breuer (1960) stated that Borel distribution gives the probability that exactly *x* members of the queue will be served before the queue vanishes.

The mean and variance of the distribution is given by

$$\mu=(1-\lambda)^{-1} \text{ and } \sigma^2=\lambda(1-\lambda)^{-3}$$

In Borel distribution the parameter λ is a fixed constant but here we have considered a problem in which λ is itself a random variable following beta distribution (BTD) of first kind with probability function given by

$$f_2 = (\lambda, \alpha, \beta) = \frac{\lambda^{\alpha-1}(1-\lambda)^{\beta-1}}{B(\alpha,\beta)}$$

2.1 DEFINITION OF BOREL – BETA DISTRIBUTION

If $X|\lambda \sim$ BD(λ) where λ is itself a random variable following beta distribution BTD(α,β)then determining the distribution that results from marginalizing over λ will be known as a compound of Borel distribution with that of Beta distribution which is denoted by BBD (α,β)

It may be noted here that the domain of the compound distribution depends upon the domain of parent distribution which is BD; therefore proposed model will be a discrete since the parent distribution is discrete.

Theorem 1: the probability function of the newly proposed distribution is given by the expression

$$f(x;\alpha,\beta) = \frac{x^{x-1}}{x!B(\alpha,\beta)}\sum_{k=0}^{\infty}(-1)^k \frac{x^k}{k!}B(\alpha+x+k-1,\beta)$$

Where $x=1,2,...,\alpha,\beta>0$

Proof: The proof of above theorem will be obtained by using definition 2.1

$$f(x,\alpha,\beta) = \int_0^1 f_1(x|\lambda)f_2(\lambda,\alpha,\beta)d\lambda$$

$$= \int_0^1 \frac{(x\lambda)^{x-1}}{x!}e^{-\lambda x}\frac{\lambda^{\alpha-1}(1-\lambda)^{\beta-1}}{B(\alpha,\beta)}d\lambda$$

$$= \frac{x^{x-1}}{x!B(\alpha,\beta)}\int_0^1 \lambda^{x+\alpha-2}(1-\lambda)^{\beta-1}e^{-\lambda x}d\lambda$$

$$= \frac{x^{x-1}}{x!B(\alpha,\beta)}\int_0^1\sum_{k=0}^{\infty}\frac{(-\lambda x)^k}{k!}x^{\alpha+x-2}(1-\lambda)^{\beta-1}d\lambda$$

$$= \frac{x^{x-1}}{x!B(\alpha,\beta)}\sum_{k=0}^{\infty}(-1)^k\frac{x^k}{k!}\int_0^1 x^{\alpha+x+k-2}(1-\lambda)^{\beta-1}d\lambda$$

$$= \frac{x^{x-1}}{x!B(\alpha,\beta)} \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!} \int_0^1 x^{(\alpha+x+k-1)-1} (1-\lambda)^{\beta-1} d\lambda$$

$$= \frac{x^{x-1}}{x!B(\alpha,\beta)} \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!} B(\alpha+x+k-1,\beta) \qquad (3)$$

where $x = 1, 2, ..., \alpha, \beta > 0$. Where B(.) refers to the beta function

$$B(r,s) = \frac{\Gamma r \, \Gamma s}{\Gamma(r+s)}; \; r,s > 0$$

From this point, we denote a random X variable following a Borel- Beta distribution as BD$(x;\alpha,\beta)$

Corollary 1.1: the probability mass function of borel- Uniform distribution is

$$f(X) = \frac{x^{x-1}}{x!} \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!(x+k)}$$

Where $x = 1, 2, ...,$

Proof: Since it is known that uniform distribution is a particular case of beta distribution, so it is obvious to prove the above result when we put $\alpha = \beta = 1$ in (3)

$$f(X) = \frac{x^{x-1}}{x!} \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!} B(x+k,1)$$

$$f(X) = \frac{x^{x-1}}{x!} \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!} \frac{\Gamma(x+k)}{\Gamma(x+k+1)}$$

$$f(X) = \frac{x^{x-1}}{x!} \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!(x+k)}$$

III.MEAN AND VARIANCE OF THE BOREL-BETA MODEL

In order to obtain the m^{th} moment of the proposed model $BBD(X,\alpha,\beta)$ about origin we need to apply the well-known results of probability theory viz

(i) Conditional expectation identity $E(X^m) = E_\lambda(X^m \mid \lambda)$ and

(ii) Conditional variance identity $V(X) = E_\lambda(Var(X \mid \lambda)) + Var_p(E(X \mid \lambda)$

$$E\left(x^m\right)=E_\lambda\left(x^m \mid \lambda\right)$$

$$=E_\lambda\left(1-\lambda\right)^{-m}$$

$$=E\left[\sum_{j=0}^{\infty}\binom{m+j-1}{j}\lambda^j\right]$$

$$=\sum_{j=0}^{\infty}\binom{m+j-1}{j}E\left(\lambda^j\right)$$

$$=\sum_{j=0}^{\infty}\binom{m+j-1}{j}\int_0^1\lambda^j\,\frac{\lambda^{\alpha-1}\left(1-\lambda\right)^{\beta-1}}{B(\alpha,\beta)}d\lambda$$

$$=\frac{1}{B(\alpha,\beta)}\sum_{j=0}^{\infty}\binom{m+j-1}{j}\int_0^1\lambda^{\alpha+j-1}\left(1-\lambda\right)^{\beta-1}d\lambda$$

$$=\frac{1}{B(\alpha,\beta)}\sum_{j=0}^{\infty}\binom{m+j-1}{j}B\left(\alpha+j,\beta\right)$$

For $m=1$, we get mean of BBD

$$E(X)=\frac{1}{B(\alpha,\beta)}\sum_{j=0}^{\infty}B\left(\alpha+j,\beta\right)$$

$$E(X)=\sum_{j=0}^{\infty}\frac{\Gamma\left(\alpha+j\right)\Gamma\left(\alpha+\beta\right)}{\Gamma(\alpha)\Gamma(\alpha+\beta+j)}$$

$$V(X)=\sum_{j=0}^{\infty}(j+1)\frac{\Gamma\left(\alpha+j\right)\Gamma\left(\alpha+\beta\right)}{\Gamma(\alpha)\Gamma(\alpha+\beta+j)}-\left(\sum_{j=0}^{\infty}\frac{\Gamma\left(\alpha+j\right)\Gamma\left(\alpha+\beta\right)}{\Gamma(\alpha)\Gamma(\alpha+\beta+j)}\right)^2$$

$$SD(X)=\sqrt{\sum_{j=0}^{\infty}(j+1)\frac{\Gamma\left(\alpha+j\right)\Gamma\left(\alpha+\beta\right)}{\Gamma(\alpha)\Gamma(\alpha+\beta+j)}-\left(\frac{\Gamma\left(\alpha+j\right)\Gamma\left(\alpha+\beta\right)}{\Gamma(\alpha)\Gamma(\alpha+\beta+j)}\right)^2}$$

IV.COMPOUNDING OF BOREL-TANNER DISTRIBUTION WITH BETA DISTRIBUTION

Since it is known that if $X_1, X_2, X_3, ..., X_n$ are iid Borel random variables then $\sum_{i=1}^{n} X_i$ ~ Borel-Tanner distribution with pmf

$$P(Y=y)=\begin{cases} \dfrac{n}{y}\dfrac{(\lambda y)^{y-n}}{(y-n)!}e^{-\lambda y}, & y=n+1,n+2,..., \\ 0 & otherwise \end{cases}$$

In order to develop a compound of Borel-Tanner distribution with beta distribution, we treat the parameter λ in Borel-Tanner distribution as a beta random variable with pdf given (2)

Theorem 2: The probability function of compound of Borel-Tanner with beta distribution is given by the expression

$$f(Y,\alpha,\beta)=\frac{n}{y}\frac{y^{y-n}}{(y-n)!B(\alpha,\beta)}\sum_{k=0}^{\infty}(-1)^k\frac{y^k}{k!}B(\alpha+\beta+k-n,\beta)$$

Where $x=1,2,...,\alpha,\beta>0$

Proof: The proof of above theorem will be obtained by using definition 2.1

$$f(Y,\alpha,\beta)=\int_0^1 P(y|\lambda)f_2(\lambda,\alpha,\beta)d\lambda$$

$$=\frac{n}{y}\frac{y^{y-n}}{(y-n)!B(\alpha,\beta)}\int_0^1 \lambda^{\alpha+y-n-1}(1-\lambda)^{\beta-1}e^{-\lambda y}d\lambda$$

$$=\frac{n}{y}\frac{y^{y-n}}{(y-n)!B(\alpha,\beta)}\sum_{k=0}^{\infty}(-1)^k\frac{y^k}{k!}\int_0^1 \lambda^{\alpha+y-n-1}(1-\lambda)^{\beta-1}d\lambda$$

$$=\frac{n}{y}\frac{y^{y-n}}{(y-n)!B(\alpha,\beta)}\sum_{k=0}^{\infty}(-1)^k\frac{y^k}{k!}B(\alpha+\beta+k-n,\beta)$$

where $x=1,2,...,\alpha,\beta>0$. Where B(.) refers to the beta function

$$E(Y^m)=E_\lambda(Y^m|\lambda)$$

$$= nE_{\lambda}\left(1-\lambda\right)^{-m}$$

$$= nE\left[\sum_{j=0}^{\infty}\binom{m+j-1}{j}\lambda^{j}\right]$$

$$= n\sum_{j=0}^{\infty}\binom{m+j-1}{j}E\left(\lambda^{j}\right)$$

$$= \frac{n}{B(\alpha,\beta)}\sum_{j=0}^{\infty}\binom{m+j-1}{j}\int_{0}^{1}\lambda^{\alpha+j-1}\left(1-\lambda\right)^{\beta-1}d\lambda$$

$$= \frac{n}{B(\alpha,\beta)}\sum_{j=0}^{\infty}\binom{m+j-1}{j}B\left(\alpha+j,\beta\right)$$

For $m=1$, we get mean of BTBD

$$E(Y) = \frac{n\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{j=0}^{\infty}\frac{\Gamma(\alpha+j)}{\Gamma(\alpha+\beta+j)}$$

$$V(Y) = \frac{n\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{j=0}^{\infty}(j+1)\frac{\Gamma(\alpha+j)}{\Gamma(\alpha+\beta+j)} - \left(n\sum_{j=0}^{\infty}\frac{\Gamma(\alpha+j)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+j)}\right)^{2}$$

$$SD(Y) = \sqrt{\frac{n\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{j=0}^{\infty}(j+1)\frac{\Gamma(\alpha+j)}{\Gamma(\alpha+\beta+j)} - \left(n\sum_{j=0}^{\infty}\frac{\Gamma(\alpha+j)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+j)}\right)^{2}}$$

V.PARAMETER ESTIMATION BOREL- BETA DISTRIBUTION

$$\pounds(X;\alpha,\beta) = \log L(X;\alpha,\beta) = \sum_{i=1}^{n}\log\left(\frac{x^{x-1}}{x!}\right) + \sum_{i=1}^{n}\log\left(\frac{1}{B(\alpha,\beta)}\right) + \sum_{i=1}^{n}\log\left(\sum_{j=0}^{x}\frac{x^{k}}{k!}(-1)^{j}\ B\left(\alpha+x+k-1,\beta\right)\right)$$

$$\frac{\partial}{\partial\alpha}\pounds(X;,\alpha,\beta) = -\frac{\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\partial}{\partial\alpha}\left(\frac{\Gamma(\alpha)}{\Gamma(\alpha+\beta)}\right) + \sum_{i=1}^{n}\left(\sum_{j=0}^{x}\frac{x^{k}}{k!}(-1)^{j}\frac{\frac{\partial}{\partial\alpha}B\left(\alpha+x+k-1,\beta\right)}{\left(\sum_{j=0}^{x}\frac{x^{k}}{k!}(-1)^{j}\ B\left(\alpha+x+k-1,\beta\right)\right)}\right)$$

$$\frac{\partial}{\partial\beta}\pounds(X;,\alpha,\beta) = -\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\frac{\partial}{\partial\beta}\left(\frac{\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right) + \sum_{i=1}^{n}\left(\sum_{j=0}^{x}\frac{x^{k}}{k!}(-1)^{j}\frac{\frac{\partial}{\partial\beta}B\left(\alpha+x+k-1,\beta\right)}{\left(\sum_{j=0}^{x}\frac{x^{k}}{k!}(-1)^{j}\ B\left(\alpha+x+k-1,\beta\right)\right)}\right)$$

the above differential equations cannot be solved analytically and therefore one needs to estimate these unknown parameters by using Newton-Raphson method which is a powerful technique for solving equations iteratively and numerically.

VI. APPLICATION

In this section we shall explore the application part of BBD by fitting the real life data set based on mean vehicle bunch sizes. The data set has been taken from Taylor et al. (1974)

Table 1**:** Bunch Size Frequency Distribution of Australian Rural highways (Taylor et al., 1974)

Bunch Size	Observed Frequency	Fitted Distribution		
		Geometric	Borel-Tanner	BBD
1	6908	6143	6740	6998
2	1676	2280	1726	1541
3	575	846	663	530
4	235	314	302	261
5	160	117	151	165
6	72	43	80	79
7	48	16	44	46
8	22	6	25	19
9	8	2	15	16
10	20	1	9	13
11+	44	*	14	19
Chi square		5635.388	115.2373	81.84943

From the table, we find that chi-square is minimum for proposed distribution in comparison to others. So it can be concluded that BBD performs excellently well

VII. CONCLUDING REMARKS

In this paper we have proposed a new discrete distribution that has been named as Borel- Beta Distribution. The proposed distribution has proved to be very flexible in terms of modeling the real life traffic data sets based on bunch vehicle sizes. The efficiency and potentiality of BBD has been tested statistically by using the chi-square goodness of fit test

REFERENCES

[1.] Aryuyuen, S., and Bodhisuwan, W. Negative Binomial-Generalized Exponential Distribution, *Applied Mathematical Sciences*, **7**(22), 2013, 1093-1105

[2.] Adil, R., and Jan, T.R. (2013). A compound of zero truncated generalized negative binomial distribution with the generalized beta distribution. *Journal of Reliability and Statistical Studies,* 6(1), 11-19.

[3.] Adil, R., and Jan, T.R. A compound of Geeta distribution with the Generalized Beta Distribution.*Journal of Modern Applied Statistical Methods,*13(1), 2014a, 278-286.

[4.] Adil, R., and Jan, T.R. A compound of Consul distribution with the Generalized Beta Distribution. *Journal of Applied probability and Statistics Methods,*9(2), 2014b, 17-24.

[5.] Adil, R., and Jan, T.R. A new three parameter Consul Kumaraswamydistribution.*International journal of Modern Mathematical Sciences,* 13(4), 2015a. 336-376.

[6.] Adil, R., and Jan, T.R. A new family of lifetime distribution with real life application.*International Journal of Mathematics and Statistics,*17(1), 2016b, 23-38.

[7.] Adil, R., Zahoor, A., and Jan, T.R. A new count data model with application in genetics and ecology.*Electronic Journal of Applied Statistical Analysis,* 9(1), 2016a, 213-226.

[8.] Adil, R., Zahoor, A., and Jan, T.R. A generalized class of compound lifetime distributions with applications in real life.*Journal of Modern Applied Statistical Methods,* in Press 2018

[9.] Adil, R., and Jan, T.R. An Advanced count data model with application in genetics.*Journal of Modern Applied Statistical Methods,*in Press 2018.

[10.] Sankaran, M. The discrete Poisson-Lindley distribution. *Biometrics*, 26, 1970) 145–149.

[11.] Gomez, D.E., Sarabia, J.M., and Ojeda, E.C. Univariate and Multivariate Versions of the Negative Binomial-Inverse Gaussian Distributions with Applications. *Insurance Mathematics and Economics* . 42, 2008 39–49

[13.] Zamani, H., and Ismail, N. Negative Binomial–Lindley Distribution And Its Application. *Journal of Mathematics and Statistics*, 1, 2010, 4–9.