International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 7, 2008, pp 31-40

A Modified Bhattacharya and Holla Distribution and It's Applications

Anwar Hassan
PG Department of Statistics
University of Kashmir
Srinagar, India
E-mail: anwar.hassan2007@gmail.com

Mehraj Ahmed Bhat
PG Department of Statistics
University of Kashmir
Srinagar, India
E-mail: mehraj_stat@yahoo.co.in

[Received August 8, 2006; Revised March 20, 2008; Accepted March 30, 2008]

Abstract

A modified Bhattacharya and Holla distribution is proposed by mixing Consul and Jain's (1973) generalized Poisson distribution with uniform distribution. The proposed model is expected to cover a very wide range of situations. We study the properties and applications of this distribution. We also obtain the estimates of the parameters of the proposed distribution. The proposed distribution is then fitted to some observed sets of data and exhibits closer fits (alternative to Consul and Jain's (1973) generalized Poisson distribution) than the Bhattacharya and Holla's (1965) distribution.

Keywords and Phrases: Generalized Poisson distribution, Uniform distribution, Moments, Estimation, Goodness of fit.

AMS Classification: 62E15.

1 Introduction and Motivation

During the past forty years or so, much work seems to have been done on the mixture probability models. Gurland (1957) studied interrelations among compound and generalized distributions. Gurland (1958) also defined a general class of contagious distributions. Patil (1964) studied on certain compound Poisson and compound binomial distributions. Bhattacharya and Holla (1965) obtained a compound distribution by mixing Poisson and uniform distribution with the help of Bayesian procedure. As regards the application, the use of models is restricted to the theory of accident proneness. But in many practical situations the Bhattacharya and Holla's (1965) distribution fails to give satisfactory fits.

Consul and Jain (1973) defined a generalized Poisson distribution (GPD) with its probability function

$$P(X=x)=\frac{\lambda_1(\lambda_1+x\lambda_2)^{x-1}e^{-(\lambda_1+x\lambda_2)}}{x!}, \lambda_1>0, |\lambda_2|<1, \mathrm{x}=0,1,2,... \tag{1}$$

and obtained the first four moments of it as

$$\mu_1'=\frac{\lambda_1}{1-\lambda_2} \tag{2}$$

$$\mu_2'=\frac{\lambda_1}{(1-\lambda_2)^3}+\frac{\lambda_1^2}{(1-\lambda_2)^2} \tag{3}$$

$$\mu_3'=\frac{\lambda_1(1+2\lambda_2)}{(1-\lambda_2)^5}+\frac{3\lambda_1^2}{(1-\lambda_2)^4}+\frac{\lambda_1^3}{(1-\lambda_2)^3} \tag{4}$$

$$\mu_4'=\frac{\lambda_1(1+8\lambda_2+6\lambda_2^2)}{(1-\lambda_2)^7}+\frac{\lambda_1^2(7+8\lambda_2)}{(1-\lambda_2)^6}+\frac{6\lambda_1^3}{(1-\lambda_2)^5}+\frac{\lambda_1^4}{(1-\lambda_2)^4} \tag{5}$$

It can be easily seen that at $\lambda_2=0$, the distribution (1) reduces to Poisson distribution. The model (1) has been found to be a member of the Consul and Shenton's (1972) family of Lagrangian distributions and also of the Gupta's (1974) modified power series distribution (MPSD). Consul and Shenton (1973) have shown that the number of customers served in any busy period of the server has Lagrangian probability distribution. In particular, the model (1) represents the distribution of the number of customers served in a busy period when the arriving customers have a Poisson distribution and the number of customers waiting, before the service begins, also has a Poisson distribution. Janardan and Schaeffer (1977) and Janardan et al (1979) showed that the GPD model provides an excellent model for explaining aggregation patterns in biology and ecology. They found numerous applications of the GPD model for the analysis of chromosomal observations in human Leukocytes. Number of properties and estimation problems for this model have been studied by many researchers like Consul and Shoukri (1984), Consul and Famoye (1988), Shoukri and Consul (1989), Consul (1989), Hassan and Harman (2003) and Hassan, et al (2004). A brief list of authors and their works can be seen in Johnson and Kotz (1969), Johnson, Kotz and Kemp (1992) and Consul and Famoye (2006).

The motivation of the proposed model is, a computer center has two computer systems labeled A and B. Incoming jobs are independently routed to system A and system B with uniform probability. The number of jobs X, arriving per unit time has GPD (1) with parameters λ_1 and λ_2 . The distribution function of the number of jobs received by system A or B is obtained as by the modified Bhattacharya and Holla distribution (MBHD). It can be applied to a computer centre having more than two computer systems. Here it may be pointed out that all computer systems are of same efficiency so that incoming jobs are uniformly distributed among the computer systems themselves.

The modified Bhattacharya and Holla distribution (MBHD) has been found to cover a very wide range of discrete distributions and so is expected to have wide spectrum of its applications. Some specific applications of the proposed distribution may be used in finance, banking, insurance, medical science, biology, agriculture etc.

In this paper we propose the model by mixing of Consul and Jain's (1973) generalized Poisson distribution having two parameters λ_1 and λ_2 with the uniform distribution and obtained a new modified Bhattacharya and Holla distribution which is more general as compared to Bhattacharya and Holla's (1965) distribution mixing Poisson distribution with uniform distribution.. We study the structural properties and applications and also obtain the estimates of the parameters of the proposed distribution. The proposed distribution is then fitted to some observed sets of data and exhibits a remarkable closer fits (alternative to Consul and Jain's (1973) generalized Poisson distribution) than the Bhattacharya and Holla's (1965) distribution.

2 Proposed Model: A Modified Bhattacharya and Holla Distribution

The proposed model is derived by mixing Consul and Jain's (1973) generalized Poisson distribution (1) and uniform distribution with the probability density function

$$g(\lambda_1) = \begin{cases} \frac{1}{b-a}, & a \leq \lambda_1 \leq b \\ 0, & otherwise \end{cases} \tag{6}$$

Multiplying (1) and (6) and integrating over the range involved, the unconditional probability distribution of X is given by

$$P_1(X = x) = \frac{1}{b-a} \left(\int_a^b \frac{\lambda_1(\lambda_1 + x\lambda_2)^{x-1} e^{-(\lambda_1+x\lambda_2)}}{x!} d\lambda_1 \right), \ \lambda_1 > 0, |\lambda_2| < 1, \ \text{x} = 0, 1, 2, \ldots \tag{7}$$

The expression (7) under integral is the pmf of (1). Put $\lambda_1 + x\lambda_2 = \theta$, we get

$$P_1(X=x)=\frac{1}{b-a}\int\limits_{r_1}^{r_2}\frac{(\theta-x\lambda_2)\theta^{x-1}e^{-\theta}}{x!}d\theta \tag{8}$$

Where $r_1 = a + x\lambda_2$ and $r_2 = b + x\lambda_2$

$$P_1(X=x)=\frac{1}{b-a}\frac{r_1^x e^{-r_1}-r_2^x e^{-r_2}}{x!}+\frac{1}{b-a}(1-\lambda_2)\int\limits_{r_1}^{r_2}\frac{\theta^{x-1}e^{-\theta}}{(x-1)!}d\theta \tag{9}$$

$$\begin{aligned}P_1(X=x)=\tfrac{1}{b-a}&\left[\tfrac{r_1^x e^{-r_1}-r_2^x e^{-r_2}}{x!}+(1-\lambda_2)\left\{e^{-r_1}\left(1+\tfrac{r_1}{1!}+\tfrac{r_1^2}{2!}+\cdots+\tfrac{r_1^{x-1}}{(x-1)!}\right)\right.\right.\\&\left.\left.-e^{-r2}\left(1+\tfrac{r_2}{1!}+\tfrac{r_2^2}{2!}+\cdots+\tfrac{r_2^{x-1}}{(x-1)!}\right)\right\}\right]\end{aligned} \tag{10}$$

This can be named as modified Bhattacharya and Holla distribution (MBHD).

Special Cases:
When $\lambda_2 = 0$ the models(7) and (10) reduces to the classical Bhattacharya and Holla's (1965) distribution.

3 Properties of Proposed Model

The first four moments of the proposed model about origin are obtained by using the relation (2) to (5) as

$$\mu_1'=\frac{1}{1-\lambda_2}\frac{b+a}{2} \tag{11}$$

$$\mu_2'=\frac{1}{(1-\lambda_2)^3}\frac{b+a}{2}+\frac{1}{(1-\lambda_2)^2}\frac{a^2+b^2+ab}{3} \tag{12}$$

$$\mu_3'=\frac{1+2\lambda_2}{(1-\lambda_2)^5}\frac{b+a}{2}+\frac{b^2+a^2+ab}{(1-\lambda_2)^4}+\frac{1}{(1-\lambda_2)^3}\frac{b^3+a^2b+ab^2+a^3}{4} \tag{13}$$

$$\begin{aligned}\mu_4'=&\tfrac{1+8\lambda_2+6\lambda_2^2}{(1-\lambda_2)^7}\tfrac{b+a}{2}+\tfrac{7+8\lambda_2}{(1-\lambda_2)^6}\tfrac{a^2+b^2+ab}{3}+\tfrac{3}{(1-\lambda_2)^5}\tfrac{a^3+b^3+a^2b+ab^2}{2}\\&+\tfrac{1}{(1-\lambda_2)^4}\tfrac{a^4+b^4+a^3b+ab^3+a^2b^2}{5}\end{aligned} \tag{14}$$

The moments about mean is derived by using the relation (11) to (14) as Variance

$$\mu_2=\frac{1}{(1-\lambda_2)^2}\frac{(b-a)^2}{12}+\frac{1}{(1-\lambda_2)^3}\frac{b+a}{2} \tag{15}$$

$$\mu_3 = \frac{1+2\lambda_2}{(1-\lambda_2)^5}\frac{b+a}{2} + \frac{1}{(1-\lambda_2)^4}\frac{(b-a)^2}{4} \tag{16}$$

$$\begin{aligned}\mu_4 = &\tfrac{1+8\lambda_2+6\lambda_2^2}{(1-\lambda_2)^7}\tfrac{b+a}{2} + \tfrac{1}{(1-\lambda_2)^6}\left[\tfrac{4a^2+4b^2+ab+2\lambda_2(b-a)^2}{3}\right] \\ &+\tfrac{1}{(1-\lambda_2)^5}\tfrac{(a+b)(b-a)^2}{4} + \tfrac{1}{(1-\lambda_2)^4}\tfrac{(b-a)^4}{80}\end{aligned} \tag{17}$$

The skewness of the proposed models is easily be obtained by using the expressions given in (15) and (16) as

$$\gamma_1 = \sqrt{\frac{108}{1-\lambda_2}}\left(\frac{2(1+2\lambda_2)(b+a)+(b-a)^2(1-\lambda_2)}{[6(b+a)+(b-a)^2(1-\lambda_2)]^{3/2}}\right) \tag{18}$$

Clearly for given values of a and b the Skewness of models (7) and (10) decreases as the value of λ_2 decreases and its Skewness increases as the value of λ_2 increases. The skewness becomes indefinitely large when the value of λ_2 is very close to unity. The skewness becomes negative when $\lambda_2 < -\frac{1}{2}$ such that

$$\frac{(b-a)^2}{b+a} < 4 - \frac{6}{1-\lambda_2}, \quad \Pr ovided\ \ b+a \neq 0 \tag{19}$$

For $\lambda_2 < -\frac{1}{2}$, the condition (19) does not hold then skew ness is always positive.

The Kurtosis of the models is obtained by using the expressions given in (15) and (17) as

$$\beta_2 = 3 + \frac{6}{5(1-\lambda_2)}\left[\frac{60(6\lambda_2^2+8\lambda_2+1)(b+a)+10(b-a)^2(1-\lambda_2)(7+8\lambda_2)-(1-\lambda_2)^3(b-a)^4}{36(b+a)^2+12(1-\lambda_2)(b+a)(b-a)^2+(1-\lambda_2)^2(b-a)^4}\right] \tag{20}$$

For given ‘a’ and ‘b’ the value of β_2 is more than 3 for $\left(\frac{1}{6}\sqrt{10}-\frac{2}{3}\right) < \lambda_2 < 1$. This reveals that MBHD models are Leptokurtic. But as the value of λ_2 goes on decreased within the interval $\left(\frac{\sqrt{10}}{6}-\frac{2}{3}\right) < \lambda_2 < 1$ the peak of the models for given values of a and b goes lower and lower and approaches the normal form. The peak will be highest when λ_2 is close to the unity. Further more for given ‘a’ and ‘b’ the value of β_2 is less than 3 for -1 $< \lambda_2 \leq -7/8$. The models are Platykurtic in this interval. However models are also Platykurtic in the interval $7/8 < \lambda_2 \leq \left(\frac{\sqrt{10}}{6}-2/3\right)$, if the condition for fixed a and b

$$\frac{60\left(6\lambda_2^2+8\lambda_2+1\right)}{(1-\lambda_2)\left[(1-\lambda_2)^2(b-a)^2-10\,(7+8\lambda_2)\right]} < \frac{(b-a)^2}{b+a}$$

is satisfied provided $a+b \neq 0$.

Further more if the above condition is not satisfied then proposed models are Leptokurtic even in the interval -7/8 $< \lambda_2 \leq \left(\frac{\sqrt{10}}{6} - 2/3\right)$.

4 Estimation of Proposed Model

The proposed models (7), and (10) is not simple to estimate its parameters through maximum likelihood estimate. So we use moments method for their estimates as $m'_r = \mu'_r$ where μ'_r are population moments about origin and m'_r are sample moments about origin. Using (11), (12) and (13) after little simplifications, we get the estimates of parameters as

$$\hat{b} = (1 - \hat{\lambda}_2)\bar{x} + \sqrt{3\left[(1 - \hat{\lambda}_2)^2 s^2 - \bar{x}\right]} \tag{21}$$

$$\hat{a} = (1 - \hat{\lambda}_2)\bar{x} - \sqrt{3\left[(1 - \hat{\lambda}_2)^2 s^2 - \bar{x}\right]} \tag{22}$$

where $\bar{x}$ and $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = m'_2 - \bar{x}^2$ are sample mean and sample variance and $m'_3 = \frac{1}{n}\sum_{i=1}^{n} x_i^3$. From (21) and (22) we have

$$(m'_3 + 2\bar{x}^3 - 3\bar{x}m'_2)\,(1-\lambda_2)^3 - 3(\mathrm{m'}_2 - \bar{x}^2)\,(1-\lambda_2) + 2\bar{x} = 0 \tag{23}$$

The estimates of above parameters can be obtained only if $(1 - \lambda_2)^2 s^2 > \bar{x}$. The three values of λ_2 can be obtained from the cubic equation (23) by using Newton Raphson method, the value of λ_2 that satisfies the following set of conditions (24) can be taken as estimate of λ_2.

$$\left.\begin{array}{rl} & (1 - \lambda_2)^2 s^2 > \bar{x} \\ \textit{and} & |\lambda_2| < 1 \end{array}\right\} \tag{24}$$

5 Goodness of Fit

The modified Bhattacharya and Holla distribution contains one additional parameter λ_2 and hence it is hoped that this form of the distribution should explain the variation in data sets a better way than the Bhattacharya and Holla's (1965) distribution.

The distribution has been fitted to a number of standard data sets which has earlier been used by others. It is encouraging to report that in almost all these cases, the modified form of the Bhattacharya and Holla distribution gives much closer fits than the present form of the Bhattacharya and Holla 's (1965) distribution. Only two of such cases being given here in Table 1 and 2.

Table 1: Absenteeism among shift-workers in steel industry: data sets of Arbous and Sichel, (1954)

Count	**Observed frequency**	**Expected frequencies**	
		Bhattacharya & Holla's (1965) distribution (BHD)	**Proposed Model Modified BHD**
0	7	9.83	8.02
1	16	15.93	15.96
2	23	19.10	19.25
3	20	19.92	19.91
4	23	19.64	20.54
5	24	18.41	19.29
6	12	16.59	16.14
7	13	14.90	14.64
8	09	13.25	13.09
9	09	11.71	11.59
10	08	10.30	9.98
11	10	9.04	9.64
12	08	7.92	7.48
13	07	6.94	7.03
14	02	6.08	6.85
15	12	5.33	6.23
16	03	4.68	4.38
17	05	4.12	3.98
18	04	3.63	3.12
19	02	3.20	3.01
20	02	2.87	2.95
21	05	2.66	2.88
22	05	2.79	2.40
23	02	1.97	2.26
24	01	1.85	2.05
25-48	16	15.34	15.33
TOTAL	**248**	**248**	**248**
Estimate		$a = 1.173$ $b = 20.251$	$a = 0.374$ $b = 12.52$ $\lambda_2 = 0.324$
χ^2		9.476284	6.445375
d.f		17	16

Table 2: Counts of the number of European red mites on apple leaves; data of Bliss (1953, table-1)

Count	Observed frequency	Expected frequencies	
		Bhattacharya & Holla's (1965) distribution (BHD)	Proposed Model Modified BHD
0	70	69.42	70.33
1	38	37.71	34.61
2	17	20.21	19.96
3	10	10.77	12.02
4	09	5.30	6.97
5	03	3.05	3.62
6	02	1.68	16.14
7	01	0.88	1.64
8	00	0.98	0.62
			0.23
TOTAL	**150**	**150**	**150**
Estimate		a = 1.214	a = 0.321
		b = 22.34	b = 11.983
			$\lambda_2 = 0.351$
χ^2		3.20782	1.70523
d.f		3	2

6 Acknowledgement

The authors thank very much the anonymous referee for his many helpful suggestions, which substantially simplified the paper.

References

[1] Bhattacharya, S. K. and Holla, M. S. (1965): On a discrete distribution with special reference to the theory of accident proneness; Journal of the American Statistical Association, 60, 1060-1066.

[2] Consul, P. C. (1989): "Generalized Poisson distribution properties and applications". John Willey, New York.

[3] Consul, P. C. and Famoye, F. (1988): "Maximum likelihood estimation for generalized Poisson distribution when the sample mean is larger than the sample variance", Comm, Statist Theory Meth 18(10), 3635-3648.

[4] Consul, P. C. and Famoye, F. (2006): Lagrangian probability distribution, Birkhäuser publications.

[5] Consul, P. C. and Jain (1973): A generalization of Poisson Distribution: Technometrics, 15(4), 791-799.

[6] Consul, P. C. and Shenton, L. R. (1972): Use of Lagrangian expansion for generating discrete generalized probability distributions, SIAM, J Appl. Math, 23(2), 239-248.

[7] Consul, P. C. and Shenton, L. R. (1973): On the probabilistic structure and properties of screte lagrangian distributions, statistical distributions in scientific work, Vol.1, Eds.

[8] Consul, P. C. and Shoukri M. M. (1984): "Maximum likelihood estimation for the generalized Poisson distribution", Comm Statist Theory, Meth, 13(12), 1533-1547.

[9] Gupta , R. C. (1974): Modified power series distribution and its application, Sankhya, Ser B, 35, 288-298.

[10] Gurland, J. (1957): Some interrelations among compound and generalized distributions, Biometriko, 44, 265-268.

[11] Gurland, J. (1958): A General class of contagious distributions, Biometrics 14, 229-249.

[12] Hassan, A. and Harman, S. (2003): "A new Bayesian estimation of generalized Poisson distribution". Szigma, xxxiv 3(4), 127-134.

[13] Hassan, A., Raja, T. A. and Patel, A. (2004): "On fitting of Poisson and Lagrangian Probability distributions to leaf spot grades in mulberry" Research Journal, Assam Statistical Review, 18(1), 1-9.

[14] Johnson, N. L. and Kotz, S. (1969): Univariate discrete distribution, John Wiley & Sons.

[15] Johnson, N. L., Kotz, S. and Kemp, A. W. (1992) Univariate discrete distribution, John Wiley & Sons.

[16] Janardan, K. G. and Schaeffer, D. J. (1977): Model for the analysis of Chromosomal aberration in human leukocytes, Biometrical J, 19, 599-612.

[17] Janardan, K. G., Kerster, W. and Schaeffer, D. J. (1979): Biological application of the Lagrangian Poisson distribution, Bio Science., 29, 599-602.

[18] Patil, G. (1964): On certain compounds Poisson and Compounds binomial distribution, Synkhya Series A, 26, 293-294.

[19] Shoukri, M. M. and Consul, P. C. (1989): "Bayesian analysis of a generalized Poisson distribution", Comm, Statist, Simula and Computa 18(4), 1465-1480.