International Journal of Scientific Research in Mathematical and Statistical Sciences
Research Paper
Volume-5, Issue-5, pp.60-64, October (2018)
E-ISSN: 2348-4519

A Bivariate Binomial Distribution

S. R. Supanekar

Department of Statistics, Balasaheb Desai College, Patan, Dist. Satara, India(MS) 415206

Corresponding Author: sup_sanjay@yahoo.com, Tel.: +91-7083881971

Available online at: www.isroset.org

Received: 01/Oct/2018, Accepted:11/Oct/2018, Online: 31/Oct/2018

Abstract— In this paper, we introduce a bivariate binomial distribution that allows positive, zero or negative correlation between two variables which depends on multiplicative factor parameter. The marginal distributions of the bivariate binomial distribution are the univariate binomial distributions. Moment estimators and maximum likelihood estimators of parameters involved in this distribution are discussed. Sample observations are simulated from bivariate binomial distribution using the conditional distribution technique and test of the goodness of fit is carried out.

Keywords— *Bivariate binomial distribution; Maximum likelihood estimation; Test of the goonness of fit*

I. INTRODUCTION

Some bivariate discrete distributions like Poisson, Geometric, and Binomial etc. play a vital role in analyzing real life situations. The different bivariate distributions are available in the literature. Numerous bivariate discrete distributions have been studied in Kocherlakota and Kocherlakota [1]. The bivariate Poisson distribution is available in the literature due to Holgate [2] and Teicher [3] considers non-trivial cases of multivariate Poisson distribution. Karlis *et al.* [4] has also mentioned the application of correlated variables in sports data. Limitations of the bivariate distributions based on trivariate reduction method and copula functions are recognized in the literature

Lakshminarayana *et al*. [5] have proposed a bivariate Poisson distribution as a product of Poisson marginals with taking an account of positive, zero or negative correlation between two variables depends on multiplicative factor parameter. They further discussed estimation based on method of moments and provided illustration. In the same line as above Famoye [6] developed a new bivariate generalized Poisson distribution using generalized Poisson marginals and Supanekar and Shirke [7] introduced a new Bivariate Generalized Power Series Distribution (BGPSD). Supanekar [8] introduced a bivariate discrete distribution from Freund bivariate exponential distribution. This distribution is constructed by transforming a bivariate continuous random vector to a bivariate discrete random vector by considering integer part of the continuous random vector.

In this paper we introduced a bivariate binomial distribution (BBD), which is particular case of BGPSD. Rest of the paper is organized as follows; the next Section II contributes to construction of bivariate binomial distribution In Section III moments of BBD are calculated . Section IV describes the moment estimators and maximum likelihood estimators of parameters involved in the BBD. In Section V sample observations are simulated from BBD and test of the goodness of fit is carried out. In the last Section VI concludes research work.

II. BIVARIATE BINOMIAL DISTRIBUTION

Definition:

Lakshminarayana *et al.* [5] and Johnson *et al.* [9] have specified joint probability mass function of bivariate distribution using the framework introduced by Sarmanov [10]. The joint probability mass function is given by,

$$P(x,y,\theta_1,\theta_2,\alpha)=P(x,\theta_1)P(y,\theta_2) \times\{1+\alpha[g_1(x)-E(g_1(X))][g_2(y)-E(g_2(Y))]\} \quad \textbf{(1)}$$

where $x \in\{0, 1, 2, \ldots\}$, $y \in\{0, 1, 2, \ldots\}$, θ_1 and θ_2 are positive real numbers, $P(x,\theta_1), P(y,\theta_2)$ are marginal probability mass functions. The functions $g_1(x)$ and $g_2(y)$ can be any arbitrary 'bounded' functions of x and y with finite mean. According to Lakshminarayana *et al.* [5] the choice of $g_1(x)$ and $g_2(y)$ are e^{-x} and e^{-y} respectively. The parameter α is a suitably chosen real number such that, $P(x,y,\theta_1,\theta_2,\alpha)$ will be non-negative for all x, y ≥ 0. Hence it satisfy the condition $1+\alpha[g_1(x)-E(g_1(X))][g_2(y)-E(g_2(Y))]\geq 0$ for all x, y ≥ 0 . Here it is possible to chose negative value of α and hence the present model allows the correlation coefficient between x and y to be negative or positive by a proper choice of the parameter α.

Using the same approach we define bivariate binomial distribution by taking univariate binomial distribution as marginal probability function. From (1) we define bivariate binomial distribution as

$$P(x,y,n_1,n_2,p_1,p_2,\alpha)=\binom{n_1}{x}\binom{n_2}{y}p_1^x p_2^y q_1^{n_1-x}q_2^{n_2-y}$$
$$\times\{1+\alpha[g_1(x)-E(g_1(X))][g_2(y)-E(g_2(Y))]\}$$

where

$p_1+q_1=1,\ p_2+q_2=1$ and x $\in\{0,1,2,\ldots n_1\}$, y $\in\{0,1,2,\ldots n_2\}$.

Let $g_1(x)=e^{-x},\ g_2(y)=e^{-y}$ and using their expectations, joint probability mass function can be rewritten as

$$P(x,y,n_1,n_2,p_1,p_2,\alpha)=\binom{n_1}{x}\binom{n_2}{y}p_1^x p_2^y q_1^{n_1-x}q_2^{n_2-y}$$
$$\times\{1+\alpha(e^{-x}-m_1)(e^{-y}-m_2)\} \qquad \textbf{(2)}$$

where $m_1=E(e^{-X})=(1-cp_1)^{n_1}$,

$m_2=E(e^{-Y})=(1-cp_2)^{n_2}$ and c = (1-1/e)

Probability mass function of bivariate binomial distribution for $n_1 = n_2 = 4$ and various values of parameters p_1, p_2, α is plotted and shown in Figure 1 to Figure 6.

III. Moments of Bivariate Binomial Distribution

$$E(X)=n_1p_1,\ \ V(X)=n_1p_1q_1,$$
$$E(Y)=n_2p_2,\ \ V(Y)=n_2p_2q_2,$$
$$Cov(X,Y)=\alpha[A-n_1p_1m_1][B-n_2p_2m_2],$$

where

$$A=E(Xe^{-X})=\frac{n_1p_1}{e}(1-cp_1)^{n_1-1},$$
$$B=E(Ye^{-Y})=\frac{n_2p_2}{e}(1-cp_2)^{n_2-1}.$$

IV. Estimation of the Parameters

Let (x_i, y_i), $i=1,2,\ldots n$ be a random sample of size n from bivariate binomial distribution.

Define

$$\bar{X}=\frac{\sum_{i=1}^n x_i}{n},\quad \bar{Y}=\frac{\sum_{i=1}^n y_i}{n},\quad c(\underline{x},\underline{y})=\frac{\sum_{i=1}^n (x_i-\bar{X})(y_i-\bar{Y})}{n}.$$

Moment estimators:

We know that

$E(X)=n_1p_1,\ \ E(Y)=n_2p_2$ and

$Cov(X,Y)=\alpha[A-n_1p_1m_1][B-n_2p_2m_2]=\alpha h_4(p_1,p_2)$.

Therefore moment estimators of p_1, p_2 and α are given by

$$\hat{p}_1=\bar{X}/n_1,\quad \hat{p}_2=\bar{Y}/n_2\quad \text{and}\quad \hat{\alpha}=\frac{c(\underline{x},\underline{y})}{h_4(\hat{p}_1,\hat{p}_2)}.$$

Maximum likelihood estimators:

The likelihood function for observed data ($\underline{x}$, $\underline{y}$) is given by

$$L(p_1,p_2,\alpha;\underline{x},\underline{y})=$$
$$=\prod_{i=1}^n\binom{n_1}{x_i}p_1^{x_i}q_1^{n_1-x_i}\prod_{i=1}^n\binom{n_2}{y_i}p_2^{y_i}q_2^{n_2-y_i}$$
$$\times\prod_{i=1}^n\{1+\alpha k_1(x_i,p_1)k_2(y_i,p_2)\}$$

where

$k_1(x_i,p_1)=(e^{-x_i}-m_1)$ and $k_2(y_i,p_2)=(e^{-y_i}-m_2)$

Now

$$Log(L)=\sum\log\binom{n_1}{x_i}+\sum x_i\log p_1+\sum(n_1-x_i)\log(1-p_1)$$
$$+\sum\log\binom{n_2}{y_i}+\sum y_i\log p_2+\sum(n_2-y_i)\log(1-p_2)$$
$$+\sum_{i=1}^n\log\{1+\alpha k_1(x_i,p_1)k_2(y_i,p_2)\}$$

The maximum likelihood estimators of p_1, p_2 and α are the solution of following likelihood equations.

$$\frac{\sum x_i}{\hat{p}_1}-\frac{\sum(n_1-x_i)}{1-\hat{p}_1}+\hat{\alpha}\sum_{i=1}^n\frac{k_1'(x_i,\hat{p}_1)k_2(y_i,\hat{p}_2)}{\{1+\hat{\alpha}k_1(x_i,\hat{p}_1)k_2(y_i,\hat{p}_2)\}}=0 \qquad \textbf{(3)}$$

$$\frac{\sum y_i}{\hat{p}_2}-\frac{\sum(n_2-y_i)}{1-\hat{p}_2}+\hat{\alpha}\sum_{i=1}^n\frac{k_1(x_i,\hat{p}_1)k_2'(y_i,\hat{p}_2)}{\{1+\hat{\alpha}k_1(x_i,\hat{p}_1)k_2(y_i,\hat{p}_2)\}}=0 \qquad \textbf{(4)}$$

$$\sum_{i=1}^n\frac{k_1(x_i,\hat{p}_1)k_2(y_i,\hat{p}_2)}{\{1+\hat{\alpha}k_1(x_i,\hat{p}_1)k_2(y_i,\hat{p}_2)\}}=0 \qquad \textbf{(5)}$$

where

$$k_1'(x_i,\hat{p}_1)=\frac{\partial k_1(x_i,\hat{p}_1)}{\partial\hat{p}_1}=n_1c(1-cp_1)^{n_1-1}$$
$$k_2'(y_i,\hat{p}_2)=\frac{\partial k_2(y_i,\hat{p}_2)}{\partial\hat{p}_2}=n_2c(1-cp_2)^{n_2-1}$$

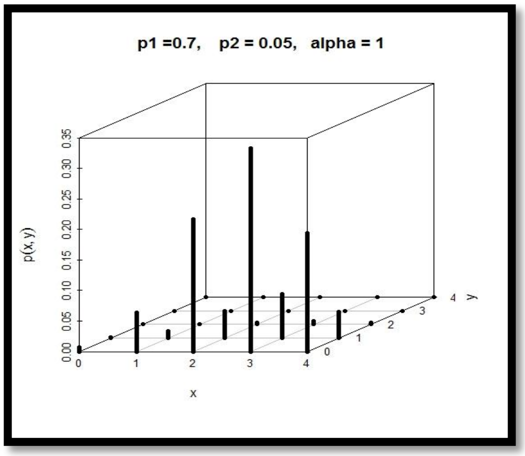

Figure (1)

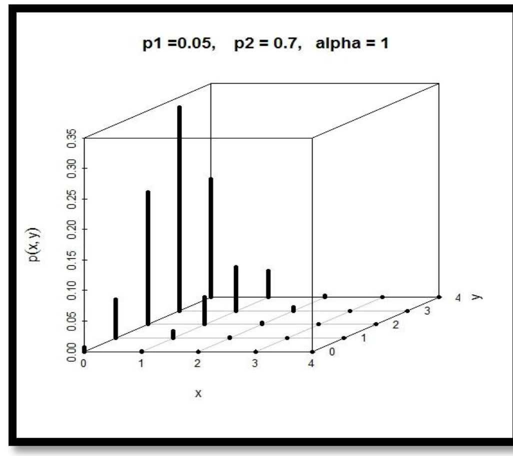

Figure (2)

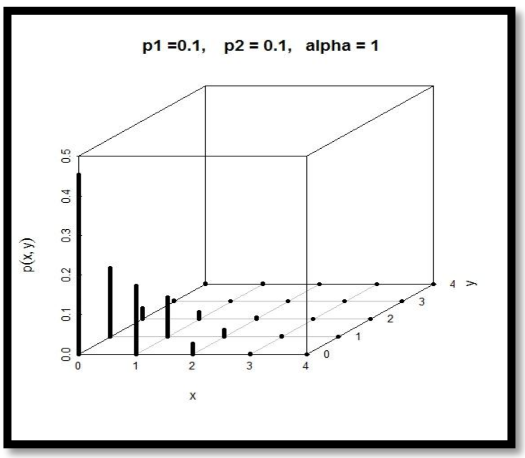

Figure (3)

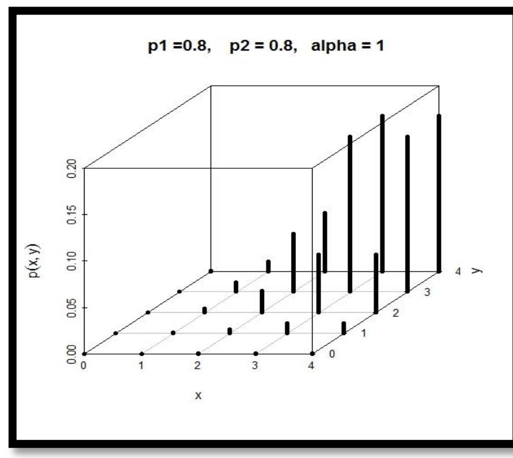

Figure (4)

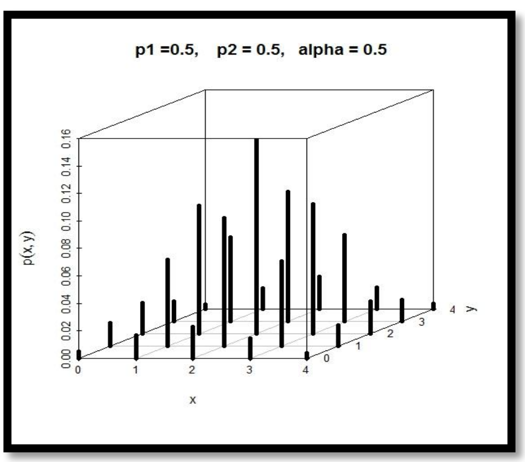

Figure (5)

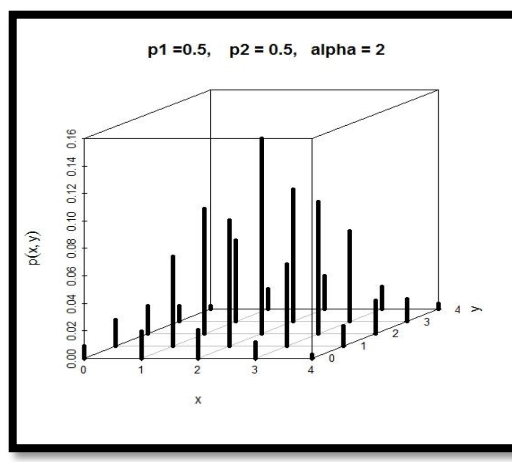

Figure (6)

V. SIMULATION FROM BIVARIATE BINOMIAL DISTRIBUTION

The marginal distribution of Y is B(n_2,p_2) and the pmf of Y is given by

$$P(y) = \binom{n_2}{y} p_2^y q_2^{n_2 - y} \qquad y = 0, 1, \ldots n_2$$

The conditional probability distribution of X given Y= y is given by

$$P(x \mid y) = \binom{n_1}{x} p_1^x q_1^{n_1 - x} \{1 + \alpha(e^{-x} - m_1)(e^{-y} - m_2)\}$$

Using the conditional distribution technique, we draw a sample of size n. In this technique, a realization of Y is generated from the distribution $P(y)$. For this simulated value y, a realization of X from $p(x|y)$ is generated; hence, the resulting pair (x, y) is an observation from the joint distribution $p(x; y; p_1, p_2, \alpha)$. This procedure is repeated n times to give a random sample of size n.

We generate 500 pairs of observations from a bivariate binomial distribution, with parameters $n_1 = 4$, $n_2 = 4$, $p_1=0.3$, $p_2 = 0.5$ and $\alpha = 1$. Table 1 gives the simulated data.

Table 1 : Simulated random sample from BBD, n=500.

X\Y	0	1	2	3	4
0	7	35	42	24	3
1	6	51	75	50	16
2	12	41	54	35	9
3	2	5	12	16	2
4	0	0	2	1	0

Table 2 : Fitting of bivariate data to BBD for $p_1 = 0.3$, $p_2 = 0.5$ and $\alpha = 1$

<table>
<tr><th>X\Y</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>0</td><td>7
(8.78)</td><td>35
(29.50)</td><td>42
(41.14)</td><td>24
(26.66)</td><td>3
(6.60)</td></tr>
<tr><td>1</td><td>6
(12.45)</td><td>51
(50.66)</td><td>75
(76.46)</td><td>50
(51.09)</td><td>16
(12.78)</td></tr>
<tr><td>2</td><td>12
(7.58)</td><td>41
(33.65)</td><td>54
(52.31)</td><td>35
(35.32)</td><td>9
(8.87)</td></tr>
<tr><td>3</td><td colspan="2">7
(12.24)</td><td>12
(15.80)</td><td colspan="2">18
(13.40)</td></tr>
<tr><td>4</td><td colspan="5">3
(4.68)</td></tr>
</table>

The distribution was fitted using maximum likelihood estimator of the parameters. The maximum likelihood estimators for the data given in Table 1 are given by $p_1=0.311$, $p_2 =0.5$ and $\alpha = 0.542$. The observed frequencies and expected frequencies in parentheses are given in the Table 2. The calculated value of chi-square is 17.4114 with 15 degrees of freedom. Since p-value is 0.2948, indicating that the fit for this distribution is good.

VI. CONCLUSION

In this Paper bivariate binomial distribution is introduced. It is a particular case of bivariate generalized power series distribution. The marginal distributions of the bivariate model are the univariate binomial distributions. The graph of probability mass function of bivariate binomial distribution for various values of parameters is plotted. Moment estimators and maximum likelihood estimators of parameters involved in this distribution are discussed. Simulation study is undertaken for this distribution and test the goodness fit for simulated data. This bivariate binomial distribution has more scope to study and use in real life situations.

REFERENCES

[1] S. Kocherlakota, K. Kocherlakota, "*Bivariate Discrete Distributions",* Marcel Dekker, New York, 1992

[2] P. Holgate, "*Estimation for the bivariate Poisson distribution"*, Biometrika, Vol.51, pp.241-245, 1964.

[3] H. Teicher, "*On the multivariate Poisson distribution",* Skandinavisk Aktuarietidskrift, Vol.37, pp.1-9, 1954

[4] D. Karlis and I. Ntzoufras, "*Analysis of sports data by using bivariate oisson models",* The Statistician, Vol.52(3), pp.381-393, 2003.

[5] J. Lakshminarayana, S. N. N. Pandit, K. Rao, Srinivasa, "*On Bivariate Poisson distribution",* Commun. Statistics-Theory Methods, Vol.28(2), pp.267-276, 1999

[6] F. Famoye, "*A new bivariate generlized Poisson distribution*", Statistica Neerlandica, Vol. 64(1), pp.112-124, 2010

[7] S. R. Supanekar and D. T. Shirke, "*A new bivariate generalized power series distributiion",* Int. J. Agricult. Stat. Sci., Vol. 10(2), pp.343-349, 2014

[8] S. R. Supanekar, "*A Bivariate Discrete Distribution from Freund Bivariate Exponential Distribution*", International Journal of Scientific Research in Mathematical and Statistical Sciences, Vol.5(2), pp.243-246, 2018

[9] N. L. Johnson, S. Kotz, and N. Balakrishnan, "*Discrete multivariate distributions",* John Wiley and Sons, New York, 1997.

[10] O. V. Sarmanov, "*Generalized normal correlation and two-dimensional Freshet classes*", Doklady (Sovyet Mathematics), Tom 168, pp.596-599, 1966

AUTHOR PROFILE

Mr. S. R. Supanekar completed M.Sc. and Ph.D. in Statistics from Shivaji University Kolhapur in 1993 & 2016. He is currently working as Assistant Professor in Department of Statistics, Balasaheb Desai College, Patan, Dist. Satara, India (MS). He has published 3 research papers in reputed international journals. His main research work focuses on discrete distributions. He has 21 years of teaching experience and 6 years of research experience.

