Canadian Studies in Population, Vol. 20(2), 1993, pp. 193-205.

On the Distribution of Births in Human Populations

Arun Kumar Sinha and **Rajiv Kumar**
Patna University
Patna, India

Abstract

Brass (1958) observed that the negative binomial distribution (NBD) describes the distribution of births in human populations fairly well. However, Onyemaekeogum (1978) has contradicted this finding. Both used data from African countries although, collected at different points of time. This paper attempts to investigate the adequacy of NBD to fit data for India. The results indicate that the model both in its complete and truncated forms provides reasonably good fit to most of the observed distributions of births. The maximum likelihood (ML) estimators of the parameters of the model have been used for the fitting.

Résumé

Brass (1958) observe que la distribution binomiale négative (NBD) décrit assez bien la distribution des naissances dans les populations humaines. Cependant, Onyemaekeogum (1978) le contredit sur ce point. Tous deux utilisent pourtant des données issues de pays africains, quoique recueillies à des moments différents. Le présent article tente d'examiner l'adéquation de la NBD ajustée aux données de l'Inde. Selon les résultats, le modèle convient raisonnablement bien à la plupart des distributions des naissances observées, à la fois dans sa forme complète et sa forme tronquée. Les estimateurs de vraisemblance maximale des paramètres du modèle ont servi à évaluer la qualité de l'ajustement.

Key Words: negative binomial distribution, births, India

Introduction

Mathematical models are often employed to describe demographic phenomena and their interrelationships. In fact, these models are invaluable tools in the study of the population dynamics of places with inadequate or defective data. Dandeker (1955) was perhaps the first to attempt to construct models that describe the distribution of births in human populations. However, neither Poisson nor binomial nor various modifications provided a satisfactory fit to the observed distributions. The data he used for this purpose were collected as a part of a socio-economic survey of Kolhapur city (Maharashtra, India) in 1945.

With the same goal, Brass (1958) was probably the first to conclude that "in suitable conditions, the number of women with no children as well as the birth distribution for mothers follows, approximately, the negative binomial form." In fact, Brass observed two distinctly different patterns of discrepancies between the observed and theoretical distributions for countries of low and high fertility. Extending the work further, Singh (1963) proposed two probability distributions to describe variations in the number of births to a couple during a given time interval. In order to test the adequacy of his model (A1) he used the two observed distributions of the number of children born to women belonging to two different age groups viz. 26-30 and 21-25. But contrary to the findings of Brass (1958) regarding populations with high fertility, Onyemaekeogum (1978) found that the negative binomial distribution (NBD) failed in describing the distribution of births in the populations comprising Ghana, Tanzania and Zambia. Nevertheless, he noticed that the pattern of discrepancy between the observed and theoretical frequencies was similar to that identified by Brass for populations with high fertility. While searching for a more suitable model to describe the completed parity for women in populations characterized by high fertility coupled with high zero parity, Golbeck (1981) suggested a negative binomial mixture distribution. The model is based on the hypothesis that human populations consist of two types of women classified according to whether they do or do not produce viable offspring. It is important to note that the model involves five parameters, whereas the NBD has only two. This, in turn, makes the fitting of the model more difficult.

Keeping in view the observations made by Onyemaekeogum (1978)—the authors of the present paper attempt to examine the adequacy of the NBD to describe Indian data on the distributions of births. First, the distribution prepared on the basis of the schedules of the demographic survey referred to by Lal (1973) is considered. The other distributions dealt with in this paper include the distributions of: male and female births; births which took place almost 20 years after the distribution was first considered; births to mothers classified as married before and after age 15; births to working and nonworking mothers and births to mothers belonging to nuclear and joint families.

It has been observed that the model provides a better fit to the distribution of male births, births to mothers who were married before 15 years of their age, to mothers belonging to nuclear families, and to working mothers in comparison to their counterpart distributions. The former distributions have higher average births than the latter. The other interesting feature of the model is that it gives a more appropriate fit to the distributions having mothers rather than women as frequencies. The parameters of the model have been estimated by the method of maximum likelihood (ML). In the following sections, the descriptions of the negative binomial distribution and the applications to Indian data on birth distributions are presented.

The Model

Following Brass (1958) we have assumed that: (i) if all women possess the constant expectation of bearing children per unit of time (r) for the same fixed period (L) exposed to the risk, the births would be distributed randomly over the period, and the distribution of women of the completed fertility by the number of births would reflect the following distribution:

$$P(X = x / r) = e^{-rL}(rL)^{x} / x!, x = 0,1,2,... \quad (1)$$

(ii) The probability differential function of r is given by

$$f(r)dr = a^{k} e^{-ar} r^{k-1} dr / (k-1)! \quad (2)$$

where a is a constant.

Thus, we obtain:

$$P(X = x) = a^{k} \int_{0}^{\infty} e^{-rL}(rL)^{x} e^{-ar} r^{k-1} dr / x!(k-1)! \quad (3)$$

or,

$$P(X = x) = (k + x - 1)(k + x - 2)...kp^{k}q^{x}/x!,$$

where $x = 0, 1, 2, \ldots, p = \frac{a}{a+L}$ and $q = 1 - p$. (4)

This is the probability mass function of the N B D with the parameters k and p. Their moment estimators are given by:

$$\hat{k} = \bar{x}^2/(s^2 - \bar{x}) \text{ and } \hat{p} = (s^2 - \bar{x})/\bar{x}$$

where

$$N = \sum f_x, \bar{x} = \sum_{x=0}^{n} x f_x / N \tag{5}$$

and

$$s^2 = \sum_{x=0}^{n} (x - \bar{x})^2 f_x / (N - 1)$$

Their maximum likelihood estimators are obtained by solving the following equations:

$$(-N\bar{x})/(1-p) - N\bar{x}/p + (kN + N\bar{x})/p = 0 \tag{6}$$

and

$$N\ell np + \sum_{x=1}^{R} f_x \sum_{i=1}^{x} \frac{1}{(k+i-1)} = 0 \tag{7}$$

In order to solve these equations, the moment estimators are used as the initial solutions and the table prepared by Sinha (1984) is helpful in reaching the solution more quickly and conveniently. Further, the probability mass function of the zero-truncated NBD is:

$$P(X = x) = (k + x - 1)(k + x - 2)\ldots kp^k q^x / [x!(1 - p^k)] \tag{8}$$
$$\text{where } x = 1, 2, \ldots$$

The ML estimators are obtained by solving the following equations:

$$(Nk)/p + (Nkp^{k-1})/(1 - p^k) - (N\bar{x})/(1 - p) = 0 \tag{9}$$

and

$$N\ell np + Np^k \ell np/(1 - p^k) + \sum_{x=1}^{R} f_x \sum_{i=1}^{x} \frac{1}{(k+i-1)} = 0 \tag{10}$$

Cohen (1965) suggested the moment estimators which may be used as the first approximation in obtaining the solution of the ML equations are:

$$-\hat{p}\ell n\hat{p}/(1-\hat{p}) = [\bar{x}/(s^2+\bar{x}^2-\bar{x})]\ell n(N\bar{x}/f_1) \tag{11}$$

and

$$\hat{k} = [\hat{p}/(1-\hat{p})][(s^2+\bar{x}^2-\bar{x})/\bar{x}]-1 \tag{12}$$

The Application

Brass (1958) and Onyemaekeogum (1978) examined the adequacy of the NBD to describe the data of the east African countries because of their high fertility and relative lack of birth control. The data used in this paper were taken from the schedule of the Demography Survey of Patna (India) conducted in 1955 and referred to by Lal (1973). The model has been fitted to all observed distributions referred to in this paper using the ML estimators.

Table 1 gives the observed and the expected distribution of births.

TABLE 1. THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS AGED 44+ YEARS BY NUMBER OF BIRTHS

Mothers	*Number of Births*							
	1	2	3	4	5	6	7	8
Observed	11	30	33	40	33	57	47	23
Expected	9.9	22.5	35.9	45.6	48.8	45.9	38.8	30.0
	9	10	11	12	13	14+	Total	
Observed	21	19	9	9	3	1	336	
Expected	21.6	14.6	9.3	5.7	3.3	4.1	336.0	

Mean births = 5.920; $P(X^2) = 0.04$; $\hat{k} = 15.143$; $\hat{p} = 0.720$

Table 1 shows that the model provides a reasonably good fit to the observed distribution, supporting the findings as well as the conjecture of Brass (1958). We have also made an attempt to investigate the adequacy of the model in describing the distributions of male and female births separately. Tables 2(a) and 2(b) present the result.

It is evident that the model gives a better fit to the distributions of male and female births compared to the distribution of births which include combined male and female births. This finding leads to a deeper understanding of the dynamics of population growth because it depends on the sex of the offspring. Moreover, the model describes the distribution of male births more successfully than that of female births. Probably since the average male births (3.122) are higher than female births (2.798). In order to examine the impact of time on the adequacy of the model describing distribution of births, we have fitted the data collected in 1975 (almost twenty years later than the data set originally used. Also, the second set of data was collected from the same geographic area as the first. Prasad and Mishra (1980) referred to this data set in Table 2.10 of their report. The observed and the expected frequencies are shown in Table 3.

TABLE 2(a). THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS AGED 44+ YEARS BY NUMBER OF MALE BIRTHS

Mothers	*Number of Male Births*						
	0	1	2	3	4	5	6
Observed	12	51	86	57	58	39	18
Expected	17.3	48.8	71.2	71.5	55.6	35.7	19.6
	7	8	9	10	11+	Total	
Observed	7	5	1	0	2	336	
Expected	9.5	4.2	1.7	0.6	0.3	336.0	

Mean births = 3.122; $P(X^2) = 0.17$; $\hat{k} = 29.160$; $\hat{p} = 0.903$

TABLE 2(b). THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS AGED 44+ YEARS BY NUMBER OF FEMALE BIRTHS

Mothers	*Number of Female Births*						
	0	1	2	3	4	5	6
Observed	27	71	70	58	49	26	25
Expected	28.8	62.4	74.6	65.1	46.4	28.5	15.7
	7	8	9+	Total			
Observed	4	3	3	336			
Expected	7.9	3.7	2.9	336.0			

Mean births = 2.798; $P(X^2) = 0.12$; $\hat{k} = 9.508$; $\hat{p} = 0.772$

After omitting the women who were not responsible for births, the zero-truncated NBD was fitted to the truncated distribution. Table 4 shows the result.

TABLE 3. THE OBSERVED AND THE EXPECTED NUMBER OF WOMEN AGED 49+ YEARS BY NUMBER OF BIRTHS

Mothers	*Number of Births*							
	0	1	2	3	4	5	6	7
Observed	1	38	58	54	57	37	39	23
Expected	14.3	36.2	53.2	59.5	56.0	46.8	35.8	25.6
	8	9	10	11	12	13+	Total	
Observed	25	4	6	3	4	1	373	
Expected	17.3	11.2	7.0	4.2	2.5	3.4	373.0	

Mean births = 4.300; $P(X^2) = 0.498$; $\hat{k} = 6.118$; $\hat{p} = 0.587$

TABLE 4. THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS AGED 49+ YEARS BY NUMBER OF BIRTHS

Mothers	*Number of Births*						
	1	2	3	4	5	6	7
Observed	38	58	54	57	37	39	23
Expected	36.7	53.6	59.5	55.9	46.6	35.6	25.5
	8	9	10	11	12	13+	Total
Observed	25	14	6	3	4	1	359
Expected	17.3	11.2	7.0	4.3	2.5	3.3	359.0

Mean births = 4.468; $P(X^2) = 0.422$; $\hat{k} = 5.976$; $\hat{p} = 0.582$

Comparing the results of Tables 1 and 4, it is seen that although the mean birth decreased from 5.920 to 4.468 in the last 220 years, the value of $P(X^2)$ increased from 0.04 to 0.422. It seems reasonable to conclude that time (duration of 20 years) has not disturbed the adequacy of the model in describing the distribution of births, at least in the prevalent conditions. The model has also been fitted to the distribution of mothers, irrespective of age, classified by number of births. Table 2.7 of Prasad and Mishra (1980) has been used for this purpose. The expected and observed frequencies and the observed ones are shown in Table 5.

TABLE 5. THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS BY NUMBER OF BIRTHS

Mothers	*Number of Births*						
	1	2	3	4	5	6	7
Observed	224	314	306	274	175	128	57
Expected	223.9	311.3	315.5	260.3	185.4	118.1	68.9
	8	9	10	11	12	13+	Total
Observed	41	19	13	3	2	1	1,557
Expected	37.4	19.1	9.3	4.3	1.9	1.6	1,557.0

Mean births = 3.601; $P(X^2) = 0.57$; $\hat{k} = 9.681$; $\hat{p} = 0.740$

TABLE 6. THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS BY NUMBER OF BIRTHS

Number of Births	*Number of Mothers Married at Age*			
	Less than 15 years		*15 or 15+ years*	
i	*Observed*	*Expected*	*Observed*	*Expected*
1	35	35.8	58	54.4
2	70	63.6	48	52.8
3	75	77.5	35	40.0
4	68	72.5	30	26.0
5	52	56.0	18	15.3
6	38	36.9	10	8.3
7	29	21.4	4	4.3
8	8	11.1	0	2.1
9	7	5.2	1	1.0
10+	2	4.0	1	0.8
Total	384	384.0	205	205.0

Mean births = 3.935; $P(X^2) = 0.554$; $\hat{k} = 34.662$; $\hat{p} = 0.900$

Mean births = 2.809; $P(X^2) = 0.505$; $\hat{k} = 4.844$; $\hat{p} = 0.670$

The model does not provide a good fit to the original distribution, that is without truncation at zero. In fact, there were 272 married women (17.47%) who were not mothers, and the inclusion of these women in the distribution makes the model inadequate. It appears that the model makes a distinction between married women and mothers, and, as such, it adequately describes the distribution of mothers.

The number of children born to a woman depends to a considerable extent on her age at the time of her marriage. Keeping this fact in mind, an attempt was made to investigate the impact of age at marriage on the distribution of mothers by the number of births so far as the description of the distribution by the model is concerned. For this purpose we have used Table 3.6 of Population Research Centre (PRC), Patna University (1977). Table 6 shows the result.

It is interesting to find that the model is more appropriate for dealing with the distribution of births to mothers of the first category, due to the mean birth of mothers married before 15 years of age being higher than that of those married at 15 years or older. Also, the probability of having three births or more is greater for the mothers of the first category than that of the second category.

Usually, two types of families, nuclear and joint, are found in India. The former includes husband, wife and their children, whereas the latter type consists of these three in addition to other members. Table 7 presents the expected and the observed number of mothers. The data mentioned in PRC. (1977) dealing with the hutment dwellers (who are mostly scheduled castes) of Patna (India) have been used. A hutment dwelling is small, temporary and made of materials other than mud and bricks.

TABLE 7. THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS OF NUCLEAR AND JOINT FAMILIES BY NUMBER OF BIRTHS

Number of Births	*Number of Mothers Married at Age*			
	Less than 15 years		*15 or 15+ years*	
	Nuclear Family		*Joint Family*	
	Observed	*Expected*	*Observed*	*Expected*
1	61	60.4	32	26.4
2	91	89.4	27	30.4
3	90	93.6	20	27.5
4	80	77.7	18	21.4
5	49	54.3	21	15.1
6	36	33.3	12	9.9
7	23	18.4	10	6.1
8	6	9.3	2	3.6
9+	8	7.6	3	4.6
Total	444	444.0	145	145.0

Mean births = 3.556

$P(X^2) = 0.774$

$\hat{k} = 15.423$; $\hat{p} = 0.820$

Mean births = 3.503

$P(X^2) = 0.06$

$\hat{k} = 4.700$; $\hat{p} = 0.595$

On comparing the values of $P(X^2)$ for nuclear and joint families, one can safely infer that the model provides a better fit to the distribution of mothers of nuclear families than that of joint families. This supports the observation of Brass (1958) because the average number of children born to a mother of a nuclear family is higher than that of a joint family. The probability of having three or more children is 0.663 for the mothers of nuclear families, whereas it is 0.608 for the mothers of joint families. The difference between the two mean births is probably due to the fact that a nuclear family amongst hutmen dwellers offers more opportunities for privacy which, in turn, results in higher fertility in the absence of contraception.

As the economic condition of a couple has a direct bearing on its fertility, the employment status of mothers was considered in order to test the adequacy of the model. On fitting the zero-truncated NBD to the distribution of births of both kinds of mothers, the results are shown in Table 8.

TABLE 8. THE OBSERVED AND THE EXPECTED NUMBER OF WORKING AND NONWORKING MOTHERS BY NUMBER OF BIRTHS

Number of Births	*Working mothers*		*Nonworking mothers*	
	Observed	*Expected*	*Observed*	*Expected*
1	21	18.9	72	69.9
2	37	36.8	80	79.2
3	46	48.3	59	67.4
4	45	48.5	52	47.7
5	39	39.6	30	29.8
6	31	27.4	17	16.9
7	21	16.5	12	8.9
8	5	8.8	3	4.4
9+	8	8.2	3	3.8
Total	253	253.0	328	328.0

Mean births = 4.170

$P(X^2) = 0.68$

$\hat{k} = 57.963$; $\hat{p} = 0.934$

Mean births = 3.067

$P(X^2) = 0.67$

$\hat{k} = 6.976$; $\hat{p} = 0.716$

The model describes the two kinds of distributions very well, but it provides a slightly better description of the distribution of births having working mothers as frequencies. This is perhaps due to the higher fertility among working mothers than among nonworking mothers. Also, the probability that a working mother has three or more births is 0.780, the value for a nonworking mother is 0.545. Like most other distributions considered earlier, the distribution of working and nonworking women that is without truncation at zero birth, fails to be described by the model. The interesting feature of the model is that it fits the distributions having mothers as frequencies fairly well. This probably happens because various kinds of women are included in the frequencies of the zero class of a distribution. Further, we have observed that the NBD is more appropriate to the distributions having higher mean births than the lower ones as envisaged by Brass (1958).

An attempt was made to compare the discrepancy between the observed and the theoretical distributions where the model does not fit well. For this, Table 1 without truncation at zero birth was considered. The result of the fitting with the observed frequencies is shown in Table 9.

TABLE 9. THE OBSERVED AND THE EXPECTED NUMBER OF MOTHERS AGED 44+ YEARS BY NUMBER OF BIRTHS

Mothers	*Number of Births*							
	0	1	2	3	4	5	6	7
Observed	16	11	30	33	40	33	57	47
Expected	5.1	16.5	30.5	41.7	47.3	46.8	42.0	34.8
	8	9	10	11	12	13	14+	Total
Observed	23	21	19	9	9	3	1	352
Expected	27.0	19.9	14.1	9.6	6.3	4.0	6.4	352.0

Mean births = 5.651; $X^2 = 49.272$; $\hat{k} = 7.734$; $\hat{p} = 0.578$

It is now evident that the observed number of mothers is higher than the expected mothers at zero birth; the expected number is greater than the observed number from the first to five births; the observed number is larger than the expected for each of the six to 10 births except at eighth birth; and for the remaining births, the expected number of mothers is greater than the observed number barring the observation at the twelfth birth. This pattern resembles that detected by Brass (1958) and supported by Onyemaekeogum (1978).

Conclusion

Using Indian data, it has been shown that the negative binomial distribution with the truncation at zero provides a good fit to the distributions of births in human populations. However, the model in its complete form sometimes fails to describe the observed distributions. Further, the pattern of discrepancy between the observed and the theoretical distributions where the model fails, tends to resemble the pattern observed by the earlier authors.

Acknowledgements

We are very grateful to the editor and the two reviewers for their suggestions which helped to improve the original version of the paper. We also express our gratitude to Dr. Ravi B. P. Verma for his constant encouragement during its preparation.

References

Brass, W. 1958. The distribution of births in human population. Population Studies 12:51-72.

Cohen, A.C., Jr.1965. Estimation in the negative binomial distribution. NASA Tech. Memo No. X-533372, Huntsville, AL: George Marshall Flight Centre.

Dandekar, V. M. 1955. Certain modified forms of binomial and Poisson distributions. Sankhya 15:237-251.

Golbeck, A.L. 1981. A probability mixture model of completed parity. Demography 18:645-658.

Lal, D.N. 1973. A demography sample survey in Patna 1955. Population Research Centre, Department of Statistics, Patna University, Patna, India.

Onyemaekeogum, G.E. 1978. The distribution of births in some African Populations: An empirical test of the adequacy of the negative binomial model. Genus 34:153-163.

P. R. C. 1977. Fertility and family planning amongst dwellers of Patna. Mimeograph No. 55, Population Research Centre, Dept. of Statistics, Patna University, Patna, India.

Prasad, R. N. and A. Mishra. 1980. Fertility and mortality in Patna. Mimeograph No. 68, Population Research Centre, Dept. of Statistics, Patna University, Patna, India.

Singh, S.N. 1963. Probability models for the variation in the number of births per couple. Journal of the American Statistical Association 58:721-727.

Sinha, A.K. 1984. The negative binomial distribution and its role in operation research and population studies. Ph. D. thesis, Patna University, Patna, India.

Received, February, 1988; revised, November, 1992.