Aligarh Journal of Statistics

Vol. 33 (2013), 1-10

A VERSION OF QUASI GEOMETRIC SERIES DISTRIBUTION HAVING FIRST TWO MOMENTS IN SIMPLE ALGEBRAIC FORMS

A. MISHRA

ABSTRACT

The estimation of the two parameters of the quasi geometric series distribution $(QGSD)$ of Mishra and Singh (2000) by the method of moments is not possible as its moments appear in terms of various series. The likelihood equations are also not directly solvable. Another version of the $(QGSD)$, the first two moments of which appear in simple algebraic forms and which has the same nature and behaviour as that of Mishra and Singh's version, has been obtained. The estimation of parameters thus becomes simple by the method of moments. This version has been fitted to some data-sets to test its goodness of fit.

1. INTRODUCTION

Mishra and Singh (2000) studied a two-parameter quasi geometric series distribution $(QGSD)$ given by its probability function

$$P_1(x;\alpha,\theta)=\frac{\alpha(1-\alpha)(\alpha+x\theta)^{x-1}}{(1+x\theta)^{x+1}}\quad;\ 0<\alpha<1;\ x=0,1,2,...... \qquad (1.1)$$

in which the probability of an event is not constant and is linearly dependent on the number of failures. It reduces to the geometric distribution with parameter α at $\theta=0$. It is a particular case of the quasi negative binomial distribution $(QNBD)$ obtained in different forms by Janardan (1975), Nandi and Das (1994), Sen and Jain (1996) and Mishra and Hassan (2006).

The mean of the $QGSD$ (1.1) appears in form of a series which does not seem possible to be summed. However, Mishra and Singh (2000) expressed it as

$$M_1^{/} = \frac{-}{\left(1-^{-}\right)} J\left(1, -2, \frac{}{\left(1- \quad\right)}\right) \tag{1.2}$$

where $J\left(s, a, b\right) = \sum_{j=0} C_j\left(s\right) a^{(j)} b^j$ is a 'factorial power series' of order s in a and b, s, being a positive integer, C_j, the coefficient of y^j in the expansion of $\left(1-y\right)^{-s}$ and $a^{(j)} = a\left(a-1\right)\left(a-2\right).....\left(a-j+1\right)$.

The expressions for the second and higher moments also come in terms of such types of series and which are in very messy forms.

As the probability of failure is not constant in the $QGSD$ and depends upon the number of previous failures, the $QGSD$ is supposed to be more realistic than the classical geometric distribution for many practical purposes. The $QGSD$ (1.1) has been found to give very close fits to the observed data-sets to which previously the geometric distribution and the beta compound of geometric distribution of Pielou (1962) have been fitted, (Mishra and Singh (2000)). But what has been most unfortunate of this distribution is that its moments do not appear in closed algebraic forms and so the method of moments fails to provide estimates of the parameters $^{-}$ and .

Moreover, the two likelihood equations are also not directly solvable and need some iteration procedure for the purpose which requires a lot of calculation and so is tedious and time taking. The difficulty in the estimation of the parameters seems to be the main reason that such a useful distribution is not being used much in practice.

In this paper, a version of the $QGSD$ has been obtained, the first two moments of which appear in simple algebraic forms which make the estimation of parameters simple by the method of moments. This version has, more or less, the same pattern of behaviour as that of $QGSD$ (1.1) and so it may be used as a better alternative to the version (1.1).The new version of the $QGSD$ has been fitted to some data-sets to which Mishra and Singh (2000) fitted the $QGSD$ (1.1) to test its goodness of fit and it has been found that the obtained version gives almost identical sets of expected frequencies as given by (1.1).

2. A NEW VERSION OF THE *QGSD*

From (1.1) we have

$$\sum_{x=0} \frac{\left(^{-} + x \quad\right)^{x-1}}{\left(1+x \quad\right)^{x+1}} = \frac{1}{^{-}\left(1-^{-}\right)} \tag{2.1}$$

$$\text{or,}\quad -\sum_{x=0} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+1}} + \sum_{x=0} x\,\frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+1}} = \frac{1}{\bar{}\left(1-\bar{}\right)} \qquad (2.2)$$

Differentiating both sides of (2.1) with respect to $\bar{}$, we get after some arrangements of terms,

$$\sum_{x=0} x\,\frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+1}} = \sum_{x=0} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+1}} - \frac{1-2\bar{}}{\bar{}^{\,2}\left(1-\bar{}\right)^2} \qquad (2.3)$$

Substituting this for the second term on the left hand side of (2.2) we get after a little simplification,

$$\sum_{x=0} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+1}} = \frac{(1-\bar{})\left(\bar{} +\ \ \right)-\bar{}}{\bar{}^{\,2}\left(1-\bar{}\right)^2\left(\bar{} +\ \ \right)} \qquad (2.4)$$

which when substituted in (2.3) gives

$$\sum_{x=0} x\,\frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+1}} = \frac{1}{\left(1-\bar{}\right)^2\left(\bar{} +\ \ \right)} \qquad (2.5)$$

For $k > 0$, let us define

$$S\left(k,\bar{},\ \ \right) = \sum_{x=0} \binom{k+x-1}{x} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+k}} \qquad (2.6)$$

$$= \sum_{x=0} \binom{k+x-1}{x} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+k+1}} + \sum_{x=0} x \binom{k+x-1}{x} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+k+1}}$$

$$= \sum_{x=0} \binom{k+x-1}{x} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+k+1}} + \sum_{x=1} \binom{k+x-1}{x-1} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+k+1}}$$

(2.7)

We have

$$\sum_{x=1} \binom{k+x-1}{x-1} \frac{\left(\bar{} + x\ \ \right)^{x-2}}{\left(1+x\ \ \right)^{x+k+1}}$$

$$= \sum_{x=0} \binom{(k+1)+x-1}{x} \frac{\left(\bar{} +\ \ + x\ \ \right)^{x-1}}{\left(1+\ \ + x\ \ \right)^{x+(k+1)+1}}$$

$$= \sum_{x=0} \binom{(k+1)+x-1}{x} (1+\sim_1)^{k+3} \frac{(\sim_1+\sim_2+x\sim_2)^{x-1}}{(1+\sim_1+\sim_2+x\sim_2)^{x+(k+1)+1}}$$

$$(1+\sim_1)^{k+3} \sum_{x=0} \binom{-(k+1)}{x} (-1)\,(-\sim_1-\sim_2-x\sim_2)^{x-1}(1+\sim_1+\sim_2+x\sim_2)^{-(k+1)-x-1}$$

(2.8)

where for convenience, we have taken

$$^{-} = \sim_1/(1+\sim_1) \text{ and } \quad = \sim_2/(1+\sim_1). \qquad (2.9)$$

Consul and Mittal (1975) obtained a three-parameter quasi binomial distribution II $(QBD\ II)$ having its probability function as

$$P_2(x;n,p_1,p_2) = \binom{n}{x} \frac{p_1(1-p_1-np_2)}{1-np_2} (p_1+xp_2)^{x-1}(1-p_1-x\,p_2)^{n-x-1} \qquad (2.10)$$

$x = 0,1,2,\ldots$

which gives

$$\sum_{x=0}^{n} \binom{n}{x} (p_1+xp_2)^{x-1}(1-p_1-x\,p_2)^{n-x-1} = \frac{1-np_2}{p_1(1-p_1-np_2)} \qquad (2.11)$$

If n is replaced by $-(k+1), p_1$ by $-(\sim_1+\sim_2)$ and p_2 by $-\sim_2$ we get the summation term in (2.8). In fact, (2.8) is associated with a $QNBD$ which is the negative analogue of QBD II (2.10).Using (2.11) under the given substitutions, we get

$$\sum_{x=1} \binom{k+x-1}{x-1} \frac{(^{-}+x\quad)^{x-2}}{(1+x\quad)^{x+k+1}} = \frac{(1-^{-})(1-^{-}-(k+1)\quad)}{(^{-}+\quad)(1-k\quad)} \cdot \left(\frac{1}{1-^{-}}\right)^{k+3} \qquad (2.12)$$

Substituting it in (2.6) and taking $k=1$, we get

$$S(1,^{-},\quad) = \sum_{x=0} \frac{(^{-}+x\quad)^{x-2}}{(1+x\quad)^{x+1}}$$

$$= \sum_{x=0} \frac{(^{-}+x\quad)^{x-2}}{(1+x\quad)^{x+2}} + \frac{(1-^{-})(1-^{-}-2\quad)}{(^{-}+\quad)(1-\quad)} \left(\frac{1}{1-^{-}}\right)^{4} \qquad (2.13)$$

Using (2.4) we get finally after some simplification

$$\sum_{x=0} \frac{(\bar{} + x\ )^{x-2}}{(1+x\ )^{x+2}} = \frac{(1-\bar{})(\bar{}+\ )(1-\bar{}-\ ) - \bar{}\ (1-\bar{}-2\ )}{\bar{}^{\,2}(\bar{}+\ )(1-\bar{})^3(1-\ )} = A(\bar{},\ )\ \text{(say)} \qquad (2.14)$$

which gives another version of $QGSD$ as

$$P_3(x;\bar{},\ ) = \frac{1}{A(\bar{},\ )} \frac{(\bar{}+x\ )^{x-2}}{(1+x\ )^{x+2}};\quad 0<\bar{}<1;\ x=0,1,2,3,...... \qquad (2.15)$$

We may call it $QGSD$ II , (1,1) being called as $QGSD$ I.

We have

$$1-P_3(0;\bar{},\ ) = \frac{\bar{}(1-\bar{})^2(\bar{}+\ )(1-\ )-\bar{}\ (1+\bar{})}{(1-\bar{})(\bar{}+\ )(1-\bar{}-\ )-\bar{}\ (1-\bar{}-2\ )} \qquad (2.16)$$

from which the zero-truncated $QGSD$ II can be obtained as

$$P_4(x;\bar{},\ ) = \frac{\bar{}(\bar{}+\ )(1-\bar{})^3(1-\ )}{(1-\bar{})^2(\bar{}+\ )(1-\ )-\ (1+\bar{})(1-\bar{}-\ )} \frac{(\bar{}+x\ )^{x-2}}{(1+x\ )^{x+2}} \qquad (2.17)$$

x = 1, 2, 3, ….

3. MOMENTS

We have the mean of the $QGSD$ II given by

$$«_1^{/} = \sum_{x=0} \frac{x}{A(\bar{},\ )} \frac{(\bar{}+x\ )^{x-2}}{(1+x\ )^{x+2}}$$

$$= \frac{1}{\ }\left[\sum_{x=0} \frac{1}{A(\bar{},\ )} \frac{(\bar{}+x\ )^{x-2}}{(1+x\ )^{x+1}} - \sum_{x=0} \frac{1}{A(\bar{},\ )} \frac{(\bar{}+x\ )^{x-2}}{(1+x\ )^{x+2}}\right] \qquad (3.1)$$

Noting from (2.15), the second term in the bracket is obviously unity. Substituting the value of the first summation in the bracket from (2.4) and that of $A(\bar{},\ )$ from (2.14), we get after a little simplification

$$«_1^{/} = \frac{\bar{}^{\,2}(1-\bar{}-2\ )}{(1-\bar{})(\bar{}+\ )(1-\bar{}-\ )-\bar{}\ (1-\bar{}-2\ )} \qquad (3.2)$$

The second moment about origin of the QGSD II is given by

$$«_2^{/} = \sum_{x=0} \frac{x^2}{A(¯ ,\ )} \frac{(¯ + x\ )^{x-2}}{(1+x\ )^{x+2}}$$

$$= \frac{1}{\ } \sum_{x=0} \frac{x((1+x\ )-1)}{A(¯ ,\ )} \frac{(¯ + x\ )^{x-2}}{(1+x\ )^{x+2}}$$

$$= = \frac{1}{\ } \left[\sum_{x=0} \frac{x}{A(¯ ,\ )} \frac{(¯ + x\ )^{x-2}}{(1+x\ )^{x+1}} - «_1^{/} \right] \tag{3.3}$$

Taking the respective values from (2.4), (2.5) and (2.14), we get

$$«_2^{/} = \frac{¯^{2}(1+¯\ )}{(1-¯\ )(¯ +\ )(1-¯ -\ )-¯\ (1-¯ - 2\ )} \tag{3.4}$$

Dividing the first two moments of the $QGSD$ II, (3.2) and (3.4) by (2.16), the first two moments about origin of the zero-truncated $QGSD$ II are obtained as

$$«_{1,T}^{/} = \frac{¯(1-¯ - 2\ )}{(1-¯\ )^2(¯ +\ )(1-\ )-\ (1+¯\ )(1-¯ -\ )} \quad (3.5$$

$$«_{2,T}^{/} = \frac{¯(1+¯\ )}{(1-¯\ )^2(¯ +\ )(1-\ )-\ (1+¯\ )(1-¯ -\ )} \tag{3.6}$$

4. GOODNESS OF FIT

The $QGSD$ I has not been used much for explaining observed data, the reason for which seems to be the difficulty in estimation of the parameters ¯ and . As even the mean of the $QGSD$ I appears in form of a series which does not seem possible to be summed, the method of moments for estimating the parameters cannot be used. The method of maximum likelihood requires much time for computation as the two likelihood equations are not easily solvable. Some iteration procedure to be adopted for solving these equations is not simple as the equation consists of terms $\sum(¯ + x\ )^{-1}$.

The basic nature of the two $QGSD$ s I and II is the same as in both the cases the probability of failure at a trial is linearly dependent upon the number of previous failures. The difference between the two is in the constant normalizing factors. On the basis of the ordinates of both the $QGSD$ s for various combinations of values of the parameters, it has been observed that both the QGSDs have more or less the same pattern of behaviour.

This suggests that any of the two *QGSD* s whose fitting is found simpler may be fitted to an observed set of data. As the first two moments of *QGSD* II are in compact forms, the method of moments for estimating the parameters ¯ and may be used.

From (3.2) and (3.4) we have

$$\frac{\text{«}_2^{/}}{\text{«}_1^{/}} = \frac{(1+\text{¯})}{(1-\text{¯}-2\ \ )} = K \text{ (say)} \qquad (4.1)$$

which gives

$$= \frac{\left[K(1-\text{¯})-(1+\text{¯})\right]}{2K} \qquad (4.2)$$

Substituting this in (3.4) we get after some simplification,

$$\text{«}_2^{/} = \frac{4K^{2}\text{¯}^{2}}{(K-1)(1-\text{¯})\left[K(1-\text{¯})+(1+\text{¯})\right]-2\text{¯}\left[K(1-\text{¯})-(1+\text{¯})\right]} \qquad (4.3)$$

Taking for convenience $\frac{\text{¯}}{(1-\text{¯})} = \sim$, we get

$$\text{«}_2^{/} = \frac{4K^2 \sim^2}{(K^2-1)+4\sim^2}$$

which gives

$$\sim\ = \frac{1}{2}\left[\frac{(K^2-1)\text{«}_2^{/}}{(K^2-\text{«}_2^{/})}\right]^{1/2} \qquad (4.4)$$

Replacing the first two population moments by their respective sample estimates we get

$$\hat{\sim} = \frac{1}{2S}\left(m_2^{/}-\overline{X}^2\right)^{\frac{1}{2}}$$

where

$S^2 = m_2^{/} - \overline{X}^2$ is the sample variance.

This gives an estimate of ¯ as

$$\overset{\Lambda}{\_} = \frac{\left(m_2'^2 - \overline{X}^2\right)^{1/2}}{2S + \left(m_2'^2 - \overline{X}^2\right)^{1/2}} \tag{4.5}$$

Substituting this in (4.2) an estimate of   can be obtained.

For the zero-truncated $QGSD$ II (2.17),   is given by the same expression as in (4.2). Substituting it in (4.1) and taking the same transformation, $^{-}/\left(1-^{-}\right) = \sim$, we get

$$\left(1+\sim\right)^2 = \frac{\left(K^2-1\right)«_2'}{4\left(K^2-«_2'\right)},$$

from which we finally get an estimate of $^{-}$ as

$$\overset{\Lambda}{\_} = 1 - \frac{2S}{\left(m_2' - \overline{X}^2\right)^{1/2}} \tag{4.6}$$

Substituting it in (4.2) an estimate of   can be obtained.

The $QGSD$ II has been fitted to all those four data-sets to which Mishra and Singh (2000) fitted the $QGSD$ I. They used the method of maximum likelihood of estimation to estimate the parameters which is much tedious and time taking. It was found that in all the cases, the $QGSD$ II gives more or less the same set of expected frequencies. As the parameters of $QGSD$ II are simple to be estimated, it can be preferred to $QGSD$ I for all practical purposes.

The fittings to two data-sets are given here. The first data-set was originally used by Pielou [962] to fit his beta compound geometric distribution and the second one by William (1944) to which he fitted both, geometric and logarithmic distributions. Pielou [1962] to the first distribution fitted both geometric and compound geometric distributions. The probability of chi-square to be exceeded by the calculated value of chi-square in case of geometric distribution was 0.114 and that in the case of compound geometric distribution it was 0.136. This probability in case of $QGSD$ II is 0.381. William [1944] fitted both geometric and logarithmic distributions to the data given in table-2. The fit by geometric distribution was found very bad as the value of chi-square was 28.21 at 4 df. The logarithmic distribution gave good fit giving the probability of chi-square to be exceeded by its calculated value as 0.190. In case of $QGSD$ II this probability is 0.44.

Table-1

Distribution of the run lengths of Entomology, zygadenus venenosus

Run length	Observed frequency	Expected frequency
1	55	49.2
2	20	25.0
3	11	14.2
4	7	8.2
5	8	
6	4	
7	2	8.3
8	2	
9	0	
10	1	
11	1	
Total	111	111.0

Λ = 0.599479   ¡ ^{2} = 3.52   d.f. = 3

Λ = 0.019017

Table-2

Publication in the review of Applied vol. 1, 1913

Number of Papers PER author	Observed frequency	Expected frequency
1	285	277.6
2	70	79.8
3	32	29.3
4	10	11.1
5	4	5.6
6	3	
7	3	
8	1	7.6
9	2	
10	1	
Total	411	411.0

Λ = 0.339679   ¡ ^{2} = 2.97   d.f.=3

Λ = 0.070766

REFERENCES

Consul, P.C and Mittal, S.P (1975): A new urn model with pre-determined strategy, *Biometrische Zeitschrift*, **17**, 67-75.

Janardan, K.G. (1975): Markov- polya urn-model with pre determined strategies, *Gujarat Statistical Review*, **2**, 17-32.

Mishra, A and Singh, S.K. (2000): On a quasi geometric series distribution, *Aligarh J. Statist.*, **20**, 45-56.

Mishra, A. and Hassan, A. (2006): On a quasi negative binomial distribution, *Aligarh J. Statist.*, **26**, 9-18.

Nandi, S. B. and Das, K. K.(1994): A family of the Abel series distributions, *Sankhya B*, **56**, 147-164.

Pielou, E. C. (1962): Runs of one species with respect to another in transects through plant population, *Biometrics*, **18**, 579-593.

Sen, K. and Jain, R. (1996): Generalised Markov-Polya urn-models with pre determined strategies, *J. Statist. Plann. Inference*, **54**, 119-133.

Williams, C. B. (1944): The number of publications written by biologists, *Annals of Eugenics*, **12**, 143-146.

Received : 01.10.2010

Revised : 10.05.2012

A. Mishra

Department of Statistics
Patna University, Patna
e-mail: mishraamar@rediffmail.com