Aligarh Journal of Statistics

Vol. 34 (2014), 1-11

MODIFIED INTERVENED GEOMETRIC DISTRIBUTION

C. Satheesh Kumar and S. Sreejakumari

ABSTRACT

In this paper, a modified version of intervened geometric distribution is considered and study some of its properties. Certain methods of estimation of the parameters of the distribution are also discussed and illustrated using a real life data set.

1. INTRODUCTION

Geometric distributions are widely used in many areas of scientific research. For details see Johnson et.al (1992). Shanmugam (1985) introduced the intervened Poisson distribution (IPD) as a modification of positive Poisson distribution. The IPD of Shanmugam (1985) has been found applications in several areas of scientific research such as reliability analysis, queuing problem, epidemiological studies etc. Interventions provides informations on effectiveness of various preventive actions taken in the areas of medical science, health etc. The IPD has been studied further by several authors such as Shanmugam (1992), Huang and Fung (1989), Dhanavanthan (1998, 2000), Scollink (2006), Kumar and Shibu (2011, 2012, 2013) etc.

Bartolucci et al. (2001) developed a geometric model suitable for cardiovascular studies through the following probability mass function which we called as the intervened geometric distribution (IGD),

$$h_x = P(X = x) = \frac{\Lambda(\rho,\theta)\,\theta^{x-1}\left(1-\rho^x\right)}{(1-\rho)}, \tag{1.1}$$

in which

$$x = 1,2,3....\quad \theta\in(0,1),\quad \rho\in[0,\infty),\quad \rho\neq 1$$

and

$$\Lambda(\rho,\theta)=(1-\theta)(1-\rho\theta)$$

In this paper we develop the modified intervened geometric distribution $(MIGD)$ in which two intervention parameters are included. If the value of the second intervention is zero then the $MIGD$ becomes the intervened geometric distribution. The paper is organized as follows. In Section 2, we present a model leading to $MIGD$ and obtain expression for its probability mass function, mean, variance and factorial moments. We also obtain a recurrence relation of probabilities of $MIGD$. In Section 3, we consider the estimation of parameters of $MIGD$ by various methods of estimation such as method of factorial moments, method of mixed moments and method of maximum likelihood. In addition the distribution is fitted to a real life data set.

We need the following series representations in the sequel.

$$\sum_{i=0}^{\infty}\sum_{r=0}^{\infty}A(i,r)=\sum_{i=0}^{\infty}\sum_{r=0}^{i}A(i-r,r) \tag{1.2}$$

$$\sum_{i=0}^{\infty}\sum_{r=0}^{\infty}B(i,r)=\sum_{i=0}^{\infty}\sum_{r=0}^{\left[\frac{i}{2}\right]}B(i-2r,r) \tag{1.3}$$

2. THE MODIFIED INTERVENED GEOMETRIC DISTRIBUTION

Let Y be a random variable having intervened geometric distribution with following probability mass function, in which

$y=1,2,3....$ $\theta\in(0,1)$, $\rho_1\in[0,\infty)$, $1\neq\rho_1>0$ $\rho_1\theta\leq 1$.

$$p(y;\theta,\rho_1)=\frac{\Lambda(\rho_1,\theta)\theta^{y-1}\left(1-\rho_1{}^{y}\right)}{(1-\rho_1)} \tag{2.1}$$

The characteristic function $\varphi_Y(t)$ of Y is the following, for any $t\in R=(-\infty,\infty)$ and $i=\sqrt{-1}$.

$$\varphi_Y(t)=\Lambda(\rho_1,\theta)e^{it}\left(1-\theta e^{it}\right)^{-1}\left(1-\rho_1\theta e^{it}\right)^{-1} \tag{2.2}$$

Let Z be a random variable having geometric distribution with following probability mass function in which

$z=0,1,2,....$ $\theta\in(0,1)$, $\rho_2\in[0,\infty)$ and $\rho_2\theta\leq 1$.

$$f(z;\theta,\rho_2)=(1-\rho_2\theta)\,(\rho_2\theta)^{z} \tag{2.3}$$

The characteristic function $\varphi_Z(t)$ of Z is the following, for any $t \in R = (-\infty, \infty)$ and $i = \sqrt{-1}$.

$$\varphi_Z(t) = (1-\rho_2\theta)\left(1-\rho_2\theta e^{it}\right)^{-1} \tag{2.4}$$

Assume that Y and Z are statistically independent. Define $X = Y + 2Z$. Then the characteristic function $\varphi_X(t)$ of X is

$$\varphi_X(t) = \varphi_Y(t)\varphi_Z(t^2)$$

$$= \gamma(\rho_1, \rho_2, \theta) e^{it}\left(1-\theta e^{it}\right)^{-1}\left(1-\rho_1\theta e^{it}\right)^{-1}\left(1-\rho_2\theta e^{2it}\right)^{-1}, \tag{2.5}$$

where

$$\gamma(\rho_1, \rho_2, \theta) = \Lambda(\rho_1, \theta)(1-\rho_2\theta).$$

The distribution of a random variable whose characteristic function is (2.5) is called the modified intervened geometric distribution or in short "the *MIGD*".

On replacing e^{it} by s in (2.5) we get the probability generating function (pgf) of X as

$$P_X(s) = \gamma(\rho_1, \rho_2, \theta)\, s(1-s\theta)^{-1}(1-s\rho_1\theta)^{-1}\left(1-s^2\rho_2\theta\right)^{-1}. \tag{2.6}$$

Clearly

$$P_X(1) = 1.$$

Result 2.1 Let X follows *MIGD* with *pgf* (2.6). Then the probability mass function $g(x; \rho_1, \rho_2, \theta)$ of *MIGD* is

$$g(x; \rho_1, \rho_2, \theta) = \Psi(\rho_1, \rho_2, \theta)\theta^{x-1}\sum_{r=0}^{\left[\frac{x-1}{2}\right]}\left(\frac{\rho_2}{\theta}\right)^r\left(1-\rho_1{}^{x-2r}\right), \tag{2.7}$$

where

$$\Psi(\rho_1, \rho_2, \theta) = \frac{\gamma(\rho_1, \rho_2, \theta)}{(1-\rho_1)}$$

and

$$\gamma(\rho_1, \rho_2, \theta)$$

is defined in (2.5).

Proof follows from expanding (2.6) in terms of s and equating co-efficient of s^x.

Remark 1. When $\rho_2 = 0$ and $\rho_1 = \rho$ (2.7) reduces to the *pmf* of the *IGD* as given in (1.1)

Result 2.2 For any positive integer k, the $k-th$ factorial moment $\mu'_{(k)}$ of *MIGD* is

$$\mu'_{(k)} = \frac{k!}{\theta(1-\rho_1)}\left[(1-\rho_1\theta)\eta(k;\delta_1,\delta_3)-(1-\theta)\eta(k;\delta_2,\delta_3)\right], \tag{2.8}$$

Where

$$\delta_1 = \theta(1-\theta)^{-1},$$

$$\delta_2 = \rho_1\theta(1-\rho_1\theta)^{-1},$$

$$\delta_3 = \rho_2\theta(1-\rho_2\theta)^{-1}$$

and for

$$q = 1,2$$

$$\eta(k;\delta_q,\delta_3) = (\delta_q)^k \sum_{j=0}^{k}\sum_{m=0}^{\left[\frac{j}{2.}\right]}(j-m)c_m(\delta_3)^{j-m}(\delta_q)^{-j}(2)^{j-2m}.$$

Proof. The factorial moment generating function $G_X(t)$ of *MIGD* with *pgf* (2.6) is

$$G_X(t) = \sum_{k=0}^{\infty}\mu'_{(k)}(t)\frac{t^k}{k!} \tag{2.9}$$

$$= P_X(1+t)$$

$$= \frac{1}{\theta(1-\rho_1)}\left[(1-\rho_1\theta)\sum_{k=0}^{\infty}\eta(k;\delta_1,\delta_3)t^k-(1-\theta)\sum_{k=0}^{\infty}\eta(k;\delta_2,\delta_3)t^k\right] \tag{2.10}$$

On equating coefficient of $\frac{t^k}{k!}$ on the right hand side expression of (2.9) and (2.10) we get (2.8)

Remark 2 For particular values of $k = 1,2,3$ in (2.8), we get the first three factorial moments of *MIGD* as

$$\mu'_{(1)} = \delta_4 + 2\delta_3$$

$$\mu'_{(2)} = 2\delta_5 + 4\delta_3\delta_4 + 2\delta_3\left(1+4\delta_3\right)$$

and

$$\mu'_{(3)} = 6\delta_6 + 12\delta_3\delta_5 + \delta_4\left(4\delta_3^{\,2} + \delta_3\right) + 24\delta_3^{\,3}\left(1+2\delta_3\right) \tag{2.11}$$

where

$$\delta_4 = \frac{1-\rho_1\theta^2}{(1-\theta)(1-\rho_1\theta)},$$

$$\delta_5 = \frac{(1+\rho_1)\theta - 3\rho_1\theta^2 + \rho_1^{\,2}\theta^4}{\left[(1-\theta)(1-\rho_1\theta)\right]^2},$$

and

$$\delta_6 = \frac{\left(1+\rho_1+\rho_1^{\,2}\right)\theta^2 - 4\rho_1(1+\rho_1)\theta^3 + 6\rho_1^{\,2}\theta^4 - \rho_1^{\,3}\theta^6}{\left[(1-\theta)(1-\rho_1\theta)\right]^3}.$$

Result 2.3 The mean and variance of *MIGD* is

$$\text{Mean} = \delta_4 + 2\delta_3 \tag{2.12}$$

and

$$\text{Variance} = 2\delta_5 + \delta_4\left(1-\delta_4\right) + 4\delta_3\left(1+\delta_3\right) \tag{2.13}$$

Proof follows from the relations

$$\text{Mean} = \mu'_{(1)}$$

and

$$\text{Variance} = \mu'_{(2)} + \mu'_{(1)} - \mu'^{\,2}_{(1)}$$

Result 2.4 The following is a simple recurrence relation for probabilities $g_x = g\left(x;\rho_1,\rho_2,\theta\right)$ of *MIGD*, for $x \geq 1$

$$x g_{x+1} = \sum_{i=0}^{x-1}\left(1+\rho_1^{\,i+1}\right)\theta^{i+1} g_{x-i} + 2\sum_{i=0}^{\left[\frac{x-2}{2}\right]}\left(\rho_2\theta\right)^{i+1} g_{x-2i-1} \tag{2.14}$$

in which

$$g_0 = 0,$$

$$g_1 = \gamma(\rho_1, \rho_2, \theta).$$

Proof. From (2.6), we have

$$P_X(s) = \sum_{x=0}^{\infty} s^x g_x \tag{2.15}$$

$$= \gamma(\rho_{1,} \rho_2, \theta) s (1 - s\theta)^{-1} (1 - s\rho_1\theta)^{-1} (1 - s^2\rho_2\theta)^{-1}. \tag{2.16}$$

On differentiating (2.15) and (2.16) with respect to s we get the following

$$\sum_{x=0}^{\infty} (x+1) s^x g_{x+1} = \frac{P_X(s)}{s} + \theta \frac{P_X(s)}{1-s\theta} + \rho_1\theta \frac{P_X(s)}{1-s\rho_1\theta} + 2\rho_2\theta \frac{sP_X(s)}{1-s^2\rho_2\theta}$$

$$= \sum_{x=1}^{\infty} s^{x-1} g_x + \theta \sum_{x=1}^{\infty}\sum_{i=0}^{\infty} \theta^i s^{x+i} g_x + \rho_1\theta \sum_{x=1}^{\infty}\sum_{i=0}^{\infty} (\rho_1\theta)^i s^{x+i} g_x$$

$$+ 2\rho_2\theta \sum_{x=1}^{\infty}\sum_{i=0}^{\infty} (\rho_2\theta)^i s^{x+2i+1} g_x$$

$$= \sum_{x=0}^{\infty} s^x g_{x+1} + \theta \sum_{x=0}^{\infty}\sum_{i=0}^{\infty} \theta^i s^{x+i+1} g_{x+1} +$$

$$\rho_1\theta \sum_{x=0}^{\infty}\sum_{i=0}^{\infty} (\rho_1\theta)^i s^{x+i+1} g_{x+1} + 2\rho_2\theta \sum_{x=0}^{\infty}\sum_{i=0}^{\infty} (\rho_2\theta)^i s^{x+2i+2} g_{x+1}$$

(2.17)

Now applying (1.2) and (1.3) in (2.17) we obtain

$$\sum_{x=0}^{\infty} (x+1) s^x g_{x+1} = \sum_{x=0}^{\infty} s^x g_{x+1} + \sum_{x=0}^{\infty} \sum_{i=0}^{x} (1+\rho_1^{i+1}) \theta^{i+1} g_{x-i+1} s^{x+1} + 2\sum_{x=0}^{\infty} \sum_{i=0}^{\left[\frac{x}{2}\right]} (\rho_2\theta)^{i+1} g_{x-2i+1} s^{x+2}$$

which implies

$$\sum_{x=0}^{\infty} x s^x g_{x+1} = \sum_{x=1}^{\infty} \sum_{i=0}^{x} (1+\rho_1^{i+1}) \theta^{i+1} g_{x-i} s^x + 2\sum_{x=2}^{\infty} \sum_{i=0}^{\left[\frac{x-2}{2}\right]} (\rho_2\theta)^{i+1} g_{x-2i-1} s^x \tag{2.18}$$

On equating coefficients of s^x on both sides of (2.18) we get (2.14).

3. ESTIMATION

Here we discuss three methods of estimation such as method of factorial moments, method of mixed moments and method of maximum likelihood.

Method of factorial moments

In method of factorial moments, equate the first three factorial moments of *MIGD* to the corresponding sample factorial moments say $m_1^/$, $m_2^/$, and $m_3^/$ and thus we obtain the following system of equations:

$$\delta_4 + 2\delta_3 = m_1^/ \tag{3.1}$$

$$2\delta_5 + 4\delta_3\delta_4 + 2\delta_3(1 + 4\delta_3) = m_2^/ \tag{3.2}$$

$$6\delta_6 + 12\delta_3\delta_5 + \delta_4(4\delta_3^2 + \delta_3) + 24\delta_3^3(1 + 2\delta_3) = m_3^/ \tag{3.3}$$

where δ_3 is given in (2.8) and $\delta_{4,}$ δ_5, δ_6 are given in (2.11). Now the parameters of *MIGD* are estimated by solving the non- linear equations (3.1), (3.2) and (3.3).

Method of mixed moments

In method of mixed moments, the parameters are estimated by using the first two sample factorial moments and the first observed frequency of the distribution. That is, the parameters are estimated by solving the following equation together with (3.1) and (3.2).

$$\gamma(\rho_1, \rho_{2,} \theta) = \frac{p_1}{N} \tag{3.4}$$

where $\gamma(\rho_1, \rho_{2,} \theta)$ is defined in (2.5), p_1 is the observed frequency of the distribution corresponding to the observation $x = 1$ and N is the total frequency.

Method of maximum likelihood

Let $a(x)$ be the observed frequency of x events, y be the highest value of x observed. Then the likelihood function of the sample is

$$L = \prod_{x=0}^{y} [g_x]^{a(x)} \quad \text{where} \quad g_x = g(x; \rho_1, \rho_2, \theta) \tag{3.5}$$

which implies

$$\log L = \sum_{x=1}^{y} a(x) \log(g_x) \tag{3.6}$$

Let $\hat{\rho}_1$, $\hat{\rho}_2$ and $\hat{\theta}$ denotes the maximum likelihood estimates of ρ_1, ρ_2 and θ respectively. Now $\hat{\rho}_1, \hat{\rho}_2$ and $\hat{\theta}$ are obtained by solving the normal equations (3.7), (3.8) and(3.9) given below.

$$\frac{\partial(\log L)}{\partial\theta} = 0$$

implies

$$\sum_{x=1}^{y} a(x)\tau(\rho_1, \rho_2, \theta) + \sum_{x=1}^{y} a(x)\frac{\xi(r, \rho_1, \rho_2, \theta)}{\zeta(r, \rho_1, \rho_2, \theta)} = \sum_{x=1}^{y} xa(x) \tag{3.7}$$

where

$$\tau(\rho_1, \rho_2, \theta) = 1 + \frac{\theta}{1-\theta} + \frac{\rho_1\theta}{1-\rho_1\theta} + \frac{\rho_2\theta}{1-\rho_2\theta},$$

$$\xi(r, \rho_1, \rho_2, \theta) = \sum_{r=0}^{\left[\frac{x-1}{2}\right]} r\left(\frac{\rho_2}{\theta}\right)^r \left(1-\rho_1^{\,x-2r}\right)$$

and

$$\zeta(r, \rho_1, \rho_2, \theta) = \sum_{r=0}^{\left[\frac{x-1}{2}\right]} \left(\frac{\rho_2}{\theta}\right)^r \left(1-\rho_1^{\,x-2r}\right)$$

$$\frac{\partial(\log L)}{\partial\rho_1} = 0$$

implies

$$\sum_{x=1}^{y} a(x)\left[\left(\frac{1}{1-\rho_1}\right) - \left(\frac{\theta}{1-\rho_1\theta}\right) - \frac{\upsilon(r, \rho_1, \rho_2, \theta)}{\zeta(r, \rho_1, \rho_2, \theta)}\right] = 0 \tag{3.8}$$

where

$$\upsilon(r, \rho_1, \rho_2, \theta) = \sum_{r=0}^{\left[\frac{x-1}{2}\right]} (x-2r)\left(\frac{\rho_2}{\theta}\right)^r \left(1-\rho_1^{\,x-2r-1}\right)$$

$$\frac{\partial(\log L)}{\partial\rho_2}=0$$

implies

$$\sum_{x=1}^{y} a(x)\frac{\xi(r,\rho_1,\rho_2,\theta)}{\zeta(r,\rho_1,\rho_2,\theta)} -- \sum_{x=1}^{y} a(x)\frac{\rho_2^{\ 2}}{1-\rho_2\theta} = 0 \tag{3.9}$$

4. CONCLUDING REMARKS

Here we have fitted both IGD and MIGD to a real life data set which is given in Table 1 by method of factorial moments, method of mixed moments and method of maximum likelihood. The data considered is related to the observed frequency of 60 cases of IPF among 208 siblings in 100 families with a COPD patient (cf. Liang et.al, 1992). By using MATHCAD software, we have also computed expected frequencies, Chi-square values and P-values. Based on Chi-square values and P-values we can observed that MIGD gives better fit than IGD.

Table 1: The observed frequencies (O_i) of 60 cases of IPF among 208 siblings in 100 families with a COPD patient (Liang et.al, 1992)

Expected frequencies by the method of							
		Factorial moments		Mixed moments		Maximum likelihood	
No of siblings	O_i	IGD	MIGD	IGD	MIGD	IGD	MIGD
1 2 3 4 5 Total	48 23 17 7 5 100	46.9 26.4 13.4 6.7 6.6 100	45.7 27.5 14.5 6.8 5.5 100	48 26.2 13.1 6.5 6.2 100	48 25.8 14.2 6.7 5.3 100	46.8 27.1 13.5 6.5 6.1 100	44.9 27.6 15.2 7 5.3 100
Estimated value of parameters		$\hat{\rho}$ =7.78 $\hat{\theta}$ =0.064	$\hat{\rho}_1$ =0.38 $\hat{\rho}_2$ =0.065 $\hat{\theta}$ =0.436	$\hat{\rho}$ =0.105 $\hat{\theta}$ =0.494	$\hat{\rho}_1$ =0.041 $\hat{\rho}_2$ =0.236 $\hat{\theta}$ =0.467	$\hat{\rho}$ =4.969 $\hat{\theta}$ =0.097	$\hat{\rho}_1$ =1.864 $\hat{\rho}_2$ =0.214 $\hat{\theta}$ =0.215
hi-square value		1.832	1.334	1.823	0.886	1.795	1.211
p-value		0.767	0.855	0.768	0.927	0.773	0.876

REFERENCES

Bartolucci, A. A., Shanmugam, R. and Singh, K. P. (2001): Development of the generalized geometric model with application to cardiovascular studies. *System Analysis, Modelling Simulation*, ***41***, 339-349.

Dhanvanthan, P. (1998): Compound intervened Poisson distributions. *Biom. J.*, ***40***, 641-646.

Dhanvanthan, P. (2000): Estimation of parameters of compound intervened Poisson distributions. *Biom. J.*, **42**, 315-320.

Haung, M. and Fun, K. Y. (1989): Intervened truncated Poisson distributions. *Sankhya Series*, **51,** 302-310.

Johnson, N. L., Kotz, S. and Kemp, A.W. (1992): *Univariate discrete distributions*. Second Edition, Wiley, New York.

Kumar, C. S. and Shibu, D. S. (2011): Modified intervened Poisson distributions. *Statistica*, **71**, 489-499.

Kumar, C. S. and Shibu, D.S. (2012): An alternative to truncated intervened Poisson distributions. *J. Stat. Appl.*, **5**, 131-141.

Kumar, C. S. and Shibu, D. S. (2013): Some aspects of intervened generalized Hermite distributions, *Metron*, **71**, 9-19.

Liang, K. Y., Zeger, S. L. and Quaqish, B. (1992): Multivariate regression analysis for categorical data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **54,** 3-40.

Scolinik, D. P. M. (2006): On intervened generalized Poisson distributions, *Comm. Statist. Theory Methods*, **35**, 953-963.

Shanmugam, R. (1985): An intervened Poisson distribution and its medical application. *Biometrics*, **41**, 1025-1029.

Shanmugam, R. (1992): An inferential procedure for the Poisson intervention parameter, *Biometrics*, **48**, 559-565.

Received : 19.12.2012

Revised : 12.08.2013

C. Satheesh Kumar

Department of Statistics
University of Kerala
Trivandrum-India
e-mail:drcsatheeshkumar@gmail.com

S.Sreejakumari

Department of Statistics.
University College,
Trivandrum,
India.
e-mail: kumari.sreeja@gmail.com