ON INTERVENED NEGATIVE BINOMIAL DISTRIBUTION AND SOME OF ITS PROPERTIES

C. Satheesh Kumar, S. Sreejakumari

1. INTRODUCTION

Since 1985, intervened type distributions have received much attention in the literature. These types of distributions provide information on the effectiveness of various preventive actions taken in several areas of scientific research. (Shanmugam, 1985) introduced an intervened Poisson distribution (*IPD*) as a replacement for positive Poisson distribution where some intervention process changes the mean of rare events. The *IPD* has been further studied by (Shanmugam, 1992), (Huang and Fung, 1989) and (Dhanavanthan, 1998, 2000). (Scollnik, 2006) developed the intervened generalized Poisson distribution and (Kumar and Shibu, 2011) considered a modified version of IPD. (Bartolucci *et al.*, 2001) studied an intervened version of geometric distribution with following probability mass function (p.m.f.).

$$p_x = P(X = x) = \begin{cases} (1-\theta)(1-\rho\theta)\left|\dfrac{\rho^x - 1}{\rho - 1}\right|\theta^{x-1} & \rho \neq 1 \\ x(1-\theta)^2\theta^{x-1} & \rho = 1\,, \end{cases} \tag{1}$$

in which $x = 1, 2, 3, \ldots,\ 0 < \theta < 1$ and $\rho > 0$.

Through this paper we develop an intervened version of negative binomial distribution, which we called as intervened negative binomial distribution (*INBD*). Negative binomial distribution have found applications in several areas of research such as accidental statistics, birth and death process, medical sciences, psychology etc. For details see (Johnson *et al.*, 2005). In most of these fields there are situations in which observations commence only when at least one event occurs due to some intervention. For this reason a distribution with support $\{1, 2, 3, \ldots\}$ is quite relevant and hence we develop a distribution which is suitable for accommodating such intervention mechanisms.

The paper is organized as follows. In section 2 we present a model leading to *INBD* and obtain expression for its probability mass function, mean, variance and factorial moments. We also obtain a recurrence relation for probabilities of *INBD* in section 2. In section 3 we consider the estimation of parameters of *INBD* by various methods of estimation such as method of factorial moments, method of mixed moments and method of maximum likelihood. A real life data is considered for illustrating these procedures.

2. INTERVENED NEGATIVE BINOMIAL DISTRIBUTION

Let Y be a random variable having zero truncated negative binomial distribution with following p.m.f., in which $0<\theta<1$, r > 0 and y = 1,2,3,... .

$$p(y;\theta,r)=P(Y=y)=\delta\binom{y+r-1}{r-1}(1-\theta)^r\theta^y, \tag{2}$$

where $\delta=[1-(1-\theta)^r]^{-1}$. The characteristic function of Y is

$$\phi_Y(t)=\delta(1-\theta)^r\,[(1-\theta\,e^{it})^{-r}-1] \tag{3}$$

Let Z be a random variable following negative binomial distribution with p.m.f.

$$f(z;\rho,\theta,r)=P(Z=z)=\binom{z+r-1}{r-1}(1-\rho\theta)^r(\rho\theta)^z, \tag{4}$$

in which z = 0,1,2,..., $\rho>0$, r > 0 and $\theta>0$ such that $0<\rho\theta<1$.

The characteristic function of Z is

$$\phi_Z(t)=(1-\rho\theta)^r(1-\rho\theta e^{it})^{-r}. \tag{5}$$

Assume that Y and Z are statistically independent. Define X = Y + Z. Then the characteristic function of X is

$$\begin{aligned}\phi_X(t)&=\phi_Y(t).\phi_Z(t)\\&=\Lambda(\theta,\rho,r)\,((1-\theta\,e^{it})^{-r}-1)\,(1-\rho\,\theta\,e^{it})^{-r}.\end{aligned} \tag{6}$$

where

$$\Lambda(\theta,\rho,r)=\delta[(1-\theta)\,(1-\rho\theta)]^r$$

The distribution of a random variable X whose characteristic function is (6) we call "the intervened negative binomial distribution" or in short "the *INBD*". On replacing e^{it} by 's' in (6) we get the probability generating function (p.g.f.) of X as

$$P_X(s) = \Lambda(\theta,\rho,r)\,[(1-s\theta)^{-r}-1]\,(1-s\rho\theta)^{-r} \tag{7}$$

Clearly $P_X(1)=1$.

Result 2.1 Let X follows *INBD* with p.g.f. (7). Then the probability mass function g_x= P(X=x) of *INBD* is the following, for $x=1,2,3,...,$ in which 0< θ <1 and ρ >0.

$$g_x = \frac{\Lambda(\theta,\rho,r)}{[\Gamma(r)]^2}\,\frac{\theta^x}{x!}\sum_{y=0}^{x-1}\binom{x}{y}\Gamma(x-y+r)\,\Gamma(y+r)\,\rho^y \tag{8}$$

Proof From (7) we have the following:

$$P_X(s) = \sum_{x=0}^{\infty} g_x s^x \tag{9}$$

$$= \frac{\Lambda(\theta,\rho)}{[\Gamma(r)]^2}\sum_{x=0}^{\infty}\frac{(s\theta)^x}{x!}\sum_{y=0}^{x-1}\binom{x}{y}\Gamma(x-y+r)\,\Gamma(y+r)\,\rho^y \tag{10}$$

On equating coefficient of s^x on right hand side expression of (9) and (10), we get (8).

Remark 2.1 When r =1, (8) reduces to the p.m.f. of intervened geometric distribution as given in (1).

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

$$\mu'_{(k)} = \frac{\delta\delta_1^{\,k}\Psi(k)}{[\Gamma(r)]^2} + \frac{(1-\delta)\delta_2^{\,k}\Gamma(r+k)}{\Gamma(r)} \tag{11}$$

in which $\delta_1 = \theta(1-\theta)^{-1}$, $\delta_2 = \rho\theta(1-\rho\theta)^{-1}$ and for $k \geq 1$,

$$\Psi(k) = \sum_{y=0}^{k}\binom{y}{k}[\Gamma(k-y+r)][\Gamma(y+r)]\left(\frac{\delta_2}{\delta_1}\right)^y.$$

Proof The factorial moment generating function $H_X(t)$ of *INBD* with p.g.f. (7) is

$$H_X(t) = \sum_{k=0}^{\infty} \mu'_{(k)}(t) \tag{12}$$

$$= P_X(1+t)$$

$$= \Lambda(\theta, \rho, r)([1-(1+t)\theta]^{-r} - 1)[1-(1+t)\rho\theta]^{-r}$$

$$= \delta(1-\delta_1 t)^{-r}(1-\delta_2 t)^{-r} + (1-\delta)\ (1-\delta_2 t)^{-r}$$

$$= \delta\left[\sum_{k=0}^{\infty}\sum_{y=0}^{k}\binom{k-y+r-1}{k-y}\binom{y+r-1}{y}(\delta_1 t)^{k}\left(\frac{\delta_2}{\delta_1}\right)^{y}\right] + (1-\delta)\sum_{k=0}^{\infty}\binom{k+r-1}{k}(\delta_2 t)^{k}. \tag{13}$$

On equating coefficient of $(k!)^{-1}t^k$ on the right hand side expressions of (12) and (13), we get (11).

Remark 2.2 Putting k=1, 2, 3 in (11) we get the first three factorial moments of *INBD* as given below.

$$\begin{aligned} \mu_{(1)}' &= r\delta_2 + r\delta\,\delta_1 \\ \mu_{(2)}' &= r\,(r+1)\,\delta_2^{\,2} + \delta\,[r(r+1)\,\delta_1^{\,2} + 2r^2\delta_1\delta_2] \\ \mu_{(3)}' &= r\,(r+1)\,(r+2)\delta_2^{\,3} + \delta\,[r\,(r+1)\,(r+2)\,\delta_1^{\,3} + 3r^2\,(r+1)\,\delta_1\delta_2(\delta_1+\delta_2)] \end{aligned} \tag{14}$$

Result 2.3 The mean $E(X)$ and variance $V(X)$ of *INBD* with p.g.f. (7) are the following:

$$E(X) = r\delta_2 + r\delta\delta_1 \tag{15}$$

and

$$V(X) = r\delta_2(1+\delta_2) + r\delta\delta_1(1+\delta_1) + \delta(1-\delta)\,(r\delta_1)^2 \tag{16}$$

Proof follows from (14) in the light of the relation

$$E(X) = \mu'_{(1)}$$

and

$$V(X) = E(X(X-1) + E(X) - [E(X)]^2$$
$$= \mu'_{(2)} + \mu'_{(1)} - [\mu'_{(1)}]^2$$

Remark 2.3 When r =1 in (15) and (16) we get the $E(X)$ and $V(X)$ of intervened geometric distribution as

$$E(X) = \delta_2 + \delta\,\delta_1$$

and

$$V(X) = \delta_2(1+\delta_2) + \delta\,\delta_1(1+\delta_1) + \delta(1-\delta)\,\delta_1^{\,2}.$$

Result 2.4 INBD is over-dispersed or under-dispersed according as

$$\delta_2^{\,2} + (r+1)\,\delta\,\delta_1^{\,2} > r(\delta\,\delta_1)^2 \quad \text{or} \quad \delta_2^{\,2} + (r+1)\,\delta\,\delta_1^{\,2} < r(\delta\,\delta_1)^2$$

Proof follows from *Result 2.3*.

Result 2.5 The following is a simple recurrence relation for probabilities of *INBD*, for x $\geq$1.

$$(x+1)g_{x+1} = r\sum_{k=0}^{x}\theta^{k+1}(1+\rho^{k+1})g_{x-k} + \xi(x;\rho,\theta,r), \tag{17}$$

in which

$$\xi(x;\rho,\theta,r) = \frac{\Lambda(\theta,\rho)}{\Gamma(r-1)}\theta^{x+1}\sum_{k=0}^{x}\frac{\Gamma(x-k+r)}{\Gamma(x-k-1)}\,\rho^{x-k}.$$

Proof From (10) we have

$$P_X(s) = \sum_{x=0}^{\infty} s^x g_x \tag{18}$$

$$= \frac{\Lambda(\theta,\rho)}{[\Gamma(r)]^2}\sum_{x=0}^{\infty}\frac{(s\theta)^x}{x!}\sum_{y=0}^{x-1}\binom{x}{y}\Gamma(x-y+r)\,\Gamma(y+r)\,\rho^y \tag{19}$$

Differentiate (18) and (7) with respect to s to obtain the following.

$$P_X{}'(s) = \sum_{x=1}^{\infty}(x+1)g_{x+1}s^x \tag{20}$$

$$= rP_X(s)\left[\left(\frac{\rho\theta}{1-s\rho\theta}\right)+\left(\frac{\theta}{1-s\theta}\right)\right]+\Lambda(\theta,\rho)\ r\theta\frac{(1-s\rho\theta)^{-r}}{(1-s\theta)}$$

$$= r\rho\theta\sum_{x=0}^{\infty}\sum_{k=0}^{\infty}(\rho\theta)^k s^{x+k} g_x + r\theta\sum_{x=0}^{\infty}\sum_{k=0}^{\infty}(\theta)^k s^{x+k} g_x$$

$$+\Lambda(\rho,\theta)\ r\theta\sum_{x=0}^{\infty}\sum_{k=0}^{\infty}\binom{x+r-1}{x}\rho^x(s\theta)^{x+k} g_x$$

$$= r\sum_{x=0}^{\infty}\sum_{k=0}^{x}\theta^{k+1}(1+\rho^{k+1})\ g_{x-k}s^x$$

$$+\Lambda(\rho,\theta)\ \ r\theta\sum_{x=0}^{\infty}\sum_{k=0}^{x}\binom{x-k+r-1}{x-k}(s\theta)^x\rho^{x-k}$$

$$= \sum_{x=0}^{\infty}\left[r\sum_{k=0}^{x}\theta^{k+1}(1+\rho^{k+1})g_{x-k}+\xi[x;\theta,\rho,r]\right]s^x \tag{21}$$

Now on equating coefficients of s^x on the right hand side expressions of (20) and (21), we get (17).

3. ESTIMATION

Here we discuss the estimation of parameters of the *INBD* by method of factorial moments, method of mixed moments and the method of maximum likelihood.

Method of factorial moment

In method of factorial moments, the first three factorial moments of the *INBD* are equated to the corresponding sample factorial moments m_1', m_2' and m_3' and thus we obtain the following system of equations:

$$r\delta_2 + r\delta\delta_1 = m_1' \tag{22}$$

$$r(r+1)\delta_2{}^2 + r(r+1)\delta\delta_1{}^2 + 2r^2\delta\delta_1\delta_2 = m_2' \tag{23}$$

and

$$r(r+1)(r+2)\delta_2{}^3 + r(r+1)(r+2)\delta\delta_1{}^3 + 3r^2(r+1)\delta\delta_1\delta_2(\delta_1+\delta_2) = m_3' \tag{24}$$

where δ_1 and δ_2 are given in (11). Now, the parameters of the *INBD* are estimated by solving the non-linear equations (22), (23) and (24) using *MATHCAD*.

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 (22) and (23).

$$[(1-\theta)(1-\rho\theta)]^r r\delta\theta = \frac{p_1}{N}, \tag{25}$$

where p_1 is the observed frequency of the distribution corresponding to the observation x=1 and N is the total observed 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 of the sample is

$$L = \prod_{x=1}^{y}[g(x)]^{a(x)},$$

which implies

$$\log \mathrm{L} = \sum_{x=1}^{y} a(x).\log[g(x)]$$

Assume that the parameter r of *INBD* is known. Let $\hat{\rho}$ and $\hat{\theta}$ denotes the maximum likelihood estimates of ρ and θ respectively. Now, $\hat{\rho}$ and $\hat{\theta}$ are obtained by solving the normal equation (26) and (27) given below.

$$\frac{\partial \log L}{\partial\theta} = 0 \quad \text{implies}$$

$$\sum_{x=1}^{y} x\, a(x) - r\sum_{x=1}^{y} a(x)\, [\delta_2 + \delta\delta_1] = 0 \tag{26}$$

$\frac{\partial \log L}{\partial \rho} = 0$ implies

$$\sum_{x=1}^{y} a(x).D(\rho,\theta) = 0 \ , \tag{27}$$

in which

$$D(\rho,\theta) = \left(\frac{\sum_{y=0}^{x-1} \binom{x}{y} .\Gamma(x-y+r).\Gamma(y+r).\, y.\rho^{y-1}}{\sum_{y=0}^{x-1} \binom{x}{y} \Gamma(x-y+r).\Gamma(y+r)\rho^{y}} \right) - \left(\frac{r\rho\theta}{1-\rho\theta} \right)$$

Here estimator of r is used for obtaining the maximum likelihood estimates $\hat{\rho}$ and $\hat{\theta}$ of ρ and θ respectively. Let $\breve{r}$ denotes the factorial moment estimator of r and $\tilde{r}$ denotes the mixed moment estimate of r. Table 1 gives the data related to the distribution of number of articles of theoretical statistics and probability for years 1940-49 and initial letters N-R of the author's name, for details, see (Kendall, 1961). Here we present the fitting of intervened geometric distribution (IGD), intervened Poisson distributions (IPD), intervened generalized Poisson distributions (IGPD) and intervened negative binomial distributions (INBD). From Table 1, it is obvious that INBD is more suitable for the data compared to IGD, IPD and IGPD.

ACKNOWLEDGMENTS

The authors would like to express their sincere thanks to the Editor and the anonymous Referee for carefully reading the paper and for valuable suggestions.

Department of Statistics
University of Kerala, Kariavattom, Trivandrum, India C. SATHEESH KUMAR

Department of Statistics
Government College, Kariavattom, Trivandrum, India S. SREEJAKUMARI

TABLE 1

Distribution of number of articles of theoretical statistics and probability for years from 1940-49 and initial letters N-R *of the author's name*

		Expected frequency by method of												
		Factorial moments				Mixed moments				Maximum likelihood				
x	Observed. freqency	IGD	IPD	IGPD	INBD	IGD	IPD	IGPD	INBD	IGD	IPD	IGPD	INBD using $\breve{r}$	INBD using $\tilde{r}$
1	213	228	200	204	213	213	213	213	213	222	191	206	214	215
2	128	107	134	137	124	108	124	120	124	113	140	134	122	123
3	37	41	50	43	45	46	48	48	45	43	48	43	45	44
4	18	15	12	12	14	19	11	15	13	15	17	13	14	15
5	3	6	3	3	3	8	3	3	3	5	3	3	4	3
6	1	3	1	1	1	6	1	1	2	2	1	1	1	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	400	400	400	400	400	400	400	400	400	400	400	400	400	400
Estimates of parameters		$\hat{\rho}=3.257$ $\hat{\theta}=0.11$	$\hat{\lambda}=0.56$ $\hat{\rho}=0.47$	$\hat{\alpha}=1.16$ $\hat{\gamma}=0.068$ $\hat{\beta}=1.01$	$\breve{r}=4.935$ $\hat{\rho}=0.301$ $\hat{\theta}=0.13$	$\hat{\rho}=0.25$ $\hat{\theta}=0.407$	$\hat{\lambda}=0.68$ $\hat{\rho}=0.39$	$\hat{\alpha}=0.43$ $\hat{\gamma}=0.068$ $\hat{\beta}=1$	$\tilde{r}=2.784$ $\hat{\rho}=2.909$ $\hat{\theta}=0.058$	$\hat{\rho}=0.972$ $\hat{\theta}=0.258$	$\hat{\lambda}=0.69$ $\hat{\rho}=0.47$	$\hat{\alpha}=0.40$ $\hat{\gamma}=0.08$ $\hat{\beta}=1.1$	$\breve{r}=4.935$ $\hat{\rho}=0.254$ $\hat{\theta}=0.135$	$\tilde{r}=2.784$ $\hat{\rho}=0.51$ $\hat{\theta}=0.172$
Chi-square value		5.665	7.49	4.83	2.44	9.131	7.03	3.65	2.44	3.193	6.14	3.27	2.196	2.224
P value based on Chi-square		0.129	0.112	0.306	0.486	0.028	0.129	0.455	0.44	0.363	0.189	0.51	0.533	0.527
Kolmogrov distance value		0.0375	0.0325	0.0225	0.01	0.05	0.0175	0.02	0.01	0.0225	0.055	0.0175	0.0125	0.01
P value based on Kolmogrov distance		0.570	0.655	0.817	0.961	0.368	0.885	0.852	0.961	0.817	0.298	0.885	0.939	0.961

REFERENCES

A. A. BARTOLUCCI, R. SHANMUGAM, K. P. SINGH (2001), *Developments of the generalized geometric model with application to cardiovascular studies*, "System Analysis Modeling Simulation", 41, pp. 339-349.

P. DHANAVANTHAN (1998), *Compound intervened Poisson distribution*, "Biometrical Journal", 40, pp. 641-646.

P. DHANAVANTHAN (2000), *Estimation of the parameters of compound intervened Poisson distribution*, "Biometrical Journal", 42, pp. 315-320.

M. HUANG, K.Y. FUNG (1989), *Intervened truncated Poisson distribution*, "Sankhya", Series B, 51, pp. 302-310.

N. L. JOHNSON, A.W. KEMP, S. KOTZ (2005), *Univariate Discrete Distributions*, John Wiley and Sons, New York.

M.G. KENDAL (1961), *Natural law in Science*, "Journal of Royal Statistical Society", Series. A, 124, pp. 1-18.

C.S. KUMAR, D.S. SHIBU (2011), *Modified intervened Poisson distribution*, "Statistica", 71, pp. 489-499.

D.P.M. SCOLLNIK (2006), *On the intervened generalized Poisson distribution*, "Communication in Statistics-Theory and Methods", 35, pp. 953-963.

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

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

SUMMARY

On intervened negative binomial distribution and some of its properties

Here we develop a new class of discrete distribution namely intervened negative binomial distribution and derive its probability generating function, mean, variance and an expression for its factorial moments. Estimation of the parameters of the distribution is described and the distribution has been fitted to a well known data set.