Extended Intervened Geometric Distribution

C. Satheesh Kumar, S. Sreejakumari

Department of Statistics, University of Kerala, Trivandrum, India

Email address:
drcsatheeshkumar@gmail.com (C. S. Kumar), kumari.sreeja@gmail.com (S. Sreejakumari)

To cite this article:
C. Satheesh Kumar, S. Sreejakumari. Extended Intervened Geometric Distribution. International Journal of Statistical Distributions and Applications. Vol. 2, No. 1, 2016, pp. 8-13.

Received: March 3, 2016; Accepted: March 29, 2016; Published: April 25, 2016

Abstract: Here we develop an extended version of the modified intervened geometric distribution of Kumar and Sreeja (The Aligarh Journal of Statistics, 2014) and investigate some of its important statistical properties. Parameters of the distribution are estimated by various methods of estimation such as the method of factorial moments, the method of mixed moments and the method of maximum likelihood. The distribution has been fitted to a real life data set for illustrating its practical relevance.

Keywords: Factorial Moments, Intervened Geometric Distribution, Method of Factorial Moments, Method of Mixed Moments, Method of Maximum Likelihood, Probability Generating Function, Probability Mass Function

1. Introduction

Intervened type distributions have found many applications in several areas such as epidemiological studies, life testing problems etc. In epidemiological studies health agencies takes various preventive actions. The information concerning the effect of such actions taken by the agencies can statistically analyzed by intervened type distributions. In life testing problem the failed items during the observational period are either replaced or repaired. This kind of actions changes the reliability of the system as only some of its components have longer life. The impact of such actions can be studied by intervened type distributions. The intervened type distributions such as intervened Poisson distribution (IPD), intervened geometric distribution (IGD) and modified intervened geometric distribution (MIGD) has been studied by several authors. For example see Shanmugan [1, 2], Huang and Fung [3], Scollink [4], Dhanavanthan [5, 6], Kumar and Shibu [7-15], Bartolucci et al [16], Kumar and Sreeja [17] etc.

Through this paper we consider a new class of intervened geometric distribution suitable for multiple intervention cases and named it as the extended intervened geometric distribution (EIGD), which contains the MIGD as its special case. The paper is organized as follows. In Section 2, we present a model leading to EIGD and obtain expression for its probability mass function, mean and variance. We also obtain a recurrence relation useful for the computation of probabilities of the EIGD. In Section 3, we consider the estimation of parameters of the EIGD by the method of maximum likelihood and the distribution has been fitted to a real life data set for highlighting the usefulness of the model.

We need the following series representation in the sequel

\[ \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} A(i, r) = \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} A(i-r, r) \]

and

\[ \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} B(i, r) = \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} B(i-\mu r, r), \]

where \([a]\) represents the integer part of “a”, for any \(a > 0\)

2. Extended Intervened Geometric Distribution

Consider a discrete random variable \(X\) having intervened geometric distribution with the following probability mass function (pmf), in which \(x = 1, 2, 3, \ldots, \theta \in (0, 1)\), \(1 \neq \rho_i > 0\) such that \(\rho_i \theta \leq 1\).

\[ h(x; \rho_i, \theta) = \frac{(1-\theta)(1-\rho_i \theta)}{1-\rho_i} \theta^{x-1} \]
The probability generating function (pgf) of $X$ is the following

$$P_X(s) = \frac{(1-\theta)(1-p_\theta)}{(1-s p_\theta)^{-1}}$$

Let $Y$ be a random variable having geometric distribution with the following pmf, in which $y = 1, 2, 3, ..., \theta \in (0, 1)$, $\rho_1 > 0$ and such that $\rho_1 \theta \leq 1$.

$$f(y; \rho_1, \theta) = \frac{1}{1 - \rho_1 \theta}$$

The pgf $Q_Y(s)$ of $Y$ is

$$Q_Y(s) = \frac{(1 - \rho_1 \theta)(1 - s p_\theta)^{-1}}{(1 - s p_\theta)^{-1}(1 - s ^\rho_1 \theta)^{-1}}$$

Define $Z = X + mY$, in which $X$ and $Y$ are assumed to be independent and $m$ is a fixed but arbitrary positive integer. Then the pgf of $Z$ is

$$G_Z(s) = \Lambda(p_1, \rho_2, \theta)$$

where

$$\Lambda(p_1, \rho_2, \theta) = \frac{(1-\theta)(1-p_\theta)}{(1-p_\theta)}$$

The distribution of a random variable whose pgf is (3) is called “the extended intervened geometric distribution” or in short “the EIGD”.

Result 2.1. Let $Z$ follows EIGD with pgf given in equation number (3). Then the pmf $g_z$ of $Z$ is the following, for $z = 1, 2, 3, ...$

$$g_z = \Lambda(p_1, \rho_2, \theta)\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (\rho_2 \theta)^{j+k} (s \rho_2 \theta)^{-k}$$

where $\Lambda(p_1, \rho_2, \theta)$ is as defined in (3), $\theta \in (0, 1)$, $0 < \rho_1 \neq 1$ such that $\rho_1 \theta \leq 1$, for $j=1, 2, 3, ...$

Proof: We have

$$G_Z(s) = \sum_{z=0}^{\infty} g_z (1-s \rho_2 \theta)^{-1}$$

$$= \Lambda(p_1, \rho_2, \theta)\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (\rho_2 \theta)^{j+k} (s \rho_2 \theta)^{-k}$$

Now by applying the results given in equations (1) and (2) in equation (6) we get

$$G_z(s) = \Lambda(p_1, \rho_2, \theta)\sum_{j=0}^{\infty} (\rho_2 \theta)^{j} (s \rho_2 \theta)^{-j}$$

On equating co efficient of $s^z$ in the right hand side expressions of (5) and (7) we get (4) when $\rho_2 \to 0$ and $\rho_1 = \rho \to 1$ reduces to the pmf of IGD defined by Bartalucci et.al [8]

Result 2.2 The first three factorial moments $\mu_1^{(1)}, \mu_1^{(2)}$ and $\mu_1^{(3)}$ of the EIGD are

$$\mu_1^{(1)} = 1 + \delta_1 + \delta_2 + m\delta_3$$

$$\mu_1^{(2)} = 2m^2\delta_2^2 + m(m+1)\delta_3 + 2\delta_1(1+\delta_1)$$

$$+ 2\delta_2(1+\delta_1) + 2m\delta_3(\delta_1 + \delta_2) + 2\delta_3$$

and

$$\mu_1^{(3)} = 6\left[m^3\delta_3^2(1+\delta_1 + \delta_2) + 6\delta_1(1+\delta_1 + \delta_2) + 6\delta_2(1+\delta_1 + \delta_2)ight]$$

$$+ 6m\delta_3(\delta_1 + \delta_2) + 6m\delta_3(\delta_1 + \delta_2) + 6m\delta_3(m+1+2m\delta_3),$$

in which $\delta_1 = \theta(1-\theta)^{-1}$, $\delta_2 = \rho_1 \theta(1-\rho_1 \theta)^{-1}$ and $\delta_3 = \rho_2 \theta(1-\rho_2 \theta)^{-1}$.

Proof follows from the fact that

$$\mu_1^{(1)} = G_z^{(1)}(s)/s = 1,$$

$$\mu_1^{(2)} = G_z^{(2)}(s)/s = 1$$

and

$$\mu_1^{(3)} = G_z^{(3)}(s)/s = 1$$

in which

$$G_z^{(r)}(s) = \frac{d^r}{ds^r}(G_z(s))$$

with $G_z(s)$ as the pgf of $Z$.

Result 2.3 The mean and variance of EIGD is

$$\text{Mean} = 1 + \delta_1 + \delta_2 + m\delta_3$$

and

$$\text{Variance} = \delta_1(1+\delta_1 + \delta_2 + m\delta_3)^2(1+\delta_3)$$

where $\delta_j$ for $j=1, 2, 3$ are as defined in Result 2.2.

Proof: On differentiating the pgf $G_z(s)$ of “$Z$” given in
equation (3) with respect to \( s \) and putting \( s=1 \), we get
\[
G^{(i)}_Z(1) = \frac{d}{ds}(G_Z(s))/s = 1.
\]

Now the mean and variance are obtained by using the following results
\[
\text{Mean} = G_Z^{(i)}(1)
\]
and
\[
\text{Variance (Z)} = G^{(i)}_Z(1) + G^{(i)}_Z(1) - \left(G^{(i)}_Z(1)\right)^2
\]

Result 2.4 For \( z \geq 1 \), the following is a simple recurrence relation for probabilities \( g_z \) of the EIGD for \( z<m \)
\[
zg_{z+1} = \sum_{i=0}^{m} (l + \rho_1^{(i)}) \theta^{z-i} g_{z-i}
\]
and for \( z \geq m \)
\[
zg_{z+1} = m \sum_{i=0}^{m} (\rho_2 \theta) \theta^{z-i} g_{z-m-i+1}
\]

Proof: From (3) we have
\[
G_z(s) = \sum_{z=0}^{\infty} s^z g_z
\]
\[
= \Lambda(\rho_1, \rho_2, \theta)
\]
\[
s(1-s\theta)^{-1}(1-s\rho_1\theta)^{-1}(1-s\rho_2\theta)^{-1}
\]

On differentiating (12) and (13) with respect to \( s \) we get the following, in the light of (3),
\[
\sum_{z=0}^{\infty} (z+1)g_{z+1}s^z = \frac{G_z(s)}{s} + \theta \frac{G_z(s)}{1-s\theta}
\]
\[
+ \rho_1 \theta \frac{G_z(s)}{1-s\rho_1 \theta} + mp_2 \theta \frac{G_z(s)^{m-1}}{1-s^m \rho_2 \theta}
\]

By applying (1) and (2) in (14) to obtain
\[
\sum_{z=0}^{\infty} (z+1)s^z g_{z+1} = \sum_{z=0}^{\infty} g_{z+1} \Lambda(\rho_1, \rho_2, \theta)
\]
\[
+ \sum_{z=0}^{\infty} \sum_{i=0}^{m-1} (l + \rho_1^{(i)}) \theta^{z-i} s^{i+1} g_{z-i+1}
\]
\[
+ \sum_{z=0}^{\infty} \sum_{i=0}^{m} (\rho_2 \theta) \theta^{z-i} s^{m} g_{z-m+i+1}
\]

which implies
\[
\sum_{z=0}^{\infty} zg_{z+1} = \sum_{z=0}^{\infty} g_{z+1} \Lambda(\rho_1, \rho_2, \theta)
\]
\[
+ \sum_{z=0}^{\infty} \sum_{i=0}^{m-1} (l + \rho_1^{(i)}) \theta^{z-i} s^{i+1} g_{z-i+1}
\]
\[
+ \sum_{z=0}^{\infty} \sum_{i=0}^{m} (\rho_2 \theta) \theta^{z-i} s^{m} g_{z-m+i+1}
\]

On equating coefficient of \( s^z \) on both sides of the expressions (15), we get (11).

3. Estimation

Here we discuss the estimation of the parameters of the EIGD by various methods of estimation such as the method of factorial moments, the method of mixed moments and the method of maximum likelihood. We assume that \( m \) is a fixed positive integer and the parameters \( \rho_1, \rho_2 \) and \( \theta \) of the EIGD are estimated for possible values of \( m \).

Method of factorial moments

In method of factorial moments, equate the first three factorial moments of the EIGD to the corresponding sample factorial moments say \( \mu_1^{'}, \mu_2^{'}, \) and \( \mu_3^{'}, \) and thereby we obtain the following system of equations:
\[
1 + \delta_1 + \delta_2 + m\delta_3 = \mu_1^{'}
\]
\[
2m^2\delta_1^2 + m(m+1)\delta_2 + 2\delta_1(1+\delta_2) + 2\delta_2(1+\delta_2) + \delta_3(1+\delta_3) = \mu_2^{'}
\]
\[
6\left[m^2\delta_1^2(1+\delta_3) + \delta_2(1+\delta_2) + \delta_3(1+\delta_3)\right] + 6m\delta_1^2(1+\delta_2) + 2\delta_1^2(1+\delta_2) + 2\delta_2(1+\delta_2) = \mu_3^{'}
\]

where \( \delta_1, \delta_2, \) and \( \delta_3 \) are given in (8). Now the parameters of EIGD are estimated by solving the non-linear equations (16), (17) and (18).

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 (16) and (17).
\[ \Lambda(p_1, p_2, \theta) = \frac{p_i}{N} \quad (19) \]

where \( \Lambda(p_1, p_2, \theta) \) is as defined in (3), \( p_i \) is the observed frequency of the distribution corresponding to the observation \( x = 1 \) and \( N \) is the total frequency.

**Method of maximum likelihood**

Let \( a(z) \) be the observed frequency of \( z \) events and \( u \) be the highest value of \( z \) observed. Then the likelihood function of the sample is

\[ L = \prod_{z=x}^n (g_z)^{a(z)} \]

which implies

\[ \log L = \sum_{z=x}^n a(z) \log (g_z) \]

Let \( \hat{p}_1, \hat{p}_2, \hat{\theta} \) denotes the maximum likelihood estimators of \( p_1, p_2, \theta \) respectively. Now \( \hat{p}_1, \hat{p}_2, \hat{\theta} \) are obtained by solving the likelihood equations (20), (21) and (22) as given below.

\[ \frac{\partial \log L}{\partial \theta} = 0 \]

implies

\[ \sum_{z=1}^n a(z) \Psi_1(p_1, p_2, \theta) + \sum_{z=1}^n a(z) \Psi_2(m, p_1, p_2, \theta) = \sum_{z=1}^n 2a(z) \quad (20) \]

where

\[ \Psi_1(p_1, p_2, \theta) = 1 + \frac{p_1 \theta}{1 - \theta} + \frac{p_2 \theta}{1 - p_1 \theta} - 1 \]

\[ \Psi_2(m, p_1, p_2, \theta) = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} (m-1)k \left( \frac{1}{p_1} \right)^j \left( \frac{p_2}{\theta^{m-1}} \right)^k \]

and

\[ E_j(m, p_1, p_2, \theta) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left( \frac{1}{p_1} \right)^j \left( \frac{p_2}{\theta^{m-1}} \right)^k \]

\[ \frac{\partial \log L}{\partial p_1} = 0 \]

implies

\[ \sum_{z=1}^n a(z)(1 + \frac{p_1 \theta}{1 - \theta}) + \sum_{z=1}^n a(z) \Psi_4(m, p_1, p_2, \theta) = 0, \quad (21) \]

where

\[ \Psi_4(m, p_1, p_2, \theta) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left( \frac{1}{p_1} \right)^j \left( \frac{p_2}{\theta^{m-1}} \right)^k \]

\[ \frac{\partial \log L}{\partial p_1} = 0 \]

implies

\[ \sum_{z=1}^n a(z) \frac{\rho_1 \theta}{1 - \rho_1 \theta} + \sum_{z=1}^n a(z) \Psi_3(m, p_1, p_2, \theta) = 0 \quad (22) \]

in which \( \Psi_3(m, p_1, p_2, \theta) \) is as defined in (20).

We present the fitting of the intervened geometric distribution (IGD) and the extended intervened geometric distribution (EIGD) for particular values of \( m \) for the data set on the count of the number of European red mites on apple leaves taken from Jani and Shah [18]. We estimate the parameters by the method of factorial moments, the method of mixed moments and the method of maximum likelihood. We have computed the values of \( \chi^2 \) statistics in the case of each model and the numerical results are summarized in Table 1, Table 2 and Table 3. From the tables it is obvious that the EIGD with \( m=3 \) gives a better fit compared to IGD as well as for the case \( m=1, m=2 \) (MIGD) and \( m=4 \).

**Table 1. (Observed frequencies \( O_j \) and Expected frequencies \( E_j \) calculated by the method of factorial moments).**

| \( x \) | \( O_i \) | \( E_i \) | Expected frequencies of EGD for different values of \( m \) |
|-------|--------|--------|----------------------------------|
|       | (IGD)  | m=1    | m=2    | m=3    | m=4    |
| 1     | 38     | 29     | 28     | 32     | 40     | 33     |
| 2     | 17     | 20     | 26     | 17     | 18     | 22     |
| 3     | 10     | 12     | 14     | 15     | 7      | 11     |
| 4     | 9      | 7      | 7      | 7      | 9      | 5      |
| 5     | 6      | 12     | 5      | 9      | 6      | 9      |
| Total | 80     | 80     | 80     | 80     | 80     |
| Estimated values of parameters | \( \hat{p}_1 = 0.17 \) | \( \hat{p}_2 = 1.21 \) | \( \hat{p}_3 = 0.31 \) | \( \hat{p}_4 = 0.427 \) | \( \hat{p}_5 = 1.05 \) |
| \( \theta = 0.596 \) | \( \theta = 0.216 \) | \( \theta = 0.416 \) | \( \theta = 0.326 \) |
| Chi- square value | 7.148 | 8.601 | 4.363 | 1.441 | 6.185 |
| p-value | 0.128 | 0.072 | 0.359 | 0.837 | 0.186 |
Table 2. (Observed frequencies \(O\), and Expected frequencies \(E\), calculated by the method of mixed moments).

| \(x\) | \(O_i\) | \(E_i\) | Expected frequencies of EGD for different values of \(m\) |
|-------|--------|--------|-----------------------------------------------------|
|       | (IGD)  | \(m=1\) | \(m=2\) | \(m=3\) | \(m=4\) |
| 1     | 38     | 27     | 28     | 26     | 40     | 29     |
| 2     | 17     | 22     | 23     | 17     | 16     | 19     |
| 3     | 10     | 14     | 14     | 15     | 6      | 10     |
| 4     | 9      | 8      | 7      | 9      | 10     | 5      |
| 5’    | 6      | 9      | 8      | 13     | 8      | 17     |
| Total | 80     | 80     | 80     | 80     | 80     | 80     |

Estimated values of parameters:
\[ \hat{\rho}_1 = 0.56 \]
\[ \hat{\rho}_2 = 0.35 \]
\[ \hat{\theta} = 0.521 \]
Chi-square value:
\[ 7.886 \]
\[ 7.351 \]
\[ 10.974 \]
\[ 3.429 \]
\[ 13.321 \]
p-value:
\[ 0.096 \]
\[ 0.118 \]
\[ 0.027 \]
\[ 0.489 \]
\[ 0.009 \]

Table 3. (Observed frequencies \(O\), and Expected frequencies \(E\), calculated by the method of maximum likelihood).

| \(x\) | \(O_i\) | \(E_i\) | Expected frequencies of EGD for different values of \(m\) |
|-------|--------|--------|-----------------------------------------------------|
|       | (IGD)  | \(m=1\) | \(m=2\) | \(m=3\) | \(m=4\) |
| 1     | 38     | 41     | 28     | 38     | 38     |
| 2     | 17     | 22     | 25     | 17     | 20     | 21     |
| 3     | 10     | 10     | 15     | 14     | 8      | 9      |
| 4     | 9      | 5      | 7      | 5      | 8      | 5      |
| 5’    | 6      | 2      | 5      | 6      | 6      | 7      |
| Total | 80     | 80     | 80     | 80     | 80     | 80     |

Estimated values of parameters:
\[ \hat{\rho}_1 = 0.275 \]
\[ \hat{\rho}_2 = 0.894 \]
\[ \hat{\theta} = 0.426 \]
Chi-square value:
\[ 12.556 \]
\[ 8.57 \]
\[ 4.343 \]
\[ 3.817 \]
p-value:
\[ 0.014 \]
\[ 0.073 \]
\[ 0.362 \]
\[ 0.898 \]
\[ 0.378 \]

Acknowledgements

Authors wish to express their sincere gratitude to the Editor-in-Chief and anonymous referees for fruitful comments.

References

[1] Shanmugam, R. (1985) An intervened Poisson distribution and its medical application, *Biometrics* 41, 1025-1029.
[2] Shanmugam, R. (1992) An inferential procedure for the Poisson intervention parameter, *Biometrics* 48, 559-565.
[3] Haung, M., Fun, K. Y. (1989) Intervened truncated Poisson distributions, *Sankhya Series*, 51, 302-310.
[4] Scollnik, D. P. M. (2006) On intervened generalized Poisson distributions, *Communication in Statistics - Theory and Methods*, 35, 953-963.
[5] Dhanavanthan, P. (1998) Compound intervened Poisson distributions, *Biometrical Journal*, 40, 641-646.
[6] Dhanavanthan, P (2000) Estimation of parameters of compound intervened Poisson distributions, *Biometrical Journal*, 42, 315-320.
[7] Kumar, C. S., Shibu, D. S. (2011) Modified intervened Poisson distributions, *Statistica*, 71, 489-499.
[8] Kumar, C. S., Shibu, D. S. (2011) Estimation of the parameters of a finite mixture of intervened Poisson distribution. *Statistical Methods in Interdisciplinary Studies*, Dept. of Statistics, Maharajas College, Cochin, 93-103.
[9] Kumar, C. S., Shibu, D. S. (2012a). Some finite mixtures of intervened Poisson distribution. *The Aligarh Journal of Statistics*, 32, 97-114.
[10] Kumar, C. S., Shibu, D. S. (2012b). Generalized intervened Poisson distribution. *Journal of Applied Statistical Sciences*, 19, 131-141.
[11] Kumar, C. S., Shibu, D. S. (2012c). An alternative to truncated intervened Poisson distribution. *Journal of Statistics and Applications*, 5, 131-141.
[12] Kumar, C. S., Shibu, D. S. (2013) On some aspects of intervened generalized Hermite distribution, *Metron*, 71, 9-19.
[13] Kumar, C. S., Shibu, D. S. (2013) Finite mixtures of extended intervened Poisson distribution, In Collection of Recent Statistical Methods and Applications, *Department of Statistics*, University of Kerala Publishers, Trivandrum, 69-82.
[14] Kumar, C. S., Shibu, D. S. (2013) An extended version of intervened Poisson distribution, *Research Journal of Fatima Mata National College Kollam*, 4, 29-40.
[15] Kumar, C. S., Shibu, D. S. (2014) On finite mixtures of modified intervened Poisson distribution and its applications, *Journal of Statistical Theory and Applications*, 13(4), 344-355.
[16] Bartolucci, A. A., Shanmugam, R., Singh, K. P. (2001) Development of the generalized geometric model with application to cardiovascular studies, *System Analysis, Modelling simulation*, 41, 339-349.

[17] Kumar, C. S., Sreejakumari. S. (2014) Modified intervened geometric distribution, *The Aligarh Journal of Statistics*, 34, 1-12.

[18] Jani, P. N., Shah, S. M. (1979). On fitting of the generalized logarithmic series distribution. *Journal of the Indian Society for Agricultural Statistics*. 30, 1-10.