

## Inflated-parameter Harris distribution

E. Sandhya<sup>a</sup>, T. Lovely Abraham<sup>b</sup>

<sup>a</sup>Department of Statistics, Prajyoti Niketan College, Pudukkad – 680301, India.

<sup>b</sup>Department of Statistics, St. Thomas College, Pala- 686574, India.

### Abstract

Inflated-parameter Harris distribution is introduced and its properties are studied. A characterization based on p.g.f is given. The maximum likelihood and moment estimators of the parameters are found out together with their standard errors. The distribution is seen to be a good fit to a real life situation concerning the published results of Kerala Public Service Commission. ©2016 All rights reserved.

*Keywords:* Harris distribution, generalized power series distribution, zero inflated distributions, estimation of parameters.

### 1. Introduction and preliminaries

To model some of real life data we need a richer class of discrete probability distributions other than the classical discrete distributions. In this aspect recently many generalized forms of discrete distributions are introduced. In the case of over dispersion and excess of zero's, zero-inflated distributions are powerful discrete distributions to catch the situation. Zero-inflated modified power series distribution is studied by Johnson et al. [3] and Gupta et al. [1]. Roski et al. [12] applied zero inflated distribution in actuarial literature. Momeni [7] had shown that in modeling insurance claims zero-inflated distributions are more appropriate than the classical discrete distributions.

Kolev et al. [4] developed a new class of discrete distributions inflated-parameter generalized power series distribution (IGPSD). It includes inflated form of binomial, Poisson, negative binomial and logarithmic series distributions. Kolev et al. [4] applied negative binomial and inflated parameter

---

*Email addresses:* [esandhya@hotmail.com](mailto:esandhya@hotmail.com) (E. Sandhya), [lovely.shiby1976@gmail.com](mailto:lovely.shiby1976@gmail.com) (T. Lovely Abraham)

Poisson to model real data. Later Minkova [6] gave some characterizations of such distribution.

In this paper we introduce inflated parameter Harris distribution (IHD). Its probability mass function, distribution function and moments are derived. Also IHD is developed by gamma mixing of appropriate Poisson distribution. It is shown that Poisson random sum of inflated-logarithmic series r.v is IHD. A characterization of probability generating function is given in Theorem 4.1. The unknown parameters are estimated using method of moments and method of maximum likelihood. We compare it using standard error of simulated samples. In section 5 we present two data sets of scores of the candidate selected for the interview in public service examination of Kerala state, where the IHD shown good fit.

1.1. Generalized power series distribution

**Definition 1.1.** A r.v  $X$  is said to have a GPSD with parameter  $\theta \geq 0$  if

$$P(X = a + nk) = \frac{c_n \theta^{a+nk}}{C(\theta)}, n = 1, 2, 3, \dots$$

Also  $c_n > 0$  and  $C(\theta) = \sum_{n=0}^{\infty} c_n \theta^{a+nk}, a > 0, k > 0$  are integer. Without loss of generality we assumes that  $C(0)=1$  dividing both numerator and denominator by  $\theta^a$  we get.

$$P(X = a + nk) = \frac{c_n (\theta^k)^n}{C(\theta)}, n = 0, 1, 2, \dots, \tag{1.1}$$

$$C(\theta) = \sum_{n=0}^{\infty} c_n (\theta^k)^n,$$

$P(0) = 0$ . The r.v  $X$  takes values on  $S = \{a, a + k, a + 2k, \dots\}, a \geq 1, k \geq 1$  integers.

1.2. Harris distribution

Harris distribution was first introduced by Harris [2]. Later it was studied by different authors. Sandhya et al [11] have shown that it is a member of GPSD. Probability generating function (p.g.f) of the distribution is

$$P_X(t) = \left( \frac{t}{m - (m - 1)t^k} \right)^{\frac{1}{k}}, k \geq 1 \text{ integer and } m > 1$$

and probability mass function (p.m.f) is given by

$$P(X = x) = \binom{\frac{x}{k}-1}{\frac{x-1}{k}-1} \left( \frac{1}{m} \right)^{\frac{1}{k}} \left( 1 - \frac{1}{m} \right)^{\frac{x-1}{k}}, x = 1, 1 + k, 1 + 2k, \dots$$

Putting  $(1 - \frac{1}{m}) = \theta^k$  and  $x = 1 + nk$ , we have,

$$P(X = 1 + nk) = \binom{\frac{1}{k}+n-1}{n} \frac{\theta^{1+nk}}{\theta (1 - \theta^k)^{-\frac{1}{k}}}, n = 0, 1, 2, \dots \tag{1.2}$$

This distribution is denoted as  $H_1\left(\theta, k, \frac{1}{k}\right)$   
 We can also consider Harris distribution with support on  $\{0, k, 2k, \dots\}$  and it has p.m.f

$$P(X = nk) = \binom{\frac{1}{k} + n - 1}{n} \frac{\theta^{nk}}{(1 - \theta^k)^{-\frac{1}{k}}}, n = 0, 1, 2, \dots$$

This satisfies (1.1) with  $C(\theta) = (1 - \theta^k)^{-\frac{1}{k}} = \sum_{n=0}^{\infty} \binom{\frac{1}{k} + n - 1}{n} \theta^{nk}$ .

From (1.1) and (1.2) it is clear that Harris distribution belongs to GPSD, where  $a = 0$ .

Now we proceed to develop inflated-parameter Harris distribution as given in Kolev et al. [4]. Let  $P_N(t)$  and  $P_X(t)$  be p.g.f of two non-negative integer valued r.v  $N$  and  $X$  respectively. Then  $S = X_1 + X_2 + \dots + X_N$  with convention that  $S = 0$  when  $N = 0$ . The p.g.f of  $S$  is

$$P_S(t) = P_N(P_X(t)). \tag{1.3}$$

If  $N$  belongs to the family of GPSD with parameter  $\theta$  defined by (1.1) and  $X$  has an arbitrary discrete distribution, then the resulting random sum  $S$  has a p.g.f given by

$$P_S(t) = \frac{C(\theta(P_X(t)))}{C(\theta)} \tag{1.4}$$

for different choices of series function  $C(\theta)$ .

## 2. Inflated parameter Harris distribution

An extension of GPSD including an additional parameter  $\rho \in (0, 1)$  is suggested in Kolev et al. [4] in connection with insurance claims data.

The following theorem gives a method of construction of inflated-parameter Harris distribution (IHD) in connection with the idea discussed in the last paragraph of Section 1.

**Theorem 2.1.** *If  $X_1, X_2 \dots$  are i.i.d geometric r.v's with parameter  $(1 - \rho)$  and with p.g.f  $P_X(t) = \frac{(1-\rho)t}{1-\rho t}$ ,  $0 < \rho < 1$  and if  $N$  has Harris distribution defined by (1.2) with series function,*

$$C(\theta) = (1 - \theta^k)^{-\frac{1}{k}} \text{ and } P_N(t) = \left( \frac{1 - (\theta t)^k}{1 - \theta^k} \right)^{-\frac{1}{k}}, 0 \leq \theta^k \leq 1$$

*then the random sum  $Z = X_1 + X_2 + \dots + X_N$  has IHD with p.g.f  $P_Z(t) = \left( \frac{1 - \left(\frac{\theta(1-\rho)t}{1-\rho t}\right)^k}{1 - \theta^k} \right)^{-\frac{1}{k}}$ ,*

$$0 \leq \theta^k \leq 1. \tag{2.1}$$

Thus we have,

$$P(Z = 0) = (1 - \theta^k)^{-\frac{1}{k}} = C_1$$

$$P(Z = k) = (1 - \theta^k)^{-\frac{1}{k}} \theta^k (1 - \rho)^k \binom{1/k}{1}$$

$$= C_1 C_2 \binom{1/k}{1}, \text{ where } C_2 = \theta^k (1 - \rho)^k$$

$$P(Z = k + i) = C_1 C_2 \binom{1/k}{1} \binom{k+i-1}{i} \rho^i, i = 1, 2, \dots, k - 1$$

$$P(Z = 2k) = C_1 \left[ C_2 \binom{1/k}{1} \binom{2k-1}{k} \rho^k + C_2^2 \binom{1/1^{+1}}{2} \right]$$

$$P(Z = 2k + i) = C_1 \sum_{j=1}^n C_2^j \binom{1/1^{+j-1}}{j} \binom{2k+i-1}{(2-j)k+i} \rho^{(2-j)k+i}, i = 1, 2, \dots, k - 1.$$

In general

$$P(Z = nk + i) = C_1 \sum_{j=1}^n C_2^j \binom{1/k^{+j-1}}{j} \binom{nk+i-1}{(n-j)k+i} \rho^{(n-j)k+i}, i = 1, 2, \dots, k - 1, n = 1, 2, \dots \quad (2.2)$$

It is to be noted that  $Z$  takes values on  $\{0, k, k + 1, \dots, 2k, 2k + 1, \dots\}$ .

Letting  $z = nk + i$  and hence  $n = \text{Quotient of } \binom{z}{k} = Q \left( \binom{z}{k} \right)$ , we can write (2.2) as

$$P(Z = 0) = (1 - \theta^k)^{-\frac{1}{k}} = C_1$$

$$P(Z = z) = C_1 \sum_{j=1}^{Q \left( \binom{z}{k} \right)} C_2^j \binom{1/k^{+j-1}}{j} \binom{z-1}{z-jk} \rho^{z-jk}, z = k, k + 1, \dots, 2k, 2k + 1, \dots \quad (2.3)$$

can be denoted as  $IHD(\theta^k, k, \rho)$ .

**Note**

- (1) When  $k=1$ ,  $IHD(\theta^k, k, \rho)$  reduces to inflated Parameter geometric distribution (IGD),  $IGD(1 - \theta, \rho)$  given in Kolev et al. [4].
- (2) When  $k = 1, \rho = 0, IHD(\theta^k, k, \rho)$  reduces to geometric distribution with parameter  $\theta$ .
- (3) When  $\rho = 0, IHD(\theta^k, k, \rho)$  reduces to  $H \left( \theta^k, k, \frac{1}{k} \right)$  given in (1.2) above.

Tables 1 and 2 give the p.m.f for different parameters  $k, \theta^k$  and  $\rho$ .

Table 1:

Z	$k = 2$ $\theta^k = 0.5$ $\rho = 0.5$	$k = 2$ $\theta^k = 0.5$ $\rho = 0.9$	$k = 2$ $\theta^k = 0.75$ $\rho = 0.9$	$k = 5$ $\theta^k = 0.5$ $\rho = 0.5$	$k = 5$ $\theta^k = 0.5$ $\rho = 0.9$	$k = 5$ $\theta^k = 0.75$ $\rho = 0.9$
0	0.7071	0.7071	0.5	0.8705	0.87055	0.75785
1	0	0	0	0	0	0
2	0.04419	0.00176	0.00187	0	0	0
3	0.04419	0.00318	0.00337	0	0	0
4	0.03728	0.00430	0.00456	0	0	0
5	0.03038	0.00517	0.00550	0.00272	$8.7055 \times 10^{-7}$	$1.13679 \times 10^{-6}$
6	0.02460	0.00585	0.00623	0.00680	$3.9174 \times 10^{-6}$	$5.11554 \times 10^{-6}$
7	0.01993	0.00635	0.00679	0.01020	0.00001	0.00001
8	0.01621	0.00672	0.00721	0.01190	0.00002	0.00002
9	0.01322	0.00698	0.00752	0.01190	0.00003	0.00005
10	0.01082	0.00714	0.00774	0.01073	0.00006	0.00008
11	0.00888	0.00723	0.00787	0.00905	0.00009	0.00012
12	0.00730	0.00725	0.00795	0.00736	0.00013	0.00017
13	0.00602	0.00723	0.00798	0.00596	0.00018	0.00024
14	0.00498	0.00716	0.00797	0.00493	0.00024	0.00031
15	0.00412	0.00706	0.00792	0.00425	0.00030	0.00039
16	0.00342	0.00694	0.00785	0.00380	0.00037	0.00048
17	0.00284	0.00680	0.00777	0.00349	0.00044	0.00058
18	0.00236	0.00664	0.00766	0.00323	0.00052	0.00068
19	0.00196	0.00648	0.00754	0.00298	0.00060	0.00079
20	0.00164	0.00630	0.00742	0.00272	0.00069	0.00090
21	0.00137	0.00612	0.00728	0.00246	0.00078	0.00102
22	0.00114	0.00594	0.00715	0.00222	0.00086	0.00113
23	0.00095	0.00576	0.00701	0.00199	0.00095	0.00124
24	0.00080	0.00557	0.00687	0.00178	0.00104	0.00136
25	0.00067	0.00539	0.00672	0.00160	0.00112	0.00146
26	0.00056	0.00521	0.00658	0.00145	0.00120	0.00157
27	0.00047	0.00503	0.00644	0.00132	0.00128	0.00167
28	0.00039	0.00486	0.00630	0.00120	0.00135	0.00177
29	0.00033	0.00469	0.00616	0.00110	0.00142	0.00186
30	0.00027	0.00452	0.00603	0.00100	0.00148	0.00194

Table 2:

$k = 10$ $\theta^k = 0.5$ $\rho = 0.5$	$k = 10$ $\theta^k = 0.5$ $\rho = 0.9$	$k = 10$ $\theta^k = 0.75$ $\rho = 0.9$
0.9330	0.93303	0.870551
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0.000045	$4.66516 \times 10^{-12}$	$6.52913 \times 10^{-12}$
0.000227	$4.19865 \times 10^{-11}$	$5.87622 \times 10^{-11}$
0.000626426	$2.07833 \times 10^{-10}$	$2.90873 \times 10^{-10}$
0.00125285	$7.48199 \times 10^{-10}$	$1.04714 \times 10^{-9}$
0.00203588	$2.18848 \times 10^{-9}$	$3.06289 \times 10^{-9}$
0.00285024	$5.51498 \times 10^{-9}$	$7.71848 \times 10^{-9}$
0.0035628	$1.24087 \times 10^{-8}$	$1.73666 \times 10^{-8}$
0.004077177	$2.55265 \times 10^{-8}$	$6.83251 \times 10^{-8}$
0.00432625	$4.88194 \times 10^{-8}$	$1.22958 \times 10^{-7}$
0.00432625	$8.78748 \times 10^{-8}$	$2.10305 \times 10^{-7}$
0.00410995	$1.50266 \times 10^{-7}$	$3.44135 \times 10^{-7}$
0.00373643	$2.4589 \times 10^{-7}$	$5.42012 \times 10^{-7}$
0.00326991	$3.87276 \times 10^{-7}$	$8.25527 \times 10^{-7}$
0.00276866	$5.89852 \times 10^{-7}$	$1.2206 \times 10^{-6}$
0.00227909	$8.72138 \times 10^{-7}$	$1.75766 \times 10^{-6}$
0.00183411	$1.25588 \times 10^{-6}$	$2.47172 \times 10^{-6}$
0.00145406	$1.76608 \times 10^{-6}$	$3.40224 \times 10^{-6}$
0.00114891	$2.43096 \times 10^{-6}$	$4.59303 \times 10^{-6}$
0.000920631	$3.28179 \times 10^{-6}$	$6.09181 \times 10^{-6}$
0.000765227	$4.35269 \times 10^{-6}$	$7.94981 \times 10^{-4}$
0.00067445	$5.68026 \times 10^{-6}$	0.0000102212
0.000637154	$7.30319 \times 10^{-6}$	0.0000129623
0.000640509	$9.26178 \times 10^{-6}$	0.0000162311
0.000671143	0.0000115974	0.000020086
0.000716191	0.0000143517	0.0000245852

Figure 1, 2 and 3 show the pictorial representation of IHD for the above values  $k, \theta^k = \eta$  and  $\rho$ . Skewness  $\beta_1$  and  $P(0)$  are given below.

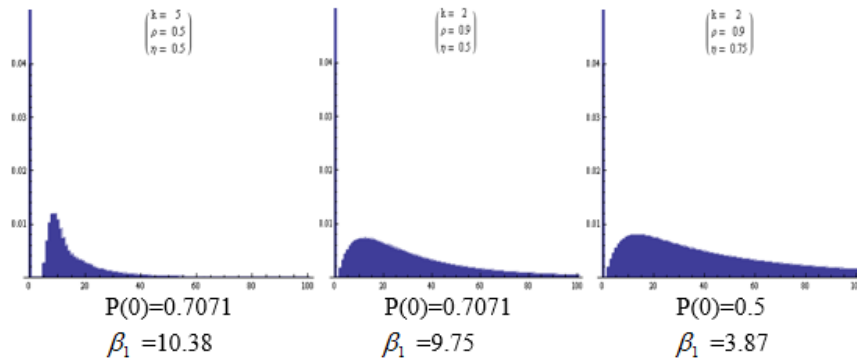


Figure 1:

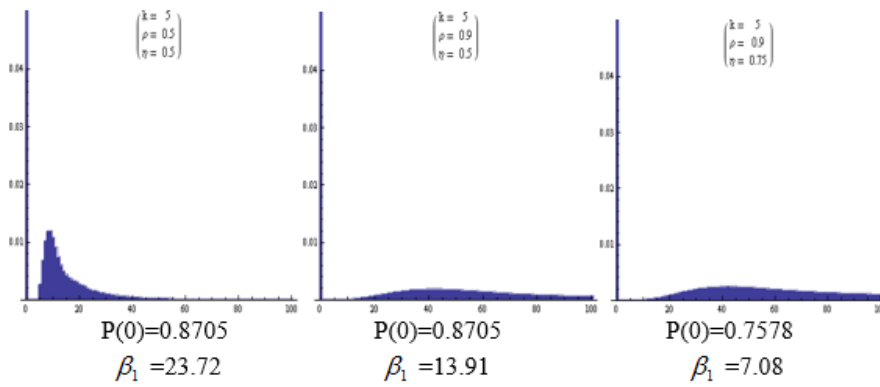


Figure 2:

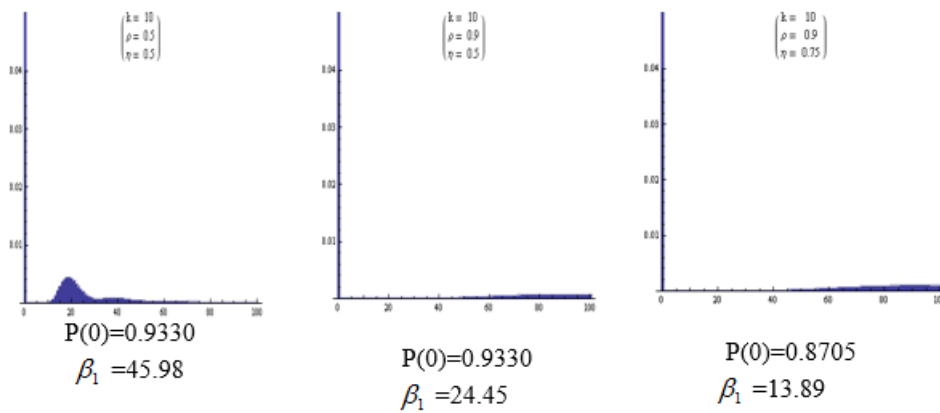


Figure 3: Sender and receiver data packet

*Remark 2.2.* For IHD there is a gap between 0 and  $k$  and has probability at all integer points after  $k$ . The nature of the distribution change as the parameters  $k$  and  $\rho$  change more than w.r.t. the change in  $\theta^k$ . This is clearly depicted in figure 2 and 3. Also it can be seen that, for fixed values of  $k$  and  $\rho$  the skewness decreases as  $\eta$  increases, and it increases for fixed values of  $\rho$  and  $\eta$  as  $k$  increases

2.1. Distribution Function

$$F(z) = P[Z \leq z] = C_1 \left\{ 1 + \sum_{j=1}^{Q(\frac{z}{k})} \binom{1/k+j-1}{j} C_2^j \left[ \sum_{i=0}^{z-jk} \binom{jk+i-1}{i} \rho^i \right] \right\}, z = 0, k, k + 1, \dots$$

2.2. Moments

1. Factorial moments

$$\begin{aligned} \mu_{(1)} &= \frac{\theta^k}{1 - \theta^k} \frac{1}{1 - \rho} \\ \mu_{(2)} &= \frac{\theta^k}{1 - \theta^k} \left[ \frac{2\rho + (k - 1)}{(1 - \rho)^2} + \frac{\theta^k}{1 - \theta^k} \frac{(k + 1)}{(1 - \rho)^2} \right] \\ \mu_{(3)} &= \frac{\theta^k}{1 - \theta^k} \left[ \frac{(k - 1)(k - 2)}{(1 - \rho)} + \frac{\rho(k + 1)(k - 1)}{(1 - \rho)^2} + \frac{\rho^2(k + 2)(k + 1)}{(1 - \rho)^3} \right] \\ &\quad + \left( \frac{\theta^k}{1 - \theta^k} \right)^2 (k + 1) \left[ \frac{(3 + \rho)(k - 1)}{(1 - \rho)^2} + \frac{3\rho(k + 1)}{(1 - \rho)^3} \right] \\ &\quad + \left( \frac{\theta^k}{1 - \theta^k} \right)^3 \frac{(1 + k)(1 + 2k)}{(1 - \rho)^3}. \end{aligned}$$

2. Ordinary moments

$$\begin{aligned} \text{Mean} = \mu'_1 &= \frac{\theta^k}{1 - \theta^k} \frac{1}{1 - \rho} \\ \mu'_2 &= \frac{\theta^k}{1 - \theta^k} \frac{1}{(1 - \rho)^2} \left[ k + \rho + \frac{\theta^k}{1 - \theta^k} (k + 1) \right] \\ \mu'_3 &= \frac{\theta^k}{1 - \theta^k} \frac{1}{(1 - \rho)^3} \left[ [(1 - \rho)(k^2 + 2\rho(k - 1))] + \rho^2(k + 1)(k + 2) \right] \\ &\quad + \left( \frac{\theta^k}{1 - \theta^k} \right)^2 \frac{(1 + k)}{(1 - \rho)^3} [3k(1 - \rho) + \rho(k + 2)] + \left( \frac{\theta^k}{1 - \theta^k} \right)^3 \frac{(1 + k)(1 + 2k)}{(1 - \rho)^3}. \end{aligned}$$

3. Central moments

$$\begin{aligned} \text{Variance} = \mu_2 &= \frac{\theta^k}{(1 - \theta^k)} \frac{1}{(1 - \rho)^2} \left[ k + \rho + \frac{k\theta^k}{(1 - \theta^k)} \right] \\ \mu_3 &= \frac{\theta^k}{1 - \theta^k} \frac{1}{(1 - \rho)^3} \left[ [(1 - \rho)(k^2 + 2\rho(k - 1))] + \rho^2(k + 1)(k + 2) \right] \\ &\quad + \left( \frac{\theta^k}{1 - \theta^k} \right)^2 \frac{1}{(1 - \rho)^3} [3k^2 - 2k^2\rho - \rho] \\ &\quad + \left( \frac{\theta^k}{1 - \theta^k} \right)^3 \frac{2k^2}{(1 - \rho)^3}. \end{aligned}$$



The expressions for  $\mu_{(4)}, \mu'_4$  and  $\mu_4$  are very complicated and not in compact form and we do not present it here.

*Remark 2.3.* When  $\rho = 0$ , all the above moments coincide with that of Harris distribution defined by (1.2).

*Remark 2.4.* When  $k = 1$ , the mean and variance reduces to that of inflated parameter geometric distribution in Kolev et al. [4].

### 3. IHD as a Randomized mixture

Let  $F$  be a distribution function depending on a parameter  $\theta^k$  and  $u$  be a probability density function. Then

$$W(x) = \int_{-\infty}^{\infty} F(x_1, \theta^k) u(\theta^k) d\theta^k$$

is monotonic function of  $x$  increasing from 0 to 1 and hence a distribution function. If  $F$  has continuing density  $f$ , then  $W$  has a density  $w$  given by

$$w(x) = \int_{-\infty}^{\infty} f(x_1, \theta^k) u(\theta^k) d\theta^k.$$

The parameter  $\theta^k$  is treated as a r.v. Then the process is called randomization and new probability density  $w(x)$  is called mixture.

Sandhya et al. [10] obtained Harris distribution as a gamma mixture of Poisson distribution. Minkova [6] generated Inflated-parameter negative binomial distribution by gamma mixing of Inflated-parameter Poisson distribution.

**Theorem 3.1.** *The inflated extended Poisson mixed with gamma distribution results in Inflated-parameter Harris distribution.*

*Proof.* The p.g.f of extended Poisson distribution is

$$e^{\lambda(s^k-1)}, \quad k > 0 \text{ integer, } \lambda > 0 \tag{3.1}$$

and p.g.f of inflated-extended Poisson distribution is

$$\exp \lambda \left( \left( \frac{(1-\rho)t}{1-\rho t} \right)^k - 1 \right). \tag{3.2}$$

Here  $\lambda$  is an outcome of a gamma distributed  $G\left(\frac{1-\theta^k}{\theta^k}, \frac{1}{k}\right)$  r. v with p.d.f

$$f(\lambda) = \frac{\left(\frac{1-\theta^k}{\theta^k}\right)^{1/k}}{\sqrt{\frac{1}{k}}} e^{-\left(\frac{1-\theta^k}{\theta^k}\right)\lambda} \lambda^{\frac{1}{k}-1}, \quad \theta^k > 0, \lambda > 0,$$

then

$$P(Z = nk) = \int_0^{\infty} P\left(Z = \frac{nk}{\lambda}\right) f(\lambda) d\lambda$$

gives the same p.m.f of (2.2).

It is well known Poisson random sum of logarithmic r.v's is negative binomial. Harris distribution had developed from modified logarithmic series distribution stopped sum by generalized Poisson random variable. (Sandhya et al. [10]). In the same way we develop IHD. Minkova [7] introduced Inflated- parameter logarithmic series distribution. Hence Inflated-parameter modified logarithmic series distribution has p.g.f.

$$P_Y(t) = \left( 1 - \frac{\alpha}{k} + \frac{\alpha \ln \left( 1 - \theta^k [P_X(t)]^k \right)}{k \ln(1 - \theta^k)} \right)^{\frac{1}{k}} \tag{3.3}$$

$0 < \alpha \leq 1, k > 0$  an integer, where  $P_X(t)$  is given in Theorem 2.1.

□

**Theorem 3.2.** *Let  $Y_1, Y_2, Y_3, \dots$  be mutually independent and identically distributed inflated – parameter modified logarithmic series r.vs ,with p.g.f (3.3) and let  $N$  follow generalized Poisson r. v with p.g.f given in (3.1). Then the random sum  $S=Y_1+Y_2+Y_3+\dots+Y_N$  has inflated–parameter Harris distribution in (2.1).*

*Proof.* From the relation (1.3) together with the p.g.fs (3.1) and (3.3)

$$\begin{aligned} P_S(t) &= \exp \lambda \left[ \left( 1 - \frac{\alpha}{k} + \frac{\alpha \ln \left( 1 - \theta^k [P_X(t)]^k \right)}{k \ln(1 - \theta^k)} \right) - 1 \right] \\ &= \exp \lambda \left( -\frac{\alpha}{k} + \frac{\alpha \ln \left( 1 - \theta^k [P_X(t)]^k \right)}{k \ln(1 - \theta^k)} \right). \end{aligned}$$

Put  $\lambda = \frac{-\ln(1-\theta^k)}{\alpha}$ , substituting and simplifying we get

$$= \exp \left\{ \ln \left( \frac{1 - \theta^k [P_X(t)]^k}{1 - \theta^k} \right)^{-\frac{1}{k}} \right\}$$

$P_S(t) = \left( \frac{(1-\theta^k [P_X(t)]^k)}{(1-\theta^k)} \right)^{-\frac{1}{k}}$ , substituting  $P_X(t)$  we get,

$$P_S(t) = \left( \frac{1 - \left( \frac{\theta(1-\rho)t}{1-\rho t} \right)^k}{1 - \theta^k} \right)^{-\frac{1}{k}},$$

which is  $IHD(\theta^k, k, \rho)$ .

□

#### 4. Characterization of inflated parameter harris distribution

The following theorem gives a characterization of IHD

**Theorem 4.1.** *A r. v  $X \sim IHD(\theta^k, k, \rho)$  iff its p.g.f  $P(t)$  satisfies the equation*

$$P'(t) = \frac{\mu t^{k-1} (1 - \rho)^{k+1}}{(1 - \rho t)^{k+1}} [P(t)]^{1+k}, \tag{4.1}$$

where  $\mu$  is the mean of  $IHD(\theta^k, k, \rho)$

*Proof.*  $P(t) = \left( \frac{1 - \left( \frac{\theta(1-\rho)t}{1-\rho t} \right)^k}{1-\theta^k} \right)^{-\frac{1}{k}}$

$$P'(t) = \theta^k (1 - \rho)^k \left( \frac{t}{1 - \rho t} \right)^{k-1} \frac{1}{(1 - \rho t)^2} \left( \frac{1 - \left( \frac{\theta(1-\rho)t}{1-\rho t} \right)^k}{1 - \theta^k} \right)^{-\frac{1}{k}} \left( 1 - \frac{\theta^k (1 - \rho)^k t^k}{(1 - \rho t)^k} \right)^{-1}.$$

Multiplying and dividing by  $\frac{(1-\rho)}{(1-\theta^k)}$  on the R.H.S we have,

$$\begin{aligned} P'(t) &= P(t) [P(t)]^k \left( \frac{t}{1 - \rho t} \right)^{k-1} \frac{(1 - \rho)^{k+1}}{(1 - \rho t)^2} \frac{\theta^k}{1 - \theta^k} \frac{1}{1 - \rho} \\ &= [P(t)]^{k+1} \frac{\theta^k}{1 - \theta^k} \frac{1}{(1 - \rho)} (1 - \rho)^{k+1} \left( \frac{t}{1 - \rho t} \right)^{k-1} \frac{1}{(1 - \rho t)^2} \\ P'(t) &= \frac{\mu t^{k-1} (1 - \rho)^{k+1}}{(1 - \rho t)^{k+1}} [P(t)]^{1+k}. \end{aligned}$$

Conversely suppose that (4.1) is true.

$$\frac{P'(t)}{[P(t)]^{1+k}} = \mu t^{k-1} \left( \frac{1 - \rho}{1 - \rho t} \right)^{k+1}.$$

Integrating

$$\begin{aligned} \int \frac{P'(t)}{[P(t)]^{1+k}} dt &= \int \mu \frac{t^{k-1}}{(1 - \rho t)^{k-1}} \frac{(1 - \rho)^{k+1}}{(1 - \rho t)^2} dt \\ &= \mu (1 - \rho)^{k+1} \int \left( \frac{t}{1 - \rho t} \right)^{k-1} \frac{1}{(1 - \rho t)^2} dt. \end{aligned}$$

$$\frac{(P(t))^{-k}}{-k} = \frac{\mu}{k} (1 - \rho)^{k+1} \left( \frac{t}{1 - \rho t} \right)^k + b, \text{ where } b \text{ is a constant.}$$

As an initial condition put  $t=1$ ,  $b = \frac{1}{k} \frac{1}{(1-\theta^k)}$ , then

$$[P_Z(t)]^{-k} = -\frac{\theta^k}{1 - \theta^k} (1 - \rho)^k \left( \frac{t}{1 - \rho t} \right)^k + \frac{1}{1 - \theta^k}$$

$$\begin{aligned}
 &= \left( \frac{1}{1 - \theta^k} - \frac{\theta^k (1 - \rho)^k t^k}{(1 - \theta^k) (1 - \rho t)^k} \right) \\
 [P_Z(t)]^{-k} &= \left( \frac{1 - \left( \frac{\theta(1-\rho)t}{1-\rho t} \right)^k}{1 - \theta^k} \right) \\
 P_Z(t) &= \left( \frac{1 - \left( \frac{\theta(1-\rho)t}{1-\rho t} \right)^k}{1 - \theta^k} \right)^{-\frac{1}{k}}.
 \end{aligned}$$

□

*Remark 4.2.* When  $\rho = 0$ , the result reduces to  $P'(t) = \mu [P(t)]^{1+k}$  (Sandhya et al. [9])

### 5. Estimation of parameters

#### 5.1. Method of moments

The mean  $\bar{x}$  and variance are given in section 2. Equating them to sample moments and  $S^2$  and solving the resulting system of equation we get,

$$\begin{aligned}
 \hat{\theta}^k &= \frac{(1 + k)\bar{x}^2}{S^2 + \bar{x}(\bar{x} + 1)}, \\
 \hat{\rho}^k &= \frac{S^2 - k\bar{x} - k\bar{x}^2}{S^2 + \bar{x} - k\bar{x}^2},
 \end{aligned}$$

assuming that  $k$  is known.

#### 5.2. Method of maximum likelihood

Given i.i.d observations  $z_1, z_2, \dots, z_n$  from  $IHD(\theta^k, k, \rho)$  the likelihood function (Sadie Beckett et al. [11])

$$L(\theta^k, \rho / \tilde{Z}) = \prod_{i=1}^n P(Z = z_i).$$

Define  $y$  = number of  $z_i$ 's taking the value 0,  $z_i \in \tilde{Z}$ .

Then

$$\begin{aligned}
 L(\theta^k, \rho / \tilde{Z}) &= (1 - \theta^k)^{\frac{y}{k}} \prod_{\substack{i=1 \\ z_i \neq 0}}^n (1 - \theta^k)^{\frac{1}{k}} \sum_{j=1}^{Q(z_i/k)} (\theta^k (1 - \rho)^k)^j \binom{1/k + j - 1}{j} \binom{z_i - 1}{z_i - jk} \rho^{z_i - jk}, \\
 L(\theta^k, \rho / \tilde{Z}) &= (1 - \theta^k)^{\frac{n}{k}} \prod_{\substack{i=1 \\ z_i \neq 0}}^n \sum_{j=1}^{Q(z_i/k)} (\theta^k (1 - \rho)^k)^j \binom{1/k + j - 1}{j} \binom{z_i - 1}{z_i - jk} \rho^{z_i - jk}.
 \end{aligned}$$

Hence the log likelihood function denoted by  $L_*$  is

$$L_* = \frac{n}{k} \ln(1 - \theta^k) + \sum_{\substack{i=1 \\ z_i \neq 0}}^n \ln \sum_{j=1}^{Q(z_i/k)} (\theta^k (1 - \rho)^k)^j \binom{1/k+j-1}{j} \binom{z_i-1}{z_i-jk} \rho^{z_i-jk}.$$

By taking partial derivatives of  $L_*$  w.r.t  $\theta^k$  and  $\rho$ , and setting them equal to zero, we get the following system of equations,

$$\frac{n}{k} \ln(1 - \theta^k) = \sum_{\substack{i=1 \\ z \neq 0}}^n \frac{\sum_{j=1}^{Q(z_i/k)} \binom{1/k+j-1}{j} \binom{z_i-1}{j k-1} \rho^{z_i-jk} j (\theta^k (1 - \rho)^k)^{j-1} (1 - \rho)^k}{\sum_{j=1}^{Q(z_i/k)} \binom{1/k+j-1}{j} \binom{z_i-1}{j k-1} \rho^{z_i-jk} (\theta^k (1 - \rho)^k)^j}, \tag{5.1}$$

$$\sum_{\substack{i=1 \\ z \neq 0}}^n \frac{\sum_{j=1}^{Q(z_i/k)} \binom{1/k+j-1}{j} \binom{z_i-1}{j k-1} \rho^{z_i-jk-1} (\theta^k)^j (1 - \rho)^{j k-1} (z_i (1 - \rho) - j k)}{\sum_{j=1}^{Q(z_i/k)} \binom{1/k+j-1}{j} \binom{z_i-1}{j k-1} \rho^{z_i-jk} (\theta^k (1 - \rho)^k)^j} = 0. \tag{5.2}$$

By simulation we solve the above equations. (All of the computations and graphical representations are done using Mathematica). The MLE's of  $\theta^k$  and  $\rho$  are hence forth denoted by  $\theta_{ml}^k$  and  $\rho_{ml}$ . The simulation study of moment and maximum likelihood estimates of  $\theta^k$  and  $\rho$  are done. The tables are appended.

### 5.3. Moment estimators (ME) and MLE by simulation

Table 3: Moment estimators of  $\theta^k$  and  $\rho$  using simulated sample of size 100 and number of repetitions 30

		$k = 2$				$k = 5$			
$\theta^k$	$\rho$	0.5		0.9		0.5		0.9	
		$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$
0.5	Estimate	0.3925	0.5249	0.8984	0.4812	0.4051	0.5224	0.8805	0.5142
	SE	0.0305	0.0142	0.007	0.0158	0.0301	0.0186	0.0086	0.0191
0.75	Estimate	0.5130	0.7411	0.8058	0.8128	0.4869	0.7484	0.7846	0.7943
	SE	0.0354	0.0128	0.0483	0.02992	0.0319	0.0133	0.1565	0.0294

Table 4: Moment estimates of  $\theta^k$  and  $\rho$  using simulated sample of size 300 and number of repetitions 30

		$k = 2$				$k = 5$			
$\theta^k$	$\rho$	0.5		0.9		0.5		0.9	
		$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$
0.5	Estimate	0.4940	0.4983	0.8974	0.5103	0.4896	0.4786	0.8712	0.532
	SE	0.0204	0.0098	0.0044	0.0593	0.0333	0.0145	0.008	0.0149
0.75	Estimate	0.4918	0.7467	0.8510	0.7590	0.4560	0.7609	0.8690	0.7505
	SE	0.0365	0.0136	0.0266	0.0153	0.0389	0.0159	0.0377	0.0294

Table 5: Moment estimates of  $\theta^k$  and  $\rho$  using simulated sample of size 500 and number of repetitions 30

		$k = 2$				$k = 5$			
$\theta^k$	$\rho$	0.5		0.9		0.5		0.9	
		$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$
0.5	Estimate	0.5124	0.5024	0.8832	0.5217	0.4258	0.5153	0.8813	0.5182
	SE	0.0143	0.008	0.0092	0.062	0.0305	0.002	0.002	0.0046
0.75	Estimate	0.4999	0.7418	0.8996	0.7296	0.4319	0.7642	0.8763	0.7516
	SE	0.0173	0.0077	0.0309	0.0135	0.0396	0.0149	0.0119	0.0218

Table 6: MLEs of  $\theta^k$  and  $\rho$  using simulated sample of size 100 and number of repetitions 30

		$k = 2$				$k = 5$			
$\theta^k$	$\rho$	0.5		0.9		0.5		0.9	
		$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$
0.5	Estimate	0.5011	0.4879	0.8925	0.5059	0.4912	0.5031	0.9016	0.488
	SE	0.138	0.0090	0.0061	0.0071	0.1200	0.009	0.0037	0.009
0.75	Estimate							0.8985	0.7403
	SE							0.0038	0.0221

Table 7: MLEs of  $\theta^k$  and  $\rho$  using simulated sample of size 300 and number of repetitions 30

			$k = 2$			$k = 5$				
$\theta^k$	$\rho$	0.5		0.9		0.5		0.9		
		$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	
0.5	Estimate	0.4995	0.5072	0.8927	0.4908	0.4933	0.5097	0.9004	0.5272	
	SE	0.0073	0.0077	0.003	0.004	0.007	0.0081	0.0018	0.019	
0.75	Estimate							0.8963	0.7492	
	SE							0.0022	0.0078	

Table 8: MLEs of  $\theta^k$  and  $\rho$  using simulated sample of size 500 and number of repetitions 30

			$k = 2$			$k = 5$				
$\theta^k$	$\rho$	0.5		0.9		0.5		0.9		
		$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	$\rho$	$\theta^k$	
0.5	Estimate	0.4928	0.5047	0.8976	0.4950	0.5021	0.4976	0.09005	0.4880	
	SE	0.0061	0.0052	0.0009	0.001	0.0049	0.0059	0.0016	0.0088	
0.75	Estimate							0.9014	0.7491	
	SE							0.0011	0.0037	

Both MLE and ME give good estimates when the sample size increases and they became approximately equal. But also from the above tables it is very clear that standard error for MLE is significantly less than that of ME, for different sample size. Hence we prefer MLEs in fitting real datasets.

### 6. Application

What is special about IHD is that, there is zero probability between zero and the next value. There are many such situations, we come across in our daily life, which shows good fit for IHD. The following dataset are from Kerala government’s recruitment section- Kerala Public Service Commission (PSC) examination results.

The first data set is PSC exam result of HSST-Physical Education Department, category no.451/07, from [www.keralapsc.gov.in/statewide-ranklist-2010/rl-740/-10,shortlist-2007/sl-49/07](http://www.keralapsc.gov.in/statewide-ranklist-2010/rl-740/-10,shortlist-2007/sl-49/07). (Table 9).

The second data set is from ranked list of the candidates for the post of microbiologist in the Health Service Department ([www.keralapsc.gov.in/statewideranklist-rl-535-12,sl-110-10,categoryno.030/2009](http://www.keralapsc.gov.in/statewideranklist-rl-535-12,sl-110-10,categoryno.030/2009)). (Table 10). Scores of the candidates, who were selected for the interview, are shown in Table 9 and Table 10.

The frequency at  $z = 0$  gives the number of candidates failed to attend the interview. Next probability is after a gap  $k$  which is the cut-off mark of the test. The estimators  $\theta_{ml}^k$  and  $\rho_{ml}$  are calculated by taking random sample of size 50 using random number table, and calculating MLE using Mathematica.

Table 9:

Z	0	58	59	60	61	62	63	64	65-80	82	82>	Total
Observed	1184	3	5	1	0	3	2	2	0	1	0	1200

Table 10:

Z	0	33	34	35	36	37	38	39	40	41	42	43
Observed	1427	1	0	1	1	1	0	0	0	2	0	2

45	46	47	48	49	50	51	52	53	54	55	Total
0	0	0	0	1	1	0	0	0	0	1	1438

The  $\chi^2$  test of goodness of fit is given in Table 11. (For tests of goodness of fit of t-distribution in similar lines see M. Maghami and M. Bahrami[6].)

Table 11:

Data set	Calculated values of $\chi^2(0.01)$	Degrees of freedom	Tabled value	Conclusion
Table 9	7.2852	2	9.21	Good fit
Table 10	6.7371	2	9.21	Good fit

### 7. Conclusion

IHD seen to be good fit to the published results of Kerala PSC. We may apply IHD to such similar situations, where there is a chance to occur large number of failures (zeros) preceding successes after a fixed interval.

### References

- [1] L. P. Gupta, R. C. Gupta, R. C. Thripati, *Inflated modified power series distributions with applications*, Communications in Statistics Theory Methods, **24** (1995), 2355–2374. 1
- [2] T. E Harris, *Annals of Mathematical Statistics*, Branching processes, **19** (1948), 474–494. 1.2
- [3] N. L. Johnson, S. Kotz , A. W. Kemp, *Univariate Discrete Distributions*, Wiley, New York, (1992). 1
- [4] N. Kolev, L. Minkova, P. Neytchev, *Inflated-parameter family of generalized power series distributions and applications in analysis of over dispersed insurance data*, ARCH Research Clearing House, **2** (2000), 295–320. 1, 1.2, 2, 2, 2.4



- 
- [5] Kerala Public Service Commission, *Statewide ranked list*, 2010,2012, category number 451/07, 030/2009. [www.keralapssc.gov.in/rl-740/10](http://www.keralapssc.gov.in/rl-740/10), cat no 451/07, rl-535/12, cat no 030/2009.
- [6] M. Meghami, M. Bahrami, *Goodness of fit test for the skewed t-distribution*, J. Math. Comput. Sci., **14**, (2015), 274–283. 1, 3, 6
- [7] L. Minkova, *A generalization of the classical distributions*, Commun. Stat. Theory Methods, **31** (2002), 871–888. 1, 3
- [8] F. Momeni, *The Generalized power series distributions and their application*, J. Math. Comput. Sci., **2** (2011), 691–697.
- [9] G. P. Patil, *Certain properties of generalized power series distribution*, Annals Institute Stat. Math., **14** (1962), Tokyo, 179–182. 4.2
- [10] E. Sandhya, S. Sherly, M. K. Jos, N. Raju, *Characterizations of the extended geometric, Harris”, negative binomial and gamma distributions*, STARS Int. Journal (Sciences), **1** (2007), 5–17. 3, 3
- [11] E. Sandhya, S. Sherly, N. Raju, *Harris family of discrete distributions. Some Innovations in Statistics*, Special volume in honor of Professor T. S. K Moothathu, University of Kerala, Trivandrum (2008), 57–72. 1.2, 5.2
- [12] S. Beckett, J. Jee, T. Nebue, S. Pompilas, Q. Washington, A. Singh , N. pal, *Zero- inflated Poisson (ZIP) distributions*, parameters estimation appl. model data natural calamities. 1
- [13] T. Rolski, H. Schmidli, V. Schmidt, J. Teugels, *Stochastic Processes for Insurance and Finance*, John Wiley & Sons, Chichester, (1999).