International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 5670-5675
© Research India Publications. http://www.ripublication.com

A comparison between ML estimators for poisson, logarithmic and Nonhomogeneous Logarithmic Distribution

Ali Sakhaei[1] and Parviz Nasiri

Department of statistics, Payame Noor University, 19395-4697, Tehran, Iran.

[1]Orcid: 00000208274853

Abstract

Nonhomogeneous logarithmic distribution is used for modeling time-dependent events. In this paper we describe the theoretical features of this distribution first. Whereas using this distribution is important in special positions, estimating the parameter function with MLE method will done. Since the domain of this distribution is zero truncated of poisson distribution, we will compare the efficiency of fitting for zero truncated Poisson and logarithmic distributions with non-homogeneous logarithmic one. We will found that, the nonhomogeneous logarithmic distribution is best fitted distribution for simulated data.

Keywords: logarithmic distribution, MLE method, Nonhomogeneous distribution, Mean Square of Error (MSE).

INTRODUCTION

The logarithmic distribution (also known as log-series distribution) is a long-tailed distribution introduced by Fisher et al. (1943) in connection with data on the abundance of individuals classified by species. Also, it is first used to investigate the distribution of butterflies in the Malayan Peninsula, and data by Williams (1947) on the number of moths of different species caught in a light-trap in a specified period. The logarithmic distribution has proved valuable as a model for many kinds of data such as diverse as ecology, marketing, linguistics and meteorology. Chatfield et al (1966) used the logarithmic distribution to represent the distribution of number of items of a product purchased by a buyer in a specified time period.

Fisher's derivation (two-parametric form) of the logarithmic distribution was not well worded. Since the late 1970 Rao, Boswell and patil have endeavored to make the one-parameter distribution. Williams (1947, 1964) has been used the logarithmic distribution extensively. Shanumugam and Singh (1984) studied some characterizations of this distribution.

For simulating from logarithmic distribution, Kemp (1981) suggested and discussed about various algorithms namely LS, LB, LBM, LK and a method based on the mixed shifted-geometric model. Shanthikumar (1985) presented two interesting method namely "discrete thinning method" (for distributions with hazard rate bounded below unity) and "dynamic thinning method" (for decreasing-failure-rate distributions). Devroye (1986) presented two other, seemingly less attractive, algorithms and showed how the discrete thinning method can be used for the logarithmic distribution.

A detailed study of MLE θ is carried out by some authors. Birch (1963) has provided a computer algorithm for estimating the parameter by Newton-Raphson iteration method. Referring to last method, Bohning (1983) found maximum likelihood estimation for the logarithmic distribution numerically. A maximum likelihood approach with using computer optimization packages is done by Kemp and Kemp (1988). This method requires bounds for the parameter estimates to be specified.

Our objective is to estimate the parameter of nonhomogeneous logarithmic distribution with maximum likelihood approach and comparing the efficiency of it with zero truncated Poisson and logarithmic distributions.

This article is organized as follows: In section 2, we introduce the nonhomogeneous logarithmic distribution. In section 3, we will find the maximum likelihood estimations for parameter of nonhomogeneous logarithmic distribution numerically and then compare its efficiency with zero truncated Poisson and logarithmic distribution fitting. Finally, the paper ends with conclusions.

NONHOMOGENEOUS LOGARITHMIC DISTRIBUTION

In probability and statistics, the logarithmic distribution is a discrete probability distribution derived from the Maclaurin series expansion for $0 < \theta < 1$ as following:

$$-\ln(1-\theta) = \theta + \frac{\theta^2}{2} + \frac{\theta^3}{3} + \cdots.$$

From this we obtain the identity

$$\sum_{x=1}^{\infty} \frac{-1}{\ln(1-\theta)} \frac{\theta^x}{x} = 1 \qquad (1)$$

(See Nasiri and Esfandyarifar (2016)).

The logarithmic distribution characterized by a parameter θ is given by

$$P(X=x)=\frac{-1}{\ln(1-\theta)}\frac{\theta^x}{x} \qquad x=1,2,3,\dots . \qquad (2)$$

It is a one-parameter generalized power series distribution (GPSD) with infinite support on the positive integers.

Figure 1 displays the pmf of this distribution with $\theta=0.8$. The tail probability for $x\geq 10$ increase rapidly as θ approaches to one.

The characteristic function is

$$\varphi(t)=\frac{\ln(1-\theta e^{it})}{\ln(1-\theta)}, \qquad (3)$$

and the probability generating function (pgf) is

$$G(z)=\frac{\ln(1-\theta z)}{\ln(1-\theta)}. \qquad (4)$$

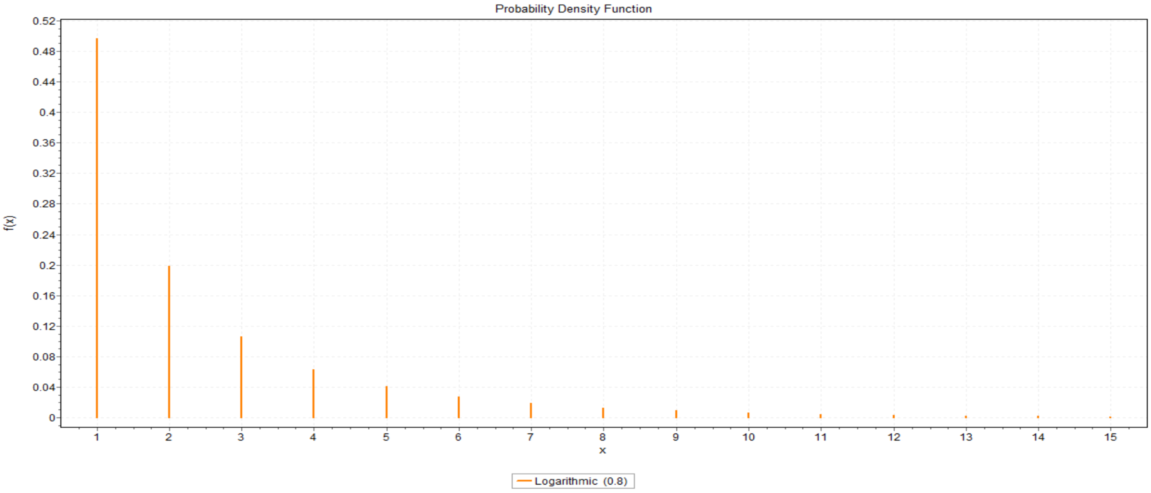

Figure 1: probability mass function of logarithmic distribution with $\theta=.8$

The mean and variance are given as

$$\mu_1'=\frac{a\theta}{1-\theta} \quad , \quad \mu_2=\frac{a\theta(1-a\theta)}{(1-\theta)^2} \qquad (5)$$

Where $a=-1/\ln(1-\theta)$.

The rth factorial moment of the logarithmic distribution is

$$\mu_r'=a\theta^r(r-1)!\,(1-\theta)^{-r}.$$

The moment generating function is

$$E[e^{tX}]=\frac{\ln(1-\theta e^t)}{\ln(1-\theta)}. \qquad (6)$$

The logarithmic distribution is a limiting form of zero-truncated negative binomial distribution. Sadinle (2008) obtained the geometric distribution with taking derivatives of the logarithmic distribution with respect to parameterθ. Following this idea, taking k derivatives at each side of probability mass fucntion and using the mathematical induction, one finds the series from which the negative binomial distribution with parameters k and θ can be obtained. The negative binomial distribution is the outcome of a number of stochastic processes, for example, the Yule–Furry process, the linear birth–death process, and the Polya process. Kendall also studied the logarithmic distribution as a limiting form of such processes, Johnson et.al.(2005).

By relaxing the assumption of homogeneity (statioanarity), the nonhomogeneous logarithmic distribution is generated. Instead of a constant parameter θ, we define$\theta(t)$, allowing the parameter to vary as a function of the variable t (e.g., time):

$$P(X=x)=\frac{-1}{\ln(1-\theta(t))}\frac{\theta(t)^x}{x} \qquad x=1,2,3,\dots \quad t\geq 0. \quad (7)$$

For practitioners, this flexibility in defining the parameter as a function of time better reflects reality. As we shall see, the sacrifice in tractability is remarkably small. $\theta(t)$ can be increasing, decreasing, bath-tub, or any other nonnegative function. This relaxation of the homogeneous data assumption allows practitioners to more accurately reflect their specific situations. For example, a restaurant could expect higher arrival rates during the lunch hour or dinner hour than during the other times of day. Figure 2 shows that the probability mass function for the Non-homogeneous logarithmic distribution with parameter $\theta(t)=\frac{1}{3}\sin\left(\frac{\pi t}{2}+\frac{3}{2}\right)$. It is observed that, when t increases, the tail of distribution will heavier than previous.

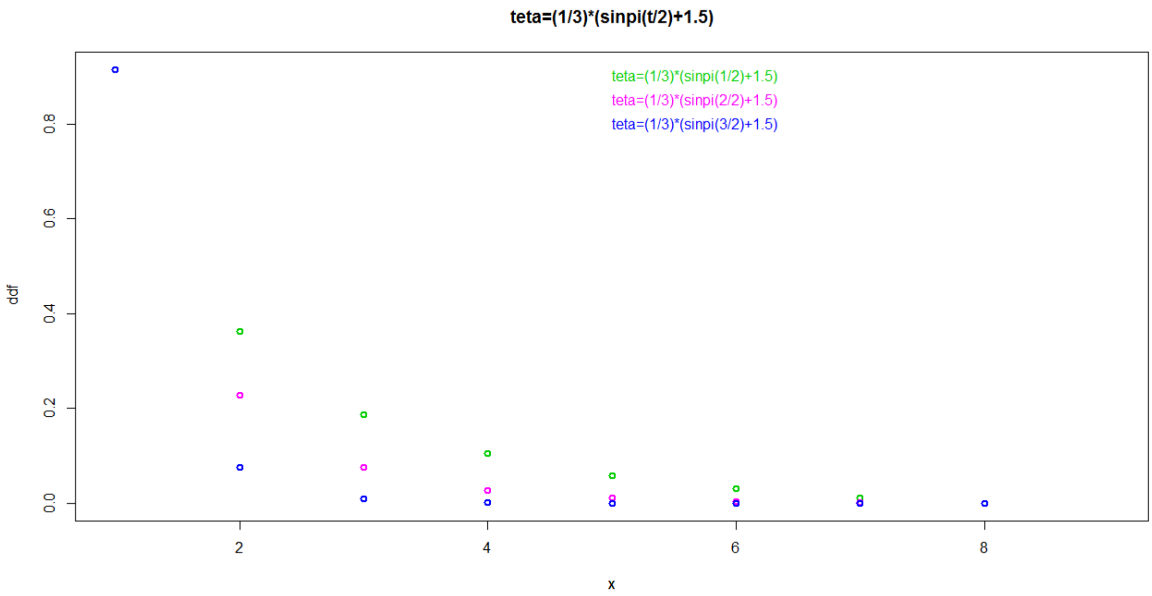

Figure 2: pmf of nonhomogeneous logarithmic distribution with $\theta(t) = \frac{1}{3}(\sin\left(\frac{\pi t}{2}\right) + \frac{3}{2})$

For generation data from this distribution for different$t$'s we simulate data from logarithmic distribution in case of homogeneity. Finally we combine the data and use them for subsequent proceedings.

MAXIMUM LIKELIHOOD ESTIMATION

There are several common accepted methods of estimating these parameters: method of moments, least squares and the method of maximum likelihood. But the last method is very popular due to favorable theoretical properties.

Suppose a random sample of size n is taken from the pmf in (2). Let the observed values be 1, 2, . . . , k with corresponding frequencies $n_1, n_2, \cdots n_k$. K is the largest observed value of x in the sample, and let $\bar{x} = \frac{\sum x\, n_x}{n}$ and $s^2 = \frac{\sum (x-\bar{x})^2 n_x}{n-1}$ be first two sample moments.

On equating the first two sample moments to the corresponding population moments, the moment estimator $\hat{\theta}$ of θ given values of n independent observations of homogeneous logarithmic distribution is given by

$$\bar{x} = \frac{\hat{\theta}}{-(1-\hat{\theta})\ln(1-\hat{\theta})} \quad (7).$$

We solve equation (7) for θ by using an iterative procedure. Because logarithmic distribution is a generalized power series distribution (PSD), maximum likelihood and moment estimation are equivalent.

For non-homogeneous case, we consider $\theta(t) = a\left(\sin\left(\frac{\pi}{2}t\right) + \frac{3}{2}\right)$ .In order to finding the ML estimate; first, we obtain the likelihood function as:

$$L(\theta(t)) = (-\ln(1-\theta(t))^{-n} \frac{\theta(t)^{\sum_i x_i}}{\prod_i x_i} x_i = 1,2,3, \ldots.$$

Hence, we have

$$\mathrm{Ln}(L(\theta(t))) = -n\ln(-\ln(1-\theta(t))) + \textstyle\sum_i x_i . \ln(\theta(t)) - \sum_i \ln(x_i).$$

Now with replacing the $\theta(t) = a(\sin\left(\frac{\pi}{2}t\right) + \frac{3}{2})$ and differentiating with respect to a and re-arranging, we have

$$\frac{n}{\left[1 - a(\sin\left(\frac{\pi}{2}t\right) + \frac{3}{2})\right][\ln(1 - a(\sin\left(\frac{\pi}{2}t\right) + \frac{3}{2}))]} = \frac{\sum x_i}{a(\sin\left(\frac{\pi}{2}t\right) + \frac{3}{2})}. \quad (8)$$

Finally with using the Newton-Raphson algorithm and considering the initial value for parameter, $\hat{a}$ will found numerically as:

$$\hat{\theta}_{n+1}(t) = \hat{\theta}_n(t) - \frac{f(\hat{\theta}_n(t))}{f'(\hat{\theta}_n(t))}. \quad (9)$$

NUMERICAL EXAMPLE

For illustrating above advices, we simulate n= 20, 50,100,200,500 and 1000 sample sizes from nonhomogeneous logarithmic distribution with$\theta(t) = \frac{1}{3}(\sin\left(\frac{\pi}{2}t\right) + \frac{3}{2})$. Then we fit them to homogeneous and nonhomogeneous logarithmic distributions. Since the data are started from 1 and poisson distribution is a good candidate for fitting these data, we included this in case of zero-truncated in our analysis.

Table 1 shows the ML estimates of θ and it's MSE for zero truncated poisson, homogeneous logarithmic and nonhomogeneous logarithmic for $t = 1$. Corresponding results for$t = 2$, $t = 3$ and $t = 4$ is appeared in Table 2, 3 and 4 respectively.

As shown in Table 1,2,3 and 4, for any t, the MSE of nonhomogeneous logarithmic for different sample sizes is lower than the MSE of two other distributions. Nevertheless, the efficiency of homogeneous logarithmic is better than zero-truncated poisson for fitting the parameter of distribution.

Table1: ML estimates and their estimated MSE for t=1 and $\theta = 0.83$

n	Zero-truncated Poisson distribution		homogeneous logarithmic distribution		nonhomogeneous logarithmic distribution	
	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\widehat{\theta(1)}$	MSE($\widehat{\theta(1)}$)
20	0.6769936	0.1342852	0.63795864	0.02730755	0.242401807	0.008503171
50	0.6758425	0.1239762	0.6486596	0.0235550	0.243956667	0.007758863
100	0.6706005	0.1182608	0.63230197	0.02841009	0.248979713	0.006716009
200	0.6593073	0.1178916	0.62773656	0.02975422	0.248111456	0.006778392
500	0.6615864	0.1105400	0.62941361	0.02913026	0.248951139	0.006600212
1000	0.6630440	0.1102951	0.62975357	0.02901535	0.248871390	0.006596173

Table 2: ML estimates and their estimated MSE for t=2 and $\theta = 0.50$

n	Zero-truncated Poisson distribution		homogeneous logarithmic distribution		nonhomogeneous logarithmic distribution	
	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\widehat{\theta(2)}$	MSE($\widehat{\theta(2)}$)
20	0.5863046	0.2116203	0.6409710	0.0455621	0.412788875	0.008847346
50	0.5792453	0.2130411	0.6412031	0.0458723	0.409097578	0.007080741
100	0.5836834	0.2103294	0.6394023	0.0337516	0.414919252	0.007729573
200	0.5901273	0.1923046	0.6412048	0.0407542	0.412943100	0.007066197
500	0.5892401	0.1928470	0.6400101	0.0301302	0.415158557	0.007335519
1000	0.5834403	0.1900260	0.6384998	0.0310153	0.414783209	0.007226907

Table 3: ML estimates and their estimated MSE for t=3 and $\theta = 0.17$

n	Zero-truncated Poisson distribution		homogeneous logarithmic distribution		nonhomogeneous distribution	
	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\widehat{\theta(3)}$	MSE($\widehat{\theta(3)}$)
20	0.2384200	0.3792710	0.4591024	0.0723127	1.20860710	0.01180436
50	0.2389015	0.3771034	0.4637222	0.0693471	1.23314239	0.02070211
100	0.2461082	0.3699123	0.4581023	0.0711212	1.23954163	0.02578102
200	0.2217029	0.3728110	0.4600128	0.0731067	1.24764575	0.01720768
500	0.2301283	0.3700028	0.4644830	0.0712239	1.24917952	0.01613925
1000	0.2371042	0.3710200	0.4599990	0.0702681	1.24336869	0.02386753

Table 4: ML estimates and their estimated MSE for t=4 and $\theta = 0.50$

n	Zero-truncated Poisson distribution		homogeneous logarithmic distribution		nonhomogeneous logarithmic distribution	
	$\hat{\theta}$	MSE($\hat{\theta}$)	$\hat{\theta}$	MSE($\hat{\theta}$)	$\widehat{\theta(4)}$	MSE($\widehat{\theta(4)}$)
20	0.3227916	0.1342852	0.8412301	0.1072254	0.245548789	0.007823795
50	0.3677813	0.1239762	0.8012634	0.1100277	0.247337412	0.007168825
100	0.3004517	0.1182608	0.7983933	0.0942615	0.248827052	0.006727498
200	0.3492833	0.1178916	0.7812942	0.0924401	0.247502310	0.006895462
500	0.3118810	0.1105400	0.8134023	0.0900147	0.247650504	0.006812923
1000	0.3623023	0.1102951	0.8204500	0.0913472	0.249167895	0.006548589

CONCLUSIONS

In this paper, we introduced the nonhomogeneous logarithmic distribution. This distribution is used for modeling the time dependent events. We calculated the likelihood function and then maximum likelihood estimates for parameter function are found numerically. In order to considering the accuracy of estimates, we compared the MSE of these estimates for zero truncated Poisson, homogeneous logarithmic and nonhomogeneous logarithmic. It is observed that for simulated data the nonhomogeneous logarithmic distribution is better fit than two other distributions.

REFERENCES

[1] Birch, M. W. (1963). An algorithm for the logarithmic series distribution, *Biometrics*, **19**, 651-652.

[2] Bӧhning, D. (1983). Maximum likelihood estimation of the logarithmic series distribution, *Statistische Hefte*, **24**, 121-140.

[3] Chatfield, C., Ehrenberg, A. S. C. & Goodhardt, G. J. (1966), Progress on a simplified model of stationary purchasing behavior (with discussion), *Journal of The Royal Statistical Society*, *Series A*, **129**, 317-367.

[4] Devroye, L. (1986). *Non-Uniform Random Variate Generation*, New York: Springer-Verlag.

[5] Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943), the relation between the number of species and the Number of individuals in a random sample of an animal population. *J. Animal Ecology,* **12**, 42-58.

[6] Johnson, N. L., Kotz, S. & Kemp, A. W. (2005), *Univariate discrete distributions*, 3rd ed. *John Wiley & Sons inc*; Hoboken, New Jersey.

[7] Kemp, A. W. (1981). Efficient generation of logarithmically distributed pseudorandom variables, *Applied Statistics*, **30**, 249‑253.

[8] Kemp, C. D., and Kemp, A. W. (1988). Rapid estimation for discrete distributions, *Statistician,* **37**, 243‑255.

[9] Nasiri, P. and Esfandyarifar, H., (2016). E-Bayesian Estimation of the Parameter of the Logarithmic Series Distribution, Journal of Modern Applied Statistical Methods, **15**, 2, 643-655**.**

[10] Sadinle, M. (2008), "Linking the negative binomial and logarithmic series distributions via their associated Series". *Revista Colombiana de Estadistica*, **31**, 311-319.

[11] Shanthikumar, J. G. (1985). Discrete random variate generation using uniformization, *European Journal of Operational Research*, **21**, 387‑398.

[12] Shanumugam, R. & Singh J. (1984), "A characterization of the logarithmic series distribution and its applications." *Communications in Statistics*, **13**, 865-875.

[13] Williams, C. B. (1947). The logarithmic series and its application to biological problems, Journal of Ecology, **34**, 253‑272.

[14] Williams, C. B. (1964). *Patterns in the Balance of Nature*, London: Academic Press.