Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/060213

Sequential Analysis of Zero Truncated Geometric Distribution

Bhagwati Devi, Rahul Gupta and Parmil Kumar*

Department of Statistics, University of Jammu, Jammu 180006, Jammu and Kashmir, India

Received: 14 Oct. 2016, Revised: 18 May 2017, Accepted: 23 May 2017
Published online: 1 Jul. 2017

Abstract: The Geometric distribution is one of the important distributions in the real life situation specifically in reliability/survival analysis and queuing theory. One of the important property of this distribution is the lack of memory property and in case of its truncation, situations arise practically in cases where the ability to record, or even to know about, occurrences is limited to values which lie above or below a given threshold or within a specified range. For example, if the date of occurrence of an infectious disease in a patient is examined, this would typically be subject to truncation relative to those of all patients suffering from that infectious disease in the area given that the doctor starts the treatment only in a given age range on a specific time. This paper aims at studying the truncated geometric distribution from sequential point of view and proposes sequential testing procedure for its parameter of interest and also study its OC and ASN functions, both theoretically and graphically.

Keywords: Probability distribution, truncation, geometric distribution, OC, ASN

1 Introduction

The Geometric distribution arises in many real life situations. Most commonly, we use this distribution to model the probability of getting first success when we perform independent Bernoulli trials. Also, it can be seen as the distribution of the total size of a population in a pure birth process and we can also find it as a limiting distribution of the size of the queue in a $M|M|1$ model. One of the important property of this distribution is the lack of memory property which plays an important part in the branch of applied probability. Many characterization of Geometric distribution can be found in the literature, mostly based on the independence of functions of order Statistics or on record values, see e.g [6], [1],[4], [10] and [5]. Xie and Goh proved that control charts based on Geometric distribution have shown to be useful when this is a better approximation of the underlying distribution than Poisson distribution [11]. They also discussed the use of Geometric distribution for process control of high yield processes. Shanbhag gave a characteristic property of the Geometric distribution based on the means of the conditional distribution [7].

In this paper, Zero truncated Geometric distribution has been studied by sequential point of view. We have developed a sequential testing procedure for the parameter θ of the Zero truncated Geometric distribution and also obtained its OC and ASN function. We have plotted the Graphs by using MatLab and analysis has been made by examining the graphs of OC and ASN.

2 Sequential Testing Procedure for Zero Truncated Geometric Distribution

Sequential analysis is a statistical analysis where the sample size is treated as a random variable and is not fixed in advance as is considered in the other testing procedures. In sequential analysis, we keep on drawing the random sample at every stage of the sampling, test our sample, draw conclusion and sampling is stopped in accordance with a pre-defined stopping rule as soon as significant results are observed. Thus, we may arrive at a conclusion at a much earlier stage than would be possible with more classical testing procedure at consequently lesser financial or human cost.

* Corresponding author e-mail: rahulgupta68@yahoo.com

A sequential test typically continues until the evidence strongly favors one of the two hypothesis. Sequential analysis was first developed during World War II by Abraham Wald with Jacob Wolfowitz, W. Allen Wallis, and Milton Friedman while at Columbia University's Statistical Research Group for using it in the field of Statistical quality control. Other contributions to this field were made by Arrow *et al.* [2]. Primarily, the purpose of adopting the sequential procedures was only to minimize the average sample number so as to save time, money and man power. Later on, researchers faced the situations that many inferential problems could not be tackled with the help of the fixed sample size procedures like the given precision problems for which only sequential methods provided solutions. These days, Sequential techniques are finding many applications in the field of Multivariate Problems, Multi hypothesis Problems, Confidence Sets and Intervals, Adaptive Non-Parametric Procedures, Ranking and Selection Problems, Secretary Problems, Decision Theory, Stochastic Approximation, Detection and Change-Point Problems, Quality Control, Clinical Trials, Reliability and Life Testing.

A random variable X is said to possess Geometric distribution if its pdf is of the form

$$f(x) = \theta(1-\theta)^x; x = 0,1,2\ldots,0 \le \theta < 1 \tag{1}$$

Then the pdf of zero truncated Geometric distribution is given by

$$g(x;\theta) = \theta(1-\theta)^{x-1}; x = 1,2,3\ldots,0 \le \theta < 1 \tag{2}$$

Consider a sequence of observations $X_1, X_2, \ldots X_n$ from (2) if we have to test the simple null hypothesis $H_0 : \theta = \theta?_0$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$ then SPRT for testing H_0 against H_1 will be

$$Z_i = ln\{\frac{p(x_i;\theta_1)}{p(x_i;\theta_0)}\}$$

$$= ln(\frac{\theta_1}{\theta_0}) + (x_i - 1)ln(\frac{1-\theta_1}{1-\theta_0}) \tag{3}$$

And we choose two numbers A and B such that $0 < B < 1 < A$. At the n^{th} stage of sampling, accept H_0 if $\sum_{i=1}^{\infty} z_i \le lnB$, reject H_0 if $\sum_{i=1}^{\infty} z_i \ge lnA$ and continue sampling by taking $(n+1)^{th}$ observation if $lnB < \sum_{i=1}^{\infty} Z_i < lnA$. If $\alpha \in (0,1)$and $\beta \in (0,1)$ are type I and type II errors respectively, then

$$A \approx \frac{1-\beta}{\alpha}$$

and

$$B \approx \frac{\beta}{1-\alpha}$$

The OC function of above SPRT is approximately given by

$$L(\theta) \approx \frac{A^h - 1}{A^h - B^h}$$

where 'h'is the non zero solution of the equation $E[e^{hz}] = 1$

$$or \quad \left(\frac{\theta_1}{\theta_0}\right)^h (\frac{1-\theta_0}{1-\theta_1})^h E[(\frac{1-\theta_1}{1-\theta_0})]^{xh} = 1 \tag{4}$$

For a real t, we have,

$$E[t]^x = \sum_x^{\infty} \theta(1-\theta)^{(x-1)} t^x$$

$$= \frac{1}{1-\theta}[\frac{\theta}{1-(1-\theta)t} - \theta]$$

$$= \frac{\theta}{1-\theta}[\frac{(1-\theta)t}{1-(1-\theta)t}] \tag{5}$$

Putting $t=(\frac{1-\theta_1}{1-\theta_0})^h$ in the above equation we have,

$$E\left[\left(\frac{1-\theta_1}{1-\theta_0}\right)^h\right]^x = \frac{\theta}{1-\theta}\left[\frac{(1-\theta)(\frac{1-\theta_1}{1-\theta_0})^h}{1-(1-\theta)(\frac{1-\theta_1}{1-\theta_0})^h}\right]$$

From (4), it can be easily deduced that ,

$$(\frac{\theta_1}{\theta_0})^h(\frac{1-\theta_0}{1-\theta_1})^h(\frac{\theta}{1-\theta})[\frac{(1-\theta)(\frac{1-\theta_1}{1-\theta_0})^h}{1-(1-\theta)(\frac{1-\theta_1}{1-\theta_0})^h}]=1 \tag{6}$$

Taking natural log on both sides of above equation, using the expression $ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\ldots\forall x\in(0,1)$ and neglecting the second and higher order terms we get,

$$h\approx\frac{-ln(\frac{\theta}{1-\theta})-ln?(1-\theta)}{ln(\frac{\theta_1}{\theta_0})+ln(\frac{1-\theta_0}{1-\theta_1})+ln(\frac{1-\theta_1}{1-\theta_0})} \tag{7}$$

The ASN function of the above SPRT is given by

$$E_\theta(N)=\frac{L(\theta)lnB+[1-L(\theta)]lnA}{E(z)}$$

where $L(\theta)$ is the OC function.

Next , we obtain the expression for E[Z]. As (5) represents probability generating function of X, Therefore,

$$\begin{aligned}E[x]&=\frac{\partial E[t^x]}{\partial t}|_{t=1}\\&=(\frac{\theta}{1-\theta})\frac{\partial}{\partial t}[\frac{\theta}{1-(1-\theta)t}-1]\\&=\frac{\theta}{\{1-(1-\theta)t\}^2}\end{aligned}$$

Putting $t=1$, we get

$$E[x]=\frac{\theta}{\{1-(1-\theta)\}^2}$$

We have from (3),

$$E[z]=ln(\frac{\theta_1}{\theta_0})+E[x]ln(\frac{1-\theta_1}{1-\theta_0})-ln(\frac{1-\theta_1}{1-\theta_0})$$

and using the value of $E[x]$, we have

$$E[z]=ln(\frac{\theta_1}{\theta_0})+\frac{\theta}{\{1-(1-\theta)\}^2}ln(\frac{1-\theta_1}{1-\theta_0})-ln(\frac{1-\theta_1}{1-\theta_0})$$

Remark : *Taking the value of $\alpha=.5$, $\beta=.6$ and for varying values of θ we have plotted the OC and ASN functions and there behavior can be seen from Figure 1 and Figure 2. It can be seen from the OC plot that as the value of θ approaches to 1 , their is a sharp decline in the associated OC function value. And in case of ASN, the value also decreases as θ approaches to 1. So, we can say that with the increasing value of the parameter, the value of OC and ASN functions of the distribution goes on decreasing. In nutshell, we can say that the probability of acceptance decreases with increasing value of parameter which infers that the procedure is not good for value of θ close to 1. But the sample number, on average, required to arrive at a right decision, will be less as θ approaches to 1.*

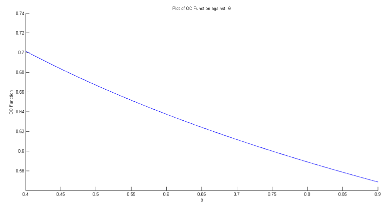

Fig. 1: Plot of OC Function Against Value of θ

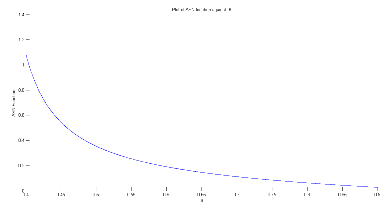

Fig. 2: Plot of ASN Function Against Value of θ

3 Conclusion

In this paper, the sequential testing procedure has been developed for the Zero truncated Geometric distribution. OC and ASN expressions have also been developed. It has been proved graphically that the Operataing Characteristics (OC) and Average Sample Number (ASN) function value decreases as the parameter value approaches to its upper limit.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] B. C Arnold, Journal of Applied Probability **17**, 570-573 (1980).
[2] B. C Arnold and M. Ghosh, Scandinavian Actuarial Journal.,**4**, 232-234 (1976).
[3] K.J Arrow, D. Blackwell and M.A Girshick, Econometrica, **17**, 213-244.
[4] E. El-Neweihi, and Z. Govindarajulu, Journal of Statistical Planning and Inference, **3**, 85-90 (1979).
[5] J. Galambos, In Statistical Distributions in Scientific Work, **3** (1975).
[6] Z. Govindarajulu , Journal of Statistical Planning and Inference, **4**, 237-247 (1980).
[7] D. N Shanbhag, Journal of American Statistical Association, **65**, 1256-1259 (1970).
[8] R. C Srivastava, Journal of American Statistical Association, **69**, 267-269 (1974).
[9] R. C Srivastava, Sankhya, Series B, **40**, 276-278 (1979).
[10] A. Wald, John Wiley and Sons, New York (1947).
[11] M. Xie and T.N Goh, International Journal of Quality and Reliability Management, **14** , 64-73 (1997).

Bhagwati Devi received the M.Phil. degree in Statistics at University of Jammu, India and currently pursuing Ph.D. from the same university. Her research interests are in the fields of Statistical Inference including Bayesian Analysis, Sequential Techniques, Inferential Reliability and Applied Statistics. She has published research articles in reputed national and international journals.

Rahul Gupta is Professor of Statistics at University of Jammu, Jammu and Kashmir, India . His areas of interest are Sequential Analysis, Bayesian Inference and Bio Statistics in which he has produced about 15 M.Phil/ Ph.D students and about 45 research publications in journals of international repute. Presently he is also Head of the Department and apart from research activities is involved in administrative work of the University of Jammu which is NAAC accredited A+ University.

Parmil Kumar is Assosciate Professor of Statistics at University of Jammu, Jammu and Kashmir, India . He received his Ph.D. degree from CCS HAU, Hisar, Haryana, India. He has published research articles in reputed national and international journals. His areas of interest are Optimization Techniques, Applied Statistics and Probability Theory. Apart from research activities, he is also involved in administrative work of the University of Jammu. At present, he is serving as Assistant Dean, Department of Student Welfare, University of Jammu. He is referee and editor of several journals.