International Journal of Advance Research, IJOAR .org
Volume 1, Issue 9, September 2013, Online: ISSN 2320-9143

A STUDY OF THE SEQUENTIAL TEST FOR THE PARAMETER OF PARETO DISTRIBUTION

Surinder Kumar1 and Mukesh Kumar

Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, India-226025
E-mail: surinderntls@gmail.com

ABSTRACT

Sequential probability ratio test is developed for testing the hypothesis regarding the parameter of Pareto distribution. The expression for the operating characteristics (OC) and average sample number (ASN) functions are derived. For the purpose of the plotting the OC and ASN functions different approaches are used. These different approaches give quite satisfactorily results.

Key words: Pareto distribution, SPRT, OC and ASN functions, Newton-Raphson method.

1. INTRODUCTION

The theory of sequential analysis is begun in about 1943 with the work of A. Wald and G.A. Barnard. Sequential analysis has been heavily dominated by the sequential probability ratio test (SPRT). SPRT for testing a simple hypothesis against a simple alternative is first time given by Wald (1947). He also obtained the expression for the Operating Characteristics (OC) and Average Sample Number (ASN) functions. The SPRT has been studied by various authors in order to test the simple and composite hypothesis, related to various probabilistic models.

For a few citations, one may refer to, Epstein and Sobel (1955) who applied the SPRT for testing the simple hypothesis regarding the mean of a one parameter negative exponential distribution, Johnson (1966) dealt with the problem of the testing hypothesis regarding the scale parameter of the Weibull distribution when the shape parameter is known, Phatarfod (1971) devolved the SPRT for testing composite hypothesis for shape distribution of gamma distribution, when the scale parameter is known, Joshi and Shah (1990) developed SPRT for testing a simple hypothesis (against a simple alternative) for the mean of an inverse Gaussian distribution, assuming the coefficient of variation (CV) to be known, Chaturvedi, Kumar and Kumar (2000) devolved sequential test of simple and composite hypothesis regarding the parameters of class of distributions representing various life testing models and Surinder and Naresh (2009) study the robustness of the sequential test for the parameters of an Inverse Gaussian distribution with known coefficient of variation .

2. SET-UP OF THE PROBLEM:

Let the random variable (r.v.) follows the Pareto distribution presented by the probabilistic density function (p.d.f.)

$$f(x,\theta) = \frac{\theta\, a^{\theta}}{x^{\theta+1}} \qquad \alpha \le x < \infty,\ \theta \ge 0, \quad \alpha \ge 0 \qquad \text{--- (2.1)}$$

Given a sequence of observation X_1, X_2, X_3… from (2.1), suppose one wish to test the simple null hypothesis $H_0 : \theta = \theta_0$against simple alternative $H_1 : \theta = \theta_1$. The expression for the OC and ASN function are obtained and their behaviour is studied by graph plotting.

3. SPRT FOR TESTING THE HYPOTHESIS REGARDING θ:

The SPRT for testing H_0 is defining as follows

$$Z_i = \ln f \left\{\frac{X_{i:}\ \theta_{1,}\ \alpha}{X_{i:}\ \theta_0,\ \alpha}\right\} \qquad \text{--- (3.1)}$$

$$Z_i = \ln\left(\frac{\theta_1}{\theta_0}\right) + (\theta_1 - \theta_0)\ln\alpha + (\theta_0 - \theta_1)\ln x \qquad \text{--- (3.2)}$$

$$e^{Z_i} = \left(\frac{\theta_1}{\theta_0}\right) \alpha^{\theta_1 - \theta_0} X^{\theta_0 - \theta_1} \qquad \text{--- (3.3)}$$

We choose two numbers A and B such that 0<B<1<A. At the nth stage, accept H_0 if $\sum_1^n Z_i \leq lnB$, reject H_0 if $\sum_1^n Z_i \leq lnA$, otherwise continue sampling by taking the (n + 1)th observation. If α ϵ (0, 1) and β ϵ (0, 1) are type 1st and type 2nd errors respectively, then according to Wald (1947), A and B are approximately given by

$$A = \frac{(1-\beta)}{\alpha}, \qquad B = \frac{\beta}{(1-\alpha)}. \qquad \text{--- (3.4)}$$

The OC function of the SPRT is given by

$$L(\theta) = \frac{A^h - 1}{A^h - B^h}, \qquad \text{--- (3.5)}$$

where h is the non –zero solution of the equation

$E\,[e^{Z_i}]$ =1,

or,

$$\int_0^\infty \left[f \left\{\frac{X_{i:}\ \theta_{1,}\ \alpha}{X_{i:}\ \theta_0,\ \alpha}\right\}\right]^h f(x,\ \theta,\ \alpha)dx = 1. \qquad \text{--- (3.6)}$$

From (2.1) and (3.3) since

$$E\,[e^{Z_i}] = \frac{\theta\left(\frac{\theta_1}{\theta_0}\right)^h}{(h\theta_1 + \theta - \mathrm{h}\theta_0)} \qquad \text{--- (3.7)}$$

The ASN function is approximately given by

$$E\,(N/\theta) = \frac{[\,L(\theta)\ln B + \{1 - L(\theta)\}\ \ln A\,]}{E(Z)}, \qquad \text{--- (3.8)}$$

Provided $E(Z) \neq 0$, where

$$E(Z) = \ln\ \left[\frac{\theta_1}{\theta_0}\right] + \frac{(\theta_0 - \theta_1)}{\theta} \qquad \text{--- (3.9)}$$

From (3.9) ASN function under H_0 and H_1 is respectively given by

$$E_0\,(N) \approx \frac{[(1-\alpha)\ \mathrm{lnB} + \alpha\ \mathrm{lnA}]}{\ln\left[\frac{\theta_1}{\theta_0}\right] + \frac{(\theta_0-\theta_1)}{\theta}}\ , \qquad \text{--- (3.10)}$$

and

$$E_1\,(N) \approx \frac{[\beta\ \mathrm{lnB} + (1-\beta)\ \mathrm{lnA}]}{\ln\left[\frac{\theta_1}{\theta_0}\right] + \frac{(\theta_0-\theta_1)}{\theta}} \qquad \text{--- (3.11)}$$

Now in order to plot OC and ASN functions some methods provided below.

4. SIMULATION METHOD:

Consider the equation (3.5) and (3.7). For any arbitrary value 'h' the point [θ, L (θ)] computed from (3.7) and will be a point on the OC function. The OC function can be draw by plotting a sufficiently large number of points [θ, L (θ)] corresponding to various value of 'h' (see Wald 1947, p.51).

Remarks 1: We considered the testing of hypothesis H_0 : θ_0 = 20 verses H_1 : θ_1 = 25 fixing α = β = .05. For equation (3.5) and (3.7) various pairs of [θ, L (θ)] are derived through varying 'h' between the intervals (-2, 2). The interval is chosen in such a ways, so that the parameter θ lies within the parametric space. The results are presented in **Tables 1**. It is evident from the table that the value of L (θ_0) and L (θ_1) are quite close to their theoretical values i.e. 0.95 and .05 respectively. The OC function curve and ASN function curve are plotted in **Fig. 1** and **Fig. 2**.

5. APPROXIMATION METHOD:

Consider the equation (3.7) and taking the logarithms of both sides, using the expression

Log (1+x), where |x|≤1, we get

$$2d^3\,h^2 - 3cd^2h + 6c^2d - 6c^3\log\,(b) = 0, \qquad \text{--- (5.1)}$$

where $= \left(\frac{\theta_1}{\theta_0}\right),\ c = \theta, d = (\theta_1 - \theta_0)$, and root are given by

$$h = \frac{3cd^2 \pm \sqrt{[48c^3d^3\log(\mathrm{b}) - 39c^2d^4]}}{4d^3} \qquad \text{--- (5.2)}$$

Remarks 2: For testing the testing of hypothesis H_0 : θ_0=20 verses H_1 : θ_1 = 25 fixing α = β = .05 for different value of θ, the real roots of 'h' obtained from (5.2). The OC and ASN functions are evaluated with the help of (3.5) and (3.8) respectively. It is interesting to note that the value of 'h' obtained though approximation are given quite satisfactory results (see **Table 1**). The OC and ASN functions are plotted in **Fig. 3** and **Fig.4**.

6. NEWTON-RAPHSON METHOD:

We use the value of ‘h’ given by (5.2) as initial values for solving (3.7) through N-R method. To apply the N-R method for solving the (3.6), we have written.

$$F = cd^{h} - hb + c \qquad \text{--- (6.1)}$$

$$FD = \log(d) - \left(\frac{\frac{b}{c}}{1+\frac{hb}{c}}\right), \qquad \text{--- (6.2)}$$

where FD is the first derivative of F.

The ASN function is approximately given by

$$E\,(N/\theta) = \frac{[L\,(\theta)\ln B + \{1 - L\,(\theta)\}\ln A\,]}{E(Z)}, \qquad \text{--- (6.3)}$$

where $E(Z) = \ln\left[\frac{\theta_1}{\theta_0}\right] + \frac{(\theta_0 - \theta_1)}{\theta}$ --- (6.4)

It is interesting to note the value of ‘h’ obtained though approximation and though N-R method (by taking the approximation value of ‘h’ as initial value for N-R method) are quit close (see **Tables 3**). This justify the use of approximation.

Table 1: Values of OC and ASN function obtained by simulation method for testing H_0 : θ = 20 against H_1 : θ = 25, α=β=.05

h	L(θ)	θ	E(N)
-2.0	0.002762	27.7778	67.8704
-1.9	0.003705	27.4917	70.8158
-1.8	0.004967	27.2075	74.0442
-1.7	0.006656	26.9252	77.5898
-1.6	0.008915	26.6446	81.4898
-1.5	0.01193	26.3659	85.7844
-1.4	0.01595	26.089	90.5149
-1.3	0.021295	25.814	95.7219
-1.2	0.028379	25.5408	101.4415
-1.1	0.037728	25.2695	107.6994
-1.0	**0.051421**	**25.0**	114.5025
-0.9	0.065989	24.7324	121.8267

-0.8	0.086626	24.4666	129.6012
-0.7	0.112935	24.2026	137.6892
-0.6	0.145958	23.9406	145.8697
-0.5	0.186606	23.6803	153.8217
-0.4	0.235452	23.422	161.1252
-0.3	0.292488	23.1655	167.2835
-0.2	0.35689	22.9108	171.7789
-0.1	0.426916	22.658	174.1567
0.1	0.573084	22.158	171.5841
0.2	0.64311	21.9108	166.7437
0.3	0.707512	21.6655	159.982
0.4	0.764548	21.422	151.8167
0.5	0.813395	21.1803	142.7949
0.6	0.854042	20.9406	133.4129
0.7	0.887065	20.7026	124.0713
0.8	0.913374	20.4666	115.0581
0.9	0.934011	20.2324	106.5585
1	**0.95013**	**20.0000**	98.6726
1.1	0.962272	19.7695	91.4387
1.2	0.971621	19.5408	84.8529
1.3	0.978705	19.314	78.8853
1.4	0.98405	19.089	73.4916
1.5	0.98807	18.8659	68.6212
1.6	0.991085	18.6446	64.222
1.7	0.993344	18.4252	60.2441
1.8	0.995033	18.2075	56.6408
1.9	0.996295	17.9917	53.3698

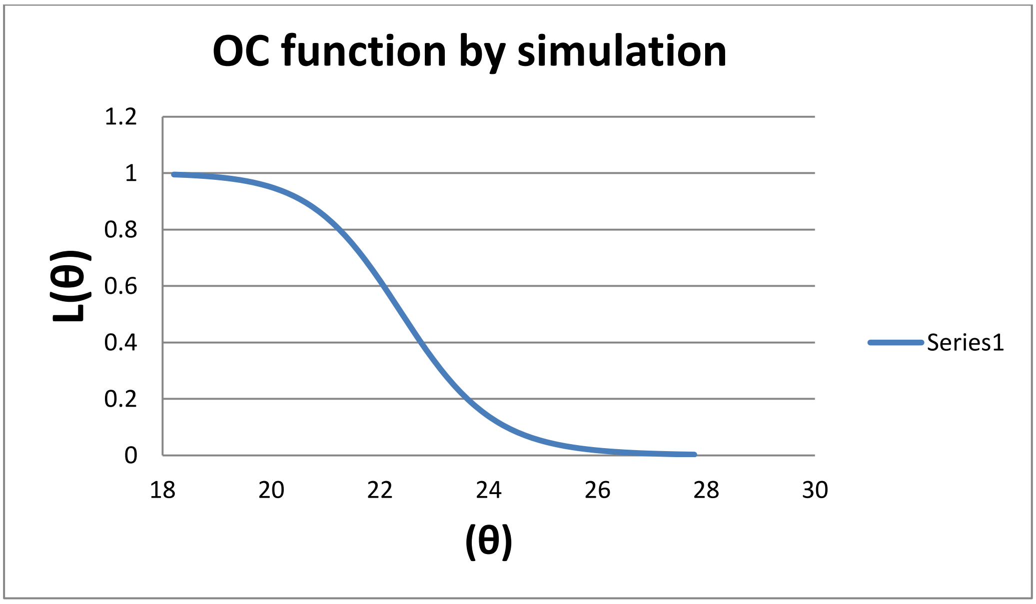

Fig: 1

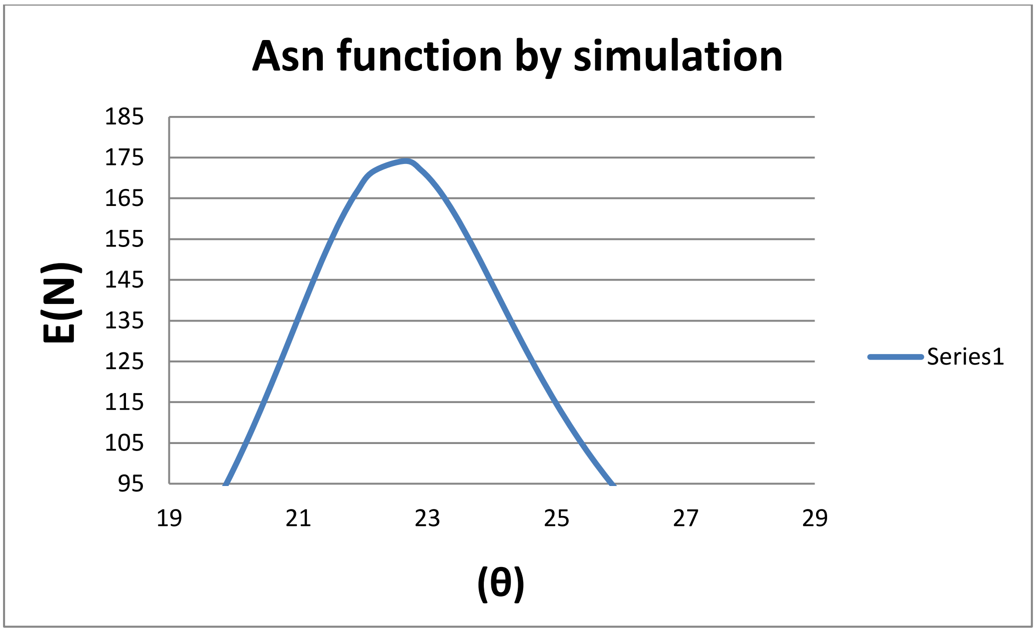

Fig: 2

Table 2: Values of OC and ASN function obtained by approximation method for testing

$H_0 : \theta = 20$ against $H_0 : \theta = 25$, $\alpha = \beta = .05$.

θ	L(θ)	θ	E(N)
19.0	0.9914	19.0	72.31
19.2	0.9873	19.2	76.99
19.4	0.982	19.4	82.07
19.6	0.9752	19.6	87.55
19.8	0.9664	19.8	93.47
20.0	**0.9552**	20.0	99.82
20.2	0.9413	20.2	106.59
20.4	0.924	20.4	113.74
20.6	0.9029	20.6	121.21
20.8	0.8774	20.8	128.9
21.0	0.847	21.0	136.66
21.2	0.8113	21.2	144.3
21.4	0.7704	21.4	151.61
21.6	0.7242	21.6	158.32

21.8	0.6732	21.8	164.15
22.0	0.6184	22.0	168.85
22.2	0.5609	22.2	172.21
22.4	0.5021	22.4	174.08
22.6	0.4436	22.6	174.39
22.8	0.3869	22.8	173.2
23.0	0.3333	23.0	170.62
23.2	0.2839	23.2	166.86
23.4	0.2393	23.4	162.14
23.6	0.1998	23.6	156.73
23.8	0.1655	23.8	150.84
24.0	0.1361	24.0	144.71
24.2	0.1112	24.2	138.49
24.4	0.0904	24.4	132.34
24.6	0.0732	24.6	126.36
24.8	0.059	24.8	120.62
25.0	**0.0474**	25.0	115.15
25.2	0.038	25.2	110
25.4	0.0304	25.4	105.17
25.6	0.0243	25.6	100.66
25.8	0.0194	25.8	96.45

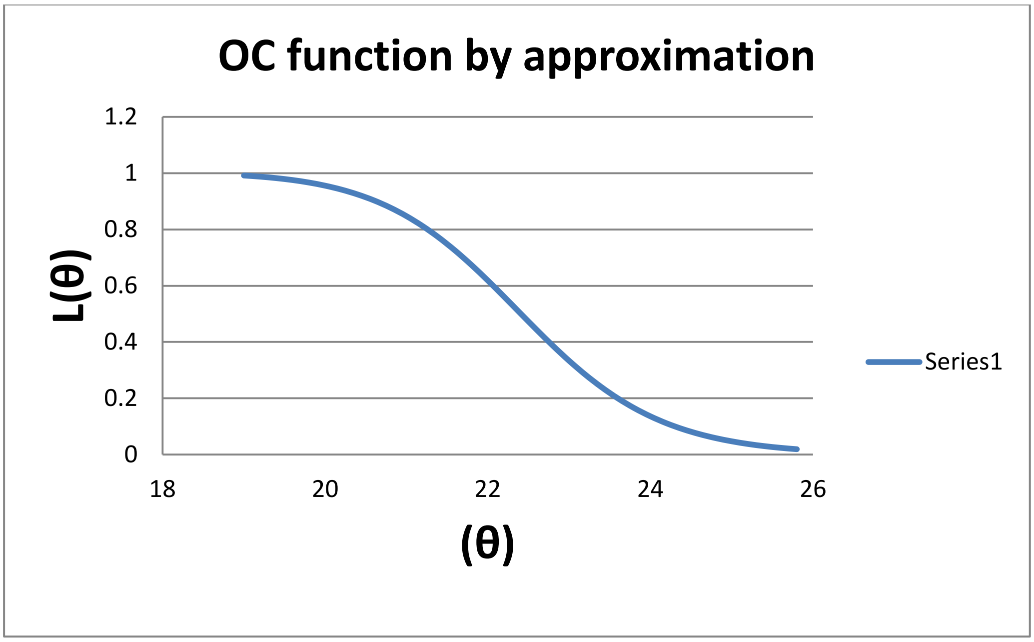

Fig: 3

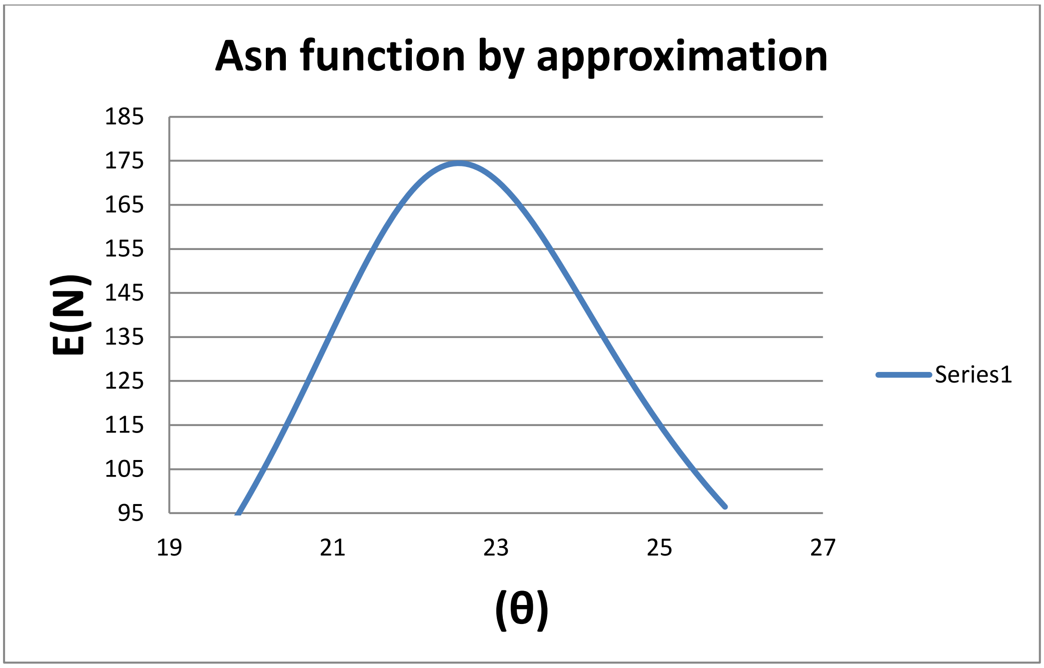

Fig: 4

Table 3: Values of OC and ASN function obtained by Newton- Raphson method

$H_0 : \theta = 20$ against $H_1 : \theta = 25, \alpha = \beta = .05$

θ	L(θ)	θ	E(N)
19.0	0.9925	19.0	72.31
19.2	0.9893	19.2	76.99
19.4	0.9979	19.4	82.07
19.6	0.9788	19.6	87.55
19.8	0.9563	19.8	93.47
20.0	**0.9552**	20.0	99.82
20.2	0.9403	20.2	106.59
20.4	0.924	20.4	113.74
20.6	0.9129	20.6	121.21
20.8	0.8874	20.8	127.9
21.0	0.885	21.0	137.66
21.2	0.8313	21.2	145.3
21.4	0.7904	21.4	152.61
21.6	0.7342	21.6	159.32
21.8	0.6832	21.8	163.15

22.0	0.6184	22.0	167.85
22.2	0.5759	22.2	171.21
22.4	0.5351	22.4	172.08
22.6	0.4436	22.6	173.39
22.8	0.3969	22.8	174.2
23.0	0.3483	23.0	171.62
23.2	0.2989	23.2	167.86
23.4	0.2563	23.4	161.14
23.6	0.1908	23.6	157.73
23.8	0.1785	23.8	155.84
24.0	0.1461	24.0	143.71
24.2	0.1212	24.2	139.49
24.4	0.0904	24.4	131.34
24.6	0.0782	24.6	127.36
24.8	0.056	24.8	120.62
25.0	**0.0464**	25.0	115.15
25.2	0.048	25.2	111
25.4	0.0314	25.4	106.17
25.6	0.0253	25.6	101.66
25.8	0.0204	25.8	97.45

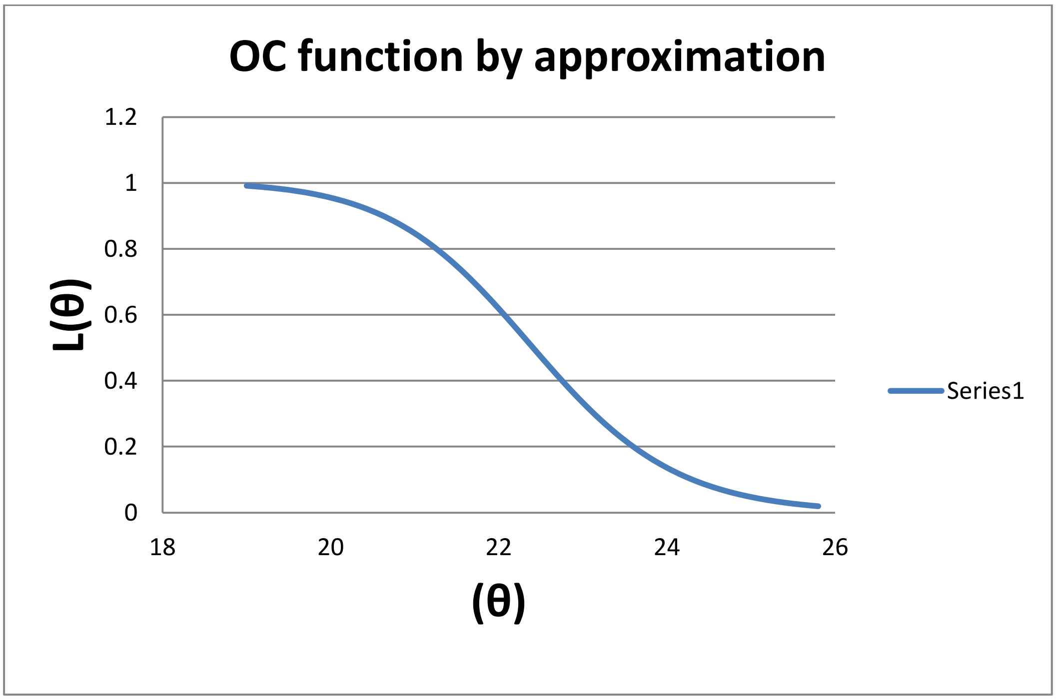

Fig. 4

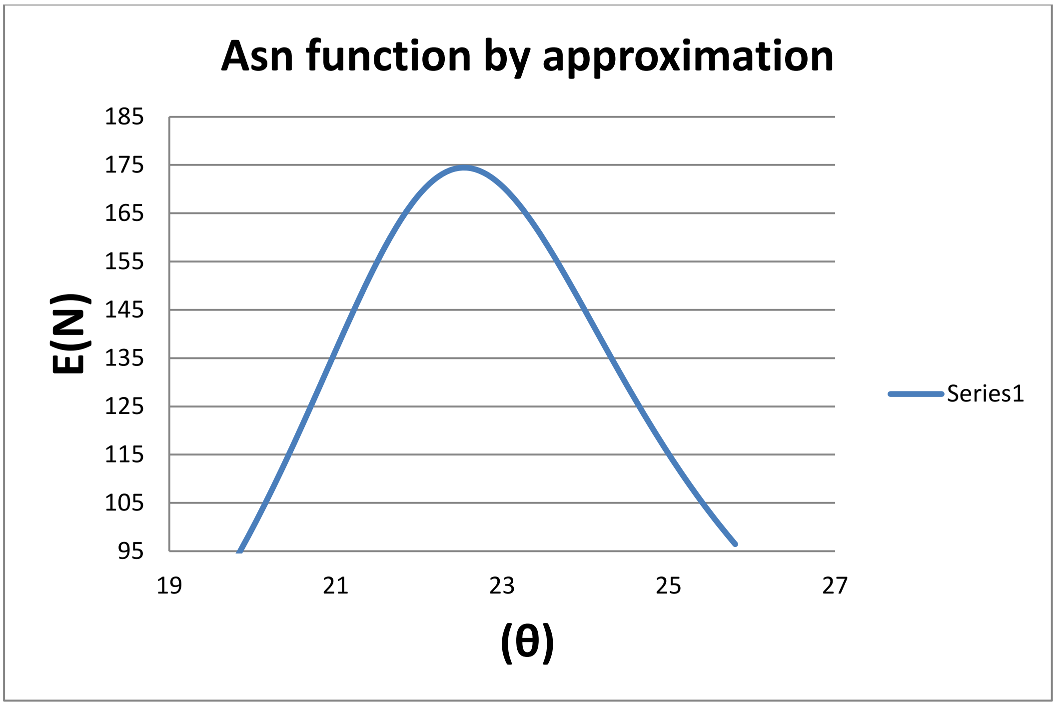

Fig. 6

References:

1. **Wald, A.** (1947): Sequential Analysis. *John Willy and sons*, New York.
2. **Chaturvedi, A., Kumar and Surinder k.** (2000): Sequential testing procedures for a class of distributions representing various life testing models. *Statistical paper,* **41**, 65-84.
3. **Epstein, B. and Sobel, M.** (1955): Sequential life test in exponential case. *Ann. Math .Statist.* **26**, 82-93.
4. **Johnson, N. L**. (1966): Cumulative sum control chart and the Weibull Distribution. *Technimetrics,* **8**, 481-491.
5. **Phatarfod, R. M.** (1971): A sequential test for Gamma distribution. *Jour. American Statist. Assoc.*, **66**, 876-878.
6. **Surinder, k. and Chandra, N.** (2009): A note on sequential testing of the mean of an inverse Gaussian distribution with known coefficient of variation. *Journal of Indian Statistical Association,* vol **47**, 151-160.