On Sample Size Estimation For Lomax Disrtibution

Abd-Elfattah, A.M.[1]

Alaboud, F.M.[2] **and Alharby ,A.H.**[2]

Abstract:

For life testing when the life times of items are continuous random variables, it is important to know the total number of individuals in the sample which is drawn from an assumed failure model, the total number of individuals may be unknown for many causes, either due to the omission in the records or perhaps because of physical conditions of the experiment, and then the sample size should be estimated. The Lomax distribution (Pareto distribution of the second kind) has, in recent years, assumed a position of importance in the field of life testing because of its uses to fit business failure data, Lomax(1954).

In this paper we consider the Lomax distribution as an important model of lifetime models and will derive the non-Bayesian and Bayesian estimators of sample size in the case of type I censored samples according to Marcus and Blumenthal (1975) approach. Numerical results for these estimators are presented in the last section of this work. An iterative procedure is used to obtain the estimators numerically.

Key Words: Conditional and Unconditional Maximum Likelhood Estimators; Bayesian Estimator; Sample Size; Censored Samples; Lomax Distribution.

1. Introduction

Many authors investigated the problem of estimating sample size. Among those who have studied non-Bayesian estimation of sample size Murphy (1948), Birnbaum and Zuckerman (1949), Johnson (1962), Harman (1967), Guenther (1970),Gokhale (1972), Sanathanan (1972, 1977) and Angers (1974).

While those who have interested with Bayesian estimation of sample size Hald (1967); Guthrie and Johns (1959) and Draper and Guttman (1971).

In Type I censored samples and exponential distribution, Marcus and Blumenthal (1975) derived Bayesian and non-Bayesian estimators for sample size. According to Marcus and Blumenthal (1975) approach, Ashour and Shalaby (1983) derived these estimators when underlying distribution is Weibull and Ashour et.al. (1996) derived these estimators in the case of Burr type XII failure model. Abd-Elfattah and Bakoban (2003) obtained the Bayesian and non-Bayesian estimators for sample size in case of generalized gamma distribution.

---

1 Cairo University, Institute of Statistical Studies and Research, Egypt.

2 King Abdulaziz University, Faculty of Sciences, Department of statistics, Saudia Arabia.

Our aim in this paper is to consider Marcus and Blumenthal (1975) approch to estimate the unknown sample size in the case when the distribution is Lomax which has the following density function:

$$f\left(x\middle|\alpha,\lambda\right)=\frac{\alpha}{\lambda}(1+\frac{x}{\lambda})^{-(\alpha+1)} \quad ; \quad x>0 \quad , \quad \alpha,\lambda>0 \tag{1}$$

and the cumulative distribution function is :

$$F\left(x\middle|\alpha,\lambda\right)=1-(1+\frac{x}{\lambda})^{-\alpha}\ ,\ x>0\ \ ,\ \alpha,\lambda>0 \tag{2}$$

where λ is a scale parameter and α is shape parameter.

In section 2, two maximum likelihood estimators of sample size (N) are derived (non-Bayesian estimators). These are based, respectively, on the conditional and unconditional distributions of the observations, and referred to as the CMLE and UMLE.

In section 3, Bayes estimates is obtained, and in section 4 a numerical investigation considered.

2. Non-Bayesian Estimators.

The conditional and unconditional maximum likelihood estimators for N if the underlying distribution is Lomax can be obtained as follows:

(i) The Unconditional Maximum Likelihood Estimators:

In type I censored sample, when N items are placed on life test and the test is terminated at time T, we will obtain the order statistics:

$$x_{(1)}\le x_{(2)}\le \ldots \le x_{(r)}\le T, \qquad r<N$$

where r is the number of failures.

By using this sample, (1) and (2) we get the likelihood function:

$$L(x|N,\alpha,\lambda)=\frac{N!}{(N-r)!}\ \alpha^{r}\lambda^{-r}H(\alpha,\lambda)\ G(\alpha,\lambda) \tag{3}$$

where

$$H(\alpha,\lambda)=\prod_{i=1}^{r}(1+\frac{x_{(i)}}{\lambda})^{-(\alpha+1)} \quad \text{and} \quad G(\alpha,\lambda)=(1+\frac{T}{\lambda})^{-\alpha(N-r)} \tag{4}$$

The unconditional maximum likelihood estimators are defined in such way that $\left(\hat{N},\ \hat{\alpha}_1,\hat{\lambda}_1\right)$ maximize (3) simultaneously with respect to N , α and λ .

Taking the natural logarithm of equation (3) and differentiating it with respect to α and λ respectively. Then we have

$$\hat{\alpha}_1=[\frac{r}{\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\hat{\lambda}_1})+(N-r)\ln((1+\frac{T}{\hat{\lambda}_1})}] \tag{5}$$

$$\hat{\lambda}_1 = r\{\left\{ r\left[\sum_{i=1}^{r} \ln(1+\frac{x_{(i)}}{\hat{\lambda}_1}) + (N-r)\ln(1+\frac{T}{\hat{\lambda}_1}) \right]^{-1} + 1 \right\} \sum_{i=1}^{r} \frac{x_{(i)}}{\hat{\lambda}_1(\hat{\lambda}_1 + x_{(i)})}$$

$$+ r(N-r)\frac{T}{\hat{\lambda}_1(\hat{\lambda}_1 + x_{(i)})}[\sum_{i=1}^{r} \ln(1+\frac{x_{(i)}}{\hat{\lambda}_1}) + (N-r)\ln(1+\frac{T}{\hat{\lambda}_1})]^{-1}\}^{-1} \tag{6}$$

Now, If we replace α, λ by $\hat{\alpha}_1, \hat{\lambda}_1$ in (3), the likelihood function becomes a function of N only as follows:

$$L(N) = \frac{N!}{(N-r)!} \quad \hat{\alpha}_1^{\ r} \ \hat{\lambda}_1^{-r} H(\hat{\alpha}_1, \hat{\lambda}_1) G(\hat{\alpha}_1, \hat{\lambda}_1) \tag{7}$$

where $H(\hat{\alpha}_1, \hat{\lambda}_1)$ and $G(\hat{\alpha}_1, \hat{\lambda}_1)$ are defined in (4) with $\alpha = \hat{\alpha}_1$ and $\lambda = \hat{\lambda}_1$

Since $\hat{N}$ maximizes (7) when it satisfies the relation:

$$L(N-1) < L(N) > L(N+1).$$

i.e.

$$\frac{L(N-1)}{L(N)} < \ 1 \ > \frac{L(N+1)}{L(N)}.$$

$$\frac{L(N+1)}{L(N)} = \frac{N+1}{N-r+1}(\frac{\hat{\alpha}_0}{\hat{\alpha}_1})^r (\frac{\hat{\lambda}_0}{\hat{\lambda}_1})^{-r} \prod_{i=1}^{r}(1+\frac{x_{(i)}}{\hat{\lambda}_0 - \hat{\lambda}_1})^{-(\hat{\alpha}_0 - \hat{\alpha}_1)} < 1$$

$$and \tag{8}$$

$$\frac{L(N-1)}{L(N)} = \frac{N-r}{N}(\frac{\hat{\alpha}_2}{\hat{\alpha}_1})^r (\frac{\hat{\lambda}_2}{\hat{\lambda}_1})^{-r} \prod_{i=1}^{r}(1+\frac{x_{(i)}}{\hat{\lambda}_2 - \hat{\lambda}_1})^{-(\hat{\alpha}_2 - \hat{\alpha}_1)} < 1$$

where $(\hat{\alpha}_0, \hat{\lambda}_0)$ are obtained by maximizing the natural logarithm of the likelihood function $L(N+1)$ with respect to (α, λ) and $(\hat{\alpha}_2, \hat{\lambda}_2)$ are obtained by maximizing the natural logarithm of the likelihood function $L(N-1)$ with respect to (α, λ) respectiviley.

i.e. the UMLE $(\hat{\alpha}_j, \hat{\lambda}_j)$ can be written as follows:

$$\hat{\alpha}_j = [\frac{r}{\sum_{i=1}^{r} \ln(1+\frac{x_{(i)}}{\hat{\lambda}_j}) + (N-r+j+1)\ln((1+\frac{T}{\hat{\lambda}_j})}], \ j = 0,1,2$$

and

$$\hat{\lambda}_j = r\{\left\{ r\left[\sum_{i=1}^{r} \ln(1+\frac{x_{(i)}}{\hat{\lambda}_j}) + (N-r+j-1)\ln(1+\frac{T}{\hat{\lambda}_j}) \right]^{-1} + 1 \right\} \sum_{i=1}^{r} \frac{x_{(i)}}{\hat{\lambda}_j(\hat{\lambda}_j + x_{(i)})}$$

$$+ r(N-r+j-1)\frac{T}{\hat{\lambda}_j(\hat{\lambda}_j + x_{(i)})}[\sum_{i=1}^{r} \ln(1+\frac{x_{(i)}}{\hat{\lambda}_j}) + (N-r+j-1)\ln(1+\frac{T}{\hat{\lambda}_j})]^{-1}\}^{-1}, \ j = 0,1,2$$

Then the UMLE of N will satisfy the following relation:

$$D(\hat{N}, x, \hat{\alpha}, \hat{\lambda}) \leq R < D(\hat{N}+1, x, \hat{\alpha}, \hat{\lambda}) \tag{9}$$

where

$$R=\frac{A}{r\,B}\quad,\qquad A=r\sum_{i=1}^{r}\frac{x_{(i)}}{\hat{\lambda}(\hat{\lambda}+x_{(i)})}\ln(1+\frac{x_{(i)}}{\hat{\lambda}})+r\,\frac{x_{(i)}}{\hat{\lambda}(\hat{\lambda}+x_{(i)})}+\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\hat{\lambda}})$$

$$B=r\ln(1+\frac{T}{\hat{\lambda}})+r\,\frac{T}{\hat{\lambda}(\hat{\lambda}+T)}[1+\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\hat{\lambda}})\,]$$

$$D\left(\hat{N},x,\hat{\alpha},\hat{\lambda}\right)=\{r[1-\left(1-\frac{r}{\hat{N}}\right)^{\frac{1}{r}}\prod_{i=1}^{r}(1+\frac{x_{(i)}}{\hat{\lambda}_2-\hat{\lambda}_1})^{-(\hat{\alpha}_2-\hat{\alpha}_1)}]^{\frac{1}{r}}\}^{-1}-1+\frac{\hat{N}}{r}\qquad(10)$$

Using numerical results and QBASIC program, (10) can be solved numerically.

(ii) The Conditional Maximum Likelihood Estimator:

The likelihood function (3) can be expressed in the form

$$L(x|N,\alpha,\lambda)=L_1(x|N,\alpha,\lambda)\,.\,L_2(x|r,\alpha,\lambda)$$

where

$$L_1(x|N,\alpha,\lambda)=\frac{N!}{r!\,(N-r)!}\ .\left[1-(1+\frac{T}{\lambda})^{-\alpha}\right]^{r}\left[1+\frac{T}{\lambda}\right]^{-\alpha(N-r)}\qquad(11)$$

and

$$L_2(x|r,\alpha,\lambda)=r!\,\alpha^{r}\lambda^{-r}\prod_{i=1}^{r}(1+\frac{x_{(i)}}{\lambda})^{-(\alpha+1)}[1-(1+\frac{T}{\lambda})^{-\alpha}\ ]^{r}\qquad(12)$$

Following Marcus and Blumenthal (1975) approach, the conditional maximum likelihood estimators $\tilde{N},\tilde{\alpha},\tilde{\lambda}$ will be obtained by first maximizing $L_2(x|r,\alpha,\lambda)$ with respect to α and λ. Thus the CMLE $\tilde{\alpha},\tilde{\lambda}$ is the solution of the following equations:

$$\frac{r}{\tilde{\alpha}}-\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\tilde{\lambda}})=0\qquad(13)$$

$$\frac{r}{\tilde{\lambda}}+\frac{(\tilde{\alpha}+1)x_{(i)}}{\tilde{\lambda}^2\sum_{i=1}^{r}(1+\frac{x_{(i)}}{\tilde{\lambda}})}-\frac{r\tilde{\alpha}(1+\frac{T}{\tilde{\lambda}})^{-(\tilde{\alpha}+1)}T}{[1-(1+\frac{T}{\tilde{\lambda}})^{-\tilde{\alpha}}\,]\tilde{\lambda}^2}=0\qquad(14)$$

Then maximizing $L_1(x|N,\tilde{\alpha},\tilde{\lambda})$ with respect to N, we get the following relations for N which satisfy:

$$\frac{L_1(N-1)}{L_1(N)}<\ 1\ >\frac{L_1(N+1)}{L_1(N)}$$

and

$$D\left(\tilde{N}\right) < \frac{\sum_{i=1}^{r} \ln(1+\frac{x_{(i)}}{\tilde{\lambda}})}{r\ln(1+\frac{x_{(i)}}{\tilde{\lambda}})} < D\left(\tilde{N}+1\right) . \tag{15}$$

where under $\tilde{N} = n$

$$D(n) = \left[-\ln\left(1-\frac{r}{n}\right)\right]^{-1} + 1 - \frac{n}{r} \tag{16}$$

The function $D(n)$ was tabulated by Marcus and Blumenthal (1975). It should be noted that, according to the theorem of Sanathanan (1972), N can be obtained from the following relation:

$$\tilde{N} = \frac{r}{1-(1+\frac{T}{\tilde{\lambda}})^{-\tilde{\alpha}}} \tag{17}$$

Numerical solution is needed to obtain the new estimators.

3. Bayesian Estimators

Suppose the conjugate prior density of α *and* λ are given by

$$g(\alpha) = \frac{\mu^{\gamma+1}}{\Gamma(\gamma+1)} \alpha^{\gamma} e^{-\mu\alpha} \quad , \quad \alpha > 0,\ \mu > 0,\ \gamma > -1$$
$$= 0 \qquad \text{other wise} \quad .$$

$$g(\lambda) = \frac{\delta^{\beta+1}}{\Gamma(\beta+1)} \lambda^{\beta} e^{-\delta\lambda} \quad , \quad \lambda > 0,\ \delta > 0,\ \beta > -1$$
$$= 0 \qquad \text{other wise}$$

and the uniform prior of N given by

$$P(N) = 1 \qquad \text{for all } N \quad .$$

Therefor, the joint prior density is given by

$$f\left(\alpha,\lambda,N\middle|x\right) \propto \frac{N!}{(N-r)!} \alpha^{(r+\gamma)} \lambda^{(r+\beta)} H(\alpha,\lambda)\exp[-\alpha\{\mu+G(\alpha,\lambda)\}-\delta\lambda] \tag{18}$$

where $H(\alpha,\lambda)$ and $G(\alpha,\lambda)$ are defined in (4). Therefore the joint mode of the posterior distribution (18) (the Bayesian estimators of α , λ , N) are given by $\alpha_1^*, \lambda_1^*, N^*$ where α_1^*, λ_1^* are obtained from the equations

$$\alpha_1^* = (r+\gamma)[\frac{1}{\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\lambda_1^*}) + (N-r)\ln(1+\frac{T}{\lambda_1^*}) + \mu}] \quad .$$

and

$$\lambda_1^* = (r+\beta)\left\{\left[(r+\beta)[\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\lambda_1^*}) + (N-r)\ln(1+\frac{T}{\lambda_1^*}) + \delta\right]^{-1} + 1\right\}\sum_{i=1}^{r}\frac{x_{(i)}}{\lambda_1^*(\lambda_1^*+x_{(i)})}$$
$$+(r+\gamma)(N-r)\frac{T}{\lambda_1^*(\lambda_1^*+x_{(i)})}[\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\lambda_1^*}) + (N-r)\ln(1+\frac{T}{\lambda_1^*}) + \delta]^{-1}\}^{-1}$$

while N^* should satisfy the following

$$\frac{f^*(N-1)}{f^*(N)} < 1 > \frac{f^*(N+1)}{f^*(N)}$$

where

$$f^*(N) \propto \frac{N!}{(N-r)!} \alpha_1^{*(r+\gamma)} \lambda_1^{*(r+\beta)} X(\alpha_1^*) \exp[-\alpha_1^*(\mu + n(\alpha_1^*) - \delta\lambda_1^*] \tag{19}$$

That is, N^* should satisfy:

$$D(N^*, x, \alpha^*, \lambda^*) \le R^* < D(N^* + 1, x, \alpha^*, \lambda^*) \tag{20}$$

where

$$R^* = \frac{A^*}{r\ B^*}$$

$$B^* = (r+\gamma)\ln(1+\frac{T}{\lambda^*}) + \frac{(r+\gamma)T}{\lambda^*(\lambda^*+T)}[1+\sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\lambda^*})] + \delta$$

$$A^* = (r+\gamma)\{\sum_{i=1}^{r}\frac{x_{(i)}}{\lambda^*(\lambda^*+x_{(i)})}\ln(1+\frac{x_{(i)}}{\lambda^*}) + (r+\gamma)\sum_{i=1}^{r}\frac{x_{(i)}}{\lambda^*(\lambda^*+x_{(i)})} + \sum_{i=1}^{r}\ln(1+\frac{x_{(i)}}{\lambda^*}) + \delta$$

under $N^* = n$, we have

$$D(n, x, \alpha^*, \lambda^*) = [r[1-\left(1-\frac{r}{n}\right)^{\frac{1}{\beta+r}} \prod (1+\frac{x_{(i)}}{\lambda_2^* - \lambda_1^*})^{-(\alpha_2^*-\alpha_1^*)}]^{\frac{1}{\beta+r}} - 1 + \frac{n}{r} \tag{21}$$

Using computer facilities and numerical techniques, the Bayesian estimates can be obtained.

4. Numerical Results

A numerical investigation will be considered in the present section using the following set of ordered observations from Lomax distribution

0.1920	0.1956	0.2362	0.2707	0.3241
0.3806	0.5959	0.6217	0.6586	1.0711

Using this set of observations we shall develop an example for finding the UMLE, CMLE and Bayesian estimator of sample size (N).

(1) The UMLE $\hat{\alpha}, \hat{\lambda}, \hat{N}$ have been obtained using an iterative procedure and QBASIC program . Using the above data with r = 8 our estimators will be

$\hat{\alpha} = 1.3064$, $\hat{\lambda} = 0.5821$ and $\hat{N} = 17$

(2) The CMLE $\tilde{\alpha}, \tilde{\lambda}, \tilde{N}$ can be obtained from the same data, so that the estimators will be $\tilde{\alpha} = 1.3064$, $\tilde{\lambda} = 0.5821$ and $\tilde{N} = 23$.

(3) The Bayesian estimator N^*.

By using the same data and different values of prior parameters we obtained the following results:

Prior information				Bayes'estimate
μ	γ	δ	β	N^*
0	0	0	0	22
2	1	2	1	27
1	3	1	3	19
0	-1	0	-1	17

From this example we note that the UMLE $(\hat{N})$ is less than the CMLE $(\tilde{N})$. Also we observe that the CMLE estimate of N is less than the Bayesian estimates and they are the same in the case of non-informative prior.

References

1) Abd-Elfattah, A. M. and Bakoban, R.A. (2003). A Study on the Estimation of Sample Size for Generalized Gamma Distribution. The Scientific Journal for Economic and Commerce, Faculty of Commerce, Ain Shames University, Vol. 2, 17-28.

2) Angers, C.(1974). A Graphical Method to Evaluate Sample Sizes for the Multinomial Distribution. Technometrics, 16,3,469-471.

3) Ashour, S. K., Abd-Elfattah, A. M. and Mahmoud, M. R. (1996). Estimation of Sample Size for the Burr Failure Model. The Annual Conference in Statistics, Computer Science and Operations Research, ISSR, Cairo Univ., 31(2), 18-26.

4) Ashour, S. K. and Shalaby, O. A. (1983). Estimating Sample Size with Weibull Failure. Math. Operationsforsch. U. Statist. Ser. Statist., Berlin, 2,263-268

5) Birnbaum, Z. W. and Zuckerman, H. S. (1949). A Graphical Determination of Sample Size for Wilks Tolerance Limits. Ann. Math. Stat. , 20, 313-316.

6) Draper, N. and Guttman, I. (1971). Bayesian Estimation of the Binomial Parameter. Tecnometrics, 13,667-673.

7) Gokhale, D. V. (1972). A Formula for Sample Size Required for Specified Tolerance. The American Statistician, 26, 21.

8) Guenther, W. C. (1970). Determination of Sample Size for Distribution Free Tolerance Limits. The American Statistician, 24, 44-45.

9) Guthrie, D. and Johns, M.V. (1959). Bayes Acceptance Sampling Procedures for Large Lots. Ann. Math. Stat., 30, 896-925.

10) Hald, A. (1967). Asymptotic Properties of Bayesian Single Sampling Plans. J. Roy. Stat. Soc., B, 29, 162-173.

11) Harman, A. J. (1967). Note on Wilks Tolerance Limit Sample Sizes. Sankhya, A, 30, 215-219.

12) Johnson, N. L. (1962). Estimation of Sample Size. Tehnometrics, 4, 59-67.
13) Lomax, H. S. (1954). Bussiness failures ; Another example of the analysis of failure data. JASA, 49, 847-852.
14) Marcus, R. and Blumenthal, S. (1975). Estimating Population Size with Exponential Failure. JASA., 70, 913-922.
15) Murphy, R. B. (1948). Non-Parametric Tolerance Limits. Ann. Math. Stat., 19, 581-589.
16) Sanathanan, L. (1972). Estimating the Size of a Multinomial Population. Ann. Math. Stat., 43, 142-152.
17) Sanathanan, L. (1977). Estimating the Size of a Truncated Sample. JASA., 72, 669-672.