University of Victoria
Department of Economics

Econometrics Working Paper EWP1104

ISSN 1485-6441

On the Bias of the Maximum Likelihood Estimator for the Two-Parameter Lomax Distribution

David E. Giles
Department of Economics, University of Victoria

Hui Feng
Department of Economics, Business & Mathematics
King's University College, University of Western Ontario
&

Ryan T. Godwin
Department of Economics, University of Victoria

April, 2011

Abstract

The Lomax (Pareto II) distribution has found wide application in a variety of fields. We analyze the second-order bias of the maximum likelihood estimators of its parameters for finite sample sizes, and show that this bias is positive. We derive an analytic bias correction which reduces the percentage bias of these estimators by one or two orders of magnitude, while simultaneously reducing relative mean squared error. Our simulations show that this analytic bias correction out-performs a correction based on the parametric bootstrap. Three examples with actual data illustrate the application of our methods.

Keywords Maximum likelihood estimator; bias reduction; Lomax distribution; Pareto II distribution; bootstrap

MSC 60E05; 62E15; 62F10; 62F40

Contact author:
David E. Giles, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: dgiles@uvic.ca; Phone: (250) 721-8540; FAX: (250) 721-6214

1. Introduction

The objectives of this paper are to explore the bias of the maximum likelihood estimators (MLEs) of the parameters of the Lomax distribution, and to compare alternative methods of reducing this bias when the sample size is relatively small. We consider both analytical and simulation-based techniques for bias-correcting the MLEs, and find that a standard second-order analytic bias correction out-performs bias correction using a parametric bootstrap, in terms of both remaining bias and mean squared error.

The Lomax (1954), or Pareto II, distribution has been quite widely applied in a variety of contexts. Although introduced originally for modelling business failure data, the Lomax distribution has been used for reliability modelling and life testing (*e.g.*, Hassan and Al-Ghamdi, 2009), and applied to income and wealth distribution data (Harris, 1968; Atkinson and Harrison, 1978), firm size (Corbelini *et al.*, 2007) and queuing problems. It has also found application in the biological sciences and even for modelling the distribution of the sizes of computer files on servers (Holland *et al*., 2006). Some authors, such as Bryson (1974), have suggested the use of this distribution as an alternative to the exponential distribution when the data are heavy-tailed.

The two-parameter Lomax distribution has the following c.d.f. and p.d.f.:

$$F(y) = 1 - [1 + y/\lambda]^{-\alpha}$$

$$f(y) = (\alpha/\lambda)[1 + y/\lambda]^{-(1+\alpha)} \qquad ; \qquad y > 0$$

where $\alpha\,(>0)$ is the shape parameter, and $\lambda\,(>0)$ is the scale parameter. In some applications it is useful to incorporate a location parameter, but we do not pursue that here.

The Lomax distribution can be motivated in a number of ways. For example, Balkema and de Haan (1974) show that it arises as the limit distribution of residual lifetime at great age; Dubey (1970) shows that it can be derived as a special case of a particular compound gamma distribution; and Tadikamalla (1980) relates the Lomax distribution to the Burr family of distributions. On the other hand, the Lomax distribution has itself been used as the basis for several generalizations. For example, Ghitany *et al.* (2007) extend it by introducing an additional parameter using the Marshal and Olkin

(1997) approach; Al-Awadhi and Ghitany (2001) use the Lomax distribution as a mixing distribution for the Poisson parameter and derive a discrete Poisson-Lomax distribution; and Punathumparambath (2011) introduced the double-Lomax distribution and applied it to IQ data. The record statistics of the Lomax distribution have been studied by Ahsanullah (1991) and by Balakrishnan and Ahsanullah (1994); the implications of various forms of right-truncation and right-censoring are discussed by Myhre and Saunders (1982), Childs *et al*. (2001), Cramer and Schmiedt (2011) and others; and sample size estimation has been discussed by Abd-Elfattah *et al.* (2006).

The rest of the paper is structured as follows. In the next section we briefly discuss maximum likelihood estimation for this distribution. A standard methodology for analytic bias approximation and bias correction is introduced in section 3; and this is then applied to the MLEs of the Lomax distribution's parameters in section 4. Section 5 describes the results of a Monte Carlo simulation experiment that we have undertaken to examine the effectiveness of the analytic bias correction, and to compare its performance with bootstrap bias correction for this problem. Some illustrative empirical applications are provided in section 6; and section 7 concludes.

2. Maximum likelihood estimation

As noted already the p.d.f. for the Lomax distribution is:

$$f(y) = (\alpha / \lambda)[1 + y/\lambda]^{-(1+\alpha)} \quad ; \quad y > 0 \tag{1}$$

where $\alpha\,(>0)$ is the shape parameter, and $\lambda\,(>0)$ is the scale parameter. Using integral 3.241, no. 4, from Gradshteyn and Ryzhik (1965), the rth central moment of the Lomax distribution is

$$E[Y^r] = \alpha\lambda^r \Gamma(r+1)\Gamma(\alpha - r)/\Gamma(\alpha+1) \;, \quad ; \quad \alpha > r \quad ; r = 1, 2, 3, \ldots.. \tag{2}$$

So, $E[Y] = \lambda/(\alpha - 1)$, for $\alpha > 1$; and $V[Y] = \alpha\lambda^2/[(\alpha-1)^2(\alpha-2)]$, for $\alpha > 2$.

We will focus on maximum likelihood estimation of this distribution's parameters, but it should be noted that other estimation procedures have been considered by Lingappaiah (1986/87) and Howlader and Hossain (2002), for example. Based on n independent observations, the log-likelihood function is:

$$l = n\log(\alpha) - n\log(\lambda) - (1+\alpha)\sum_{i=1}^{n}\log(1+y_i/\lambda)\ , \tag{3}$$

and

$$\partial l/\partial\lambda = -(n/\lambda) + [(1+\alpha)/\lambda]\sum_{i=1}^{n}[y_i/(\lambda+y_i)] \tag{4}$$

$$\partial l/\partial\alpha = (n/\alpha) - \sum_{i=1}^{n}\log(1+y_i/\lambda) \quad . \tag{5}$$

Note that there is no closed-form solution to the likelihood equations based on (4) and (5), and a suitable numerical algorithm must be used to obtain the MLEs of λ and α. In practice, this poses little difficulty as the log-likelihood in (3) is strictly concave. This can be verified from the second-order derivatives that are derived in section 4. Although we cannot obtain exact closed-form expressions for the MLEs, $\hat{\lambda}$ and $\hat{\alpha}$, analytic expressions for the $O(n^{-1})$ biases of these MLEs *can* be derived and then used to reduce the finite-sample biases of these estimators, as we now discuss.

3. Bias approximation to $O(n^{-1})$

If $l(\theta)$ is the (total) log-likelihood based on a sample of n observations, with p-dimensional parameter vector, θ, and $l(\theta)$ is regular with respect to all derivatives up to and including the third order, then the joint cumulants of the derivatives of $l(\theta)$ are defined as

$$k_{ij} = E(\partial^2 l/\partial\theta_i\partial\theta_j) \quad ; \quad i, j = 1, 2, \ldots, p \tag{6}$$

$$k_{ijl} = E(\partial^3 l/\partial\theta_i\partial\theta_j\partial\theta_l) \quad ; \quad i, j, l = 1, 2, \ldots, p \tag{7}$$

$$k_{ij,l} = E[(\partial^2 l/\partial\theta_i\partial\theta_j)(\partial l/\partial\theta_l)] \quad ; \quad i, j, l = 1, 2, \ldots, p \tag{8}$$

The derivatives of these cumulants are:

$$k_{ij}^{(l)} = \partial k_{ij}/\partial\theta_l \quad ; \quad i, j, l = 1, 2, \ldots, p \tag{9}$$

and all of the expressions in (6) to (9) are assumed to be $O(n)$.

Cox and Snell (1968) proved that for sample data that are independent, but not necessarily identically, distributed the bias of the sth element of the MLE of θ is:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{si}k^{jl}[0.5k_{ijl} + k_{ij,l}] + O(n^{-2})\ ;\ \ s = 1, 2, \ldots, p \tag{10}$$

where k^{ij} is the $(i,j)^{\text{th}}$ element of the inverse of the information matrix, $K = \{-k_{ij}\}$. Cordeiro and Klein (1994) subsequently showed that (10) still holds if the data are non-independent, and that it can be re-expressed as:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^{p} k^{si} \sum_{j=1}^{p}\sum_{l=1}^{p}[k_{ij}^{(l)} - 0.5k_{ijl}]k^{jl} + O(n^{-2})\ ;\ s = 1, 2, \ldots, p. \tag{11}$$

The bias expression in (11) is generally easier to evaluate than (10), as it does not involve terms of the form given in (8).

Defining $a_{ij}^{(l)} = k_{ij}^{(l)} - (k_{ijl}/2)$, for $i, j, l = 1, 2, \ldots, p$; and gathering up terms into matrices:

$$A^{(l)} = \{a_{ij}^{(l)}\}\ ; \quad i, j, l = 1, 2, \ldots, p$$

$$A = [A^{(1)} \mid A^{(2)} \mid \ldots\ldots \mid A^{(p)}],$$

the $O(n^{-1})$ bias of the MLE of θ in (11) can be re-written in the convenient form:

$$Bias(\hat{\theta}) = K^{-1}Avec(K^{-1}) + O(n^{-2})\,.$$

Then, a "bias-corrected" MLE for θ can then be constructed as:

$$\tilde{\theta} = \hat{\theta} - \hat{K}^{-1}\hat{A}\, vec(\hat{K}^{-1}),$$

where $\hat{K} = (K)|_{\hat{\theta}}$ and $\hat{A} = (A)|_{\hat{\theta}}$, and numerical values for the elements of $\hat{\theta}$ can be obtained by numerically solving the likelihood equations based on (4) and (5). The bias of $\tilde{\theta}$ is $O(n^{-2})$.

This approach to analytically bias-adjusting MLEs has been widely and successfully used, and some recent applications are reported by Cribari-Neto and Vasconcellos (2002), Giles and Feng (2009), Giles (2011) and Giles *et al*. (2011).

4. Bias reduction and the Lomax distribution

In what follows, we will require the following higher-order derivatives of the log-likelihood function:

$$\partial^2 l/\partial\alpha^2 = -(n/\alpha^2)$$

$$\partial^3 l/\partial\alpha^3 = (2n/\alpha^3)$$

$$\partial^2 l/\partial\lambda^2 = (n/\lambda^2) - [(1+\alpha)/\lambda^2]\sum_{i=1}^{n} y_i/(\lambda+y_i) - [(1+\alpha)/\lambda]\sum_{i=1}^{n}[y_i/(\lambda+y_i)^2]$$

$$\partial^3 l/\partial\lambda^3 = -(2n/\lambda^3) + 2[(1+\alpha)/\lambda^3]\sum_{i=1}^{n} y_i/(\lambda+y_i) + 2[(1+\alpha)/\lambda^2]\sum_{i=1}^{n}[y_i/(\lambda+y_i)^2]$$
$$+ 2[(1+\alpha)/\lambda]\sum_{i=1}^{n}[y_i/(\lambda+y_i)^3]$$

$$\partial^2 l/\partial\lambda\partial\alpha = (1/\lambda)\sum_{i=1}^{n}[y_i/(\lambda+y_i)]$$

$$\partial^3 l/\partial\lambda\partial\alpha^2 = 0$$

$$\partial^3 l/\partial\lambda^2\partial\alpha = -(1/\lambda^2)\sum_{i=1}^{n}[y_i/(\lambda+y_i)] - (1/\lambda)\sum_{i=1}^{n}[y_i/(\lambda+y_i)^2]$$

To evaluate the expectations of these derivatives we will use the following result:

<u>*Theorem*</u>

Let Y follow a Lomax distribution with parameters α and λ. Then

$$E[Y(\lambda+Y)^{-r}] = \alpha\lambda^{1-r}/[(\alpha+r)(\alpha+r-1)] \quad ; \quad r = 1, 2, 3, \ldots\ldots \quad (12)$$
$$\alpha + r > 1$$

<u>*Proof*</u>

$$E[Y(\lambda+Y)^{-r}] = \int_0^\infty y(\lambda+y)^{-r}(\alpha/\lambda)(1+y/\lambda)^{-(\alpha+1)}dy$$
$$= \alpha\lambda^\alpha \int_0^\infty y(\lambda+y)^{-(r+\alpha+1)}dy$$

Applying the change of variable, $x = \lambda + y$:

$$E[Y(\lambda+Y)^{-r}] = \alpha\lambda^\alpha \int_\lambda^\infty (x-\lambda)x^{-(r+\alpha+1)}dx$$
$$= \alpha\lambda^\alpha[\int_\lambda^\infty x^{-(r+\alpha)}dx - \lambda\int_\lambda^\infty x^{-(r+\alpha+1)}dx]$$
$$= \alpha\lambda^\alpha\{[x^{1-(r+\alpha)}(1-r-\alpha)^{-1}]_\lambda^\infty - \lambda[x^{-(r+\alpha)}(r+\alpha)^{-1}]_\lambda^\infty\}$$
$$= \alpha\lambda^\alpha\{\lambda^{1-r-\alpha}(r+\alpha-1)^{-1} - \lambda^{1-r-\alpha}(r+\alpha)^{-1}\}$$
$$= \alpha\lambda^{1-r}/[(\alpha+r)(\alpha+r-1)]$$

and this expression is positive only if $(\alpha+r) > 1$. □

We now have the following results for the joint cumulants of the derivatives of the log-likelihood function:

$$k_{11} = E[\partial^2 l / \partial\lambda^2] = -n\alpha / [\lambda^2(\alpha+2)]$$

$$k_{12} = E[\partial^2 l / \partial\lambda\partial\alpha] = n / [\lambda(\alpha+1)]$$

$$k_{22} = E[\partial^2 l / \partial\alpha^2] = -n / \alpha^2$$

$$k_{111} = E[\partial^3 l / \partial\lambda^3] = 4n\alpha / [\lambda^3(\alpha+3)]$$

$$k_{112} = E[\partial^3 l / \partial\lambda^2\partial\alpha] = -2n / [\lambda^2(\alpha+2)]$$

$$k_{222} = E[\partial^3 l / \partial\alpha^3] = 2n / \alpha^3$$

$$k_{122} = E[\partial^3 l / \partial\lambda\partial\alpha^2] = 0 \qquad .$$

In addition,

$$k_{11}^{(1)} = \partial k_{11} / \partial\lambda = 2n\alpha / [\lambda^3(\alpha+2)]$$

$$k_{12}^{(1)} = \partial k_{12} / \partial\lambda = -n / [\lambda^2(\alpha+1)]$$

$$k_{22}^{(1)} = \partial k_{22} / \partial\lambda = 0$$

$$k_{11}^{(2)} = \partial k_{11} / \partial\alpha = -2n / [\lambda^2(\alpha+2)^2]$$

$$k_{12}^{(2)} = \partial k_{12} / \partial\alpha = -n / [\lambda(\alpha+1)^2]$$

$$k_{22}^{(2)} = \partial k_{22} / \partial\alpha = 2n / \alpha^3 \qquad .$$

So,

$$a_{11}^{(1)} = k_{11}^{(1)} - 0.5k_{111} = 2n\alpha / [\lambda^3(\alpha+2)(\alpha+3)]$$

$$a_{12}^{(1)} = k_{12}^{(1)} - 0.5k_{121} = k_{12}^{(1)} - 0.5k_{112} = -n / [\lambda^2(\alpha+1)(\alpha+2)]$$

$$a_{22}^{(1)} = k_{22}^{(1)} - 0.5k_{221} = k_{22}^{(1)} - 0.5k_{122} = 0$$

and,

$$a_{11}^{(2)} = k_{11}^{(2)} - 0.5k_{112} = n\alpha / [\lambda^2(\alpha+2)^2]$$

$$a_{12}^{(2)} = k_{12}^{(2)} - 0.5k_{122} = -n / [\lambda(\alpha+1)^2]$$

$$a_{22}^{(2)} = k_{22}^{(2)} - 0.5k_{222} = n / \alpha^3 \qquad .$$

The information matrix is (see, also, Brazouskas, 2003, p. 321; Arnold, 1983, p. 210):

$$K = \{-k_{ij}\} = n\begin{bmatrix} \alpha/[\lambda^2(\alpha+2)] & -1/[\lambda(\alpha+1)] \\ -1/[\lambda(\alpha+1)] & 1/\alpha^2 \end{bmatrix} \quad , \tag{13}$$

which is clearly positive-definite (so $-K$ is negative-definite and the log-likelihood is strictly concave).

Defining $A^{(q)} = \{a_{ij}^{(q)}\}$, $q = 1, 2$; and $A = [A^{(1)} \mid A^{(2)}]$, we have:

$$A = n\begin{bmatrix} \dfrac{2\alpha}{\lambda^3(\alpha+2)(\alpha+3)} & \dfrac{-1}{\lambda^2(\alpha+1)(\alpha+2)} & \dfrac{\alpha}{\lambda^2(\alpha+2)^2} & \dfrac{-1}{\lambda(\alpha+1)^2} \\ \dfrac{-1}{\lambda^2(\alpha+1)(\alpha+2)} & 0 & \dfrac{-1}{\lambda(\alpha+1)^2} & \dfrac{1}{\alpha^3} \end{bmatrix}$$

Using Cordeiro and Klein's (1994) modification of the Cox and Snell (1968) result,

$$Bias\begin{pmatrix} \hat{\lambda} \\ \hat{\alpha} \end{pmatrix} = K^{-1}Avec(K^{-1}),$$

and bias-adjusted estimators can be obtained as $\begin{pmatrix} \tilde{\lambda} \\ \tilde{\alpha} \end{pmatrix} = \hat{K}^{-1}\hat{A}vec(\hat{K}^{-1})$, where

$\hat{K} = K\big|_{\lambda=\hat{\lambda};\alpha=\hat{\alpha}}$; and $\hat{A} = A\big|_{\lambda=\hat{\lambda};\alpha=\hat{\alpha}}$.

5. A simulation experiment

The above bias expressions are valid to $O(n^{-1})$. We have compared the actual bias and mean squared error (MSE) of the MLEs and bias-corrected MLEs in a small Monte Carlo experiment. The MLEs were obtained using the Newton-Raphson algorithm in the *maxLik* package (Toomet and Henningsen, 2008) for the *R* statistical software environment (R, 2008). In addition to $\hat{\alpha}$, $\tilde{\alpha}$, $\hat{\lambda}$ and $\tilde{\lambda}$, we have also considered bias-correcting the estimators using a parametric bootstrap. In the case of α (for example) this is obtained as $\breve{\alpha} = 2\hat{\alpha} - (1/N_B)[\sum_{j=1}^{N_B}\hat{\alpha}_{(j)}]$, where $\hat{\alpha}_{(j)}$ is the MLE of α obtained from the jth of the N_B bootstrap samples. These bootstrap-bias-corrected estimators are also unbiased to $O(n^{-1})$, but this comes at the expense of increased variance.

Each part of the experiment uses 10,000 Monte Carlo replications. In the case of the bootstrap-corrected estimators we use 1,000 bootstrap samples *per* replication. The results that are reported in Tables 1 are *percentage* biases and MSEs. The latter are defined as $[100 \times \text{MSE}(\hat{\alpha}) / \alpha^2]$, *etc*., and are reported in square brackets below the corresponding percentage biases.

The scale parameter, λ, is set to unity in this experiment. Some additional experimentation indicated that the orders of magnitude of the percentage biases and MSEs are relatively invariant to the value of this parameter. Various values of the shape parameter, α, have been considered, including ones that are consistent with some of the empirical studies discussed in section 1. For example, Corbellini *et al*. (2007) obtained estimates ranging from 0.96 to 1.06; and Holland *et al.* (2006) reported estimates of α in the range 0.5 to 3.3. We report results based on $\alpha = 1.1$ and 2.1. The latter value ensures that the first two moments of the Lomax distribution exist, while the former value ensures that the first moment is finite. The minimum sample sizes that are considered are determined by factors that are discussed at the end of this section.

The simulation results in Table 1 are strikingly clear. We see that the MLEs, $\hat{\alpha}$ and $\hat{\lambda}$ exhibit sizeable positive percentage bias when the sample size is small. The biases and mean squared errors of these MLEs fall quickly as the sample size increases, reflecting the consistency of the estimators. The analytic bias adjustment, leading to $\tilde{\alpha}$ and $\tilde{\lambda}$, is extremely successful, reducing the (absolute) percentage bias by one or even two orders of magnitude, regardless of the sample size. In addition, this success comes with an improvement in precision – the percentage MSEs of both $\tilde{\alpha}$ and $\tilde{\lambda}$ are less than those of $\hat{\alpha}$ and $\hat{\lambda}$. For small sample sizes, this reduction in %MSE can be more than 50%. Of course, for very large samples the bias adjustment becomes irrelevant, and the biases and MSEs of the unadjusted and adjusted MLEs converge. Another important feature of these results is that bias adjusting reverses the sign of the bias, for all but the very largest sample size considered (when the bias is negligible in any case). This "over-correction" has been reported in similar studies involving the Cox-Snell methodology, including those of Cribari-Neto and Vasconcellos (2002) for the beta distribution, Giles (2011) for the half-logistic distribution, and Giles *et al*. (2011) for the generalized Pareto distribution. A practical implication of our results is that researchers will know the direction of the biases in the estimators in small samples.

Table 1. Percentage biases and MSE's

n	$\%Bias(\hat{\alpha})$ [$\%MSE(\hat{\alpha})$]	$\%Bias(\tilde{\alpha})$ [$\%MSE(\tilde{\alpha})$]	$\%Bias(\breve{\alpha})$ [$\%MSE(\breve{\alpha})$]	$\%Bias(\hat{\lambda})$ [$\%MSE(\hat{\lambda})$]	$\%Bias(\tilde{\lambda})$ [$\%MSE(\tilde{\lambda})$]	$\%Bias(\breve{\lambda})$ [$\%MSE(\breve{\lambda})$]
			α = 1.1 ; λ = 1.0			
35	18.343	-15.354	-24.697	34.891	-27.267	-43.520
	[27.453]	[5.668]	[13.673]	[96.910]	[17.494]	[45.277]
40	18.171	-11.283	-20.185	32.871	-20.571	-36.336
	[27.141]	[5.331]	[11.777]	[87.710]	[16.273]	[36.800]
45	18.489	-8.066	-15.954	33.118	-14.930	-29.003
	[27.919]	[5.447]	[10.924]	[87.205]	[15.878]	[33.513]
50	18.138	-5.429	-12.878	31.902	-10.517	-23.887
	[26.643]	[5.790]	[9.785]	[81.205]	[16.783]	[29.026]
55	15.823	-4.353	-9.239	28.758	-7.835	-16.242
	[23.002]	[5.883]	[10.765]	[71.779]	[17.001]	[33.387]
60	14.396	-3.334	-8.158	25.747	-6.138	-14.538
	[20.386]	[6.087]	[9.376]	[61.617]	[17.101]	[28.277]
75	11.060	-1.501	-5.460	19.690	-2.865	-9.889
	[13.480]	[5.857]	[5.971]	[40.495]	[16.442]	[17.399]
100	8.298	-0.298	-2.527	14.289	-1.006	-4.987
	[9.111]	[5.205]	[4.670]	[26.242]	[14.385]	[12.936]
125	6.163	-0.195	-1.473	10.677	-0.631	-2.908
	[5.569]	[3.813]	[3.440]	[16.561]	[10.895]	[9.897]
150	5.053	-0.070	-0.866	8.999	-0.113	-1.543
	[4.391]	[3.262]	[3.044]	[12.778]	[9.172]	[8.524]
175	4.026	-0.242	-0.767	7.183	-0.392	-1.331
	[3.565]	[2.801]	[2.683]	[10.294]	[7.861]	[7.510]
200	3.520	-0.148	-0.519	6.211	-0.288	-0.953
	[2.852]	[2.315]	[2.248]	[8.264]	[6.556]	[6.354]
250	2.985	0.101	-0.120	5.196	0.095	-0.298
	[2.257]	[1.911]	[1.879]	[6.401]	[5.320]	[5.232]
300	2.258	-0.098	-0.240	4.037	-0.128	-0.378
	[1.753]	[1.539]	[1.524]	[5.016]	[4.331]	[4.289]
400	1.638	-0.097	-0.170	2.945	-0.118	-0.248
	[1.250]	[1.137]	[1.132]	[3.536]	[3.175]	[3.163]
500	1.430	0.052	0.007	2.440	0.011	-0.068
	[0.992]	[0.917]	[0.916]	[2.783]	[2.551]	[2.548]
750	0.879	-0.026	-0.048	1.545	-0.048	-0.086
	[0.630]	[0.599]	[0.598]	[1.802]	[1.704]	[1.702]
1000	0.667	-0.007	-0.017	1.165	-0.022	-0.042
	[.463]	[0.446]	[0.446]	[1.318]	[1.264]	[1.264]
2000	0.345	0.0115	0.011	0.657	0.069	0.066
	[0.229]	[0.225]	[0.225]	[0.653]	[0.639]	[0.640]

Table 1. (continued)

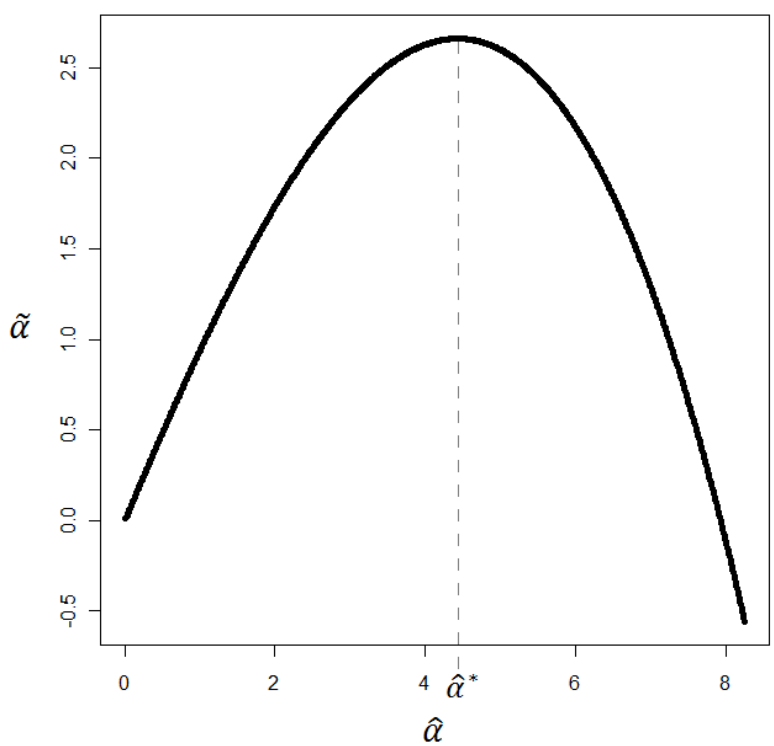
n	$\%Bias(\hat{\alpha})$ [$\%MSE(\hat{\alpha})$]	$\%Bias(\tilde{\alpha})$ [$\%MSE(\tilde{\alpha})$]	$\%Bias(\breve{\alpha})$ [$\%MSE(\breve{\alpha})$]	$\%Bias(\hat{\lambda})$ [$\%MSE(\hat{\lambda})$]	$\%Bias(\tilde{\lambda})$ [$\%MSE(\tilde{\lambda})$]	$\%Bias(\breve{\lambda})$ [$\%MSE(\breve{\lambda})$]
			α = 2.1 ; λ = 1.0			
80	16.886	-16.414	-18.095	24.193	-23.319	-25.732
	[25.696]	[5.329]	[11.936	[52.511]	[11.100]	[23.212]
90	16.587	-12.587	-13.661	23.617	-17.953	-19.477
	[24.429]	[4.741]	[13.561]	[49.415]	[9.856]	[26.502]
100	16.194	-9.797	-11.680	22.910	-14.079	-16.718
	[23.765]	[4.642]	[12.524]	[47.639]	[9.341]	[24.646]
110	15.725	-7.508	-11.888	22.266	-10.846	-17.003
	[22.633]	[4.699]	[9.764]	[45.661]	[9.354]	[19.367]
125	14.754	-5.291	-7.523	20.828	-7.728	-10.793
	[21.720]	[5.012]	[11.493]	[43.667]	[9.725]	[23.121]
150	11.963	-3.131	-5.994	17.174	-4.318	-8.382
	[16.684]	[5.227]	[8.448]	[33.268]	[10.170]	[16.489]
175	10.379	-1.559	-4.500	14.590	-2.319	-6.503
	[12.745]	[5.199]	[6.106]	[24.906]	[10.034]	[11.680]
200	8.781	-1.009	-3.533	12.289	-1.568	-5.142
	[10.008]	[4.820]	[5.031]	[19.567]	[9.291]	[9.692]
250	6.993	-0.271	-1.878	9.759	-0.514	-2.785
	[7.073]	[4.132]	[4.072]	[13.901]	[8.006]	[7.905]
300	5.228	-0.426	-1.490	7.384	-0.604	-2.108
	[4.941]	[3.418]	[3.113]	[9.730]	[6.683]	[6.082]
400	4.006	-0.034	-0.549	5.611	-0.088	-0.814
	[3.390]	[2.621]	[2.489]	[6.612]	[5.076]	[4.820]
500	2.984	-0.319	-0.428	4.119	-0.278	-0.687
	[2.531]	[2.087]	[2.032]	[4.870]	[3.996]	[3.890]
750	1.974	-0.031	-0.146	2.779	-0.045	-0.205
	[1.517]	[1.342]	[1.328]	[2.984]	[2.629]	[2.604]
1000	1.432	-0.044	-0.102	2.005	-0.073	-0.155
	[1.083]	[0.990]	[0.987]	[2.123]	[1.937]	[1.931]
2000	0.752	0.031	0.021	1.097	0.083	0.066
	[0.519]	[0.496]	[0.497]	[1.021]	[0.974]	[0.975]

The bootstrap bias-adjustment generally exhibits a stronger tendency to over-correct, than does the analytic adjustment. For small enough sample sizes, the bootstrap's over-correction can result in higher (absolute) biases than those for the unadjusted MLEs. This can be seen for n = 35, 40 when α = 1.1, and for n = 80 when α = 2.1, so the bootstrap is definitely not recommended in such cases Both sets of bias-adjusted MLEs have similar percentage MSEs, though the analytically corrected

estimators are more precise than the bootstrap-corrected MLEs for the smaller sample sizes considered. Overall, the superiority of the Cox-Snell / Cordeiro-Klein methodology over the parametric bootstrap in this context is very clear. In addition, given the bias expressions that we have derived, the analytical bias adjustment may also be preferred on the grounds of computational simplicity. It is almost costless to implement in practice.

As a further practical matter, care needs to be taken with the decision to bias-adjust $\hat{\alpha}$ using the analytical formula if n is relatively small. Specifically, depending on the true value of the shape parameter, it is possible for $\tilde{\alpha}$ or $\breve{\alpha}$ to be negative, which is clearly nonsensical. This may be a consequence of using a bias correction that is valid only to $O(n^{-1})$. In practice, negative estimates would not be accepted, and such rare cases have been discarded in our simulation experiment. Figure 1 shows a representative relationship between our recommended $\tilde{\alpha}$, and $\hat{\alpha}$. This relationship depends on n, but is invariant to the value of $\hat{\lambda}$. It is a trivial arithmetic matter to avoid a negative estimate of α. However, recalling from Table 1 that the bias of $\hat{\alpha}$ is always positive, while that of $\tilde{\alpha}$ is always negative, we can see from Figure 1 that if an estimate based on $\hat{\alpha}$ exceeds $\hat{\alpha}^*$, then there is an increased likelihood of the *absolute* bias of $\tilde{\alpha}$ exceeding that of $\hat{\alpha}$.

Figure 1. Relationship between $\hat{\alpha}$ and $\tilde{\alpha}$: n = 100



So, to assist practitioners, Figure 1 suggests that a *conservative* rule of thumb would be to analytically bias-adjust $\hat{\alpha}$ to $\tilde{\alpha}$ only if $\hat{\alpha} < \hat{\alpha}^*$. This also ensures a one-to-one relationship between $\hat{\alpha}$ and $\tilde{\alpha}$. It should be noted that negative values of $\tilde{\lambda}$ can also arise, but it can be shown that this cannot occur if our rule of thumb is applied, so we don't need to consider this estimator further in the present discussion. A table of the exact values of $\hat{\alpha}^*$, for n = 21(1)500, can be downloaded from http://web.uvic.ca/~dgiles/downloads/data. The value of $\hat{\alpha}^*$ increases monotonically with n, and by way of illustration, its values are 1.49 and 4.43 for n = 20, 100 respectively. The following (regression-based) formula can be used to determine $\hat{\alpha}^*$ accurately for various values of n:

$$\begin{aligned}\hat{\alpha}^* = &-0.991404 + 0.027773n - 0.0000328n^2 + 0.64541ln(n) - 3.617656D \\ &- 0.014898(D \times n) + 0.0000293(D \times n^2) + 1.036946(D \times ln(n))\end{aligned} \tag{14}$$

Here, D is a dummy variable, defined as $D = 0$ if $20 \le n \le 150$; $D = 1$ if $151 \le n \le 500$.

6. Illustrative applications

We have considered three simple applications, the first two of which involve data that have been used previously with the Lomax distribution. The third application uses new data that is of a type found previously to be modelled better by a Lomax distribution than by various competitors. In each case, the sample size is quite modest. We have fitted the Lomax distribution to each data-set using MLE, using the Marquardt algorithm, and then bias-adjusted the parameter estimates using the analytic bias formulae derived in section 4. Given the relative inferiority of the bootstrap bias correction noted already, we do not report those results here.

The first application involves American Insurance Association data relating to insurance losses in excess of $5million (in 1981 dollars) due to major hurricanes between 1949 and 1980. The data are provided by Hogg and Klugman (1983, p.92), and we have subtracted 5million from each of their sample values, so allowing for the scale of the data reported by those authors, our first datum is 1766.0, *etc.* The second set of data consists of observations for precipitation in a Florida meteorological study by Simpson (1972), and further analyzed by Bryson (1974). Finally, we consider a sample of computer file sizes (in bytes) for all 269 files with the .ini extension on the first author's Windows-based personal computer. These data can be downloaded from http://web.uvic.ca/~dgiles/downloads/data. Previous work by Holland *et al.* (2006) has demonstrated

the superiority of the Lomax distribution over several other competitors for modelling such file sizes. Those authors also provide technical information suggesting that in this context the distribution should have infinite variance (*i.e.*, $\alpha < 2$).

Table 2 reports the MLEs and the bias-adjusted MLEs when the Lomax distribution is fitted to each of these three data-sets. Each of the estimates based on $\hat{\alpha}$ are less than the values of $\hat{\alpha}^*$ for the respective sample sizes, so our rule of thumb from section 5 suggests that the bias-adjusted estimates based on $\tilde{\alpha}$ and $\tilde{\lambda}$ should be used. In each case the estimated values of α imply that the distributions have infinite variance, and for the third data-set the distribution has an infinite mean. For the other two data-sets the MLEs of α exceed unity before bias-adjustment (although the excess is not statistically significant), implying a finite mean. However, this is reversed by bias-adjusting the estimates. So, focusing on the point estimates of α in these first two cases, bias-adjusting the MLEs induces an important change to the characteristics of the Lomax distribution, even though the numerical differences may not seem large. Interestingly, the values of $\hat{\alpha}$ and $\hat{\lambda}$ associated with computer file data are very similar to those obtained by Holland *et al.* (2006) using a sample of 410 *.ini files. Their estimates were $\hat{\alpha} = 0.556$ and $\hat{\lambda} = 119.091$. They did not report standard errors, and of course, they did not bias-adjust their estimates. We can see that for sample size associated with the third data-set, bias-adjusting the MLEs has negligible impact on the point estimates.

Table 2. MLEs for empirical applications

Data-set	***n***	$\hat{\lambda}$	$\hat{\alpha}$	$\tilde{\lambda}$	$\tilde{\alpha}$
1	35	78670.267	1.157	51701.358	0.926
		(64882.53)	(1.78)		
2	26	137.972	1.097	84.616	0.856
		(100.06)	(0.58)		
3	269	128.307	0.498	124.032	0.492
		(33.70)	(0.06)		

Notes: Asymptotic standard errors appear in parentheses. The data-sets are 1: Hurricane losses; 2: Precipitation; 3: Computer file sizes.

The fitted c.d.f.'s based on the first two sets of MLEs are shown in Figures 1 and 2, over the ranges of the samples. (The fitted c.d.f.'s for the third data-set are indistinguishable.) For the hurricane loss data (Figure 1), the effect of the bias adjustment is to increase the intercept of the Lomax *density* at the origin and slightly "thicken" the extreme right tail. The effect on the fitted Lomax distribution for the precipitation data (Figure 2) is just the opposite, so that the probability of extreme values is somewhat reduced. However, it should be noted that for the parameter estimates in Table 2, neither the skewness nor kurtosis coefficients are defined.

7. Conclusions

In this paper we have evaluated some of the small-sample properties of the maximum likelihood estimator for the two-parameter Lomax (Pareto II) distribution. This distribution has been applied in a broad range of settings, and information about the bias and mean squared errors of the parameter estimators is of practical importance. Although the maximum likelihood estimators of the shape and scale parameters of the Lomax distribution cannot be expressed in closed form, we have used the methods of Cox and Snell (1968) to derive analytic second-order bias expressions which can then be used to bias-adjust the numerical point estimates.

Our simulation results show that this procedure is very effective. The maximum likelihood estimators of both parameters are positively biased in small samples. Applying the analytic bias correction results in a negative bias, but the absolute value of this bias is one or two orders of magnitude smaller than that of the original estimators. As an added bonus, the bias correction also has the effective of sharply reducing the mean squared errors of the estimators in small samples.

We have also considered using (parametric) bootstrap bias correction for this problem. This is found to be less effective than the analytic correction in terms of bias reduction, and also in terms of mean squared error.

The analytic bias correction that is derived in this paper is very simple to apply in practice, and is recommended over the bootstrap alternative. Except in quite extreme cases, when the sample size is very small, this bias correction is strongly recommended over the choice of no correction at all. We provide precise guidelines to enable a practitioner to determine when the correction should not be applied.

Figure 2. Fitted Lomax c.d.f.'s
(Hurricane loss data)

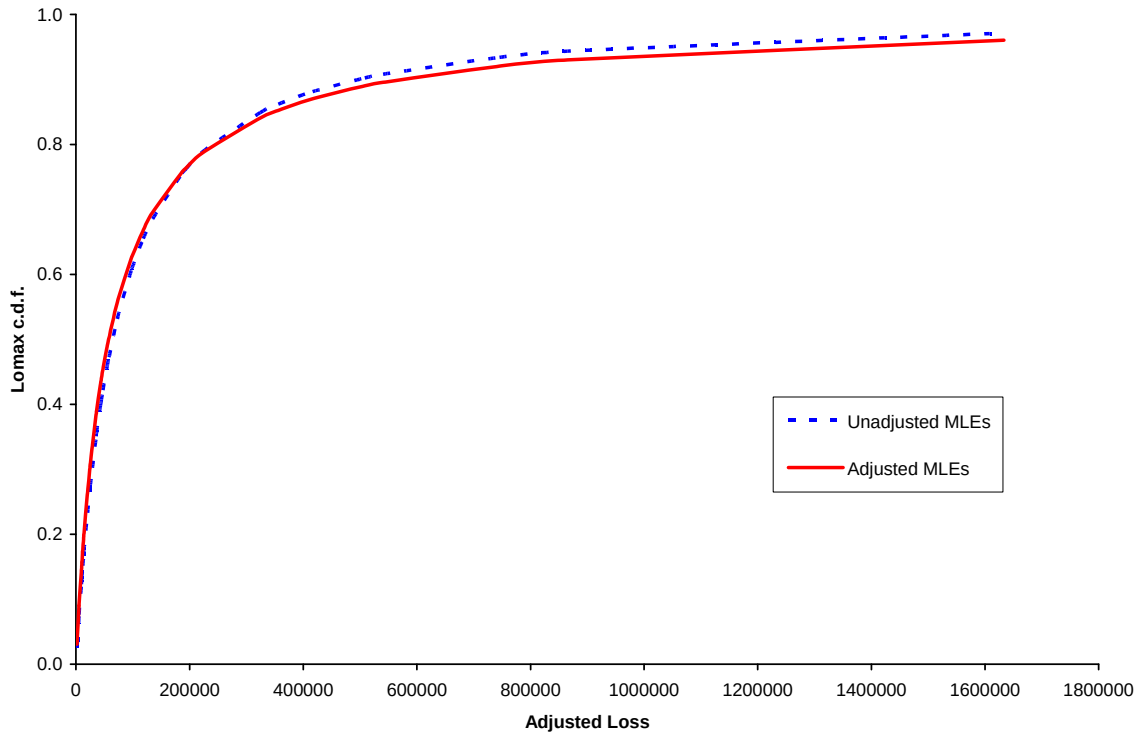
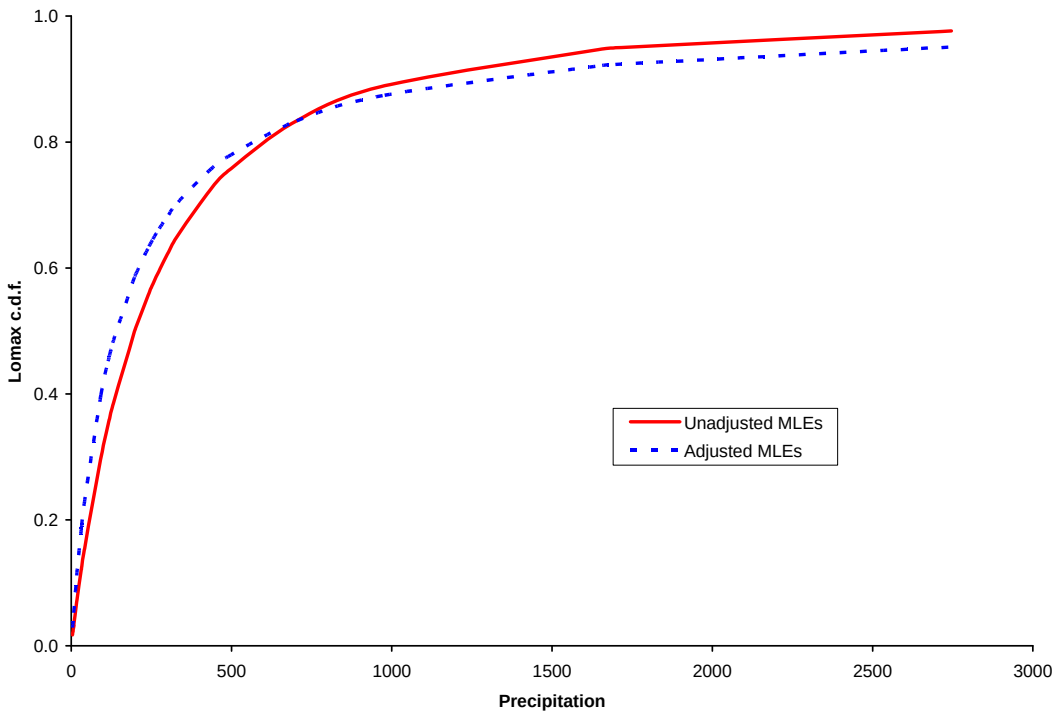


Figure 3. Fitted Lomax c.d.f.'s
(Precipitation data)



Acknowledgments

We are grateful to Jacob Schwartz for several helpful discussions and comments, and to Lief Bluck for providing access to the additional computing facilities needed to complete the simulation experiments. We also thank Matthew Giles for writing the Python script that was used to harvest the data on computer file sizes. The second author acknowledges financial support from King's University College at the University of Western Ontario.

References

Abd-Elfattah, A. M., F. M. Alaboud and A. H. Alharby (2006). On sample size estimation for Lomax distribution. Mimeo., Institute for Statistical Studies and Research, Cairo, Egypt.

Ahsanullah, M. (1991). Record values of Lomax distribution, *Statistica Nederlandica*, 41, 21–29.

Al-Awadhi, S. A. and M. E. Ghitany (2001). Statistical properties of Poisson–Lomax distribution and its application to repeated accidents data. *Journal of Applied Statistical Science,* 10, 365–372.

Arnold, B. C. (1983). *Pareto Distributions*. International Cooperative Publishing House, Fairland MD.

Atkinson, A. B. and A. J. Harrison(1978). *Distribution of Personal Wealth in Britain*. Cambridge University Press, Cambridge

Balakrishnan, N. and M. Ahsanullah (1993). Relations for single and product moments of record values from exponential distribution. *Communication in Statistics-Theory and Methods*, 23, 2841-2852.

Balkema, A. A. and L. de Haan (1974). Residual life at great age. *Annals of Probability*, 2**,** 792-804.

Brazouskas, V. (2003). Information matrix for Pareto (IV), Burr and related distributions. *Communications in Statistics*, 32, 315-325.

Bryson, M. C. (1974). Heavy-tailed distributions: Properties and tests. *Technometrics*, 16, 61-68.

Childs, A., N. Balakrishnan, and M. Moshref. (2001). Order statistics from non-identical right truncated Lomax random variables with applications. *Statistical Papers*, 42, 187–206.

Corbellini, A., L. Crosato, P. Ganugi and M. Mazzoli (2007). Fitting Pareto II distributions on firm size: Statistical methodology and economic puzzles. Paper presented at the International Conference on Applied Stochastic Models and data Analysis, Chania, Crete.

Cordeiro, G. M. and R. Klein (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, 19, 169-176.

Cox, D. R. and E. J. Snell (1968). A general definition of residuals. *Journal of the Royal Statistical Society, B*, 30, 248-275.

Cramer, E. and A.B. Schmiedt (2011). Progressively type-II censored competing risks data from Lomax distributions. *Computational Statistics and Data Analysis,* 55, 1285–1303.

Cribari-Neto, F. and K. L. P. Vasconcellos (2002). Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, 72, 107-118.

Dubey, S. D. (1970). Compound gamma, beta and F distributions. *Metrika*, 16, 27-31.

Ghitany, M. E., F. A. Al-Awadhi and L. A. Alkhalfan (2007). Marshal-Olkin extended Lomax distribution and its application to censored data. *Communications in Statistics – Theory and Methods*, 36, 1855-1866.

Giles, D. E. (2011). Bias reduction for the maximum likelihood estimators of the parameters in the half-logistic distribution. *Communications in Statistics – Theory and Methods*, forthcoming.

Giles, D. E., H. Feng (2009). Bias of the maximum likelihood estimators of the two-parameter gamma distribution revisited. Econometrics Working Paper EWP0906, Department of Economics, University of Victoria.

Giles, D. E., H. Feng and R. T. Godwin (2011). Bias - corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. Econometrics Working Paper EWP1105, Department of Economics, University of Victoria.

Gradshteyn, I. S. and I. W. Ryzhik (1965), *Table of Integrals, Series and Products*, 4th ed., translation, A. Jeffrey (ed.). Academic Press, New York.

Harris, C. M. (1968). The Pareto distribution as a queue service discipline. *Operations Research***,** *16,* 307-313.

Hassan, A. S. and A. S. Al-Ghamdi (2009). Optimum step stress accelerated life testing for Lomax distribution. *Journal of Applied Sciences Research*, 5, 2153-2164.

Hogg, R. V. and S. A. Klugman (1983). On the estimation of long-tailed skewed distributions with actuarial applications. *Journal of Econometrics*, 23, 91-102.

Holland, O., A. Golaup and A. H. Aghvami (2006). Traffic characteristics of aggregated module downloads for mobile terminal reconfiguration. *IEE Proceedings – Communications*, 135, 683-690.

Howlader, H. A. and A. M. Hossein (2002). Bayesian survival estimation of Pareto distribution of the second kind based on failure-censored data. *Computatational Statistics and Data Analysis*, 38, 301–314.

Lingappaiah, G. S. (1986/87). On the Pareto distribution of the second kind (Lomax distribution). *Revista de Mathemática e Estatística*, 4/5, 63–68.

Lomax, H. S. (1954). Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, 49, 847-852.

Marshall, A.W. and I. Olkin (1997). A new method of adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84, 641–652.

Myhre, J. and S. Saunders (1982). Screen testing and conditional probability of survival. In: Crowley, J., Johnson, R. A., eds. *Survival Analysis*. Lecture Notes-Monograph Series. Institute of Mathematical Statistics, 2, 166–178.

Punathumparambath, B. (2011). Estimation of $P(X > Y)$ for the double Lomax distribution. *ProbStat Forum*, 4, 1-11.

R (2008), The R Project for Statistical Computing, http://www.r-project.org

Simpson, J. (1972). Use of the gamma distribution in single-cloud rainfall analysis. *Monthly Weather Review*, 100, 309-312.

Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review*, 48, 337-344.

Toomet, O. and A. Henningsen (2008), maxLik: Maximum likelihood estimation, http://CRAN.R-project.org ; http://maxLik.org