LEAST SQUARES ESTIMATION BASED ON ORDER STATISTICS IN LOCATION–SCALE FAMILIES OF DISTRIBUTIONS

Robert Schall[1] and Ruaan van Zyl[2]
[1] University of the Free State, Bloemfontein, South Africa
[2] Corresponding author: PAREXEL International, Bloemfontein, South Africa,
email: ruaanvz@gmail.com

ABSTRACT

Maximum likelihood estimation for location-scale families of distributions works well, but may require advanced programming, especially for censored samples. We propose a general framework for least squares estimation using order statistics, which can handle certain censored samples. Our simulation study suggests the best least squares estimators are competitive with MLEs.

Keywords**:** censored samples, covariance, Pareto, Weibull

1. Introduction

In location-scale families of distributions, the estimation of location and scale parameters is required. General estimation methods like maximum likelihood have been developed and work well for many distributions, but in some cases maximum likelihood estimation is difficult and requires advanced programming methods. For censored samples maximum likelihood estimation is not trivial, even for otherwise simple cases like the normal distribution.

In this paper we propose a general framework for least squares estimation based on order statistics in location-scale families of distributions. We consider both the estimation of the location and scale parameters of the distribution as well as the estimation of quantiles. We apply the general approach to the Weibull and Pareto distributions and compare the mean squared error of various least squares estimators and of the maximum likelihood estimator.

2. Location-scale families of distributions

We consider a random variable Y whose distribution depends on at most two parameters $\alpha = [\alpha_1, \alpha_2]'$. We assume that there exists a reparametrization of α to parameters $\beta = [\beta_1, \beta_2]'$ and a transformation G(Y) of Y, such that the distribution of

$$Z = \frac{G(Y) - \beta_1}{\beta_2} \qquad 1$$

does not depend on the parameters β. Y is then considered a member of a location-scale family (Mann, Schafer and Singpurwalla, 1974).

Example 1: Two Parameter Weibull Distribution

Let Y follow a two parameter Weibull(λ, k) distribution with cdf

$$F_\beta(Y) = 1 - \exp\left[-\left(\frac{Y}{\lambda}\right)^k\right] \qquad 2$$

Then

$$Z = \log\left[\left(\frac{Y}{\lambda}\right)^k\right] = \frac{\log(Y) - \log\lambda}{\frac{1}{k}} = \frac{G(Y) - \beta_1}{\beta_2} \qquad 3$$

follows the distribution of the logarithm of a standard Weibull(1,1) variate (which is the distribution of the logarithm of a standard Exponential(1) variate). Thus, in terms of (1), for the Weibull(λ, k) distribution we have $G(Y) = \log(Y)$, $\beta_1 = \log\lambda$ and $\beta_2 = \frac{1}{k}$.

Alternatively Z can be written as

$$Z = G\{F^{-1}[F_\beta(Y)]\} = \log\{F^{-1}[F_\beta(Y)]\} = \log\{-\log[1 - F_\beta(Y)]\} \qquad 4$$

where $G(Y) = \log(Y)$, and $F^{-1}(U) = -\log(1 - U)$ is the inverse of the cdf of the standard $\mathrm{Weibull}(1,1)$.

Example 2: Two Parameter Pareto Distribution

Let Y follow a $\mathrm{Pareto}(\alpha, x_m)$ distribution with cdf

$$F_\beta(Y) = 1 - \left(\frac{x_m}{Y}\right)^\alpha \qquad 5$$

The Pareto distribution can be written in the form of (1) such that $G(Y) = \log(Y)$, $\beta_1 = \log x_m$ and $\beta_2 = \frac{1}{\alpha}$.

Alternatively Z can be written as

$$Z = G\{F^{-1}[F_\beta(Y)]\} = \log\{F^{-1}[F_\beta(Y)]\} = -\log[1 - F_\beta(Y)] \qquad 6$$

where $G(Y) = \log Y$ and $F^{-1}(U) = (1 - U)^{-1}$ is the inverse of the cdf of the standard $\mathrm{Pareto}(1,1)$ distribution.

3. A general class of least squares estimators for location-scale families

3.1 Least Squares Estimators

Solving equation (1) for $G(Y)$ we have

$$G(Y) = \beta_1 + \beta_2 \cdot Z \qquad 7$$

Now let $y_1, y_2, \dots, y_n$ be an i.i.d. random sample from the distribution of Y, and let

$$z_i = \frac{G(y_i) - \beta_1}{\beta_2}, i = 1,2, \dots, n$$

be the transformed and standardized variates, in the manner of (1). Furthermore, let $y = [y_{(1)}, y_{(2)}, \dots, y_{(n)}]'$ be the vector of order statistics of the sample $y_1, y_2, \dots, y_n$ and similarly let $z = [z_{(1)}, z_{(2)}, \dots, z_{(n)}]'$ be the vector of order statistics of the standardized variates $z_1, z_2, \dots, z_n$. Then we can write

$$G(y) = \beta_1 + \beta_2 \cdot z \qquad 8$$

The expectation and covariance matrix of $G(Y)$ are respectively given by

$$E[G(y)] = \beta_1 + \beta_2 \cdot E(z) \qquad 9$$

and

$$Cov[G(y)] = \beta_2^2 \cdot Cov(z) = \beta_2^2 \cdot V \qquad 10$$

where $V = \mathrm{Cov}(z)$ is the covariance matrix of the order statistics of the standard variates $z_1, z_2, \dots, z_n$.

Writing $G(y) = E[G(y)] + e$, and using (9), we have the following general linear model for $G(y)$:

$$G(y) = \beta_1 + \beta_2 \cdot E(z) + e; \;\; Cov(e) = \beta_2^2 \cdot V \qquad 11$$

In matrix notation model (11) can be written as

$$G(y) = X\beta + e; \;\; Cov(e) = \beta_2^2 \cdot V \qquad 12$$

where the $n \times 2$ matrix X is given by

$$X = [1_n : E(\boldsymbol{z})] = \begin{pmatrix} 1 & E(z_{(1)}) \\ \vdots & \vdots \\ 1 & E(z_{(n)}) \end{pmatrix} \qquad 13$$

Under model (12) the generalized least squares (GLS) estimator $\hat{\beta}$ of β is given by

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}G(y) \qquad 14$$

Clearly $\hat{\beta}$ in (14) is the best (minimum variance) linear unbiased estimator for β (Mann, Schafer and Singpurwalla, 1974: 96-97). We call (14) a least squares estimator for the parameters in the β parameterization of the distribution.

Generalized least squares estimation of β, as described above, requires the calculation of the expectation and covariance matrix of the vector of order statistics $z = [z_{(1)}, z_{(2)}, \dots, z_{(n)}]'$. While these can be determined quite straightforwardly through simulation, certain approximations can obviate the calculation of E(z), or of V, or both. Thus, a general class of least squares estimators for β can be written as

$$\hat{\beta}[z_A, V_A] = (X_A' V_A^{-1} X_A)^{-1} X_A' V_A^{-1} G(y) \qquad 15$$

where V_A is some approximation of $V = Cov(z)$,

$$X_A = [1_n : z_A] = \begin{pmatrix} 1 & z_{A,1} \\ \vdots & \vdots \\ 1 & z_{A,n} \end{pmatrix} \qquad 16$$

and z_A is some approximation of E(z).

Often, approximations for z_A are derived from the expected values of the order statistics of a uniform distribution, given by

$$\begin{aligned} E(U) &= E\{F_\beta(y)\} = m \\ &= [m_1, m_2, \ldots, m_n]' \\ &= \left[\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1}\right]' \end{aligned} \qquad 17$$

(Balakrishnan and Cohen, 1990: 31). For example, one approximation z_A of E(z) for the Weibull and Pareto distributions can be obtained by substituting m as defined in (17) into $F_\beta(Y)$ in equations (4) and (6), respectively. An example of the use of such an approximation is the Bergman estimator for the Weibull distribution (Bergman, 1986).

Examples of approximations V_A for V are the identity matrix, which generates a simple least squares estimator $\hat{\beta}[z_A, I]$, and the diagonal of $V = Cov(z)$, which generates a weighted least squares estimator.

3.2 Inverse Least Squares Estimators

Solving for E(z) in the linear model (11) we obtain

$$E(z) = -\frac{\beta_1}{\beta_2} + \frac{1}{\beta_2} \cdot G(y) - \frac{1}{\beta_2} \cdot e \qquad 18$$

Equation (18) is a "linear model" for E(z) which, in matrix notation, can be written as

$$E(z) = X_I \gamma + \tilde{e}; \;\; Cov(\tilde{e}) = V \qquad 19$$

where $\gamma = [\gamma_1, \gamma_2]'$, with $\gamma_1 = \frac{-\beta_1}{\beta_2}$, $\gamma_2 = \frac{1}{\beta_2}$, $\tilde{e} = -\frac{e}{\beta_2}$, and

$$X_I = [1_n : G(y)] = \begin{pmatrix} 1 & G(y_{(1)}) \\ \vdots & \vdots \\ 1 & G(y_{(n)}) \end{pmatrix} \qquad 20$$

By analogy with the GLS estimator (14), equation (19) suggests the following "generalized least squares" estimator for γ, namely

$$\hat{\gamma} = (X_I' V^{-1} X_I)^{-1} X_I' V^{-1} E(z) \qquad 21$$

Of course (19) does not satisfy the requirements for a linear model for the "dependent" variable E(z): In particular, $E[E(z)] = E(z) \neq X_I\gamma$, and $Cov(E(z)) = 0 \neq V$. Thus the least squares estimator $\hat{\gamma}$ in (19) is not unbiased in general.

We call (21) a least squares estimator for the parameters in the γ parameterization of the distribution. Similarly to the general class of least squares estimators (15) for β, a general class of least squares estimators for γ can be written as

$$\hat{\gamma}[z_A, V_A] = (X_I' V_A^{-1} X_I)^{-1} X_I' V_A^{-1} z_A \qquad 22$$

Here X_I is defined as in (20), and as before in (15), V_A is some approximation of $V = Cov(z)$ and z_A is some approximation of E(z).

3.3 Least Squares Estimation for Censored Samples

An important advantage of using least squares estimators based on order statistics is that certain types of censoring, namely left censoring, right censoring and random censoring where the order positions of the censored observations are known, can be easily handled. In these cases of censoring the general least squares estimator (15) can quite simply be modified as follows: the censored observations are removed from $G(y)$, the corresponding rows are removed from the matrix X_A, and the corresponding rows and columns are removed from the matrix V_A. The general inverse least squares estimator (21) is modified similarly for censored samples.

4. Application to the Weibull and Pareto distributions

4.1 Weibull Distribution

Let $y_1, y_2, \dots, y_n$ be an i.i.d. sample from the Weibull(λ, k) distribution. Using result (4) the general linear model (11) for the vector of log-transformed order statistics $G(y) = \log(y) = [\log(y_{(1)}), \log(y_{(2)}), \dots, \log(y_{(n)})]'$ is

$$\log(y) = \beta_1 + \beta_2 \cdot E(z) + e;\ \ Cov(e) = \beta_2^2 V \qquad 23$$

where $\beta_1 = \log\lambda$, $\beta_2 = \frac{1}{k}$, and $z = k\log(y) - k\log\lambda = \frac{\log(y)-\beta_1}{\beta_2}$.

Furthermore, solving equation (23) for $E(z)$ we obtain the "inverse" linear model for $E(z)$

$$E(z) = \gamma_1 + \gamma_2 \cdot \log(y) + \tilde{e};\ \ Cov(\tilde{e}) = V \qquad 24$$

where $\gamma_1 = -\frac{\beta_1}{\beta_2} = -k\cdot\log\lambda$, $\gamma_2 = \frac{1}{\beta_2} = k$, and $\tilde{e} = -\frac{e}{\beta_2} = k\cdot e$. Thus the parameters of the Weibull distribution can also be determined through "inverse" least squares regression of $E(z)$ against $\log(y)$.

Quantile Estimation

We consider the problem of estimating $\log(q_{1-\alpha,\beta}) = \log[F_\beta^{-1}(1-\alpha)]$, which is the logarithm of the $1-\alpha$ quantile $q_{1-\alpha,\beta}$ of the Weibull distribution with parameters $\beta = [\beta_1, \beta_2]'$. From (2) we have

$$\log(q_{1-\alpha,\beta}) = \beta_1 + \beta_2 \cdot \log[-\log(\alpha)]$$
$$= \beta_1 + \beta_2 \cdot \log(q_{1-\alpha})$$

where $q_{1-\alpha} = -\log(\alpha)$ is the $1-\alpha$ quantile of the standard Weibull distribution with parameters $\beta = [0,1]'$ (that is, of the standard exponential distribution). Thus the logarithm of the $1-\alpha$ quantile of the Weibull distribution with parameters $\beta = [\beta_1, \beta_2]'$ is a linear combination of the parameters $\beta_1$and β_2. We estimate $\log(q_{1-\alpha,\beta})$ as

$$\log(\hat{q}_{1-\alpha,\beta}) = \log(q_{1-\alpha,\hat{\beta}}) = \hat{\beta}_1 + \hat{\beta}_2 \cdot \log(q_{1-\alpha}) \qquad 25$$

where the parameters β_1 and $\beta_2$are estimated using an estimator from the general class (15) of least squares estimators.

4.2 Pareto Distribution

Similar to the Weibull distribution the general linear model for the Pareto is

$$\log(y) = \beta_1 + \beta_2 \cdot E(z) + e;\ \ Cov(e) = \beta_2^2 V \qquad 26$$

where $\beta_1 = \log x_m$, $\beta_2 = \frac{1}{\alpha}$ and $z = \alpha\log(y) - \alpha\log x_m = \frac{\log(y)-\beta_1}{\beta_2}$.

Furthermore, solving equation (26) for $E(z)$ we obtain the "inverse" linear as

$$E(z) = \gamma_1 + \gamma_2 \cdot \log(y) + \tilde{e};\ \ Cov(\tilde{e}) = V \qquad 27$$

where $\gamma_1 = -\frac{\beta_1}{\beta_2} = -\alpha\cdot\log x_m$, $\gamma_2 = \frac{1}{\beta_2} = \alpha$, and $\tilde{e} = -\frac{e}{\beta_2} = \alpha\cdot e$. Thus the parameters of the Pareto distribution can also be determined through "inverse" least squares regression.

Quantile Estimation

The same problem as for the Weibull distribution is being considered. From (5) we have

$$log(q_{1-\alpha,\beta}) = \beta_1 + \beta_2 \cdot \log(\alpha)$$
$$= \beta_1 + \beta_2 \cdot log(q_{1-\alpha})$$

where $q_{1-\alpha} = \frac{1}{\alpha}$ is the $1-\alpha$ quantile of the standard Pareto distribution with parameters $\beta = [0,1]'$. Thus the logarithm of the $1-\alpha$ quantile of the Pareto distribution with parameters $\beta = [\beta_1, \beta_2]'$ is a linear combination of the parameters $\beta_1$and β_2. We estimate $\log(q_{1-\alpha,\beta})$ similarly to the quantile estimate of the Weibull distribution, given in (25).

5. Simulation study

A simulation study to compare different least squares estimators and the maximum likelihood estimator was performed. The root mean square error (RMSE) for each estimator was determined.

5.1 Estimators Studied

The full simulation study involved 13 least squares estimators and the maximum likelihood estimator. Only the following seven estimators are reported here:

a) Generalized Least Squares (GLS) = $\hat{\beta}[\mathrm{E(z)}, V]$.
b) Simple Least Squares (SLS) = $\hat{\beta}[\mathrm{E(z)}, I_n]$.
c) The Best Linear Invariant (BLI) estimator as given by Mann, Schafer and Sinpurwalla (1974, p. 98).
d) Approximate Generalized Least Squares (AGLS) = $\hat{\beta}[z_A, V]$, where

$$X_A = [1_n : z_A] = \begin{pmatrix} 1 & z_{A,1} \\ \vdots & \vdots \\ 1 & z_{A,n} \end{pmatrix}$$

z_A for the Weibull distribution and Pareto distribution was determined by substituting m as defined in (17) into $F_\beta(Y)$ into (4) and (6) respectively.

e) Inverse Generalized Least Squares (IGLS) = $\hat{\gamma}[\mathrm{E(z)}, \mathrm{V}]$.
f) Inverse Simple Least Squares (ISLS) = $\hat{\gamma}[E(z), I_n]$.
g) Maximum Likelihood Estimator (MLE).

5.2 Sample Sizes and Parameter Combinations Studied

For each distribution a small sample size (n=15), moderate sample sizes (n=30 and n=50), and a large sample size (n=100) were studied.

Four different sets of parameters of the Weibull and Pareto distributions where studied, namely $\beta = [0,1]'$, $\beta = [0,4]'$, $\beta = [4,4]'$ and $\beta = [4,1]'$.

5.3 Simulation Results

Simulation results for the Weibull$(\lambda = 1, k = 1)$ and Pareto$(x_m = 1, \alpha = 1)$ distributions are being presented.

TABLE 1
Weibull Distribution: RMSE of Various Least Squares Estimators $(\mathbf{n = 30})$

Parameters	Estimators						
	GLS	BLI	SLS	AGLS	IGLS	ISLS	MLE
$\beta_1 = 0$	0.190	0.190	0.193	0.190	0.314	0.196	0.191
$\beta_2 = 1$	0.147	0.146	0.184	0.169	0.660	0.205	0.146
$\gamma_1 = 0$	0.196	0.200	0.202	0.184	0.187	0.196	0.201
$\gamma_2 = 1$	0.157	0.165	0.196	0.151	0.368	0.185	0.167
Quantiles							
$\log(q_{1-\alpha}) = 1.305$	0.228	0.227	0.288	0.248	0.632	0.327	0.229

$\boldsymbol{\alpha = 0.025}$ $\boldsymbol{\lambda = 1, k = 1}$ RMSE: Root Mean Squared Error

TABLE 2
Pareto Distribution: RMSE of Various Least Squares Estimators Pareto $(\mathbf{n = 30})$

Parameters	Estimators						
	GLS	BLI	SLS	AGLS	IGLS	ISLS	MLE
$\beta_1 = 0$	0.034	0.034	0.159	0.034	0.048	0.190	0.047
$\beta_2 = 1$	0.186	0.183	0.244	0.205	1.051	0.271	0.183
$\gamma_1 = 0$	0.035	0.036	0.166	0.033	0.024	0.157	0.051
$\gamma_2 = 1$	0.204	0.219	0.276	0.191	0.470	0.256	0.219
Quantiles							
$\log(q_{1-\alpha}) = 3.689$	0.682	0.670	0.805	0.749	3.843	0.871	0.671

$\boldsymbol{\alpha = 0.025}$ $\boldsymbol{x_m = 1, \alpha = 1}$ RMSE: Root Mean Squared Error

6. Discussion

6.1 Weibull Distribution

Comparing the different estimators with respect to RMSE, the GLS and BLI estimators are best for estimating β_1 and β_2. The BLI estimator has a slight advantage over the GLS estimator, largely associated with the estimate of β_2. The RMSEs for the BLI and maximum likelihood (ML) estimators are very similar. The "natural" scale for quantile estimation is the logarithmic scale, in the sense that $\log(q_{1-\alpha})$ is a linear combination of β_1 and β_2. Furthermore, the sampling distribution of the estimate of $\log(q_{1-\alpha})$ is more symmetric than that of the estimate of $q_{1-\alpha}$, so that the RMSE seems more suitable for judging the quality of an estimator on the log-quantile scale than on the quantile scale itself. Since $\log(q_{1-\alpha})$ is a linear combination of β_1 and β_2, it is not surprising that the GLS and BLI estimators proved to be best for estimating the log-quantile. The inverse least squares estimators directly provide estimates for the parameters γ_1 and γ_2. However, perhaps somewhat surprisingly, not one of the "inverse" least squares estimators, but the Approximate Generalized Least Squares (AGLS) estimator performed best for estimating γ_1 and γ_2.

6.2 Pareto Distribution

Similar to the results for the Weibull distribution, the GLS and BLI estimator are best for estimating β_1 and β_2. Again, BLI has a slight advantage over GLS, largely associated with the estimate of β_2. Interestingly the MLE for β_1 is not competitive with the BLI and GLS estimators, which is in contrast to the Weibull distribution. However, the BLI and MLE estimators have similar RMSE for estimating β_2. As for the Weibull $\log(q_{1-\alpha})$ is a linear combination of β_1 and β_2, It is therefore not surprising that GLS and BLI proved to be the best estimators for the log-quantile. Similarly to the Weibull distribution, AGLS is the best estimator for γ_1 and γ_2.

7. Conclusions

Our investigation shows that which estimator is best depends on which parameterization is chosen for estimation (β or γ parameterization), or on which scale relevant quantities such as quantiles are estimated. However, for a *given* parameterization or scale, the best estimators are the same for the two distributions investigated, and the simulation results are consistent across the distribution parameters and sample sizes studied. In summary:

- The GLS and BLI estimators are best for estimation of the β parameters, and for quantile estimation, for both the Weibull and Pareto distribution.
- The AGLS estimators is best for estimation of the γ parameters for both the Weibull and Pareto distribution..

Overall the study results suggest that the use of the exact covariance structure of the order statistics, that is, use of GLS (or BLI) estimation, increases the performance of least squares estimators and resulted in better estimates. In all cases, the best performing least squares estimators were competitive with, or slightly superior to, the maximum likelihood estimator.

8. References

Balakrishnan, N. and Cohen, A.C. (1990) Order Statistics and Inference: Estimation Methods, 1st edition, Academic Press, Inc.

Bergman, B. (1986) 'Estimation of Weibull parameters using a weight function', Journal of Materials Science Letters, no. 5, pp. 611-614.

Mann, N.R., Schafer, R.E. and Singpurwalla, N.D. 1974. *Methods for Statistical Analysis of Reliability & Life Data*, 1st edition, Wiley-Interscience

Zhang, L.F., Xie, M. and Tang, L.C. 2006. 'A Study of Two Estiamtion Approaches for Parameters of Weibull Distribution Based on WPP', *Reliability Engineering and System Safety*, no. 92, May, pp. 360-368.