JSS 31 (1988) 151-158

# FITTING THE TRUNCATED PARETO DISTRIBUTION TO LOSS DISTRIBUTIONS

By Albert V. Boyd

## 1. INTRODUCTION

Hogg and Klugman[1] use the truncated Pareto distribution with probability density function

$$f(x;\alpha,\lambda)=\frac{\alpha(\lambda+\delta)^{\alpha}}{(x+\lambda)^{\alpha+1}},\ (\delta<x<\infty),$$

where $\delta\geqslant 0$ is specified and $\alpha>0$ and $\lambda>0$ are unknown parameters, to describe insurance claims. This is fitted first of all by the method of moments, using the estimators

$$\tilde{\alpha}=\frac{2s^2}{s^2-(\bar{x}-\delta)^2}$$

and

$$\tilde{\lambda}+\delta=\frac{(\bar{x}-\delta)\{s^2+(\bar{x}-\delta)^2\}}{s^2-(\bar{x}-\delta)^2}$$

where $\bar{x}$ is the mean of a simple random sample, and the (biased) variance

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

The authors then suggest, on pp. 113–16, that these estimates be used as starting values in a Newton iteration to get the maximum likelihood estimates of the parameters, but this technique can fail as a result of convergence problems. The object of this note is to show that this has led Hogg and Klugman to underestimate seriously the area in the tail of a fitted loss distribution, and to discuss a method of circumventing this difficulty.

## 2. EXISTENCE OF A MAXIMUM LIKELIHOOD SOLUTION

For a simple random sample $x_1, x_2, \ldots, x_n$ the likelihood

$$L=\alpha^n(\lambda+\delta)^{n\alpha}\prod_{i=1}^{n}(\lambda+x_i)^{-\alpha-1}=(\alpha\theta^{\alpha})^n\prod_{i=1}^{n}(\theta+x_i-\delta)^{-\alpha-1}$$

satisfies

$$g_1(\alpha, \theta) \equiv \frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + n \ln \theta - \sum_{i=1}^{n} \ln(\theta + x_i - \delta) \tag{1}$$

$$g_2(\alpha,\theta) \equiv \frac{\partial \ln L}{\partial \theta} = \frac{n\alpha}{\theta} - (\alpha + 1) \sum_{i=1}^{n} \frac{1}{\theta + x_i - \delta} \tag{2}$$

where $\theta = \lambda + \delta$, and we need to solve simultaneously the equations $g_1 = 0$ and $g_2 = 0$.

If we first fix $\theta > 0$ then

$$\frac{\partial g_1}{\partial \alpha} = -\frac{n}{\alpha^2} < 0;$$

$$g_1(\alpha, \theta) \to \infty \text{ and } \ln L \to -\infty \text{ as } \alpha \to 0+; \text{ and}$$

$$\ln L = n \ln \alpha - \sum_{i=1}^{n} \ln(\theta + x_i - \delta) + \alpha \sum_{i=1}^{n} \ln \frac{\theta}{\theta + x_i - \delta}$$

$$\to -\infty \text{ as } \alpha \to \infty, \text{ since } x_i > \delta \text{ for each } i.$$

It follows that then $\ln L$ has exactly one relative maximum for some positive $\alpha$.

Next, fix $\alpha > 0$. Then

$$\text{as } \theta \to 0+, \text{ so } g_2(\alpha, \theta) \sim \frac{n\alpha}{\theta} \to +\infty;$$

$$\text{as } \theta \to +\infty, \text{ so } g_2(\alpha, \theta) \sim \frac{n\alpha}{\theta} - (\alpha + 1)\frac{n}{\theta} \to 0-;$$

Since $g_2$ is a continuous function of $\theta$ for $\theta > 0$, therefore $g_2 = 0$ for some $\theta > 0$; and if $\theta_0$ is the least such value of $\theta$ then, for $\theta > \theta_0$,

$$\begin{aligned} g_2(\alpha,\theta) &= \frac{\alpha + 1}{\theta} \left\{ \frac{n\alpha}{\alpha + 1} - \sum_{i=1}^{n} \frac{1}{1 + \frac{x_i - \delta}{\theta}} \right\} \\ &< \frac{\alpha + 1}{\theta} \left\{ \frac{n\alpha}{\alpha + 1} - \sum_{i=1}^{n} \frac{1}{1 + \frac{x_i - \delta}{\theta_0}} \right\} \\ &= \frac{\theta_0}{\theta} g_2(\alpha,\theta_0) = 0. \end{aligned}$$

Hence for each fixed $\alpha > 0$ there is only one positive solution of $g_2(\alpha,\theta) = 0$ and, as $g_2$ is changing sign from positive to negative at this point then $\ln L$, as a function of $\theta$, has a relative maximum there.

In practical applications when $\ln L$ is plotted as a function of $\alpha$ and $\theta$ it is found that the loci of maxima of $\ln L$ for fixed $\theta$ and for fixed $\alpha$ are usually of the form shown in Figure 1, where they correspond to curves lying on a long ridge of the surface $\ln L(\alpha,\theta)$ as a function of $\alpha$ and $\theta$, and these curves intersect at a point where $\alpha > 0$, $\theta > 0$ and $L$ has a relative maximum there. (For an example where this is not the case see the illustration $\delta = 1$, $x_1 = x_2 = 2$, $x_3 = 3$ discussed below.)

Hogg and Klugman warn readers that to reach the peak by Newton's successive approximation technique it is important to have good preliminary guesses of $\alpha$ and $\theta$, and they suggest that $\tilde{\alpha}$ and $\tilde{\theta} = \tilde{\lambda} + \delta$ are often convenient starting values.

In their example (p. 64) the simple random sample is

| $x_i$, *loss (in \$10⁶) due to wind-related catastrophes* | $f_i$, *frequency in 1977* | $x_i$ | $f_i$ |
|---|---|---|---|
| 2 | 12 | 17 | 1 |
| 3 | 4 | 22 | 1 |
| 4 | 3 | 23 | 1 |
| 5 | 4 | 24 | 2 |
| 6 | 4 | 25 | 1 |
| 8 | 2 | 27 | 1 |
| 9 | 1 | 32 | 1 |
| 15 | 1 | 43 | 1 |

The method of moments estimators are $\tilde{\alpha} = 4{\cdot}809$, $\tilde{\lambda} = 27{\cdot}921$ and, with $\delta = 1{\cdot}5$, this makes $\tilde{\theta} = \tilde{\lambda} + 1{\cdot}5 = 29{\cdot}421$. Hogg and Klugman discuss (pp. 115–16) the maximum likelihood procedure, starting from the moments estimators $\alpha$ and $\lambda$, and give $\alpha = 5{\cdot}084$ and $\theta = 30{\cdot}498$; but at this point $g_2 = -{\cdot}043$ which is not close enough to 0, and the Newton iteration diverges when started from $\tilde{\alpha}$ and $\tilde{\lambda}$.

By using an alternative optimization technique such as the method of Nelder and Mead discussed on pp. 81–4[2], and for which Bunday[3] has provided a BASIC program, it can be found that $\ln L$ attains its maximum value of $-117{\cdot}7359858$ at $\hat{\alpha} = 1{\cdot}455688$ and $\hat{\lambda} = 3{\cdot}613672$. For comparison we mention that $\ln L = -119{\cdot}54605$ at $\tilde{\alpha} = 4{\cdot}809$ and $\tilde{\theta} = 29{\cdot}421$, while at $\alpha = 5{\cdot}084$ and $\theta = 30{\cdot}498$ the value of $\ln L$ is $-119{\cdot}58179$.

The discrepancy in the values of $\ln L$ might not appear to be large, but in applications it can be serious. Thus Hogg and Klugman use the fitted truncated Pareto distribution to estimate the probability of getting a loss exceeding \$29,500,000, and find this to be

$$h(\alpha,\lambda) = Pr(X > 29{\cdot}5) = \left(\frac{\lambda + 1{\cdot}5}{\lambda + 29{\cdot}5}\right)^{\alpha}.$$

With the incorrect values $\alpha = 5{\cdot}084$ and $\lambda = 28{\cdot}998$ this gives a point estimate of $h$ as ${\cdot}036$ and an approximate 95% confidence interval as 0 to ${\cdot}084$. With the correct

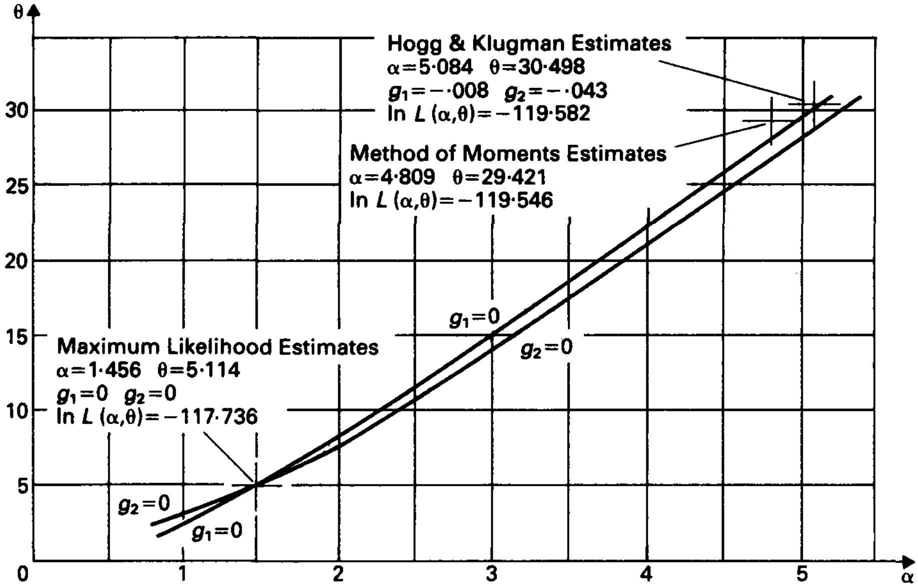


Figure 1. *Projection of the Log Likelihood Function ln* L *(α,θ) onto the θ,α plane showing the locus of* $g_1 = g_2 = O$ *and the positions of alternative estimates.*

maximum likelihood estimates, however, $h(\hat{\alpha},\hat{\lambda}) = {\cdot}0659$ and making the relevant changes to the argument on pp. 116–18[1] gives an approximate 95% confidence interval for $h$ as ·002 to ·130.

Because of the appreciable discrepancy between the two sets of estimators it is desirable to have a better method of starting the search for the maximum likelihood estimators. Two such methods will now be considered.

## 3. OBTAINING FIRST APPROXIMATIONS TO MAXIMUM LIKELIHOOD ESTIMATES

### *Method A*

On equating $g_1(\alpha,\theta)$ and $g_2(\alpha,\theta)$ from (1) and (2) to zero we get

$$\frac{1}{\alpha} = -\ln\theta + \frac{1}{n}s_2(\theta) \text{ and } \alpha = \frac{s_1(\theta)}{\dfrac{n}{\theta} - s_1(\theta)}$$

$$\text{where } s_1(\theta) = \sum_{i=1}^{n} \frac{1}{x_i + \theta - \delta} \text{ and } s_2(\theta) = \sum_{i=1}^{n} \ln(x_i + \theta - \delta).$$

Eliminating $\alpha$ leads to $F(\theta) = 0$ where

$$F(\theta) \equiv \frac{n}{\theta} - s_1(\theta)\left\{1 + \frac{s_2(\theta)}{n} - \ln\theta\right\}$$

$$= \frac{n}{\theta} - \frac{1}{\theta}\left\{\sum_{i=1}^{n} \frac{1}{1 + \frac{x_i - \delta}{\theta}}\right\}\left\{1 + \frac{1}{n}\sum_{i=1}^{n} \ln\left(1 + \frac{x_i - \delta}{\theta}\right)\right\}$$

$$= \frac{n}{\theta} - \frac{1}{\theta}\left\{n - \frac{1}{\theta}\Sigma(x_i - \delta) + \frac{1}{\theta^2}\Sigma(x_i - \delta)^2 + \text{terms in } \frac{1}{\theta^3} \text{ etc}\right\}$$

$$\times \left\{1 + \frac{1}{n\theta}\Sigma(x_i - \delta) - \frac{1}{2n\theta^2}\Sigma(x_i - \delta)^2 + \text{terms in } \frac{1}{\theta^3} \text{ etc}\right\}$$

$$= \frac{-1}{2\theta^3}\left\{\Sigma(x_i - \delta)^2 - \frac{2}{n}[\Sigma(x_i - \delta)]^2\right\} + \text{terms in } \frac{1}{\theta^4} \text{ etc}$$

$$= -\frac{n}{2\theta^3}\left\{s_x^2 - (\bar{x} - \delta)^2\right\} + \text{terms in } \frac{1}{\theta^4}$$

and $s_x^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$.

Hence if $s_x > |\bar{x} - \delta| = \bar{x} - \delta$ then $F(\theta) \to 0-$ as $\theta \to \infty$. As $\theta \to 0+$, so $s_1(\theta) \to$ constant and constant $\times |\ln \theta|$ tends to infinity less rapidly than $1/\theta$; and so

$$F(\theta) \sim \frac{n}{\theta} \to +\infty.$$

In the case $s_x > \bar{x} - \delta$ the graph of $F(\theta)$ must therefore cross the $\theta$-axis for some $\theta_0 > 0$; since

$$\sum \frac{1}{x_i + \theta_0 - \delta} < \sum \frac{1}{\theta_0}$$

it then follows that

$$\alpha = \frac{s_1(\theta_0)}{\frac{n}{\theta_0} - s_1(\theta_0)} > 0.$$

When $s_x \leqslant \bar{x} - \delta$ the above analysis does not guarantee the existence of a positive solution of $F(\theta) = 0$ and, in fact, the concentration of the $x$-values about their mean suggests that a heavy-tailed distribution, such as the Pareto, no longer provides a suitable description of the data.

For a simple illustration of this, one may take the values $\delta=1$, $x_1=x_2=2$, $x_3=3$. In the plane of $\alpha$ and $\theta$ the locus of maxima of $\ln L$ for fixed $\theta$ can be shown to approach $\theta=4\alpha/3-3/4$ asymptotically as $\theta$ increases; the locus of maxima of $\ln L$ for fixed $\alpha$ approaches $\theta=4\alpha/3-1/6$ asymptotically as $\alpha$ increases; and on both these asymptotes $\ln L$ increases to the limiting value $3 \ln 3/4-3$ as $\theta$ or $\alpha$ tends to infinity, so that there are no finite maximum likelihood estimates of the parameters for the truncated Pareto distribution.

In this case, if $k$ is any constant,

$$f\left(x;\alpha,\frac{4\alpha}{3}+k\right)=\frac{3}{4}\cdot\frac{\left(1+\dfrac{k+1}{4\alpha/3}\right)^{\alpha}}{\left(1+\dfrac{k+x}{4\alpha/3}\right)^{\alpha+1}}\quad\text{for } x>1$$

$$\sim\frac{3}{4}e^{-\frac{3}{4}(x-1)}\quad\text{as } \alpha\to\infty,$$

which suggests that the Pareto distribution should be replaced by an exponential one. It is easily checked that, when the density function is taken as $ce^{-c(x-1)}$ for $x>1$, then the maximum likelihood estimate of $c$ is $3/4$.

*Method B*

If $t=1/\theta$ and $h(t)=-F(1/t)/s_1(1/t)$ then it is easily verified that

(i) solving the equation $F(\theta)=0$ is equivalent to solving $h(t)=0$ where

$$h(t)=1+\frac{1}{n}s_2\left(\frac{1}{t}\right)+\ln t-\frac{nt}{s_1(1/t)};$$

(ii) for small values of $|t|$ the Maclaurin expansion of $h(t)$ is

$$h(t)=\frac{t^2}{2n}\left\{\sum_{i=1}^{n}(c_i-\bar{c})^2-n\bar{c}^2\right\}+\text{higher powers of } t,$$

where $c_i=x_i-\delta$; and

(iii) when $t\to+\infty$, $h(t)\sim\dfrac{-nt}{\sum_{i=1}^{n}1/c_i}$.

It follows that if $s_x>\bar{x}-\delta$ then

$$\sum_{i=1}^{n} (c_i - \bar{c})^2 - n\bar{c}^2 > 0,$$

so that in the neighbourhood of $t=0$ the graph of $h(t)$ behaves like a parabola with a minimum turning point at the origin; and as $t \to \infty$ so $h(t) \to -\infty$. There is therefore a positive solution $t=t_0$ of $h(t)=0$, and hence a solution $\hat{\theta}=1/t_0$ of the maximum likelihood equations, $\hat{\alpha}$ can then be found from $s_1(\hat{\theta})$ and $s_2(\hat{\theta})$ as in Method A.

If $s_x \leqslant \bar{x}-\delta$ then, as with the function $F(\theta)$ in Method A, we do not necessarily get a solution of the maximum likelihood equations, and some other form of distribution should be fitted to the data.

## 4. COMPARISON OF METHODS A AND B

As will be seen from Figures 2 and 3 which correspond to the data of Hogg and Klugman's example, both methods are suitable for attack by the Newton–Raphson technique for a single variable with a suitable starting value, since $s_x = 10{\cdot}108 > 7{\cdot}725 = \bar{x}-\delta$.

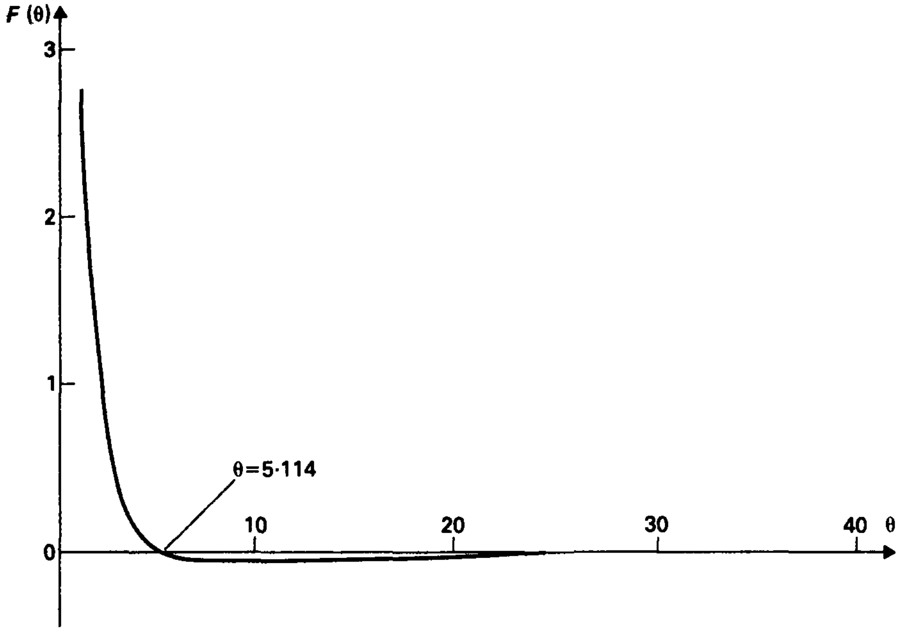


Figure 2. *Method A—Plot of* F($\theta$) *against* $\theta$ *showing the maximum likelihood estimate of* $\theta$.

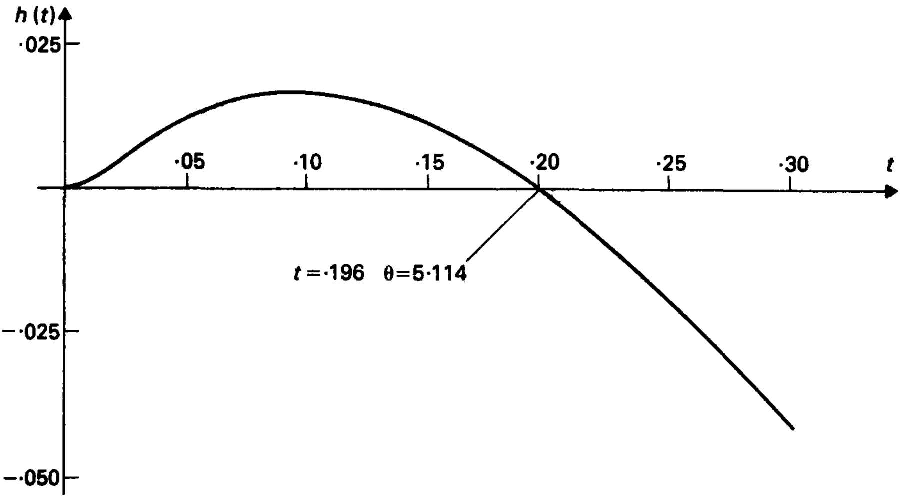


Figure 3. *Method B—Plot of* h(t) *against t showing the maximum likelihood estimate of $\theta$.*

In the case of method A any value of $\theta$ for which $F(\theta) > 0$, and certain values for which $F(\theta) < 0$ and $F'(\theta) < 0$, could usefully be taken as starting values; for method B any value of $t$ for which $h'(t) < 0$ will lead to convergence of the process. Numerical evidence suggests that the convergence is sometimes slightly faster with method A, but that with method B it is a bit easier to hit on a suitable starting value when the estimators given by the method of moments are used to initiate a search.

## 5. REFERENCES

(1) Hogg, R. V. and Klugman, S. A. (1984) *Loss Distributions*. Wiley.
(2) Walsh, G. R. (1975) *Methods of Optimisation*. Wiley.
(3) Bunday, B. D. (1984) *Basic Optimisation Methods*. Arnold, London.