Estimation of generalized Pareto distribution

Joan del Castillo, Jalila Daoudi

Accepted Manuscript

Estimation of generalized Pareto distribution

Joan del Castillo, Jalila Daoudi

PII: S0167-7152(08)00498-7
DOI: 10.1016/j.spl.2008.10.021
Reference: STAPRO 5250

To appear in: *Statistics and Probability Letters*

Received date: 2 October 2008
Accepted date: 20 October 2008

Please cite this article as: del Castillo, J., Daoudi, J., Estimation of generalized Pareto distribution. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.10.021

ESTIMATION OF GENERALIZED PARETO DISTRIBUTION

JOAN DEL CASTILLO

ABSTRACT. Research partially supported by the Spanish Ministerio de Educacion y Ciencia, grant MTM2006-01477.: This paper provides precise arguments to explain the anomalous behavior of the likelihood surface when sampling from the generalized Pareto distribution for small or moderate samples. The behavior of the profile-likelihood function is characterized in terms of the empirical coefficient of variation. A sufficient condition is given for global maximum of the likelihood function of the Pareto distribution to be at a finite point.

Keywords: Heavy-tailed inference. Extreme value theory. The coefficient of variation.

1. Introduction

The Pareto distribution has long been used as a model for the tails of another long-tailed distribution, see Arnold (1983). Applications to risk management in finance and economics are now of increasing importance. Since Pickands (1975), it has been well known that the conditional distribution of any random variable over a high threshold is approximately generalized Pareto distribution (GPD), which includes the Pareto distribution, the exponential distribution and distributions with bounded support. These distributions are closely related to the extreme value theory (Coles, 2001, and Embrechts et al. 1997).

The GPD has been used by many authors to model excedances in several fields such as hydrology, insurance, finance and environmental science, see Finkenstadt and Rootzén (2003), Coles (2001) and Embrechts et al. (1997). In general, GPD can be applied to any situation in which the exponential distribution might be used but in which some robustness is required against heavier tailed or lighter tailed alternatives, see Van Montford and Witter (1985). The asymptotic behavior of maximum likelihood estimators was studied by Davison (1984) and Smith (1985). Nevertheless, there is evidence that numerical techniques for maximum likelihood estimation do not work well in small samples and other estimation methods have been proposed, see Castillo and Hadi (1997) and Hosking and Wallis (1987).

The paper provides precise arguments to explain the anomalous behavior of the likelihood surface when sampling from the GPD distribution (Davison and Smith, 1990, and Castillo and Hadi, 1997). In Section 2, Theorem 1 proves that the behavior of the profile-likelihood for GPD is characterized by the sign of the empirical coefficient of variation of the sample. Corollary 1 proves that the maximum of the likelihood function for the Pareto distribution is at a finite point, for samples in which the coefficient of variation is larger than 1. An example in the Appendix shows that a local maximum for the likelihood function of the GPD could not exist when the condition is not fulfilled.

Monte Carlo simulation in Section 3 raises the problem of mis-specification for small or moderate samples in GPD. It can be also explained from the difference between the theoretical and the empirical coefficients of variation for small samples. The practical relevance of these results is also discussed.

2. Main Results

The cumulative distribution function of GPD is

$$F(x) = 1 - (1 + \xi x/\psi)^{-1/\xi}, \tag{2.1}$$

where $\psi > 0$ and ξ are scale and shape parameters. For $\xi > 0$ the range of x is $x > 0$ and the GPD is just one of several forms of the usual *Pareto family of distribution* often called the Pareto distribution. For $\xi < 0$ the range of x is $0 < x < \psi/|\xi|$, then GPD have bounded support. The limit case $\xi = 0$ corresponds to the exponential distribution.

An alternative parameterization is $\sigma = \psi/|\xi|$ and $|\xi| = s\ \xi$, where $s = sign(\xi)$. Then, the probability density function for GPD is given by

$$f(x; \sigma, \xi) = \frac{1}{\sigma|\xi|}\left(1 + s\ \frac{x}{\sigma}\right)^{-(1+\xi)/\xi}, \tag{2.2}$$

for $\xi < 0$, the range of x is now $0 < x < \sigma$. Using this notation the two families of distributions corresponding to $\xi > 0$ and $\xi < 0$ can be studied at the same time.

Given a sample $\{x_i\}$ of size n, the log-likelihood function for GPD distribution, divided by the sample size, is

$$l(\sigma, \xi) = -\log(s\ \xi\sigma) - \frac{1+\xi}{\xi\ n}\sum_{i=1}^{n}\log(1 + s\ x_i/\sigma). \tag{2.3}$$

If $\xi < 0$ it is assumed $\sigma > M = \max\{x_i\}$, otherwise the likelihood is zero. In this case the likelihood may be made arbitrarily large as σ tends to M, so the maximum likelihood estimators are taken to be the values which yield a local maximum of (2.3), that often appears.

Maximum likelihood estimation of generalized Pareto parameters was discussed by Davison (1984) and Smith (1985). In particular, for large samples, maximum likelihood estimator exist and is asymptotically normal and efficient, provided that $-0.5 < \xi$. The restriction $-0.5 < \xi < 0.5$ is usually assumed for both practical and theoretical reasons, since GPD with $\xi < -0.5$ have finite end points and the probability density function is strictly positive at each endpoint, and GPD with $\xi > 0.5$ have infinite variance. When GPD is used as an alternative to the exponential distribution, values of ξ near 0 will be of greatest interest, because the exponential distribution is a GPD with $\xi = 0$.

For moderate or small samples, anomalous behavior of the likelihood surface can be encountered when sampling from the GPD distribution (Davison and Smith, 1990, and Castillo and Hadi, 1997). This will be explained in this paper from the coefficient of variation of the sample. For instance, the coefficient of variation for Pareto distribution ($\xi < 0.5$) is given by

$$\zeta = \sqrt{1/(1-2\xi)} > 1, \tag{2.4}$$

but for small samples the *empirical* coefficient of variation

$$cv = \sqrt{m_2 - m_1^2}/m_1, \tag{2.5}$$

where $m_k = \sum x_i^k/n$ are the sample moments, may be lower than 1. Theorem 1 and Corollary 1 below in this Section provide more precise arguments.

Equating to zero the derivative of $l(\sigma, \xi)$ in(2.3), with respect to ξ, we find $\hat{\xi} = \xi(\sigma)$, where

$$\xi(\sigma) \equiv \xi(\sigma, s) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + s\ x_i/\sigma). \tag{2.6}$$

The profile-likelihood is given by

$$l_p(\sigma, s) = -\log[s\ \xi(\sigma)\,\sigma] - \xi(\sigma) - 1. \tag{2.7}$$

Proposition 1. *The following limits hold:*

$$\lim_{\sigma\to\infty} \sigma\ \log(1 + x/\sigma) = x, \qquad \lim_{\sigma\to\infty} \sigma x/(\sigma + x) = x,$$
$$\lim_{\sigma\to\infty} \sigma^2\ (\log(1 + x/\sigma) - x/(\sigma + x)) = x^2/2.$$

Proof. It is an elementary exercise in calculus, using series expansion. ∎

Proposition 2. *Let $l_p(\sigma, s)$ be the profile-likelihood, defined by* (2.7) *and let $\overline{x}$ be the sample mean, then*

$$\lim_{\sigma\to\infty} \xi(\sigma) = 0, \qquad \lim_{\sigma\to\infty} \sigma\xi(\sigma) = s\ \bar{x}.$$
$$l_0 \equiv \lim_{\sigma\to\infty} l_p(\sigma, s) = -\log(\overline{x}) - 1.$$

Proof. The first limit is straightforward. From Proposition 1 it follows:

$$\lim_{\sigma\to\infty} \sigma\xi(\sigma) = \lim_{\sigma\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sigma \log(1 + s\ x_i/\sigma) = s\ \bar{x}$$

This prove the second limit and hence, since $s^2 = 1$, the last limit follows. ∎

Remark 1. *The limit of the profile-likelihood, $l_p(\sigma, s)$, as σ tends to infinity, in Proposition* 2*, corresponds to the log-likelihood of the exponential distribution for the same sample. More precisely, the log-likelihood function for the exponential distribution, $\sigma e^{-\sigma x}$, divided by the sample size n, is $l(\sigma, 0) = \log\sigma - \sigma\ \overline{x}$ and the maximum likelihood estimator is $\widehat{\sigma} = 1/\overline{x}$ then,*

$$l(\widehat{\sigma}, 0) = l_0 = -\log(\overline{x}) - 1. \tag{2.8}$$

Theorem 1. *For the Pareto distribution $(\xi > 0)$, if the empirical coefficient of variation is $cv > 1$ then $l_p(\sigma, 1)$ is a monotonous decreasing function for sufficiently large σ, and if $cv < 1$ it is monotonous increasing. For the distributions with bounded support in GPD $(\xi < 0)$, if $cv > 1$ then $l_p(\sigma, -1)$ is a monotonous increasing function for sufficiently large σ, and if $cv < 1$ it is monotonous decreasing.*

Proof. The derivative of (2.7) with respect to σ is given by

$$-s\ l_p'(\sigma, s) = (\xi(\sigma) + \sigma\ \xi'(\sigma) + \sigma\ \xi(\sigma)\,\xi'(\sigma)) / (\sigma\ |\xi(\sigma)|).$$

and the sign of $-s\ l_p'(\sigma, s)$ is the same as the sign of $num(\sigma) = \xi(\sigma) + \sigma\ \xi'(\sigma) + \sigma\ \xi(\sigma)\,\xi'(\sigma)$, since $\sigma\ |\xi(\sigma)| > 0$.

Taking derivative with respect to σ in (2.6) it follows

$$\sigma\ \xi'(\sigma) = -\frac{1}{n}\sum_{i=1}^{n} s\ x_i/(\sigma + s\ x_i),$$

hence

$$\begin{aligned} num(\sigma) &= \frac{1}{n}\sum_{i=1}^{n}\left(\log\left(1 + s\ x_i/\sigma\right) - s\ x_i/(\sigma + s\ x_i)\right) - \\ &\left(\frac{1}{n}\sum_{i=1}^{n}\log\left(1 + s\ x_i/\sigma\right)\right)\left(\frac{1}{n}\sum_{i=1}^{n} s\ x_i/(\sigma + s\ x_i)\right). \end{aligned}$$

From Proposition 1, we have

$$\lim_{\sigma\to\infty}\left\{\sigma^2\ num(\sigma)\right\} = \frac{1}{2n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

Finally, note that $\frac{1}{2n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 > 0$ is equivalent to $\left(m_2 - m_1^2\right) > m_1^2$ and equivalent to $cv > 1$. ∎

A first consequence of Theorem 1 is obtained immediately. $l_p(\sigma, s)$ tends to l_0, from Proposition 2, and $l_p(\sigma, s)$ is a monotonous function for sufficiently large σ (Theorem 1) then these facts determine whether $l_p(\sigma, s)$ is greater or less than l_0. If $cv > 1$, $l_p(\sigma, 1)$ is a monotonous decreasing function and $l_p(\sigma, -1)$ is a monotonous increasing function, for sufficiently large σ, then

$$l_p(\sigma, 1) > l_0 > l_p(\sigma, -1)\,. \tag{2.9}$$

In the same way, if $cv < 1$ then

$$l_p(\sigma, 1) < l_0 < l_p(\sigma, -1)\,, \tag{2.10}$$

for sufficiently large σ.

Remark 2. *From* (2.6)*, as σ tends to infinite $|\xi(\sigma)|$ tends to zero. Hence, for ξ in a neighborhood of zero, the inequalities* (2.9) *and* (2.10) *show that if $cv > 1$ the Pareto distribution is more likely than the exponential distribution and if $cv < 1$ a bounded support distribution in GPD is more likely than the exponential distribution. These facts are numerically relevant for an algorithm to obtain the maximum likelihood estimator in GPD.*

Corollary 1. *Given a sample $\{x_i\}$ of positive numbers with an empirical coefficient of variation $cv > 1$, the likelihood function for the Pareto distribution has a global maximum at a finite point and the maximum is higher than the maximum for the likelihood function of the exponential distribution.*

Proof. For small values of σ we write

$$l_p(\sigma, 1) = -\left(\log\sigma + \xi(\sigma)\right) - \log\left[\xi(\sigma)\right] - 1,$$

$(\log\sigma + \xi(\sigma))$, tends to $\frac{1}{n}\sum_{i=1}^{n}\log(x_i)$ and $\log[\xi(\sigma)]$ goes to ∞. This proves

$$\lim_{\sigma\to 0} l_p(\sigma, 1) = -\infty.$$

Since $l_p(\sigma, 1)$ is a continuous and monotonous decreasing function for sufficiently large σ (Theorem 1) the last limit prove that a global maximum exists. Inequality (2.9) shows that it is up the maximum for the exponential distribution. ∎

When the coefficient of variation of the sample is $cv < 1$ there may be no maximum likelihood estimator for the Pareto distribution and neither is there a local maximum for the parameter space of the bounded support distributions in the GPD. In the Appendix we give a simple example of this situation.

If $l_p(\sigma, -1)$ has a local minimum, as is usual, and $cv < 1$ then Theorem 1 proves that there is a local maximum for the likelihood function, since $l_p(\sigma, -1)$ increase on the right side of the minimum and is monotonous decreasing for large values of σ.

3. Discussion

Analytical maximization of the log-likelihood for GPD is not possible, so numerical techniques are required taking care to avoid numerical instabilities when ξ near zero (Coles, 2001, pp 81). Theorem 1 clarifies the behavior of the likelihood function in terms of the coefficient of variation of the sample. If $cv > 1$ the Pareto distribution is more likely than a bounded support distribution in a neighborhood of zero, and if $cv < 1$ a bounded support distribution is more likely than a Pareto distribution. The numerical algorithms have to consider the cv sign of the sample

Corollary 1 proves that the likelihood function for the Pareto distribution has a global maximum at a finite point for samples in which $cv > 1$, so it is extremely simple to find it numerically. This is a sufficient condition and we also believe, from numerical experiments, that it is necessary, although we are not able to prove it.

Hosking and Wallis (1987, pp 343) say "we conclude that the vast majority of failures of the algorithms are caused by the nonexistence of a local maximum of the likelihood function rather than by failure of our algorithm to find a local maximum". We agree with them. Moreover, the example given in the Appendix proves the nonexistence of a local maximum for a particular sample. Now it is clear that the nonexistence of a local maximum it is possible. This problem increases when ξ decreases, specially for the bounded support distributions in GPD, as Hosking and Wallis (1987) showed.

Sampling from Pareto distribution in GPD shows another problem. A simulation experiment was run to compute mis-specification for sample sizes $n = 15, 25, 50, 100$ and shape parameter $\xi = 0.1, 0.2, 0.3, 0.4$. The scale parameter σ was set to 1, since the model (2.2) is invariant under scale changes. For each combination of values of n and ξ, $50,000$ random samples were generated from the Pareto distribution and the number of times the parameter estimates $\hat{\xi}$ to be positive, negative or that the algorithm does not converge are reported. It is noted from (2.4) that the theoretical coefficient of variation for Pareto distribution is $\zeta > 1$, but for small samples the empirical coefficient of variation, cv, may be lower than 1. Table 1 shows that for $\xi = 0.3$ (a distribution with infinite kurtosis) and sample size $n = 15$, 29% of cases that lead to a wrong decision, $\hat{\xi} < 0$ (bounded support distribution), while in 4.8% cases the algorithm does not converge. The problem remains for larger samples. For $\xi = 0.1$ and sample size $n = 100$, 24.9% of cases have $\hat{\xi} < 0$. However, if we assume Pareto distribution without considering the global GPD model (with the bounded support distributions), then samples with $cv > 1$ lead to Pareto distribution and samples with $cv < 1$ lead to the exponential distribution, $\xi = 0$, in both cases the support for the distribution is $(0, \infty)$.

In the context of heavy-tailed inference assuming that the true distribution has support in $(0, \infty)$, an alternative model for samples with $cv < 1$ may be truncated normal distribution. Castillo (1994) shows that the likelihood equations for truncated normal distribution have a solution if and only if the empirical coefficient of variation is $cv < 1$. Pareto distribution and truncated normal distribution are two complementary families of distributions the former with theoretical coefficient of variation $\zeta > 1$ and the latter with $\zeta < 1$. In both cases the exponential distribution is the limit distribution as ζ tends to 1.

ξ	0.4			0.3			0.2			0.1		
n	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC
15	73.4	23.1	3.5	66.2	29.0	4.8	57.1	36.4	6.5	46.8	44.5	8.7
25	84.5	15.3	0.2	77.2	22.4	0.3	66.9	32.6	0.5	53.5	45.7	0.9
50	95.8	4.2	0.0	90.7	9.3	0.0	80.5	19.5	0.0	63.8	36.2	0.0
100	99.6	0.4	0.0	98.0	2.0	0.0	92.3	7.7	0.0	75.1	24.9	0.0

Table 1. Random samples generated from the Pareto distribution for sample sizes $n = 15, 25, 50, 100$ and shape parameters $\xi = 0.1, 0.2, 0.3, 0.4$. The frequency with which the parameter estimate $\hat{\xi}$ is positive, negative or the algorithm does not converge (NC), are reported.

4. Bibliography

(1) Arnold, B.(1983). *Pareto distributions.* International Co-operative Publishing House. Fairland, Maryland.
(2) Castillo, E. and Hadi, A. (1997). "Fitting the Generalized Pareto Distribution to Data". *Journal of the American Statistical Association*, 92, 1609-1620.
(3) Castillo, J. (1994). "The Singly Truncated Normal Distribution, a Non-Steep Exponential Family". *Annals of the Institute of Mathematical Statistics.* 46, 57-66.
(4) Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values.* Springer, London.
(5) Davison, A (1984). Modelling Excesses Over High Thresholds, with an Application, in *Statistical Extremes and Applications*, ed. J.Tiago de Oliveira, Dordrecht: D.Reidel,pp. 461-482.
(6) Davison, A. and Smith, R. (1990). "Models for Exceedances over High Thresholds". *J.R. Statist. Soc. B*, 52, 393-442.
(7) Embrechts, P. Klüppelberg, C. and Mikosch, T. (1997). *Modeling Extremal Events for Insurance and Finance.* Springer, Berlin.
(8) Finkenstadt, B.and Rootzén, H. (edit) (2003). *Extreme values in Finance, Telecommunications, and the Environment.* Chapman & Hall .
(9) Hosking, J. and Wallis, J. (1987). "Parameter and quantile estimation for the generalized Pareto distribution". *Technometrics* 29 , 339–349.
(10) Pickands, J. (1975). "Statistical inference using extreme order statistics". *The Annals of Statistics* 3, 119-131.
(11) Smith, R. (1985). "Maximum likelihood estimation in a class of nonregular cases". *Biometrika* 72, 67–90.
(12) Van Montford, M. and Witter, J. (1985). Testing Exponentiality Against Generalized Pareto Distribution. *Journal of Hydrology*, 78, 305-315.

5. Appendix

Let us examine the sample of size two given by $\{x_1, x_2\} = \{1, 2\}$ for the GPD. First, we will show that the profile-likelihood $l_p(\sigma, 1)$, given by (2.7), is a monotonous increasing function for $\sigma > 0$ and, hence, there is no maximum likelihood estimator for the Pareto distribution with this sample.

The derivative $l_p'(\sigma, 1)$ is given by

$$num(\sigma) = 8 + 6\sigma - \sigma(3 + 2\sigma)(\log(1 + 1/\sigma) + \log(1 + 2/\sigma)),$$

divided by a positive function. Hence, $l_p'(\sigma, 1)$ and $num(\sigma)$ have equal sign and we will see that it is positive.

The following results hold for the function $num(\sigma)$ and for its derivatives:

$$\lim_{\sigma\to\infty} num(\sigma) = 4, \qquad \lim_{\sigma\to\infty} num'(\sigma) = 0, \qquad \lim_{\sigma\to\infty} num''(\sigma) = 0, \tag{5.1}$$

$$num'''(\sigma) = -\frac{48 + 152\sigma + 162\sigma^2 + 69\sigma^3 + 9\sigma^4}{\sigma^2(2 + 3\sigma + \sigma^2)^3} < 0. \tag{5.2}$$

Then, $num''(\sigma)$ is a monotonous decreasing function and, from the limit zero property, $num''(\sigma) > 0$. Then, $num'(\sigma)$ is monotonous increasing and, hence, $num'(\sigma) < 0$. Therefore, $num(\sigma)$ is a monotonous decreasing function, hence $num(\sigma) > 4$, and its sign is always positive, as we said.

We will also show that the profile-likelihood $l_p(\sigma, -1)$ with the same sample, $\{1, 2\}$, is a monotonous decreasing function for $\sigma > 2$ and, hence, does not exist a local maximum for the parameter space of the bounded support distributions in the GPD.

The derivative $l_p'(\sigma, -1)$ is given by

$$nu(\sigma) = 8 - 6\sigma + \sigma(3 - 2\sigma)(\log(1 - 1/\sigma) + \log(1 - 2/\sigma)),$$

divided by a negative function. We will see that the sign of $nu(\sigma)$ is positive for σ greater than 2 .

The following results hold for the function $nu(\sigma)$ and for its derivatives:

$$\lim_{\sigma\to\infty} nu(\sigma) = 4, \qquad \lim_{\sigma\to\infty} nu'(\sigma) = 0, \qquad \lim_{\sigma\to\infty} nu''(\sigma) = 0, \tag{5.3}$$

$$nu'''(\sigma) = \frac{48 - 152\sigma + 162\sigma^2 - 69\sigma^3 + 9\sigma^4}{\sigma^2(2 - 3\sigma + \sigma^2)^3}. \tag{5.4}$$

If $\sigma > 2$, the denominator of $nu'''(\sigma)$ is positive; the greater real root of the numerator is at $\sigma_1 = 4.36556$, then $nu'''(\sigma) > 0$, for greater values of σ. Therefore, for $\sigma > \sigma_1$, $nu''(\sigma)$ is a monotonous increasing function and, from the limit zero property, $nu''(\sigma) < 0$. Then, $nu'(\sigma)$ is monotonous decreasing and, hence, $nu'(\sigma) > 0$, for $\sigma > \sigma_1$. Finally, $nu(\sigma)$ is a monotonous increasing function for $\sigma > \sigma_1$. It is also clear that $\lim_{\sigma\downarrow 2} nu(\sigma) = \infty$.

For $2 < \sigma < \sigma_1$, $nu(\sigma)$ has a minimum at $\sigma_0 = 2.89221$ and $nu(\sigma_0) = 3.53456 > 0$. Then, we conclude that the sign of $nu(\sigma)$ is positive for $\sigma > 2$, $l_p(\sigma, -1)$ is a monotonous decreasing function for $\sigma > 2$ and does not exist a local maximum for the likelihood function of the GPD model with this sample.

Servei d'Estadística Universitat Autò noma de Barcelona, Jalila Daoudi, Departament de Matemtiques, Universitat Autònoma de Barcelona