Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution

Scott D. Grimshaw
College of Business and Management
University of Maryland
College Park, MD 20742

The generalized Pareto distribution (GPD) is a two-parameter family of distributions that can be used to model exceedances over a threshold. Maximum likelihood estimators of the parameters are preferred, since they are asymptotically normal and asymptotically efficient in many cases. Numerical methods are required for maximizing the log-likelihood, however. This article investigates the properties of a reduction of the two-dimensional numerical search for the zeros of the log-likelihood gradient vector to a one-dimensional numerical search. An algorithm for computing the GPD maximum likelihood estimates based on this dimension reduction and properties are given.

KEY WORDS: Statistical computing.

The traditional method for modeling extreme-value data is based on the extreme-value limiting distributions originally introduced by Fisher and Tippett (1928). Recently, alternative approaches have been studied. One such methodology is to look at exceedances over high thresholds rather than maxima over fixed time periods. The generalized Pareto distribution (GPD) was shown by Pickands (1975) to be a stable distribution for excesses over thresholds. Using those values that exceed a high threshold in an annual flood record, the GPD can be used for estimating extreme floods. R. L. Smith (1989) applied these ideas to the study of ozone levels in the upper atmosphere.

In this article, the GPD is used to model tensile-strength data from a random sample of nylon carpet fibers. The tensile strength of the fibers is not only an important product property to customers, but it is also important to production. In nylon carpet yarn production, the molten polymer is prepared and then forced through several very-small-diameter holes, which produce the fibers. These fibers then enter into the spinning process, which combines several fibers to form the carpet yarn and wind the yarn onto bobbins. If any one fiber is weak, the tension in the spinning process will break the fiber, and production on this particular bobbin is halted until a new fiber strand replaces the broken strand. This replacement is very costly because of the time required for an operator to retread a new fiber in the spinning process.

There are two main reasons for analyzing this tensile strength data as exceedances over a threshold. First, it is common to measure the properties of the nylon fiber after it has gone through the production process and been spun into carpet yarn. Therefore, if a fiber had a tensile strength below the tension threshold of the spinning process it would have broken and been replaced. Hence a fiber in a tested yarn sample has a tensile strength exceeding the spinning-process threshold.

Second, to measure the tensile strength of a randomly selected fiber, a small section of the fiber is placed in a measuring device that successively increases longitudinal stress on the fiber. The point at which the fiber breaks is the measured value of the fiber's tensile strength. These measuring devices often have either a minimum threshold at which testing can begin or the test is initiated at an operator-specified threshold that the fiber is believed to exceed. Any nylon fiber with a tensile strength below this minimum threshold, however, will break at test commencement. Therefore, the tensile-strength data from a random sample of nylon fibers contains only those measurements in which the fiber exceeds the testing threshold.

The values given in Table 1 are the results of tensile-strength testing on a random sample of nylon carpet fibers. This data set will be used to demonstrate the GPD probability model and the details of the problem of computing the maximum likelihood parameter estimates. The measuring device had a high testing threshold, and the 15 observations correspond to those fibers whose tensile strength exceeded this threshold. The tabulated values are the exceedances of the threshold for each fiber in kilograms/
Table 1. Exceedance (in kg/mm²) of the Testing Threshold in Tensile-Strength Tests for a Random Sample of n = 15 Nylon Carpet Fibers

<table>
<thead>
<tr>
<th>Value</th>
<th>0.051</th>
<th>0.268</th>
<th>0.011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.140</td>
<td>0.194</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>0.365</td>
<td>0.100</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>0.561</td>
<td>0.876</td>
<td>0.338</td>
</tr>
<tr>
<td>.030</td>
<td>.268</td>
<td>.184</td>
<td>.092</td>
</tr>
<tr>
<td>.200</td>
<td>.010</td>
<td>.561</td>
<td>.056</td>
</tr>
</tbody>
</table>

NOTE: The measures of center are mean = 0.25267 kg/mm², median = 0.16200 kg/mm². The measures of dispersion are standard deviation = 0.24360 kg/mm², interquartile range = 0.29000 kg/mm².

square millimeters (kg/mm²)—that is, fiber tensile strength minus testing threshold. The known value of the testing threshold is not given for proprietary reasons. Figure 1 contains a nonparametric density estimate for the exceedances formed with a bandwidth of .18 using the Epanechnikov kernel with boundary kernel modifications given by Gasser and Müller (1979). The boundary kernel modifications are necessary since the exceedances must be greater than 0. The 15 observations are plotted along the y = 0 line.

1. GENERALIZED PARETO DISTRIBUTION

A random variable X is defined to have a generalized Pareto distribution (GPD), with parameters k and \( \alpha \) such that \(-\infty < k < \infty, \alpha > 0\), if the cumulative distribution function is given by

\[
F(x; \theta, \alpha) = 1 - \left( 1 - \frac{kx}{\alpha} \right)^{1/k}, \quad k \neq 0
\]

\[
= 1 - e^{-x/\alpha}, \quad k = 0.
\]

The range for \( x > 0 \) for \( k \leq 0 \) and \( 0 < x < \alpha/k \) for \( k > 0 \). The quantile function, \( Q(u) = F^{-1}(u) \), is given by \( Q(u; k, \alpha) = -\alpha \cdot g(1 - u; k) \), where \( g(\cdot) \) is the power transformation (also called the Box-Cox transformation), defined for \( z > 0 \) by

\[
g(z; \lambda) = \frac{z^\lambda - 1}{\lambda}, \quad \lambda \neq 0
\]

\[
= \ln z, \quad \lambda = 0.
\]

Notice that the GPD can also be naturally referred to as the power uniform distribution, since it can be formed by taking the power transformation of a Uniform (0, 1) random variable; that is, if \( U \) has a Uniform (0, 1) distribution, then \( X = -\alpha \cdot g(U; k) \) is GPD(\( k, \alpha \)).

Pickands (1975) introduced the GPD as a two-parameter family of distributions for exceedances over a threshold. The parameters of the GPD are \( \alpha \), the scale parameter, and \( k \), the shape parameter. See the article by Hosking and Wallis (1987) for examples of the GPD for different parameter values. Three special cases of the GPD are (1) if \( k = 1 \), the distribution is uniform (0, \( \alpha \)), (2) if \( k = 0 \), the distribution is exponential (1/\( \alpha \)), and (3) if \( k < 0 \), the distribution is Pareto. Notice that, if \( k > 0 \), the GPD has finite support indicating a finite maximum value for \( X \) given by \( Q(1) = \alpha/k \).

Maximum likelihood estimation of the parameters \( (k, \alpha) \) was considered by DuMouchel (1983), Davidson (1984), R. L. Smith (1984, 1987), J. A. Smith (1986), and Joe (1987). When \( k \leq 0 \), R. L. Smith (1984) showed that, under certain regularity conditions, the maximum likelihood estimators are asymptotically normal and asymptotically efficient. If \( \hat{k}, \hat{\alpha} \) denote the maximum likelihood estimators, then as \( n \to \infty \),

\[
\hat{k}, \hat{\alpha} \to \frac{1}{\alpha} N \left( k, \frac{1}{\alpha} \right).
\]

When \( k > 0 \), R. L. Smith (1985) identified the problem as nonregular, which alters the rate of convergence of the maximum likelihood estimators and possibly their existence, and identifies the special procedures required for the problem. This case rarely occurs in statistical applications, however, since for \( k > 1 \) the GPD has finite endpoints and \( f(x) > 0 \) at each endpoint.

In addition to parameter estimates, a maximum likelihood quantile estimator is defined by substituting the estimators \( k \) and \( \hat{\alpha} \) into the quantile function to obtain \( \hat{Q}(u; \hat{k}, \hat{\alpha}) = -\hat{\alpha} \cdot g(1 - u; \hat{k}) \) for a given value of \( u \) such that \( 0 < u < 1 \). It is easily shown that

\[
\hat{Q}(u) = AN \left( Q(u), \frac{\alpha (1 - k)^2}{n} \left[ (1 - k)g'(1 - u; k)^2 \right] + 2[g(1 - u; k) \cdot g'(1 - u; k) + 2g(1 - u; k)]^2 \right).
\]
where \( g'(z; \lambda) = [-g(z; \lambda) + z^\lambda \ln z]/\lambda \). An estimator of the upper bound \( Q(1) \), when it is finite, is a special case in which

\[
\hat{Q}(1) = \frac{\alpha}{k} k^2 \left( 1 - 2k \right)
\]

These results can be used to form approximate confidence intervals for \( k, \alpha, Q(\alpha), \) or any function of these estimators. The unknown parameters in the asymptotic variance of the estimator can be replaced by the maximum likelihood estimators to obtain a consistent estimator of the variance. The approximate 100(1 - \( \alpha \))% confidence interval can be constructed using maximum likelihood estimator

\[
\pm z_{\alpha/2} \sqrt{\text{estimated variance}}
\]

where \( z_{\alpha/2} \) denotes the \( 1 - \alpha/2 \) quantile of the standard normal distribution.

The maximum likelihood estimates must be derived numerically because the minimal sufficient statistics for the GPD are the order statistics and there is no obvious simplification of the nonlinear likelihood equation. Hosking and Wallis (1987) proposed a modified Newton-Raphson algorithm to find the maximum of the log-likelihood. They also proposed method of moments and method of probability-weighted moments as alternative parameter estimators for the GPD when a reduction of the parameter space to \(-1 < k < 1\) is reasonable. These alternative estimators are easy to compute, and in a study of finite sample properties, Hosking and Wallis (1987) determined that very large samples (in some cases as large as 500) are required before the maximum likelihood estimators clearly demonstrate their efficiency.

In this article an algorithm for computing the maximum likelihood estimates is presented. The two-dimensional numerical search for the zeros of the gradient of the GPD log-likelihood is reduced to a one-dimensional numerical search. This simplification is due to a reparameterization pointed out by Davison (1984).

2. COMPUTING MAXIMUM LIKELIHOOD PARAMETER ESTIMATES

Suppose that \( X = \{X_1, \ldots, X_n\} \) is a random sample from the GPD with largest value \( X_{(n)} \). The log-likelihood is given by

\[
L(k, \alpha; X) = -n \ln \alpha + \left( \frac{1}{k} - 1 \right) \sum_{i=1}^{n} \ln \left( 1 - \frac{kX_i}{\alpha} \right), \quad k \neq 0,
\]

\[
L(k, \alpha; X) = -n \ln \alpha - \frac{1}{\alpha} \sum_{i=1}^{n} X_i, \quad k = 0.
\]

The range for \( \alpha \) is \( \alpha > 0 \) for \( k \leq 0 \) and \( \alpha > kX_{(n)} \) for \( k > 0 \).

If \( k > 1 \), there is no maximum likelihood estimate since, for any \( k > 1 \),

\[
\lim_{\alpha \to X_{(n)}} L(k, \alpha; X) = \infty.
\]

To obtain a finite maximum of the GPD log-likelihood, the constraint \( k \leq 1 \) must be imposed.

Consider also the special case of \( k = 0 \). It can be shown that the gradient vector at \( k = 0 \) has elements

\[
\frac{\partial L(k, \alpha; X)}{\partial k} = \frac{\sum_{i=1}^{n} X_i}{2\alpha^2} - \frac{\sum_{i=1}^{n} X_i}{\alpha} \quad k \neq 0,
\]

\[
\frac{\partial L(k, \alpha; X)}{\partial \alpha} = 1 \left( \frac{\sum_{i=1}^{n} X_i}{\alpha} - n \right) \quad k = 0,
\]

which are equal to 0 iff \( (1/n) \sum_{i=1}^{n} X_i^2 = 2\overline{X}^2 \). Therefore, if this condition is not satisfied, then the case \( k = 0 \) can also be eliminated from consideration.

Therefore, computing the GPD maximum likelihood estimates is an optimization on the constrained space \( A = \{k < 0, \alpha > 0\} \cup \{0 < k \leq 1, \alpha/k > X_{(n)}\} \).

There are two values of \( (k, \alpha) \) that must be investigated to compute the GPD maximum likelihood estimate. The first is the local maximum of the log-likelihood on the space \( A \). The second is at the boundary of \( A \), where \( k = 1 \).

2.1 Local Maximum on \( A \)

To compute the local maximum on the space \( A \), consider the gradient vector of the GPD log-likelihood given in the Appendix. The solution to the simultaneous equations may be simplified and written as

\[
\begin{align*}
\frac{\partial L(k, \alpha; X)}{\partial k} = 0 & \Rightarrow \left\{ \begin{array}{l}
n(k - 1) = \sum_{i=1}^{n} \ln \left( 1 - \frac{kX_i}{\alpha} \right) \\
\frac{n}{k} - 1 = \sum_{i=1}^{n} \left( 1 - \frac{kX_i}{\alpha} \right)^{-1} \end{array} \right.
\end{align*}
\]

\[
\frac{\partial L(k, \alpha; X)}{\partial \alpha} = 0 \Rightarrow \left\{ \begin{array}{l}
\left( \frac{n}{k} - 1 \right) = \sum_{i=1}^{n} \ln \left( 1 - \frac{kX_i}{\alpha} \right) \\
\frac{n}{k} - 1 = \sum_{i=1}^{n} \left( 1 - \frac{kX_i}{\alpha} \right)^{-1} \end{array} \right.
\]

\[
\begin{align*}
\left( \frac{n}{k} - 1 \right) \sum_{i=1}^{n} \ln \left( 1 - \frac{kX_i}{\alpha} \right) & = 1 \\
\left( \frac{n}{k} - 1 \right) \sum_{i=1}^{n} \left( 1 - \frac{kX_i}{\alpha} \right)^{-1} & = 1
\end{align*}
\]

The bivariate search for the zeros of the gradient vector over \( A \) can be reduced to a univariate search because the second equation is a closed-form representation for the estimate of \( k \) given the ratio \( k/\alpha \), and the first equation depends only on \( k/\alpha \). This reduction was pointed out by Davison (1984).

Therefore, to compute the maximum likelihood estimates, consider the reparameterization \( (k, \alpha) \) to
(k, \theta), where \theta = k/a, which is one-to-one in the space \mathcal{A}. Then substitute

\hat{k} = -1/n \sum_{i=1}^{n} \ln(1 - \theta X_i),

an explicit function of the data and \theta, into the GPD log-likelihood for (k, \theta) so that the log-likelihood depends only on \theta. This is called the profile log-likelihood of \theta. The value of \theta which corresponds to the local maximum of this profile log-likelihood, given by

L(\theta; X) = -n - \sum_{i=1}^{n} \ln(1 - \theta X_i)

- n \ln \left[ -\frac{1}{n} \sum_{i=1}^{n} \frac{1}{\theta} \ln(1 - \theta X_i) \right], \quad (2.1)

must be computed numerically on the space \mathcal{B} = \{ \theta < 1/X_{(n)}, \theta \neq 0 \}.

It is important to emphasize that the local maximum of the profile log-likelihood corresponds to the local maximum of the GPD log-likelihood, but reducing to the profile log-likelihood introduces two complications to the numerical computation. First, it can be shown that, for the profile log-likelihood,

\lim_{\theta \to 1/X_{(n)}} L(\theta; X) = \infty,

but in the GPD log-likelihood, \theta \to 1/X_{(n)} is equivalent to \alpha/k \to X_{(n)}^+, and

\lim_{\alpha \to X_{(n)}^+} L(k; a; X) = -\infty.

Therefore, an algorithm must be designed to avoid convergence to the upper limit, 1/X_{(n)}, as well as avoiding convergence to 0, which is not in the domain of the profile log-likelihood. Second, because of the behavior of the profile log-likelihood near 1/X_{(n)}, there will exist a local minimum for every local maximum. Therefore, the first derivative of the profile log-likelihood will have multiple roots on \mathcal{B}, each of which must be found by an algorithm, and every zero must be evaluated to determine if it corresponds to a local maximum of the profile log-likelihood.

Consider the first derivative of the profile log-likelihood given in the Appendix. Solving \frac{dL(\theta; X)}{d\theta} = 0 is equivalent to finding the zeros of the function

h(\theta) = \left[ 1 + (1/n) \sum_{i=1}^{n} \ln(1 - \theta X_i) \right]

\cdot \left[ 1 + (1/n) \sum_{i=1}^{n} (1 - \theta X_i)^{-1} \right] - 1 \quad (2.2)

on \mathcal{B}.

The following theorem states several properties of \( h(\theta) \) that are useful in formulating an algorithm for determining zeros of \( h(\theta) \).

**Theorem.** Consider the function \( h(\theta) \) given in (2.2) defined on the space \mathcal{B}. Then

(1) \lim_{\theta \to 1/X_{(n)}} h(\theta) = -\infty,

(2) \( h(\theta) < 0 \) for all \( \theta < \theta_1 \), where

\[ \theta_1 = \frac{2[X_{(1)}] - \bar{X}}{[X_{(1)}]^2}, \]

(3) \( h'(\theta) = \frac{1}{\theta} \left\{ \left[ (1/n) \sum_{i=1}^{n} (1 - \theta X_i)^{-1} \right]^2 - \left[ (1/n) \sum_{i=1}^{n} \ln(1 - \theta X_i) \right] \right. \]

\[ - \left. \left[ (1/n) \sum_{i=1}^{n} (1 - \theta X_i)^{-1} \right] \right. \]

- \left. \left[ (1/n) \sum_{i=1}^{n} (1 - \theta X_i)^{-2} \right] \right\}, \quad (2.3)

(4) \lim_{\theta \to 0} h'(\theta) = 0,

and

(5) \lim_{\theta \to 0} h''(\theta) = (1/n) \sum_{i=1}^{n} X_i^2 - 2\bar{X}^2.

**Proof.** Results (1) and (3)–(5) are straightforward. The proof of (2) follows by noticing that, by Jensen’s inequality,

\[ \frac{1}{n} \sum_{i=1}^{n} \ln(1 - \theta X_i) \leq \ln(1 - \theta \bar{X}) \]

for \( \theta < 1/X_{(n)} \) and, since \( X_{(1)} \leq X_i \) for all \( i \),

\[ \frac{1}{n} \sum_{i=1}^{n} (1 - \theta X_i)^{-1} \leq [1 - \theta X_{(1)}]^{-1} \]

for \( \theta < 0 \). It follows then that \( h(\theta) \leq [1 + \ln(1 - \theta \bar{X})] \cdot [1 - \theta X_{(1)}]^{-1} - 1 \) for \( \theta < 0 \). Suppose that \( \theta < \theta_2 \). Then \( 1 - \theta \bar{X} < 1 + [\theta X_{(1)}] + \left( -\theta X_{(1)} \right)^2 < e^{-\theta X_{(1)}} \), and hence \( 1 + \ln(1 - \theta \bar{X}) < 1 - \theta X_{(1)} \). Therefore \( h(\theta) < 0 \) for all \( \theta < \theta_2 \).

Result (1) indicates an upper bound, \( \theta_U = 1/X_{(n)} \), for any zero of \( h(\theta) \). Since this is a limiting result, the algorithm will use \( \theta_U = \varepsilon \) for some \( \varepsilon > 0 \) as the upper bound. Result (2), provided by an anonymous referee, provides a lower bound, \( \theta_L \), for any zero of \( h(\theta) \). Coupling these two results with the fact that \( \mathcal{B} \) does not contain zero, an algorithm must numerically search for zeros of \( h(\theta) \) on both \( (\theta_L, 0) \) and \( (0, \theta_U) \). Because all bounds are known, modifications of the Newton–Raphson zero-search algorithms can be made that limit step size so that iterative solutions remain within the known boundaries.

Result (3) is the derivative of \( h(\theta) \), required for the Newton–Raphson algorithm to search for zeros of \( h(\theta) \). Result (4) indicates the importance of forcing
the algorithm to search on the bounded intervals, 
since otherwise it may converge to zero, which is not in \( \mathbb{B} \).

Result (5) can be used to help orient an algorithm for searching around the bisection point zero. If \( \lim_{\theta \to 0} h'(\theta) > 0 \) then there are \( j_n \) roots on \((\theta_L, 0)\), where \( j_n \) is an odd integer, and there are \( j_p \) roots on \((0, \theta_U)\), where \( j_p \) is an odd integer. This follows, since \( \lim_{\theta \to 0} h'(\theta) > 0 \) implies that, for some \( \varepsilon > 0 \), \( h(\theta - \varepsilon) > 0 \), and since \( h(\theta_U) < 0 \), then the number of zeros on \((\theta_U, 0)\) given by \( j_n \) must be an odd integer. The argument for \( j_p \) odd is similar. In the data sets used investigating the GPD maximum likelihood estimates, it appears that \( j_n = j_p = 1 \).

If \( \lim_{\theta \to 0} h'(\theta) < 0 \), then there are \( j_n \) roots on \((\theta_L, 0)\), where \( j_n \) zero or an even integer and there are \( j_p \) roots on \((0, \theta_U)\), where \( j_p \) is zero or an even integer. This follows since \( \lim_{\theta \to 0} h'(\theta) < 0 \) implies that for some \( \varepsilon < 0 \), \( h(\theta - \varepsilon) < 0 \), and since \( h(\theta_U) < 0 \), then either there is no zero of \( h(\theta) \) on \((\theta_U, 0)\) or the number of zeros on \((\theta_U, 0)\) given by \( j_n \) must be an even integer. The argument for \( j_p \) zero or an even integer is similar. In the data sets used investigating the GPD maximum likelihood estimates, it appears that in many cases \( j_n = j_p = 0 \). This result agrees with the finding of Hosking and Wallis (1987) indicating that in many cases with \( k > 0 \) and \( n < 25 \) the GPD maximum likelihood estimates do not exist. The remaining data sets in the investigation indicated that either \( j_n = 0, j_p = 2 \), or \( j_n = 2, j_p = 0 \) or \( j_n = 2, j_p = 2 \).

The possible existence of multiple zeros of \( h(\theta) \) on \( \mathbb{B} \) complicates the numerical search, but the algorithm given in Section 3 is designed to find these multiple zeros.

Each zero of \( h(\theta) \) indicates a candidate for the local maximum of the GPD log-likelihood. For each of the \( j_n + j_p \) zero(s), denoted by \( \theta^{(i)} \), compute

\[
\begin{align*}
k_i &= -\left(1/n\right) \sum_{i=1}^{n} \ln(1 - \theta_i^{(i)} X_i) \\
\alpha_i &= \frac{k_i}{\theta_i^{(i)}}.
\end{align*}
\]

Evaluate the GPD log-likelihood at each of these points to determine the local maximum. Let \((k_m, \alpha_m)\) denote the local maximum of the GPD log-likelihood on \( \mathbb{A} \).

2.2 Boundary Maximum on \( \mathbb{A} \)

Any local maximum of the GPD log-likelihood on the domain \( \mathbb{A} \) must exceed the GPD log-likelihood evaluated at the boundary to be the maximum likelihood estimate. Hence the second value that must be investigated is at the boundary of \( \mathbb{A} \), where \( k = 1 \). Given \( k = 1, \alpha > X_{(n)} \), then \( L(k, \alpha; X) = -n \ln \alpha \). Therefore, the boundary maximum, denoted by \((k_b, \alpha_b)\), is given by \( k_b = 1 \) and \( \alpha_b = X_{(n)} \). The problem is complicated by the optimization being taken over an open set, but it is treated as a maximum taken over a closed set.

The GPD maximum likelihood estimates, denoted by \((\hat{k}, \hat{\alpha})\), is then given by the local maximum \((k_m, \alpha_m)\) if \( L(k_m, \alpha_m; X) > -n \ln X_{(n)} \) and is given by the boundary maximum \((k_b, \alpha_b)\) if \( L(k_b, \alpha_b; X) < -n \ln X_{(n)} \).

If no local maximum is found, then there is no GPD maximum likelihood estimate and the alternative estimators given by Hosking and Wallis (1987) are recommended.

3. PROPOSED ALGORITHM FOR THE GPD MAXIMUM LIKELIHOOD ESTIMATES

An algorithm that computes the GPD maximum likelihood estimates using the technical details discussed in Section 2 is given by the following:

1. Choose an \( \varepsilon \) such that for numerical purposes \( \theta_1 = \theta_2 \) if \( |\theta_1 - \theta_2| < \varepsilon \). For example, let \( \varepsilon = 10^{-6}X \), since \( \theta = k/\alpha \) is not independent of the measurement units.

2. Compute the lower and upper bounds for zeros of \( h(\theta) \) given by

\[
\begin{align*}
\theta_L &= \frac{2[X_{(1)} - X]}{[X_{(1)}]^2}, \\
\theta_U &= \frac{1}{X_{(n)} - \varepsilon}.
\end{align*}
\]

3. Compute \( \lim_{\theta \to 0} h'(\theta) = (1/n) \sum_{i=1}^{n} X_i^2 - 2X^2 \).

4. If \( \lim_{\theta \to 0} h'(\theta) > 0 \), then there exists at least one zero of \( h(\theta) \) on \((\theta_L, 0)\) and at least one zero of \( h(\theta) \) on \((0, \theta_U)\). In the data sets used investigating the GPD maximum likelihood estimates, it appears that there exists one zero on each interval.

A. To determine the zero of \( h(\theta) \) on the bounded interval \((\theta_L, -\varepsilon)\), use the modified Newton–Raphson algorithm given by Press, Flannery, Teukolsky, and Vetterling (1989), which limits step size to remain within the interval. The initial value for the algorithm is \( \theta_L \). Denote this zero by \( \theta_L^{(0)} \).

B. To determine the zero of \( h(\theta) \) on the bounded interval \((\varepsilon, \theta_U)\), use the modified Newton–Raphson algorithm. The initial value for the algorithm is \( \theta_U \). Denote this zero by \( \theta_U^{(0)} \).

5. If \( \lim_{\theta \to 0} h'(\theta) < 0 \), then there exists either no zero or an even integer number of zeros of \( h(\theta) \) on \((\theta_L, 0)\) and either no zeros or an even integer number of zeros of \( h(\theta) \) on \((0, \theta_U)\). In the data sets used investigating the GPD maximum likelihood estimates, it appears that there exists either no zero or two zeros on each interval.

A. To determine the first zero of \( h(\theta) \) on the bounded interval \((\theta_L, -\varepsilon)\), use the modified
Newton–Raphson algorithm. The initial value for the algorithm is \( \theta_1 \). Denote this zero by \( \theta_1^{(0)} \). If no zero of \( h(\theta) \) exists on this interval, the modified Newton–Raphson algorithm will converge to the boundary value \( -\varepsilon \) and it is unnecessary to search for a second zero on this interval.

B. Compute \( h'(\theta_1^{(0)}) \), where \( h'(\theta) \) is given by (2.3).

C. If \( h'(\theta_1^{(0)}) > 0 \), the second zero of \( h(\theta) \) on the bounded interval \( (\theta_1, -\varepsilon) \) lies on the interval \( (\theta_1^{(0)}, -\varepsilon) \). Use the bisection algorithm to compute this second root and denote it by \( \theta_2^{(0)} \).

D. If \( h'(\theta_1^{(0)}) < 0 \), the second zero of \( h(\theta) \) on the bounded interval \( (\theta_1, -\varepsilon) \) lies on the interval \( (\theta_1, \theta_1^{(0)}) \). Use the bisection algorithm to compute this second root and denote it by \( \theta_2^{(0)} \).

E. To determine the first zero of \( h(\theta) \) on the bounded interval \( (\varepsilon, \theta_2) \), use the modified Newton–Raphson algorithm. The initial value for the algorithm is \( \theta_2 \). Denote this zero by \( \theta_2^{(0)} \). If no zero of \( h(\theta) \) exists on this interval, the modified Newton–Raphson algorithm will converge to the boundary value \( \varepsilon \), and it is unnecessary to search for a second zero on this interval.

F. Compute \( h'(\theta_2^{(0)}) \).

G. If \( h'(\theta_2^{(0)}) > 0 \), the second zero of \( h(\theta) \) on the bounded interval \( (\varepsilon, \theta_2) \) lies on the interval \( (\theta_2^{(0)}, \varepsilon) \). Use the bisection algorithm to compute this second root and denote it by \( \theta_3^{(0)} \).

H. If \( h'(\theta_2^{(0)}) < 0 \), the second zero of \( h(\theta) \) on the bounded interval \( (\varepsilon, \theta_2) \) lies on the interval \( (\varepsilon, \theta_2^{(0)}) \). Use the bisection algorithm to compute this second root and denote it by \( \theta_3^{(0)} \).

6. For each \( \theta_i^{(0)} \), compute \( k_i \) and \( \alpha_i \) given by (2.4) and evaluate the GPD log-likelihood \( L(k_i, \alpha_i; X) \).

7. Let \( (k_m, \alpha_m) \) denote the local maximum of the GPD log-likelihood on \( \Theta \). If there is no local maximum, then there is no GPD maximum likelihood estimate.

8. The GPD maximum likelihood estimates, denoted by \( (\hat{k}, \hat{\alpha}) \), is given by the local maximum on \( \Theta \), \( (k_m, \alpha_m) \), if \( L(k_m, \alpha_m; X) > -n \ln X(\alpha) \) and is given by the boundary maximum \( (k_b, \alpha_b) \), where \( k_b = 1 \) and \( \alpha_b = X(\alpha) \) if \( L(k_b, \alpha_b; X) < -n \ln X(\alpha) \).

4. GPD MODELING OF MOTIVATING EXAMPLE

In this section, the maximum likelihood parameter estimates are computed for the data discussed in the introduction. Treating the data in Table 1 as a random sample of \( n = 15 \) observations from a GPD \( (k, \alpha) \) population, the algorithm in Section 3 is used to compute the GPD maximum likelihood estimates. The algorithm uses \( \varepsilon = 10^{-6}/\bar{X} \).

To begin the search for zeros of \( h(\theta) \), compute the bounds given by \( \theta_1 = -3994.490358 \) and \( \theta_1 = 1.141552 \). The algorithm will search for two zeros on \( (\theta_1, -\varepsilon) \) and two zeros on \( (\varepsilon, \theta_2) \), since \( \lim_{\theta \to 0} h'(\theta) = -0.008455 < 0 \).

On the interval \( (\theta_2, -\varepsilon) \), the modified Newton–Raphson algorithm, using an initial value of \( \theta_2 \), converged to \( -\varepsilon \) indicating that no \( 0 \) of \( h(\theta) \) exists on this interval. On the interval \( (\varepsilon, \theta_2) \), the modified Newton–Raphson algorithm, using an initial value of \( \theta_2 \), converged to \( \theta_2^{(0)} = 1.141551 \). Since \( h'(\theta_2^{(0)}) < 0 \), the second zero on this interval is bounded by \( (\varepsilon, \theta_2^{(0)}) \). The bisection algorithm on \( (\varepsilon, \theta_2^{(0)}) \) converged to \( \theta_2^{(0)} = 0.415834 \).

The zeros \( \theta_2^{(0)} \) and \( \theta_2^{(0)} \) correspond to (a) the values \( (k_3 = 1.111357, \alpha_3 = .973553) \) with \( L(k_3, \alpha_3; X) = 2.072396 \) and (b) the values \( (k_4 = .117698, \alpha_4 = .283040) \) with \( L(k_4, \alpha_4; X) = 5.697968 \).

Since \( L(k_4, \alpha_4; X) = 5.697968 > L(k_3, \alpha_3; X) = 2.072396 \), the local maximum of the GPD log-likelihood on \( \Theta \) is given by \( (k_m = .117698, \alpha_m = .283040) \).

The GPD maximum likelihood estimates for the data given in Table 1 is \( (k = .117698, \alpha = .283040) \). This follows because \( L(k_m, \alpha_m; X) = 5.697968 \) exceeds the GPD log-likelihood evaluated at the boundary maximum, \( (k_b = 1, \alpha_b = .876000) \), given by \( L(k_b, \alpha_b; X) = -n \ln X(\alpha) = 1.985838 \).

The mechanism for measuring tensile strength has a region of operation. As previously discussed, there is a lower testing threshold. There is also an upper testing threshold, which should exceed the tensile strength of all sample fibers; that is, it should be greater than \( Q(1) \). If the testing mechanism has an upper testing threshold less than any sample fiber's tensile strength, the sample may be right-censored, resulting in a reduction in the sample information. If the testing mechanism has an upper testing threshold much larger than any sample fiber's tensile strength, the measurement precision may be sacrificed because the increments of the successive increases in longitudinal stress may be unnecessarily large, since the increments are usually a function of the difference of the upper and lower testing thresholds. From the data, an estimate of \( Q(1) \) can be obtained that can be used to choose an appropriate upper testing threshold. The point estimate is \( \hat{Q}(1) = \hat{k}/\hat{\alpha} = 2.404804 \) and an approximate 95% upper confidence bound for \( Q(1) \) is given by 9.099524.

Figure 2 demonstrates how well the GPD density estimate models the data in Table 1. The parameters in the GPD density estimate are the maximum likelihood estimates, \( (k = .117698, \alpha = .283040) \). The GPD density estimate has been overlaid on the non-parametric density estimate from Figure 1 for comparison. Notice that the GPD model has more area in the tail, denoting the likelihood of much higher...
tensile strengths for fibers than those obtained in the small sample. The 15 observations are plotted along the $y = 0$ line as circles.

ACKNOWLEDGMENTS

I thank William P. Alexander for providing MATRIX, a matrix language environment for the IBM family of personal computers, which was used to develop the algorithm given in this article, perform the calculations for the example, and produce the graphs herein. Gratitude is also expressed to the DuPont Quality Management and Technology Center for the opportunity to learn about the nylon production process while I was an intern and for providing the data used in this example. I also thank the editors and an anonymous referee who provided many important insights and comments that led to a much improved article. This work was supported by an ARO grant while I was a graduate student at Texas A&M University.

APPENDIX: DERIVATIVES OF LOG-LIKELIHOOD AND PROFILE LOG-LIKELIHOOD

The gradient vector of the GPD log-likelihood has elements

\[
\frac{\partial L(k, \alpha; X)}{\partial k} = \frac{n}{k} \left( \frac{1}{k} - 1 \right) - \frac{n}{k} \sum_{i=1}^{n} \ln \left( 1 - \frac{kX_i}{\alpha} \right) - \frac{1}{k} \left( \frac{1}{k} - 1 \right) \sum_{i=1}^{n} \left( 1 - \frac{kX_i}{\alpha} \right)^{-1}
\]

\[
\frac{\partial L(k, \alpha; X)}{\partial \alpha} = - \frac{n}{k \alpha} + \frac{1}{\alpha} \left( \frac{1}{k} - 1 \right) \sum_{i=1}^{n} \left( 1 - \frac{kX_i}{\alpha} \right)^{-1}.
\]

The first derivative of the profile log-likelihood for \( \theta \) is given by

\[
\frac{dL(\theta; X)}{d\theta} = \frac{1}{\theta} \sum_{i=1}^{n} \left( 1 - \theta X_i \right)^{-1} - \frac{n}{\theta} \left[ \sum_{i=1}^{n} \frac{1}{\left( 1 - \theta X_i \right)^{\alpha}} \right] 
\cdot \left[ n - \sum_{i=1}^{n} \left( 1 - \theta X_i \right)^{-1} \right].
\]