Goodness-of-Fit Tests for the Generalized Pareto Distribution

V Choulakian & M. A Stephens

Goodness-of-Fit Tests for the Generalized Pareto Distribution

V. Choulakian

Department of Mathematics and Statistics
Université de Moncton
Moncton, N.B. E1A 3E9
Canada
(*choulav@umoncton.ca*)

M. A. Stephens

Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, B.C. V5A 1S6
Canada
(*stephens@cs.sfu.ca*)

Tests of fit are given for the generalized Pareto distribution (GPD) based on Cramér–von Mises statistics. Examples are given to illustrate the estimation techniques and the goodness-of-fit procedures. The tests are applied to the exceedances over given thresholds for 238 river flows in Canada; in general, the GPD provides an adequate fit. The tests are useful in deciding the threshold in such applications; this method is investigated and also the closeness of the GPD to some other distributions that might be used for long-tailed data.

KEY WORDS: Exceedances; Extreme values; Hydrology; Threshold selection.

The generalized Pareto distribution (GPD) has the following distribution function:

$$F(x) = 1 - (1 - kx/a)^{1/k}, \qquad (1)$$

where a is a positive scale parameter and k is a shape parameter. The density function is

$$f(x) = (1/a)(1 - kx/a)^{(1-k)/k}; \qquad (2)$$

the range of x is $0 \le x < \infty$ for $k \le 0$ and $0 \le x \le a/k$ for $k > 0$. The mean and variance are $\mu = a/(1+k)$ and $\sigma^2 = a^2/\{(1+k)^2(1+2k)\}$; thus the variance of the GPD is finite only for $k > -0.5$. For the special values $k = 0$ and 1, the GPD becomes the exponential and uniform distributions, respectively. The name generalized Pareto was given by Pickands (1975); the distribution is sometimes called simply Pareto when $k < 0$. In this case, the GPD has a long tail to the right and has been used to model datasets that exhibit this form in several areas of applied statistics.

In particular, the GPD is used to model extreme values. This application was discussed by many authors—for example, Hosking and Wallis (1987), Smith (1984, 1989, 1990), Davison (1984), and Davison and Smith (1990). Smith (1990) gave an excellent review of the two most widely used methods in this field, based on generalized extreme value distributions and on the GPD. In hydrology, the GPD is often called the "peaks over thresholds" (POT) model since it is used to model exceedances over threshold levels in flood control. Davison and Smith (1990) discussed this application in their section 9, using river-flow exceedances for a particular river over a period of 35 years. The authors fit the GPD for exceedances over a series of thresholds and also calculated the Kolmogorov–Smirnov and Anderson–Darling statistics to test fit. In the article, they mentioned the lack of tests for the GPD [echoing a remark made earlier by Smith (1984)]. In the absence of such tests, they used tables for testing exponentiality, which, as the authors pointed out, give too high critical values. This dataset is tested for GPD in Section 3.

In this article, goodness-of-fit tests are given for the GPD, based on the Cramér–von Mises statistic W^2 and the Anderson–Darling A^2. We concentrate on the most practical case in which the parameters are not known. Estimation of parameters is discussed in Section 1, and the goodness-of-fit tests are given in Section 2. In Section 3, they are applied to exceedances for a Canadian river, and it is shown how the tests can be used to help select the threshold in the POT model, following ideas suggested by Davison and Smith (1990). Moreover, the Davison–Smith example is revisited, using the GPD test and the Anderson–Darling statistic. The tests are then used to model the exceedances over thresholds of 238 Canadian river-flow series; they indicate the adequacy of the GPD fit. The technique of choosing the threshold is based on a stability property of the GPD; the efficacy of the technique when the true exceedance distribution is not GPD is examined in Section 4. In Section 5, we investigate the versatility of the GPD as a distribution that might be used for many types of data with a long tail and no mode in the density. Finally, the asymptotic theory of the tests is given in Section 6.

1. ESTIMATION OF PARAMETERS

Before the test of fit can be made, unknown parameters in (1) must first be estimated. The asymptotic theory of the test statistics requires an efficient method of estimation; we shall use maximum likelihood. Although it is theoretically possible to have datasets for which no solution exists to the likelihood equations, in practice this appears to be extremely rare. For example, in our examination of the 238 Canadian rivers, maximum likelihood estimates of the parameters could be found in every case. Thus we shall assume that such estimates exist for the dataset under test. Hosking and Wallis (1987), Castillo and Hadi (1997), and Dupuis and Tsao (1998) studied other methods of estimating the parameters. One of these methods is the use of probability-weighted moments; although there

appear to be some advantages in terms of bias, Chen and Balakrishnan (1995) and Dupuis (1996) later showed that this technique is not always feasible.

We now discuss estimation by maximum likelihood. Suppose that $x_1, \ldots, x_n$ is a given random sample from the GPD given in (1), and let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ be the order statistics. We consider three distinct cases—Case 1, in which the shape parameter k is known and the scale parameter a is unknown; Case 2, in which the shape parameter k is unknown, and the scale parameter is known; Case 3, in which both parameters a and k are unknown. Case 3 is the most likely situation to arise in practice. The log-likelihood is given by

$$L(a, k) = -n \log a - (1 - 1/k)$$
$$\times \sum_{i=1}^{n} \log(1 - k x_i/a) \quad \text{for} \quad k \neq 0$$
$$= -n \log a - \sum_{i=1}^{n} x_i/a \quad \text{for} \quad k = 0. \quad (3)$$

The range for a is $a > 0$ for $k \leq 0$ and $a > k x_{(n)}$ for $k > 0$. When $k < 1/2$, Smith (1984) showed that, under certain regularity conditions, the maximum likelihood estimators are asymptotically normal and asymptotically efficient. When $0.5 \leq k < 1$, Smith (1984) identified the problem as nonregular, which alters the rate of convergence of the maximum likelihood estimators. For $k \geq 1$, and as $n \to \infty$, the probability approaches 1 that the likelihood has no local maximum. We now consider Cases 1 to 3 separately.

Case 1 (k known, a unknown). For this case, we have the following result.

Proposition 1. For any known k with $k < 1$, the maximum likelihood estimate of a exists and is unique.

Proof. For $k = 0$ (the exponential distribution), the result is well known. Suppose that $k \neq 0$; then $\hat{a}$, the maximum likelihood estimate of a, will be a solution $\hat{a}$ of $\partial L(a, k)/\partial a = 0$, which may be simplified to $L_1(a) = 0$, where $L_1(a) = n - (1 - k) \sum_{i=1}^{n} x_i/(a - k x_i)$. The value of $\partial^2 L(a, k)/\partial a^2$ at $a = \hat{a}$ is $-(1 - k) \sum_{i=1}^{n} x_i(\hat{a} - k x_i)^{-2}/\hat{a} < 0$, which implies that at $a = \hat{a}$, the likelihood function attains its maximum value. Moreover, the function $L_1(a)$ is an increasing function on the range of a because $\partial L_1(a)/\partial a = (1 - k) \sum_{i=1}^{n} x_i/(a - k x_i)^2 > 0$; the function can take negative and positive values; thus it cuts the a axis exactly at one point. Hence, $\hat{a}$ is unique.

Case 2 (a known, k unknown). In this situation, there may or may not exist a maximum likelihood estimate for k. To see this, consider the likelihood function $L(a, k)$ given in (3) for $-\infty < k \leq a/x_{(n)}$. Since a is known, the likelihood will be regarded as a function $L(k)$ of k only. Set $k^* = a/x_{(n)}$; then

$$\lim_{k \to \infty} L(k) = -\infty, \quad (4)$$
$$\lim_{k \to k^*} L(k) = -\infty \quad \text{if} \quad k^* < 1, \quad (5)$$

and

$$\lim_{k \to k^*} L(k) = +\infty \quad \text{if} \quad k^* > 1. \quad (6)$$

Note that the value of k is not restricted to be less than or equal to 1. From (4) and (5), it follows that, if $k^* < 1$, there is at least one maximum. For a fixed sample size n, $\Pr(k^* <$

$1) = 1 - [1 - (1 - k)^{1/k}]^n > 1 - (3/4)^n$. Similarly, if $k^* > 1$, it follows from (4) and (6) that there may or may not exist a local maximum.

Case 3 (both parameters unknown). Maximum likelihood estimation of the parameters a and k when both are unknown was discussed in detail by Grimshaw (1993). This case is similar to Case 2: In principle, maximum likelihood estimates for a and k may not exist. However, as stated previously, this is unlikely in practical applications, especially when the sample size is reasonably large. To find a solution, Davison (1984) pointed out that, by a change of parameters to $\theta = k/a$ and $k = k$, the problem is reduced to a unidimensional search; we search for θ, which gives a local maximum of the profile log-likelihood (the log-likelihood maximized over k). This is

$$L^*(\theta) = -n - \sum_{i=1}^{n} \log(1 - \theta x_i)$$
$$- n \log \left[-(n\theta)^{-1} \sum_{i=1}^{n} \log(1 - \theta x_i) \right] \quad (7)$$

for $\theta < 1/x_{(n)}$. Suppose that a local maximum $\hat{\theta}$ of (7) can be found; then

$$\hat{k} = -(n^{-1}) \sum_{i=1}^{n} \log(1 - \hat{\theta} x_i) \quad (8)$$

and

$$\hat{a} = \hat{k}/\hat{\theta}. \quad (9)$$

2. GOODNESS-OF-FIT TESTS

In this section, the Cramér–von Mises statistic W^2 and the Anderson–Darling statistic A^2 are described. The Anderson–Darling statistic is a modification of the Cramér–von Mises statistic giving more weight to observations in the tail of the distribution, which is useful in detecting outliers. The null hypothesis is H_0: the random sample $x_1, \ldots, x_n$ comes from Distribution (1).

When parameters a and k are known in (1), the GPD is completely specified; we call this situation Case 0. Then the transformation $z_i = F(x_i)$ produces a z sample that will be uniformly distributed between 0 and 1 under H_0. Many tests of the uniform distribution exist, including Cramér–von Mises tests (see Stephens 1986), so we shall not consider this case further. In the more common Cases 1, 2, and 3, when one or both parameters must be estimated, the goodness-of-fit test procedure is as follows:

1. Find the estimates of unknown parameters as described previously, and make the transformation $z_{(i)} = F(x_{(i)})$, for $i = 1, \ldots, n$, using the estimates where necessary.
2. Calculate statistics W^2 and A^2 as follows:

$$W^2 = \sum_{i=1}^{n} \{z_{(i)} - (2i - 1)/(2n)\}^2 + 1/(12n)$$

and

$$A^2 = -n - (1/n) \sum_{i=1}^{n} (2i - 1)[\log\{z_{(i)}\} + \log\{1 - z_{(n+1-i)}\}].$$

Table 1. Upper-Tail Asymptotic Percentage Points for W^2(normal type) and for A^2 (bold), Cases 1 and 2; p is $Pr(W^2 \geq z)$, or $Pr(A^2 \geq z)$, Where z Is the Table Entry

$k\backslash p$	0.500	0.250	0.100	0.050	0.025	0.010	0.005	0.001
			Case 1: k known, a unknown					
−0.90	0.056	0.083	0.120	0.149	0.178	0.219	0.250	0.324
−0.90	**0.429**	**0.615**	**0.858**	**1.047**	**1.242**	**1.508**	**1.713**	**2.206**
−0.25	0.064	0.098	0.145	0.181	0.219	0.271	0.311	0.406
−0.25	**0.456**	**0.665**	**0.945**	**1.166**	**1.394**	**1.706**	**1.947**	**2.522**
0.00	0.074	0.116	0.174	0.222	0.271	0.338	0.390	0.513
0.00	**0.496**	**0.736**	**1.062**	**1.321**	**1.590**	**1.958**	**2.244**	**2.922**
0.25	0.090	0.148	0.232	0.301	0.373	0.471	0.547	0.728
0.25	**0.581**	**0.886**	**1.310**	**1.651**	**2.006**	**2.493**	**2.870**	**3.765**
0.50	0.119	0.209	0.347	0.461	0.581	0.742	0.869	1.168
0.50	**0.774**	**1.248**	**1.933**	**2.492**	**3.077**	**3.878**	**4.379**	**5.969**
			Case 2: a known, k unknown					
−0.90	0.097	0.171	0.287	0.382	0.482	0.618	0.723	0.972
−0.90	**0.608**	**0.981**	**1.540**	**2.001**	**2.484**	**3.144**	**3.649**	**4.865**
−0.25	0.105	0.186	0.311	0.415	0.523	0.669	0.785	1.056
−0.25	**0.652**	**1.059**	**1.664**	**2.162**	**2.684**	**3.397**	**3.957**	**5.258**
0.00	0.109	0.193	0.323	0.430	0.542	0.693	0.813	1.093
0.00	**0.680**	**1.104**	**1.732**	**2.248**	**2.789**	**3.528**	**4.101**	**5.457**
0.25	0.114	0.201	0.335	0.446	0.562	0.719	0.842	1.132
0.25	**0.718**	**1.164**	**1.819**	**2.356**	**2.918**	**3.686**	**4.277**	**5.692**
0.50	0.119	0.209	0.347	0.461	0.581	0.742	0.869	1.168
0.50	**0.774**	**1.248**	**1.933**	**2.492**	**3.077**	**3.878**	**4.379**	**5.969**

Tables 1 and 2 give upper-tail asymptotic percentage points for the statistics W^2 and A^2, for values of k between −0.9 and 0.5, and for Cases 1, 2, and 3. In Cases 2 and 3, where k must be estimated, the appropriate table should be entered at $\hat{k}$, and in all tables, if k or $\hat{k}$ is greater than 0.5, the table should be entered at $k = 0.5$. Critical points for other values of k can be obtained by interpolation; linear interpolation in Table 2 for A^2, Case 3, gives a maximum error in α from 0.0011 to 0.0003 as α moves through the important range R; $0.10 \geq \alpha \geq 0.005$ and is less than 0.003 for the two values S; $\alpha = 0.5$ and 0.25; for W^2 the corresponding figures are 0.0025 to 0.0004 over R and less than 0.007 for S. For the much less important cases 1 and 2, linear interpolation in Table 1 gives maximum error for A^2, Case 1, from 0.009 to 0.0017 over R, and 0.014 for S;

for W^2, Case 1, the figures are 0.0055 to 0.0015 over R and 0.008 for S. For Case 2, the maximum errors are smaller by a factor of approximately 10.

The points in these tables were calculated from the asymptotic theory to be presented in Section 6. Points for finite n were generated from Monte Carlo samples, using 10,000 samples for each combination of n and k. The results show that these points converge quickly to the asymptotic points for all three cases, a feature of statistics W^2 and A^2 found also in many other applications. The asymptotic points can then be used with good accuracy for, say, $n \geq 25$. More extensive versions of Tables 1 and 2, and also tables showing the convergence to the asymptotic points, were given by Choulakian and Stephens (2000).

Table 2. Case 3: Both k and a Unknown: Upper-Tail Asymptotic Percentage Points for W^2(normal type) and for A^2 (bold); p is $Pr(W^2 \geq z)$, or $Pr(A^2 \geq z)$, Where z Is the Table Entry

$k\backslash p$	0.500	0.250	0.100	0.050	0.025	0.010	0.005	0.001
−0.90	0.046	0.067	0.094	0.115	0.136	0.165	0.187	0.239
−0.90	**0.339**	**0.471**	**0.641**	**0.771**	**0.905**	**1.086**	**1.226**	**1.559**
−0.50	0.049	0.072	0.101	0.124	0.147	0.179	0.204	0.264
−0.50	**0.356**	**0.499**	**0.685**	**0.830**	**0.978**	**1.180**	**1.336**	**1.707**
−0.20	0.053	0.078	0.111	0.137	0.164	0.200	0.228	0.294
−0.20	**0.376**	**0.534**	**0.741**	**0.903**	**1.069**	**1.296**	**1.471**	**1.893**
−0.10	0.055	0.081	0.116	0.144	0.172	0.210	0.240	0.310
−0.10	**0.386**	**0.550**	**0.766**	**0.935**	**1.110**	**1.348**	**1.532**	**1.966**
0.00	0.057	0.086	0.124	0.153	0.183	0.224	0.255	0.330
0.00	**0.397**	**0.569**	**0.796**	**0.974**	**1.158**	**1.409**	**1.603**	**2.064**
0.10	0.059	0.089	0.129	0.160	0.192	0.236	0.270	0.351
0.10	**0.410**	**0.591**	**0.831**	**1.020**	**1.215**	**1.481**	**1.687**	**2.176**
0.20	0.062	0.094	0.137	0.171	0.206	0.254	0.291	0.380
0.20	**0.426**	**0.617**	**0.873**	**1.074**	**1.283**	**1.567**	**1.788**	**2.314**
0.30	0.065	0.100	0.147	0.184	0.223	0.276	0.317	0.415
0.30	**0.445**	**0.649**	**0.924**	**1.140**	**1.365**	**1.672**	**1.909**	**2.475**
0.40	0.069	0.107	0.159	0.201	0.244	0.303	0.349	0.458
0.40	**0.468**	**0.688**	**0.985**	**1.221**	**1.465**	**1.799**	**2.058**	**2.674**
0.50	0.074	0.116	0.174	0.222	0.271	0.338	0.390	0.513
0.50	**0.496**	**0.735**	**1.061**	**1.321**	**1.590**	**1.958**	**2.243**	**2.922**

3. EXAMPLES

In this section we consider two specific examples and then discuss the overall results when the GPD is applied to 238 Canadian rivers.

In the first example, the data are $n = 72$ exceedances of flood peaks (in m^3/s) of the Wheaton River near Carcross in Yukon Territory, Canada. The initial threshold value τ is 27.50 (how this was found will be explained later), and the 72 exceedances, for the years 1958 to 1984, rounded to one decimal place, were 1.7, 2.2, 14.4, 1.1, 0.4, 20.6, 5.3, 0.7, 1.9, 13, 12, 9.3, 1.4, 18.7, 8.5, 25.5, 11.6, 14.1, 22.1, 1.1, 2.5, 14.4, 1.7, 37.6, 0.6, 2.2, 39, 0.3, 15, 11, 7.3, 22.9, 1.7, 0.1, 1.1, 0.6, 9, 1.7, 7, 20.1, 0.4, 2.8, 14.1, 9.9, 10.4, 10.7, 30, 3.6, 5.6, 30.8, 13.3, 4.2, 25.5, 3.4, 11.9, 21.5, 27.6, 36.4, 2.7, 64, 1.5, 2.5, 27.4, 1, 27.1, 20.2, 16.8, 5.3, 9.7, 27.5, 2.5, 27. The maximum likelihood estimates of the parameters are $\hat{k} = -0.006$ and $\hat{a} = 12.14$. We now apply the tests presented in Section 2; the values of the test statistics are $W^2 = 0.2389$ (p value < 0.025) and $A^2 = 1.452$ (p value < 0.010), so the GPD does not fit the dataset well at the threshold value of 27.50. Next we follow the technique of Davison and Smith (1990), who suggested raising the threshold until the GPD fits the new values of the remaining exceedances. Here the threshold was raised successively by the value of the smallest order statistic, which is then deleted, until the p values of the A^2 and W^2 statistics exceeded 10%. This happens when six order statistics are deleted; the threshold value is then $27.50 + 0.60 = 28.10$. The GPD now fits the data quite well: Details are given in Table 3. This example shows how the tests may help in choosing a threshold value in the POT model.

In this connection, we next revisit the example given by Davison and Smith (1990). Their tables 5 and 6 give the parameter estimates (Case 3) and the values of the Anderson–Darling statistic for a range of thresholds from 140 down to 70 at intervals of 10 units. The values of the Anderson–Darling statistic were compared with 1.74, the 5% point for a test for exponentiality since no test for GPD was available, and the statistics were not nearly significant by this criterion. Now Table 2 can be used to give the asymptotic 5% critical values z_5 for A^2 corresponding to the estimate of the parameter k. The first two and last two threshold results are given in Table 4. Davison and Smith pointed out the sudden increase in the value of A^2 at threshold level 70; against the exponential

Table 4. Results for Various Thresholds

Threshold	140	130	80	70
$\hat{k}$	0.24	0.14	−0.34	−0.32
z_5	1.10	1.04	0.865	0.872
A^2	0.42	0.36	0.37	0.87

point this value is still not significant, but using Table 2, it falls just at the critical 5% level.

The Wheaton River in the previous example is one of 238 Canadian rivers for which we have similar data. We now examine how well the GPD fits the exceedances for the other rivers. First one must decide the threshold level for a given river flow. The first threshold estimate, $\hat{\tau}$, was chosen so that the number of exceedances per year could be modeled by a Poisson distribution; see, for instance, Todorovic (1979). This was done by taking $\hat{\tau}$ such that, if $N(\hat{\tau})$ is the number of exceedances, the mean of $N(\hat{\tau})$ divided by its variance was approximately 1. The Poisson assumption will be tested more rigorously later. After the threshold level was chosen, the maximum likelihood estimates of the parameters k and a were calculated and the W^2 and A^2 tests applied. Table 5 gives the frequencies of p values for the 238 rivers. It is clear that the GPD fits quite well; using W^2, only 9 gave p values less than 0.01, and using A^2, there were only 15. At the 0.05 level, these figures are 34 using W^2 and 49 using A^2. The results demonstrate also that A^2 is more sensitive (and therefore more powerful) than W^2 against possible outliers in the tail, as was suggested in Section 2. More details were given by Choulakian and Stephens (2000). For the 49 "rejections" using A^2, the threshold was increased, as described in the first example, by deleting the smallest order statistics until the p value became larger than 0.10. Then only 10 of the 49 sets were still rejected as GPD by A^2. Finally, for these 10 rejected river flows, the Poisson assumption was tested by the Cramér–von Mises tests given by Spinelli and Stephens (1997). Only one dataset rejected the Poisson assumption.

Another result of interest is that 229 of the 238 values of $\hat{k}$ were between −0.5 and 0.4; this confirms the findings of Hosking and Wallis (1987), who restricted attention to $-0.5 < k < 0.5$, where the great majority of values fall.

Table 3. Maximum Likelihood Estimates of k and a, the Values of the Statistics W^2 and A^2 and Their p Values When the Smallest Order Statistics of the Wheaton River Dataset Are Deleted One by One; n Is the Sample Size

Deleted observation	n	$\hat{k}$	$\hat{a}$	W^2 (p value)	A^2 (p value)
None	72	−0.006	12.14	0.2309 (< 0.01)	1.452 (< 0.01)
0.1	71	0.019	12.63	0.208 (< 0.025)	1.317 (< 0.025)
0.3	70	0.046	13.15	0.187 (< 0.025)	1.197 (< 0.025)
0.4	69	0.047	13.34	0.162 (< 0.05)	1.039 (< 0.05)
0.4	68	0.077	13.93	0.143 (< 0.10)	0.967 (≈ 0.05)
0.6	67	0.078	14.13	0.125 (≈ 0.10)	0.853 (< 0.10)
0.6	66	0.112	14.8	0.112 (> 0.10)	0.804 (> 0.10)

Table 5. Frequencies of the p Values of W^2 and of A^2 for 238 Canadian Rivers

p value	No. of W^2	No. of A^2
< 0.001	1	1
$0.001 < p < 0.005$	2	7
$0.005 < p < 0.01$	6	7
$0.01 < p < 0.025$	12	19
$0.025 < p < 0.05$	13	15
$0.05 < p < 0.1$	28	20
$0.1 < p < 0.15$	21	21
$0.15 < p < 0.25$	31	31
$0.25 < p < 0.5$	51	42
> 0.5	73	75
Total	238	238

Table 6. Power Study for the Method of Choosing Threshold

Distribution	n	Power	Mean K	Std. dev.	Max K	Q1	Q2	Q3
Weibull(0.75)	100	50.9	8.169	9.176	52	2	5	11
	50	27.8	4.525	4.782	26	1	3	5
	30	18.6	2.925	2.231	11	1	2	4
Weibull(1.25)	100	23.6	4.767	9.926	48	1	2	5
	50	10.1	3.564	4.568	28	1	2	3
	30	4.7	2.872	2.651	11	1	2	3
Lognormal	100	83.5	2.728	4.185	43	1	1	3
	50	38.5	1.662	1.734	15	1	1	2
	30	20.7	1.502	1.202	9	1	1	1
Gamma(2,1)	100	89.7	5.118	6.358	47	1	3	6
	50	42.9	2.359	2.446	19	1	1	3
	30	19.8	1.717	1.710	13	1	1	2
$Z(-0.4, 0.4)$	100	99.6	10.121	8.838	55	3	8	14
	50	78.0	3.506	3.428	20	1	2	5
	30	43.9	1.991	1.737	13	1	1	2
$Z(-0.4, 0.0)$	100	91.9	7.476	8.503	66	2	4	10
	50	55.3	2.901	3.116	21	1	2	3
	30	27.9	2.018	1.672	11	1	1	2
$Z(0.25, 0.25)$	100	94.7	6.337	6.887	46	2	4	8
	50	54.9	2.537	2.750	26	1	1	3
	30	23.0	1.600	1.277	13	1	1	2

NOTE: Power is the percentage of 1,000 samples of size n, from the distribution given, rejected by statistic A^2 using a 5% test for GPD. Columns 4 through 9 give statistics on K, the number of low-order statistics that must be rejected, (zeros excluded), to pass the test. Q1, Q2, and Q3 are the quartiles of K. Weibull(k) refers to the standard Weibull distribution with parameter k. The label $Z(r, s)$ means that the distribution is an equiweighted mixture of a GPD(r, 1) and a GPD(s, 1).

4. POWER OF THE FAILURE-TO-REJECT METHOD

The method of choosing the threshold by using as many exceedances as possible subject to passing a test for GPD can be investigated for power as follows. The data analyst wishes to fit a GPD to the exceedances far enough in the tail. She\he therefore starts with as many as possible, say 100, and tests; if the test fails, she/he omits the smallest values one by one and tests again until the test yields acceptance for the GPD. The efficiency of this procedure can now be investigated when the true distribution is *not* GPD. Suppose, for instance, the true exceedance distribution is Weibull with parameter 0.75, and suppose, starting with sample size 100, the sample fails the GPD test until K lower-order statistics have been removed; then it passes. Statistic K will vary from sample to sample, but its mean will give a measure of the power of the test. The standard deviation and other statistics from the distribution of K are also of interest. This has been investigated for various alternative distributions for the true exceedances; Table 6 provides these results. They are based on 1,000 Monte Carlo samples of initial size $n = 100$, 50, or 30. Only statistic A^2 is reported because A^2 outperforms W^2 at detecting tail discrepancies. Column 3 gives the initial rejection rate for a GPD test with parameters estimated, at test size 0.05; this is also the power in a conventional study of power against the specified alternative. Subsequent columns give the mean, standard deviation, the three quartiles, and the maximum values for K. Thus for the Weibull(0.75) alternative, 509 of the 1,000 samples of size 100 were rejected as GPD initially; *of these rejected samples*, the mean number of order statistics to be deleted before acceptance was 8.169 with standard deviation 9.178. The distribution of K is long-tailed; the quartiles are 2, 5, and 11, respectively, but one sample needed 52 order statistics removed to achieve acceptance by the GPD test.

In Table 6, the distributions were chosen to give a reasonable starting rejection rate. Several other distributions could be supposed as alternatives (e.g., the half-normal, half-Cauchy, or gamma distributions), but these can be made very close to the GPD by suitable choice of parameters, so the initial power is small, and very few low-order statistics, if any, need be deleted to achieve GPD acceptance (cf. the numerical results in Sec. 5). In their section 9, Davison and Smith suggested that the exceedances for their example might possibly be truly modeled by a mixture of two populations, and in Table 6 we have included three mixtures of GPD as possible true models. The test statistic distinguishes quite effectively between these mixtures and a single GPD.

For all alternative distributions considered in Table 6, it may be seen that, as n increases, the mean, standard deviation, and maximum of K all increase; the results also show the increase in the power of the test as the sample size increases.

5. VERSATILITY OF THE GPD

As was suggested previously, some distributions often fitted to long-tailed distributions may be brought close to a GPD by suitable choice of k and a. This closeness was investigated, and results are given in Table 7. In this table, for example, the standard half-normal and standard half-Cauchy distributions (respectively, the distribution of $|X|$ when X has the standard normal or the standard Cauchy distribution) are compared with the GPD(0.293,1.022) for the first and GPD(-0.710, 1.152) for the second; the standard lognormal is compared with GPD(-0.140, -1.406). The choice of the GPD parameters can be made by equating the first two moments of the GPD and those of the distribution compared, but this does not produce as good a fit as the following maximum likelihood procedure.

Table 7. Percentiles (Q%) of the GPD(k,a) Distribution Matched to
Alternative Distributions

Q%	1	5	10	50	90	95	99
GPD(0.293, 1.022)	0.010	0.052	0.106	0.641	1.711	2.038	2.583
Half-normal	0.015	0.065	0.125	0.675	1.645	1.955	2.575
GPD(−0.14, 1.406)	0.014	0.072	0.149	1.023	3.820	5.233	9.093
Lognormal	0.098	0.193	0.277	1.005	3.615	5.181	10.227
GPD(−0.71, 1.152)	0.012	0.060	0.126	1.032	6.699	11.990	41.054
Half-Cauchy	0.016	0.079	0.158	1.000	6.314	12.706	63.657
GPD(0.251, 2.455)	0.025	0.125	0.255	1.562	4.293	5.170	6.702
Gamma(2,1)	0.149	0.355	0.532	1.678	3.890	4.744	6.638
GPD(0.314,1.152)	0.012	0.059	0.119	0.718	1.888	2.237	2.805
Weibull(1.5)	0.047	0.138	0.223	0.783	1.744	2.078	2.768
GPD(0.199,1.111)	0.011	0.057	0.116	0.719	2.052	2.507	3.350
Weibull(1.25)	0.025	0.093	0.165	0.746	1.949	2.405	3.393
GPD(−0.375, 0.788)	0.008	0.041	0.085	0.624	2.882	4.361	9.715
Weibull(0.75)	0.002	0.019	0.050	0.613	3.041	4.319	7.662
GPD(−1.414, 0.322)	0.003	0.017	0.037	0.379	5.680	15.514	153.02
Weibull(0.5)	0.000	0.003	0.011	0.480	5.302	8.974	21.208

A sample of 500 values was taken from the half-normal distribution, and a GPD fit was made, estimating the parameters by maximum likelihood, first solving (7) for θ, and then obtaining k, a from (8) and (9). This was repeated 100 times, and the average values of the k, a values were used as the parameters in the GPD. Although the matches are clearly not perfect, the error in α, the probability level of a given percentile, is small when one distribution is used instead of another. Similar comparisons are given for several Weibull distributions, and a gamma distribution. The exponential (Weibull with parameter 1) is omitted because it is a special case of the GPD with $k = 0$. Again the two distributions are quite close for the Weibull parameter greater than 1 (where the Weibull has a mode), but the match is less good for Weibull with parameter less than 1—for example, the Weibull(0.5), or the Weibull(0.75)—where the density rises to infinity at $X = 0$.

Overall, the results suggest that a GPD could often be used as a model for data with a long tail when neither a mode nor an infinite density is suggested by the nature of the variables or by the data themselves.

6. ASYMPTOTIC THEORY OF THE TESTS

In this section we summarize the asymptotic theory of Cramér–von Mises tests. The calculation of asymptotic distributions of the statistics follows a procedure described, for instance, by Stephens (1976). It is based on the fact that $y_n(z) = \sqrt{n}\{F_n(z) - z\}, 0 \le z \le 1$, where $F_n(z)$ is the empirical distribution function of the z set, it tends to a Gaussian process $y(z)$ as $n \to \infty$ and the statistics are functionals of this process. The mean of $y(z)$ is 0: We need the covariance function $\rho(s, t) = E\{y(s)y(t)\}, 0 \le s, t \le 1$. When all the parameters are known, this covariance is $\rho_0(s, t) = \min(s, t) - st$. When parameters are estimated, the covariance will depend in general on the true values of the estimated parameters. However, if the method of estimation is efficient, the covariance will not depend on the scale parameter a but will depend on the shape parameter k. We illustrate for Case 3 only. As the sample size $n \to \infty$, the maximum likelihood estimators $(\hat{a}, \hat{k})$ have a bivariate normal

distribution with mean (a, k) and variance-covariance matrix Σ, where

$$\Sigma = n^{-1} \begin{pmatrix} 2a^2(1-k) & a(1-k) \\ a(1-k) & (1-k)^2 \end{pmatrix}. \quad (10)$$

When both parameters a and k are estimated, the covariance function of $y(z)$ becomes

$$\rho_3(s, t) = \rho_0(s, t) - \{g(s)\}' \Sigma g(t), \quad (11)$$

where $s = F(x)$, and $g(s) = (g_1(s), g_2(s))'$ is a vector having coordinates

$$g_1(s) = \partial F / \partial a = (1-s)\{1 - (1-s)^{-k}\}/(ak)$$

and

$$g_2(s) = \partial F / \partial k = (1-s)k \log(1-s) - 1 + (1-s)^{-k}/k^2.$$

When Σ and $g(s)$ are inserted into (11), $\rho_3(s, t)$ will be independent of a. When $k \ge 0.5$, the maximum likelihood estimates of a and k are superefficient in the sense of Darling (1955), and then the covariance and the resulting asymptotic distributions will be the same as for $k = 0.5$. Thus, if $0.5 \le \hat{k} \le 1$, the table should be entered at $\hat{k} = 0.5$, as described in Section 2. In Case 1 the covariance of $y(z)$ becomes

$$\rho_1(s, t) = \rho_0(s, t) - (1 - 2k)g_1(s)g_1(t) \quad (12)$$

and for Case 2 it becomes

$$\rho_2(s, t) = \rho_0(s, t) - (1-k)(1 - 2k)g_2(s)g_2(t)/2. \quad (13)$$

In both these cases, at $k = 0.5$, the asymptotic covariance becomes $\rho_0(s, t)$. This is the same as for a test for Case 0, when both a and k are known, and the asymptotic points are the same as for such a test (see Stephens 1986).

The Cramér–von Mises statistic W^2 is based directly on the process $y(z)$, while A^2 is based on the process $w(z) = y(z)/\{z(1 - z)\}^{1/2}$; asymptotically the distributions of W^2 and A^2 are those of $W^2 = \int_0^1 y^2(z)dz$ and $A^2 = \int_0^1 w^2(z)dz$. The asymptotic distributions of both statistics are a sum of weighted independent χ_1^2 variables; the weights for W^2

must be found from the eigenvalues of an integral equation with the appropriate $\rho_j(s, t)$ for Case j as kernel. For A^2, $\rho_A(s, t)$, the covariance of the $w(z)$ process, is $\rho_j(s, t)/\{st(1-s)(1-t)\}^{1/2}$, and this is the kernel of the integral equation. Once the weights are known, the percentage points of the distributions can be calculated by Imhof's method. For details of these procedures, see, for example, Stephens (1976).

ACKNOWLEDGMENTS

REFERENCES

Castillo, E., and Hadi, A. S. (1997), "Fitting the Generalized Pareto Distribution to Data," *Journal of the American Statistical Association*, 92, 1609–1620.

Chen, G., and Balakrishnan, N. (1995), "The Infeasibility of Probability Weighted Moments Estimation of Some Generalized Distributions," in *Recent Advances in Life-Testing and Reliability*, ed. N. Balakrishnan, London: CRC Press, pp. 565–573.

Choulakian, V., and Stephens, M. A. (2000), "Goodness-of-Fit Tests for the Generalized Pareto Distribution," research report, Simon Fraser University, Dept. of Mathematics and Statistics.

Darling, D. (1955), "The Cramér–von Mises Test in the Parametric Case," *The Annals of Mathematical Statistics*, 26, 1–20.

Davison, A. C. (1984), "Modeling Excesses Over High Thresholds, With an Application," in *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, Dordrecht: D. Reidel, pp. 461–482.

Davison, A. C., and Smith, R. L. (1990), "Models for Exceedances Over High Thresholds" (with comments), *Journal of the Royal Statistical Society*, Ser. B, 52, 393–442.

Dupuis, D. J. (1996), "Estimating the Probability of Obtaining Nonfeasible Parameter Estimates of the Generalized Pareto Distribution," *Journal of Statistical Computer Simulation*, 54, 197–209.

Dupuis, D. J., and Tsao, M. (1998), "A Hybrid Estimator for Generalized Pareto and Extreme-Value Distributions," *Communications in Statistics—Theory and Methods*, 27, 925–941.

Grimshaw, S. D. (1993), "Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution," *Technometrics*, 35, 185–191.

Hosking, J. R. M., and Wallis, J. R. (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, 29, 339–349.

Pickands, J. (1975), "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3, 119–131.

Smith, R. L. (1984), "Threshold Methods for Sample Extremes," in *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, Dordrecht: Reidel, pp. 6211–6638.

———(1989), "Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-level Ozone," *Statistical Science*, 4, 367–393.

———(1990), "Extreme Value Theory," in *Handbook of Applicable Mathematics* (Vol. 7), ed. W. Ledermann, Chichester, U.K.: Wiley.

Spinelli, J. J., and Stephens, M. A. (1997), "Test of Fit for the Poisson Distribution," *The Canadian Journal of Statistics*, 25, 257–268.

Stephens, M. A. (1976), "Asymptotic Results for Goodness-of-Fit Statistics With Unknown Parameters," *The Annals of Statistics*, 4, 357–369.

———(1986), "Tests Based on EDF Statistics," in *Goodness-of-fit Techniques*, eds. R. B. D'Agostino and M. A. Stephens, New York: Marcel Dekker, pp. 97–122.

Todorovic, P. (1979), "A Probabilistic Approach to Analysis and Prediction of Floods," in *Proceedings of the International Statistical Institute, Buenos Aires* (Vol. 1), pp. 113–124.

Book Reviews

This section will review those books whose content and level reflect the general editorial policy of *Technometrics*. Publishers should send books for review to Eric R. Ziegel, BP Naperville Complex, Mail Code C-7, 150 West Warrenville Road, Naperville, IL 60563-8460. (*ziegeler@bp.com*).

The opinions expressed in this section are those of the reviewers. These opinions do not represent positions of the reviewer's organization and may not reflect those of the editors or the sponsoring societies. Listed prices reflect information provided by the publisher and may not be current.

The book purchase program of the American Society for Quality can provide some of these books at reduced prices for members. For information, contact the American Society for Quality, 1-800-248-1946.