Journal of Modern Applied Statistical Methods

Volume 17 | Issue 2 Article 11

3-22-2019

Maximum Likelihood Estimation for the Generalized Pareto Distribution and Goodness-Of-Fit Test with Censored Data

Minh H. Pham
University of South Florida, minhpham@mail.usf.edu

Chris Tsokos
University of South Florida, ctsokos@usf.edu

Bong-Jin Choi
North Dakota State University, bongjin.choi@ndsu.edu

Recommended Citation
Pham, M. H., Tsokos, C., & Choi, B.-J. (2018). Maximum likelihood estimation for the generalized Pareto distribution and goodness-of-fit test with censored data. *Journal of Modern Applied Statistical Methods, 17*(2), eP2608. doi: 10.22237/jmasm/1553261471

Maximum Likelihood Estimation for the Generalized Pareto Distribution and Goodness-Of-Fit Test with Censored Data

Erratum

In the original published version of this article, the affiliation for the third author was incorrectly given as "University of North Carolina at Chapel Hill" instead of "North Dakota State University". This has been corrected.

Journal of Modern Applied Statistical Methods
November 2018, Vol. 17, No. 2, eP2608
doi: 10.22237/jmasm/1553261471

ISSN 1538 − 9472

EMERGING SCHOLARS

Maximum Likelihood Estimation for the Generalized Pareto Distribution and Goodness-of-Fit Test with Censored Data

Minh H. Pham
University of South Florida
Tampa, FL

Chris Tsokos
University of South Florida
Tampa, FL

Bong-Jin Choi
North Dakota State University
Fargo, ND

The generalized Pareto distribution (GPD) is a flexible parametric model commonly used in financial modeling. Maximum likelihood estimation (MLE) of the GPD was proposed by Grimshaw (1993). Maximum likelihood estimation of the GPD for censored data is developed, and a goodness-of-fit test is constructed to verify an MLE algorithm in R and to support the model-validation step. The algorithms were composed in R. Grimshaw's algorithm outperforms functions available in the R package 'gPdtest'. A simulation study showed the MLE method for censored data and the goodness-of-fit test are both reliable.

Keywords: Computational statistics, survival analysis, generalized Pareto distribution, maximum likelihood estimation, censored data, goodness-of-fit test

Introduction

The generalized extreme value distribution (GEVD) is a family of distributions that are usually used to model the maxima of long sequences of random variables. The GEVD is useful when the data contain a finite set of maxima (Embrechts, Klüppelberg, & Mikosch, 2012). One particularly useful GEVD distribution is the generalized Pareto distribution (GPD), which was introduced by Pickands (1975) to model excess over thresholds instead of maxima. GPD was then broadly applied to many topics such as environmental (Hosking & Wallis, 1987), engineering (Castillo, 2012; Holmes & Moriarty, 1999), and health data (Cebrián, Denuit, & Lambert, 2003).

The GPD is a two-parameter probability distribution. The cumulative probability distribution function is given by

doi: 10.22237/jmasm/1553261471 | Accepted: September 10, 2017; Published: March 22, 2019.
Correspondence: Minh H. Pham, minhpham@mail.usf.edu

$$\mathrm{F}(x;k,\alpha)=\begin{cases}1-\left(1-\dfrac{kx}{\alpha}\right)^{1/k}, & k\neq 0\\ 1-\mathrm{e}^{-\frac{x}{\alpha}}, & k=0\end{cases}, \tag{1}$$

where k is the shape parameter and α is the scale parameter. Uniform, Pareto, and exponential distributions are special cases of the GPD; the GPD becomes the exponential distribution if $k = 0$, the uniform distribution if $k = 1$, and the Pareto distribution if $k < 0$.

Hosking and Wallis (1987) discussed the estimation by the method of moments (ME). Their estimations were

$$\hat{\alpha}_{\mathrm{ME}}=\frac{\bar{X}\left(\dfrac{\bar{X}^2}{s^2}+1\right)}{2} \quad \text{and} \quad \hat{k}_{\mathrm{ME}}=\frac{\dfrac{\bar{X}^2}{s^2}-1}{2}, \tag{2}$$

where $\bar{X}$ and s^2 are the sample mean and variance, respectively. In the same study, they also considered the probability-weighted moment (PWM) estimation method, and their results are given by

$$\hat{\alpha}_{\mathrm{PWM}}=\frac{2\gamma\bar{x}}{\bar{x}-2\gamma} \quad \text{and} \quad \hat{k}_{\mathrm{PWM}}=\frac{\bar{x}}{\bar{x}-2\gamma}-2, \tag{3}$$

where

$$\gamma=n^{-1}\sum_{i=1}^{n}\frac{n-i}{n-1}x_i$$

Grimshaw (1993) published an algorithm for computing the maximum likelihood estimation (MLE) of the parameters of the GPD. Juárez and Schucany (2004) proposed the minimum probability density power divergence method, which allows control over efficiency and robustness. When efficiency is maximized, this method is equivalent to the MLE method. Zhang (2010) proposed an improved maximum likelihood estimation using the empirical Bayesian method (Zhang, 2007). Zhang's estimation was found to be better than other procedures in terms of efficiency and bias.

According to Zhang (2007), there were problems associated with all of these methods. The PWM estimators do not exist asymptotically if $k \leq -1$. The ME estimators are not asymptotically consistent if the simulated data has $k \leq -1/2$. Both the ME and PWM estimators have low asymptotic efficiencies. MLE estimators are asymptotically efficient, but it is difficult to compute them and MLEs do not exist for $k \geq 1$.

The aim of the present study is to develop an estimation algorithm for right-censored survival data using the MLE method. The package *gPdtest*, by Gonzalez Estrada and Villasenor Alva (2012), includes the function `gpd.fit()` that calculates the estimation of the parameters. This program uses the MLE method and the combined method proposed by the authors. The MLE method of this function did not perform well in the simulation of this study.

Mathematical Approach

Likelihood Function

Let δ be the right-censoring indicator, with value 1 being an observation and value 0 being a censored point. Klein and Moeschberger (2003) described the likelihood function

$$L = \prod_{i=1}^{n} \left[\mathrm{f}\left(X_i \mid \theta \right)^{\delta_i} \cdot \mathrm{S}\left(X_i \mid \theta \right)^{1-\delta_i} \right]. \tag{4}$$

The likelihood function and log-likelihood function for the generalized Pareto distribution can be written as

$$L = \prod_{i=1}^{n} \left[\frac{1}{\alpha^{\delta_i}} \left(1 - \frac{k}{\alpha} X_i \right)^{\delta_i \left(\frac{1}{k} - 1 \right)} \cdot \left(1 - \frac{k}{\alpha} X_i \right)^{(1-\delta_i)/k} \right] \tag{5}$$

and

$$\ln L = \sum_{i=1}^{n} \left[-\delta_i \ln \alpha + \left(\frac{1}{k} - \delta_i \right) \ln \left(1 - \frac{k}{\alpha} X_i \right) \right]. \tag{6}$$

To estimate the local maximum of ln$L$, we have to solve the following system of equations:

$$\begin{cases} \dfrac{\partial \ln L}{\partial \alpha} = 0 \\ \dfrac{\partial \ln L}{\partial k} = 0 \end{cases} \Leftrightarrow \begin{cases} n = \sum_{i=1}^{n} \left\{ \left[1 - \delta_i \cdot \dfrac{-1}{\sum_{j=1}^{n} \delta_j} \cdot \sum_{k=1}^{n} \ln\left(1 - \dfrac{k}{\alpha} X_k\right) \right] \left(1 - \dfrac{k}{\alpha} X_i\right)^{-1} \right\} \\ k = -\dfrac{1}{\sum_{j=1}^{n} \delta_j} \sum_{i=1}^{n} \ln\left(1 - \dfrac{k}{\alpha} X_i\right) \end{cases}. \quad (7)$$

Order the data so all of the observations are placed before the censored values. Let r be the number of observations in the data. From this arrangement, we have

$$\delta_1 = \delta_2 = \ldots = \delta_r = 1 \quad \text{and} \quad \delta_{r+1} = \ldots = \delta_n = 0\,.$$

The percentage of censorship in the data is r / n.

Let $\theta = k / \alpha$. The simultaneous equations (7) can be rewritten as the following equations (8) and (9):

$$\frac{1}{nr} \sum_{i=1}^{n} \ln(1 - \theta X_i) \cdot \sum_{i=1}^{n} (1 - \theta X_i)^{-1} + \frac{1}{n} \sum_{i=1}^{n} (1 - \theta X_i)^{-1} - 1 = 0\,, \quad (8)$$

$$k = -\frac{1}{r} \sum_{i=1}^{n} \ln(1 - \theta X_i). \quad (9)$$

This format is similar to Grimshaw's (1993) pair of equations. If $r = n$, which means there is no censorship in the data, this pair of equations become the equations that were presented by Grimshaw. Thus, similar to Grimshaw's work, the left-hand side of equation (8) is the univariate function given by

$$\mathrm{h}(\theta) = \frac{1}{nr} \sum_{i=1}^{n} \ln(1 - \theta X_i) \cdot \sum_{i=1}^{n} (1 - \theta X_i)^{-1} + \frac{1}{n} \sum_{i=1}^{n} (1 - \theta X_i)^{-1} - 1. \quad (10)$$

Finding solutions for this function will easily lead to the solutions for system (7). A closed-form solution for this function is not known. Using some mathematical characteristics of the function $\mathrm{h}(\theta)$ presented in Appendix A, the following algorithm can be used to estimate the solutions:

The Structure of the Algorithm

1. Let $\epsilon = 10^{-4} / \bar{X}$. For numerical purposes, $\theta_1 = \theta_2$ if $|\theta_1 - \theta_2| < \epsilon$.
2. The lower and upper bounds for solution of $h(\theta)$ are calculated to be

$$\theta_L = \frac{2\left(X_{(1)} - \bar{X}\right)}{\left(X_{(1)}\right)^2}; \quad \theta_U = \frac{1}{X_{(n)}} - \epsilon$$

3. $\lim_{\theta \to 0} h''(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{r} X_i$ (proof in Appendix A). If $\lim_{\theta \to 0} h''(\theta) > 0$ then there exists at least one solution of $h(\theta)$ on $(\theta_L, 0)$ and at least one zero of $h(\theta)$ on $(0, \theta_U)$.
   a. Use the Newton-Raphson algorithm with initial θ_L to determine the solution θ_1 on $(\theta_L, 0)$.
   b. Use the Newton-Raphson algorithm with initial θ_U to determine the solution θ_2 on $(\theta_U, 0)$.
4. If $\lim_{\theta \to 0} h''(\theta) < 0$ then there exists no solution or an even number of solutions on each of the intervals $(\theta_L, 0)$ and $(0, \theta_U)$.
   a. To determine the first solution θ_1 of $h(\theta)$ on $(\theta_L, 0)$, use the Newton-Raphson algorithm with initial value θ_L. If the Newton-Raphson algorithm does not converge, there is no solution on $(\theta_L, 0)$. If θ_1 exists, calculate $h'(\theta)$ using equation (A3) in Appendix A. If $h'(\theta_1) > 0$, the second solution is on $(\theta_1, 0)$; otherwise, the second solution is on (θ_L, θ_1). We can use the bisection algorithm on the appropriate interval to determine the second solution and denote it θ_2.
   b. To determine the solution θ_3 of $h(\theta)$ on $(0, \theta_U)$, use the Newton-Raphson algorithm with initial value θ_U. If the Newton-Raphson algorithm does not converge, there is no solution on $(\theta_L, 0)$. If θ_3 exists, calculate $h'(\theta)$ using equation (A3) in Appendix A. If $h'(\theta_3) > 0$, the second solution is on $(0, \theta_3)$; otherwise, the second solution is on (θ_3, θ_U). We can use the bisection algorithm on the appropriate interval to determine the second solution and denote it θ_4.
5. For each θ_i available, calculate k_i and α_i using equation (9) and the log-likelihood $\ln L_i$ using equation (6). The pair (k_i, α_i) that generates the local maximum of $\ln L_i$ is the final estimate of our algorithm, as presented in Figure 1.

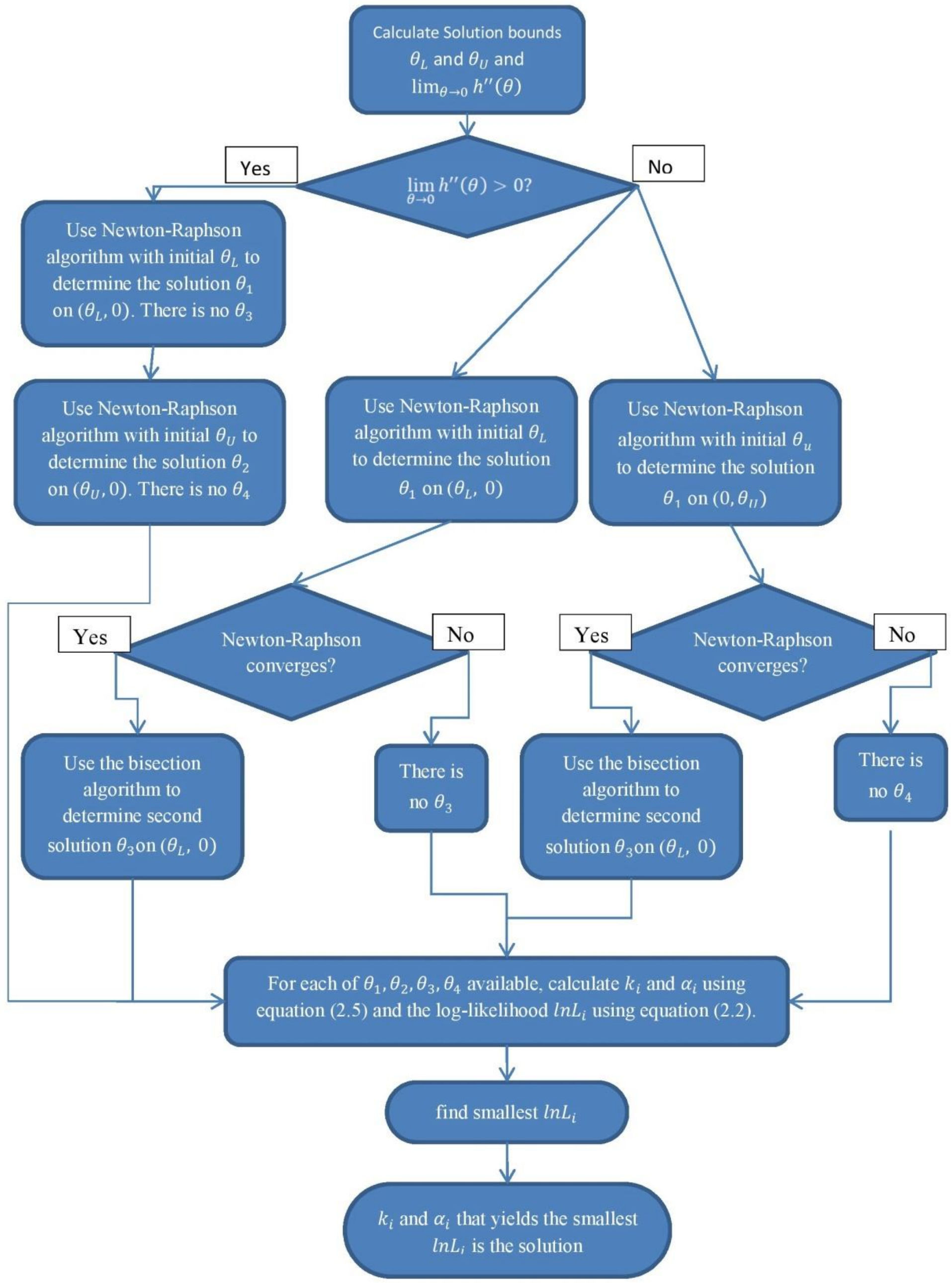

Figure 1. Solution process

With the existence of the censoring weight r, the program is generalized. In the case of no censorship, $r = 0$ and all the likelihood functions, $h(\theta)$, $h'(\theta)$, and $\lim_{\theta \to 0} h''(\theta)$, reduce to the ones proposed by Grimshaw (1993). The algorithm was written in R. Maximum likelihood estimation for right-censored data is created using the function `mle.gpd(time, censor)`, where `time` indicates the survival time vector and `censor` indicates the censoring vector (1 = observation, 0 = censored).

Program Validation by Simulations for Non-Censored Data

The performance of the algorithm is now tested when there is no censor ($r = n$). In this case, the algorithm is identical to the classical MLE proposed by Grimshaw (1993), which has been tested by others. The focus of this simulation is to compare the quality of MLE with the `gpd.fit` function in the R package gPdtest. This function has two separate methods that were proposed by Villaseñor-Alva and González-Estrada (2009), namely asymptotic maximum likelihood (AMLE) and combined.

When $k \le -0.5$, the GPD has infinite variance; when $k > 1$, maximum likelihood estimation has been proven to not exist (Castillo & Hadi, 1997). This simulation considers $-0.5 \le k \le 1$. More specifically, k will assume the values −0.4, −0.2,… 1. The results do not vary with respect to α (Hosking & Wallis, 1987). Thus, we set $\alpha = 1$. For each combination of k and α, we generate 10,000 random samples and calculate the average root mean square error (RMSE) for each method. The results are given in Table 1 (for k) and Table 2 (for α) below.

Table 1. Root mean square error of k for each estimator

	Method		
k	**MLE**	**AMLE**	**Combined**
−0.4	8.600E-04	4.644E-03	6.975E-03
−0.2	1.523E-03	5.373E-03	5.049E-03
0.0	1.280E-04	6.550E-03	1.830E-03
0.2	6.190E-04	7.253E-03	1.059E-03
0.4	3.280E-05	7.930E-03	2.750E-04
0.6	2.410E-04	8.528E-03	1.810E-04
0.8	1.760E-05	8.890E-03	3.680E-04
1.0	6.570E-04	9.230E-03	5.030E-04
Average	5.100E-04	7.300E-03	2.030E-03

Table 2. Root mean square error of α for each estimator

	Method		
k	**MLE**	**AMLE**	**Combined**
−0.4	6.2600E-04	1.9630E-03	4.2650E-03
−0.2	7.5000E-04	3.0100E-03	3.4470E-03
0.0	3.6100E-05	4.2210E-03	1.6390E-03
0.2	3.3500E-04	5.2180E-03	8.6100E-04
0.4	6.1700E-05	6.6740E-03	2.6900E-04
0.6	1.5100E-04	7.8880E-03	1.5800E-04
0.8	2.7100E-05	9.5720E-03	2.7100E-04
1.0	6.4200E-04	1.1188E-02	4.8700E-04
Average	3.2900E-04	6.2170E-03	1.4250E-03

The MLE performs better than the AMLE and combined methods in the gPdtest package. On average, the RMSE of k is 0.0005, 0.0073, and 0.002 for MLE, AMLE, and combined, respectively. The RMSE of α is 0.0003, 0.006, and 0.0014 for MLE, AMLE, and combined, respectively. MLE's RMSE is 93% lower than that of AMLE and 75% lower than that of combined for k, and 95% lower than that of AMLE and 79% lower than that of combined for α. This proves that Grimshaw's (1993) MLE algorithm has higher accuracy than the current methods existing in R.

Goodness-of-Fit Test for Censored Data

Testing the algorithm on censored data is challenging. Let T_i denote the time to failure and C_i denote the time to termination of the subject of study. The observed time will be $X_i = \min(T_i, C_i)$. Right censoring happens when termination time comes before failure time, i.e. $C_i < T_i$. In reality, a goodness-of-fit test for censored data is challenging because little is known about termination times. Apply the goodness-of-fit testing method proposed by Bagdonavičius and Nikulin (2011b), then test the null hypothesis that the simulated censored data follows the GPD with the set of parameters fitted by the MLE algorithm described earlier.

The chi-squared goodness-of-fit test for the hypothesis H_0 that the data X_i with status δ_i comes from the GPD with the estimated parameters $\hat{k}, \hat{\alpha}$ is performed as follows:

Divide the interval $[0, X_{(n)}]$ into $k > 2$ subintervals $I_j = (a_{j-1}, a_j]$. The a_j are determined to be

$$a_j = \Lambda^{-1}\left(\frac{E_j - \sum_{l=1}^{i-1}\Lambda\left(X_{(l)},\hat{k},\hat{\alpha}\right)}{n-i+1},\hat{k},\hat{\alpha}\right); \quad a_1 = 0; \quad a_k = X_{(n)},$$

where

$$\Lambda\left(\mathrm{x},\hat{k},\hat{\alpha}\right) = \frac{1}{k}\ln\left(\frac{\hat{\alpha}}{\hat{\alpha}-\hat{k}x}\right), \quad \Lambda^{-1}\left(\mathrm{x},\hat{k},\hat{\alpha}\right) = \frac{\hat{\alpha}}{\hat{k}}\left(1-\mathrm{e}^{-kx}\right)$$

$$E_k = \sum_x \Lambda\left(x,\hat{k},\hat{\alpha}\right), \quad E_j = \frac{j}{k}E_k$$

For each interval I_j, calculate U_j and e_j by

$$U_j = \sum_{i:X_i\in \mathrm{I}_j}\delta_i \quad \text{and} \quad e_j = \frac{E_k}{k}$$

In order to calculate the test statistic Y^2, calculate the matrices **Z**, **C**, **A**, **I**:

(1) $\mathbf{Z} = [Z_j]_{k\times 1}$, where $Z_j = \frac{1}{\sqrt{n}}\left(U_j - e_j\right)$.

(2) $\mathbf{C} = [C_{lj}]_{2\times k}$, where

$$C_{1j} = \frac{1}{n}\sum_{i:\,X_i\in \mathrm{I}_j}\delta_i\,\frac{X_i}{\hat{\alpha}-\hat{k}X_i}$$

and

$$C_{2j} = \frac{1}{n}\sum_{i:\,X_i\in \mathrm{I}_j}\delta_i\,\frac{-1}{\hat{\alpha}-\hat{k}X_i}.$$

(3) $\mathbf{A} = [A_{ij}]_{k\times k}$, where $A_{jj} = U_j / n$ and $A_{jj} = 0$ for $i \neq j$.

(4) $\mathbf{I} = [A_{lh}]_{2\times 2}$, where

$$I_{11} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\,\frac{X_i^2}{\left(\hat{\alpha}-\hat{k}X_i\right)^2}, \quad I_{22} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\,\frac{1}{\left(\hat{\alpha}-\hat{k}X_i\right)^2}, \quad \text{and}$$

$$I_{12} = I_{21} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{-X_i}{\left(\hat{\alpha} - \hat{k}X_i\right)^2}$$

The test statistic given by

$$Y^2 = \sum_{j=1}^{k} \frac{\left(U_j - e_j\right)^2}{U_j} + \mathbf{W}^{\mathrm{T}}\mathbf{G}^{-}\mathbf{W},$$

where $\mathbf{W} = \mathbf{CA}^{-1}\mathbf{Z}$, $\mathbf{G} = \mathbf{I} - \mathbf{CA}^{-1}\mathbf{C}^{\mathrm{T}}$. Y^2 follows the chi-square distribution with degrees of freedom given by $r = \mathrm{rank}((\mathbf{A} - \mathbf{C}^{\mathrm{T}}\mathbf{I}^{-1}\mathbf{C})^{-})$. The p-value is given by $\Pr\left(\chi_r^2 > Y^2\right)$. We reject the null hypothesis if the p-value is larger than a significance level.

Simulation Study to Validate the Goodness-of-Fit Test

In order to check the sensitivity and specificity of the proposed test, a simulation study was carried out to investigate the frequency of type I and type II errors at a 5% level of significance.

Sensitivity Test

For the sensitivity test, the data sets were simulated as follows: For each k in the set $\{-0.4, -0.2, \ldots, 1\}$ and for $\alpha = 1$, we generate 10,000 random samples to generate failure times T_i that follow the GPD with parameters k and α. Termination times C_i were generated by $C_i = \mathrm{Q}_3(T_i) + s\mathrm{U}$ where $\mathrm{Q}_3(T_i)$ and s are the third quartile and standard deviation of T_i, respectively, and U is the standard uniform distribution. This was done to target a censoring rate of about 15% of the data. The observed time is $X_i = \min(T_i, C_i)$ and status is $\delta_i = 1_{\{T_i < C_i\}}$.

The parameters $\hat{\alpha}$ and $\hat{k}$ were estimated using the proposed MLE algorithm discussed above. The goodness-of-fit test was used to test the hypothesis that failure times T_i follow the GPD with parameters $\hat{\alpha}$ and $\hat{k}$. There were 1000 samples for each k, and the number of false rejections were recorded and presented in Table 3. The results show that, at a 5% level of significance, the probability of a type I error is about 2% and, therefore, sensitivity is 98%.

Specificity Test

For the specificity test, the data sets were simulated as follows: simulate event time T_i from a gamma distribution with shape parameter k and scale parameter α. k was set to be 1.1, 1.2, 1.3, 1.4, or 1.5, and α was set at 2. Termination times C_i were generated by $C_i = Q_3(T_i) + sU$, where $Q_3(T_i)$ and s are the third quartile and standard deviation of T_i, respectively, and U in the standard uniform distribution. The observed time is $X_i = \min(T_i, C_i)$ and status is $\delta_i = 1_{\{T_i < C_i\}}$. With this design, the test is expected to fail to reject more frequently when k gets closer to 1 because the gamma distribution approaches the exponential distribution, which is also a special case of the GPD.

There were 1000 samples for each k and the number of correct rejections were recorded and presented in Table 4. This is similar to the expected outcome. Specificity is 98.7% when $k = 1.5$ and 7.2% when $k = 1.1$, which makes the GPD almost an exponential distribution.

Table 3. Count of rejections for each value of k ($\alpha = 1$)

k	**−0.4**	**−0.2**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
Count of rejections	15	20	25	19	25	29	18	19

Table 4. Count of rejections ($\alpha = 2$)

k	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**
Count of rejections	72	291	654	925	987

		Data simulated		
		From GPD	Not from GPD	
Hypothesis test that simulated data is from a GPD population	Not reject	7830	2071	Positive predicted value = 79.1%
	Reject	170	2929	Negative predicted value = 94.5%
		97.9%	58.6%	

Figure 2. Contingency table on simulation study (significance level 0.05)

The results of the simulation test can be summarized in the contingency table in Figure 2. This shows that for right-censored data, the overall sensitivity and specificity of our algorithm is 58.6% and 97.9%, respectively.

Discussion

The methods to fit censored data into the generalized Pareto distribution (GPD) were examined. The result was satisfying, with sensitivity when the probability of not rejecting the correct null hypothesis is 97% and higher. Specificity is 98.7% when the gamma distribution is used with shape parameter $k = 1.5$. As k approaches 1, specificity reduces significantly, being 7.2% when $k = 1.1$. This is acceptable because the gamma distribution becomes the exponential distribution when $k = 1$, which is also a special case of the GPD. These results indicate that our proposed methods are reliable.

References

Bagdonavičius, V., & Nikulin, M. (2011b). Chi-squared tests for general composite hypotheses from censored samples. *Comptes Rendus Mathematique, 349*(3-4), 219-223. doi: 10.1016/j.crma.2011.01.007

Castillo, E. (2012). *Extreme value theory in engineering*. San Diego, CA: Academic Press, Inc.

Castillo, E., & Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, *92*(440), 1609-1620. doi: 10.1080/01621459.1997.10473683

Cebrián, A. C., Denuit, M., & Lambert, P. (2003). Generalized Pareto fit to the Society of Actuaries' large claims database. *North American Actuarial Journal, 7*(3), 18-36. doi: 10.1080/10920277.2003.10596098

Embrechts, P., Klüppelberg, C., & Mikosch, T. (2012). *Modelling extremal events: For insurance and finance*. New York, NY: Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-33483-2

Gonzalez Estrada, E. G., & Villasenor Alva, J. A. (2012). gPdtest: Bootstrap goodness-of-fit test for the generalized Pareto distribution [R package, version 0.4]. Retrieved from https://cran.r-project.org/package=gPdtest

Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics, 35*(2), 185-191. doi: 10.1080/00401706.1993.10485040

Holmes, J. D., & Moriarty, W. W. (1999). Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics, 83*(1-1), 1-10. doi: 10.1016/S0167-6105(99)00056-2

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics, 29*(3), 339-349. doi: 10.1080/00401706.1987.10488243

Juárez, S. F., & Schucany, W. R. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes, 7*(3), 237-251. doi: 10.1007/s10687-005-6475-6

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). New York, NY: Springer-Verlag New York. doi: 10.1007/b97377

Pickands, J., III (1975). Statistical inference using extreme order statistics. *The Annals of Statistics, 3*(1), 119-131. Available from https://projecteuclid.org/euclid.aos/1176343003

Villaseñor-Alva, J. A., & González-Estrada, E. (2009). A bootstrap goodness of fit test for the generalized Pareto distribution. *Computational Statistics & Data Analysis, 53*(11), 3835-3841. doi: 10.1016/j.csda.2009.04.001

Zhang, J. (2007). Likelihood moment estimation for the generalized Pareto distribution. *Australian & New Zealand Journal of Statistics, 49*(1), 69-77. doi: 10.1111/j.1467-842X.2006.00464.x

Zhang, J. (2010). Improving on estimation for the generalized Pareto distribution. *Technometrics, 52*(3), 335-339. doi: 10.1198/TECH.2010.09206

Appendix A: Mathematical Proof for the Algorithm

Grimshaw (1993) presented five properties of the function h(θ) that were used to structure the algorithm for equation (6). In this study, the function h(θ) contains the censoring information r. Therefore, those five properties need to be revised in accordance with the new function.

Following Grishaw's (1993) approach, the following properties (A1) to (A5) of h(θ) are important to structure the algorithm.

$$\lim_{\theta \to 1/X_{(n)}^{-}} \mathrm{h}(\theta) \to -\infty \tag{A11}$$

According to Jensen's inequality, we can write the following:

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} \ln(1-\theta X_i) \le \ln(1-\theta X_i), \\ \frac{1}{r}\sum_{i=1}^{r} (1-\theta X_i)^{-1} \le \frac{1}{r}\sum_{i=1}^{r} \left(1-\theta X_{(1)}\right)^{-1} = \left(1-\theta X_{(1)}\right)^{-1}, \\ \frac{1}{n}\sum_{i=1}^{n} (1-\theta X_i)^{-1} \le \left(1-\theta X_{(1)}\right)^{-1} \end{cases}$$

which implies

$$\mathrm{h}(\theta) \le \left[1+\ln\left(1-\theta \bar{X}\right)\right]\left(1-\theta X_{(1)}\right)^{-1} - 1$$

and

$$\mathrm{h}(\theta) < 0 \quad \text{for} \quad \theta < \frac{2\left(X_{(1)} - \bar{X}\right)}{\left(X_{(1)}\right)^2} = \theta_L . \tag{A12}$$

Also,

$$\begin{aligned}\mathrm{h}'(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\frac{-X_i}{1-\theta X_i}\cdot\frac{1}{r}\sum_{i=1}^{r}\left(1-\theta X_i\right)^{-1} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\ln\left(1-\theta X_i\right)\cdot\frac{1}{r}\sum_{i=1}^{r}\left[X_i\left(1-\theta X_i\right)^{-2}\right] + \frac{1}{n}\sum_{i=1}^{n}\left[X_i\left(1-\theta X_i\right)^{-2}\right]\end{aligned}$$

(A13)

$$\lim_{\theta\to 0}\mathrm{h}'(\theta) = 0 \tag{A14}$$

$$\begin{aligned}\mathrm{h}''(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\frac{-X_i^2}{\left(1-\theta X_i\right)^2}\cdot\frac{1}{r}\sum_{i=1}^{r}\left(1-\theta X_i\right)^{-1} + \frac{1}{n}\sum_{i=1}^{n}\frac{-X_i^2}{1-\theta X_i}\cdot\frac{1}{r}\sum_{i=1}^{r}X_i\left(1-\theta X_i\right)^{-2} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{-X_i}{1-\theta X_i}\cdot\frac{1}{r}\sum_{i=1}^{r}X_i\left(1-\theta X_i\right)^{-2} + \frac{1}{n}\sum_{i=1}^{n}\ln\left(1-\theta X_i\right)\cdot\frac{1}{r}\sum_{i=1}^{r}2X_i^2\left(1-\theta X_i\right)^{-3} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}2X_i^2\left(1-\theta X_i\right)^{-3}\end{aligned}$$

Finally,

$$\lim_{\theta\to 0}\mathrm{h}''(\theta) = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - 2\bar{X}\bar{X}' \tag{A15}$$

where

$$\bar{X}' = \frac{1}{r}\sum_{i=1}^{r}X_i$$

The algorithm structure proposed by Grimshaw (1993) is maintained and modified according to the change of the 5 properties presented above.