Minimum Distance Estimation for the Generalized Pareto Distribution

The generalized Pareto distribution (GPD) is widely used for extreme values over a threshold, for instance, high-tide water levels to design dikes and return series for financial risks. To handle exceedances (difference between high value and threshold) GDP is one of the most well-recognized distributions. The cdf of GDP is defined as-

$$F_\theta(x) = \begin{cases} 1-(1-kx/\sigma)^{1/k}, & k \neq 0 \\ 1-\exp(-x/\sigma), & k = 0 \end{cases}$$

where k is the shape parameter and $\sigma > 0$ is the scale parameter. The range of x is $x > 0$ for $k \leq 0$ and $0 < x < \frac{\sigma}{k}$ for $k > 0$. If $X \sim \mathrm{GPD}(k, \sigma)$, then $(X - t) \sim \mathrm{GPD}(k, \sigma - kt)$ given $X > t$.

1 Review of Existing Methods:

Let $X \sim \mathrm{GPD}(k, \sigma)$. Most existing methods for parameter estimation either perform unsatisfactorily When the shape parameter k is larger than 0.5, or for a large sample size. The MLE of $\theta = k/\sigma$ is the solution of

$$1 - \frac{n}{\sum_i (1-\theta x_i)^{-1}} + \frac{\sum_i \log(1-\theta x_i)}{n} = 0; \qquad \theta < 1/x_{(n)}. \tag{1}$$

Then we calculate the MLE of k, σ as-

$$\hat{k}_{\mathrm{ML}} = -1/n \sum \log\left(1 - \hat{\theta}_{\mathrm{ML}} x_i\right), \quad \hat{\sigma}_{\mathrm{ML}} = \hat{k}_{\mathrm{ML}}/\hat{\theta}_{\mathrm{ML}}.$$

But here arises a lot of deficiencies for $k > 0.5$. To overcome these situations Methods of Moments(MOM) and Probability-Weighted Moments(PWM) can be used. PWM only makes use of the first moment. The estimators are

$$\hat{k}_{\mathrm{MOM}} = \left(\bar{x}^2/s^2 - 1\right)/2, \quad \hat{\sigma}_{\mathrm{MOM}} = \bar{x}\left(\bar{x}^2/s^2 + 1\right)/2,$$

$$\hat{k}_{\mathrm{PWM}} = \bar{x}/(\bar{x} - 2u) - 2, \quad \hat{\sigma}_{\mathrm{PWM}} = 2\bar{x}u/(\bar{x} - 2u),$$

where u can be $\sum_i \dfrac{n-i}{n-1} x_{(i)}$. The application of MOM and PWM is restricted to $k > -1$ and $k > -0.5$ respectively. The other estimator for θ based on Bayesian perspective is defined as

$$\hat{\theta}_{\mathrm{ZJ}} = \sum_{j=1}^{m} w_j \theta_j \tag{2}$$

where $\theta_j = \dfrac{n-1}{n+1} x_{(n)}^{-1} - \dfrac{\sigma^*}{k^*}\left[1 - \left(\dfrac{j-0.5}{m}\right)^{k^*}\right]$, $w_j = 1/\sum_{k=1}^{m} \exp\left[l(\theta_k) - l(\theta_j)\right]$. m, k^* and σ^* are some prefixed integer and $l(\theta)$ is log-likelihood function based on θ. This ZJ method works well when $k < 0.5$. Outside this range r, its performance deteriorates, and uncertain asymptotic properties.

Elemental Percentile Method (EPM) extracts two distinct order statistics by solving

$$F_\theta\left(x_{(i)}\right) = i/(n+1) \text{ and } F_\theta\left(x_{(j)}\right) = j/(n+1).$$

Then the estimator is calculated as the median of all $n(n-1)/2$ estimators. But this needs more improvement to estimate the shape parameter due to inconsistent asymptotic properties and huge computational cost.

2 Proposed Methods:

Two new estimators for the GPD parameters are proposed based on the M-estimation in the linear regression and minimum distance estimation.

2.1 M-Estimation in Linear Regression:

Consider a linear regression model $y_i = \mathbf{x}_i^T\boldsymbol{\beta} + e_i, 1 \leq i \leq n$, where $y_i \in \mathbb{R}$ is the response variable, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T \in \mathbb{R}^p$ are the explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T \in \mathbb{R}^p$ are the unknown parameters and e_i is the zero-mean error term. Let $r_i(\boldsymbol{\beta}) = y_i - \mathbf{x}_i^T\boldsymbol{\beta}$ be the residuals. Then the least-square (LS) estimator of β is defined as- $\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^n [r_i(\boldsymbol{\beta})]^2$.

LS estimator is sensitive to outliers and if the distribution of the error terms is not exactly normal, the estimator may be far from optimal. To deal with the difficulty M-estimator seems to be a popular robust method for its high efficiency. The M-estimator is-

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^n \rho\left[r_i(\boldsymbol{\beta})\right]. \tag{3}$$

Here $[r_i(\boldsymbol{\beta})]^2$ is replaced by a less rapidly increasing function ρ of the residuals. Tukey biweight function is one of the popular choices of ρ that is

$$\rho_c(u) = \begin{cases} \frac{u^2}{2}\left(1 - \frac{u^2}{c^2} + \frac{u^4}{3c^4}\right) & |u| \leq c, \\ \frac{c^2}{6}, & |u| > c, \end{cases}.$$

Set $c = 4.6851$ as a tuning parameter. The efficiency of 95%, consistency, and asymptotic normality hold for independent the M-estimator with these set-ups under Gaussian errors.

If e_i's are heteroscedastic and suppose the variance of e_i is proportional to w_i^2, the weighted M-estimator defined as-

$$\tilde{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^n \rho\left[r_i^*(\boldsymbol{\beta})\right] \tag{4}$$

where $r_i^*(\boldsymbol{\beta}) = r_i(\boldsymbol{\beta})/w_i$.

2.2 New Estimators for the GPD:

Consider an iid sample $x_1, \ldots, x_n$ from GPD(k_0, σ_0) and $\boldsymbol{\theta}_0 = (k_0, \sigma_0)$. Let $F_n(x)$ be the corresponding empirical distribution function. At the discontinuity points $x_{(1)}, ..., x_{(n)}$, we define $F_n\left(x_{(i)}\right) = (i - 0.5)/n \ \forall i$, as a "continuity correction" to $F_n(x(i))$. Similarly, the residuals are defined as $r_i(\theta) = F_n(x_i) - F_\theta(x_i), \quad i = 1, \ldots, n$, Now, we can estimate θ minimizing a distance measure $\rho(.)$ of $r_i(\boldsymbol{\theta})$.

Observe that $r_i(\theta_0)$'s are asymptotically normal and have different asymptotic variances, that is,

$$\sqrt{n} r_i\left(\theta_0\right)=\sqrt{n}\left[F_n\left(x_i\right)-F_{\theta_0}\left(x_i\right)\right] \xrightarrow{d} \mathrm{N}\left[0, F_{\theta_0}\left(x_i\right)\left(1-F_{\theta_0}\left(x_i\right)\right)\right].$$

Therefore we can construct a weighted M-estimator in which the appropriate weight can be $w_i(\boldsymbol{\theta}) = \sqrt{F_\theta\left(x_i\right)\left(1-F_\theta\left(x_i\right)\right)}$. For computational ease first, we can calculate $\tilde{\boldsymbol{\theta}}_n$ and apply them to calculate weights $w_i\left(\tilde{\boldsymbol{\theta}}_n\right)=\sqrt{F_{\boldsymbol{\theta}_\pi}\left(x_i\right)\left(1-F_{\boldsymbol{\theta}_n}\left(x_i\right)\right)}$ and thus $\tilde{\boldsymbol{\theta}}_n^*$ can be regarded as a two-stage estimator.

Still due to the complicated form of $\rho(.)$ it may not be easy to obtain $\tilde{\boldsymbol{\theta}}_n$ and $\tilde{\boldsymbol{\theta}}_n^*$. In such cases, the iterative reweighed least squares (IRLS) method can be used. Solve

$$\sum_{i=1}^n \phi\left[r_i(\boldsymbol{\theta})\right] r_i(\boldsymbol{\theta}) \frac{\partial r_i(\boldsymbol{\theta})}{\partial \theta_j}=0; \quad j=1,2 \tag{5}$$

where $\theta_1 = k$ and $\theta_2 = \sigma$ is equivalent to solve Eq(4). In each iteration of the IRLS algorithm, we solve -

$$\tilde{\boldsymbol{\theta}}_n^{(p)}=\arg \min_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^n \phi\left[r_i\left(\tilde{\boldsymbol{\theta}}_n^{(p-1)}\right)\right]\left[r_i(\boldsymbol{\theta})\right]^2$$

using Gauss-Newton algorithm where p is the iteration number.

2.3 Confidence Interval:

A resampling-based method such as the bootstrap-t is used here for interval estimation. First, estimate the parameters for a given dataset. Next, Generate B bootstrap samples each with size n from GPD$(\hat{k}, \hat{\sigma})$. For each bootstrap sample, estimate k denoted as $\hat{k}_b$ using the proposed method and compute $t_b = \left(\hat{k}_b-\hat{k}\right)/\widehat{\mathrm{se}}_{\hat{k}_b}$, where $\widehat{\mathrm{se}}_{\hat{k}_b}$ is the estimated standard error of $\hat{k}_b$. Finally, the equal-tailed $100(1-\alpha)\%$ confidence interval for k is $\left(\hat{k}-t_{1-\alpha/2}\widehat{\mathrm{se}}_{\hat{k}}, \hat{k}-t_{\alpha/2}\widehat{\mathrm{se}}_{\hat{k}}\right)$ where t_α is the α percentile point of t_b's.

3 Result:

A simulation study is conducted to compare the proposed methods with the ML method, the EPM, and the ZJ method setting $\sigma = 1$ (due to invariant results). From fig:2 of the original paper we can observe the weighted M-estimator performs slightly better between the two proposed methods and these methods have a much better performance for estimating the scale parameter σ than the ZJ method.

The coverage probabilities by two proposed estimators and EPM seem to work uniformly well improving with the increasing sample size. Moreover, for generated type II censoring data the proposed methods have a better performance than the EPM when $k > 0.5$.

Overall, the first estimator minimizes the distance between the empirical distribution and the family of GPDs, while the second one is a weighted version of the first one, where the weight is computed based on the first estimator. Both of the two estimators are consistent. The proposed methods are comparable to the existing methods for $k < 0.5$ while performing much better for $k > 0.5$. The effect of outliers can be greatly reduced and these are more robust than the existing methods.