MTH598A

Parameter Estimation of Generalized Pareto Distribution

Aindrila Garai (221258)
M.Sc. Statiistics, IIT Kanpur
Supervisor: Prof. Debasis Kundu

November, 2023

Objective

*Statistical modeling of the **extreme values** of certain natural phenomena has a major importance in various applications in our real life such as design of dikes, dams, financial risks etc.*

- The conventional approach analysising of extreme values is based on the Generalized Extreme Value Distribution (GEVD).
- GEDV is a limiting distribution for extreme values, comprising the Gumbel, Frechet, and Weibull distributions.
- It fits the data on maxima that leads to loss of information.

Peaks-Over-Threshold

In the POT method, several of the largest-order statistics are used instead of the maxima-only considering all values larger than a given threshold (exceedances over the threshold).

Generalized Pareto Distribution

GPD was explicitly introduced by Pickands in 1975.

GPD is one of the most well-recognized two-parameter families of distributions that deal with addressing the estimation problem of exceedances.

The CDF of GDP is defined as,

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

- If $k = 1$ GPD becomes $U(0, \sigma)$.
- If $k = 0$, the GPD reduces to $Exp(mean = \sigma)$.
- When $k < 0$, it becomes Pareto distribution.
- $E(X) = \sigma/(1+k)$ for $k > -1$.
- $Var(X) = \sigma^2/\left[(1+k)^2(1+2k)\right]$ for $k > -0.5$.
- If $X \sim \text{GPD}(k, \sigma)$, then $(X-t) \sim \text{GPD}(k, \sigma - kt)$ given $X > t$.

Generalized Pareto Distribution

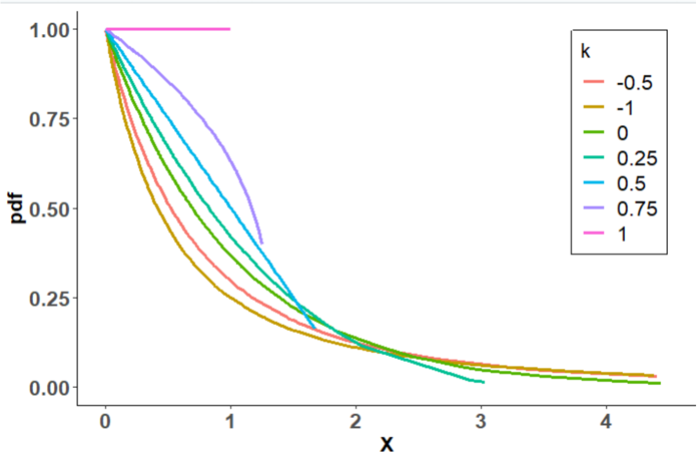

Figure: PDF of GPD for different values of k fixing $\sigma = 1$.

- A small value of the tail index $(-1/k)$ indicates that the events associated with large values occur with high probability.
- For the smaller value of k the tail of the distribution becomes heavy.

Review of Different Methods

- Method of Moment (MOM): Hosking and Wallis in 1987.
- Probability-Weighted Moment (PWM): Hosking and Wallis in 1987.
- Maximum Likelihood Estimator (MLE): Smith in 1984.
- Element Percentile Method (EPM): Castillo and Hadi in 1997.
- ZJ method: Zhang in 2010.
- M-Estimation: Piao Chen, Zhi-Sheng Ye in 2017.
- Weighted M-Estimation: Piao Chen, Zhi-Sheng Ye in 2017.
- Efficient Likelihood Method (ELM): Hideki Nagatsuka, N. Balakrishnan in 2020.

MOM & PWM

- MOM and PWM are based on the first two moments, the **sample mean ($\bar{X}$)** and the **sample variance (s^2)**.

Method of Moment

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

- MOM is restricted to k > -0.5 due to the non-existence of variance.

Probability-Weighted Moment

$$\hat{k}_{\text{PWM}} = \bar{X}/(\bar{X} - 2u) - 2, \quad \hat{\sigma}_{\text{PWM}} = 2\bar{X}u/(\bar{X} - 2u);$$

$$u = \frac{1}{n}\sum_i \frac{n-i}{n-1} X_{(i)}$$

- PWM exists only when $k > -1$ as expectation does not exist.

Maximum Likelihood Estimator:

- First, we will reparametrize (σ, k) of the GPD to (θ, k), where $\theta = k/\sigma$.
- The log-likelihood of (θ, k) is,

$$l^*(\theta, k) = n\log(\theta/k) + (1/k - 1)\sum_{i=1}^{n}\log\left(1 - \theta X_i\right).$$

- The profile likelihood of θ is defined as,

$$l(\theta) = 1 - \frac{n}{\sum_i (1 - \theta X_i)^{-1}} + \frac{\sum_i \log\left(1 - \theta X_i\right)}{n} = 0; \theta < 1/X_{(n)}.$$

- After maximizing,

$$\hat{k}_{\mathrm{ML}} = -1/n\sum_i \log\left(1 - \hat{\theta}_{\mathrm{ML}} X_i\right), \quad \hat{\sigma}_{\mathrm{ML}} = \hat{k}_{\mathrm{ML}}/\hat{\theta}_{\mathrm{ML}} \qquad (1)$$

- This technique is only acceptable when $k < 0.5$ due to consistency and efficiency and under this region.

Element Percentile Method:

EPM makes full use of the information of the order statistics by equating the cdf at the observed points to their corresponding percentile values.

- Compute C_i and C_j, where $C_i = \ln\left(1 - p_{i:n}\right) < 0$, $p_{(i)n} = \dfrac{i-\gamma}{n+\beta}$ for any two selected $x_{(i)} < x_{(j)}$.
- Calculate $d = C_j x_{(i)} - C_i x_{(j)}$.
  - If $d = 0$, then let $\hat{\delta}(i,j) = \pm\infty, \hat{k}(i,j) = 0$
- Compute $\delta_0 = x_{(i)} x_{(j)} \left(C_j - C_i\right)/d$.
- Solve $C_i \ln\left(1 - x_{j:n}/\delta\right) = C_j \ln\left(1 - x_{i:n}/\delta\right)$ to get $\hat{\delta}(i,j)$
  - If $\delta_0 > 0$, use the bisection method on $\left[x_{j:n}, \delta_0\right]$.
  - Otherwise, on the interval $[\delta_0, 0]$ apply bisection method.
- Using Eq (1) to compute $\hat{k}(i,j)$ and $\hat{\sigma}(i,j)$.
- Compute $\hat{k}(i,j)$ and $\hat{\sigma}(i,j)$ $\forall$ distinct pair $x_{(i)} < x_{(j)}$.
- Finally,

$$\hat{k}_{\text{EPM}} = \text{median}(\hat{k}(1,2), \hat{k}(1,3), \ldots, \hat{k}(n-1,n))$$

$$\hat{\sigma}_{\text{EPM}} = \text{median}(\hat{\sigma}(1,2), \hat{\sigma}(1,3), \ldots, \hat{\sigma}(n-1,n)).$$

ZJ Method

ZJ method was proposed to estimate the parameter $\theta = k/\sigma$ from a Bayesian perspective.

Our final estimate is,

$$\hat{\theta}_{\text{ZJ}} = \sum_{j=1}^{m} w_j \theta_j,$$

$$\theta_j = \frac{n-1}{n+1} x_{(n)}^{-1} - \frac{\sigma^*}{k^*}\left[1 - \left(\frac{j-0.5}{m}\right)^{k^*}\right]; \forall j = 1, \ldots, m,$$

$$w_j = 1/\sum_{k=1}^{m} \exp\left[l\left(\theta_k\right) - l\left(\theta_j\right)\right],$$

where, $l(\theta)$ is the log-likelihood of θ and m is a pre-fixed number. We get the estimated shape and scale parameter from θ_j.

- This method works well when $k < 0.5$.
- This ZJ estimator performs slightly poor $k \geq 0.5$.

M-Estimation

Then the least-square (LS) estimator of β is defined as,

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} [r_i(\boldsymbol{\beta})]^2 .$$

The M-estimator will be-

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \rho\,[r_i(\boldsymbol{\beta})]\,;$$

where $\mathrm{r}_i(\theta) = F_n(x_i) - F_\theta(x_i), F_n(x_i)$ is the empirical distribution.

- $[r_i(\boldsymbol{\beta})]^2$ is replaced by ρ of the residuals.
- Tukey biweight function is used as ρ,

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

c is a tuning parameter.

Weighted M-Estimation

The Weighted M-Estimator is,

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

where weights are $w_i(\boldsymbol{\theta}) = \sqrt{F_\theta (x_i) (1 - F_\theta (x_i))}$ and $r_i^*(\boldsymbol{\theta}) = r_i(\boldsymbol{\theta})/w_i$.

- Both the method M-estimation and Weighted M-Estimation are performs well for all over ranges of k.
- Here we use optim to minimise the equations.
- For M-estimation we use the ZJ estimator as starting value.
- For Weighted M-Estimation the initial value can be the M-Estimator.

Efficient Likelihood-based Method

- We work with $S_{(i)}^{(j)} = X_{(i)}/X_{(j)}, i \neq j, 1 \leq i \leq n, 1 \leq j \leq n$.
- For any fixed j,

  the joint density of $S_n^{(j)} = \left(S_{(1)}^{(j)}, \ldots, S_{(j-1)}^{(j)}, S_{(j+1)}^{(j)}, \ldots, S_{(n)}^{(j)}\right)$ is,

$$\phi\left(s_n^{(j)}; k\right) = \begin{cases} n! \int_{\chi_k} \frac{1}{|k|} \left(\frac{u}{k}\right)^{n-1} \prod_{i=1}^{n} (1 - u s_i)^{1/k-1} du, & k \neq 0, \\ \dfrac{n!(n-1)!}{\left(\sum_{i=1}^{n} s_i\right)^n}, & k = 0, \end{cases}$$

$$s_1 \leq \cdots \leq s_{j-1} \leq 1 \leq s_{j+1} \leq \cdots \leq s_n,$$

  where $\chi_k = \{u : -\infty < u < 0, \text{ if } k < 0, \text{ or, } 0 < u < 1/s_n, \text{ if } k > 0\}$.
- The likelihood function for k based on $S_n^{(j)}$ is $l\left(k; s_n^{(j)}\right) = \phi\left(s_n^{(j)}; k\right)$.
- Compute the MLE of k based on $S_n^{(j)}$ maximizing $l\left(k; \mathbf{s}_n^{(j)}\right)$.

Remark

- $S_n^{(j)}$ does not depend on σ.
- The MLE of k does not depend on j, for any j.
- It always has a unique solution concerning k.

A Simulation Study

A simulation study is conducted to compare the effectiveness of different estimation methods for the GPD shape parameters.

- We implemented all the techniques for $k = -4, -3.9, -3.8, \ldots, 3.9, 4$ of sample sizes $n = 20, 50$, and 100.

Model Performance

Method	Performs well	Moderate Performance	Poor Performance
MOM	-	$k > -0.5$	Does not exist
PWM	-	$k > -1$	Does not exist
MLE	$-0.5 < k < 0.5$	$-0.5 \geq k$	Does not exist
EPM	$-1.5 < k < 1.5$	$-2.5 < k < -1.5$ & $1.5 < k < 2.5$	$-2.5 \geq k$ & $k \geq 2.5$
ELM	$\forall k$	-	-
ZJ	$k < 0.5$	$k \geq 0.5$	-
M-est	$\forall k$	-	-
WM-est	$\forall k$	-	-

Table: Performance of different methods for the shape parameter.

Comparison of Method

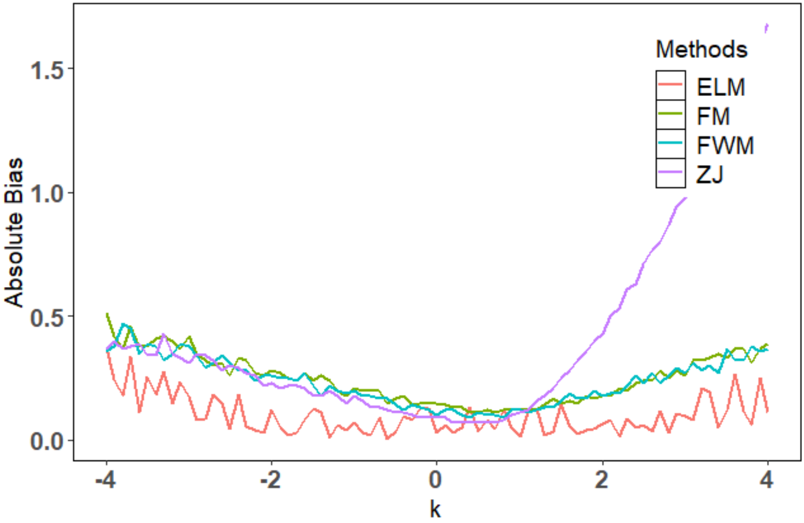

Figure: Comparision of Absolute Bias for $n = 100$.

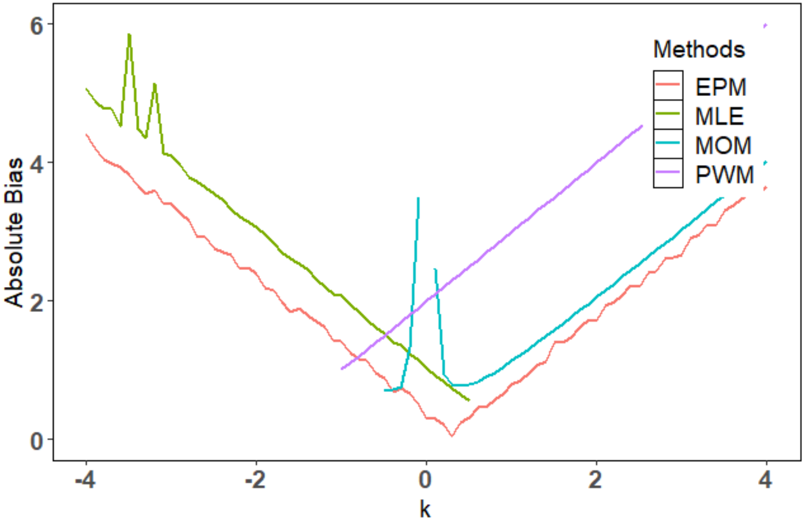

Figure: Comparision of Absolute Bias for $n = 100$.

Comparison of Method

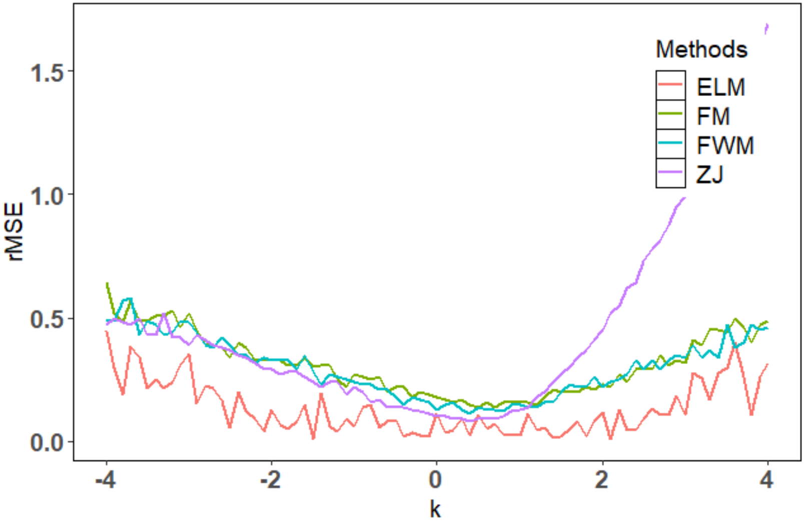

Figure: Comparision of rMSE for $n = 100$.

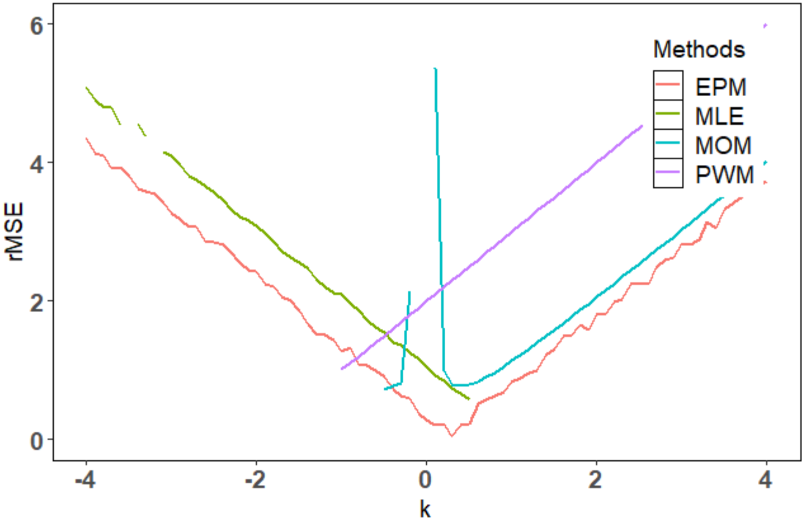

Figure: Comparision of rMSE for $n = 100$.

Conclusion

THANK YOU!