Change Point Analysis of Extreme Values

Goedele Dierckx
Economische Hogeschool Sint Aloysius, Brussels, Belgium
e-mail: goedele.dierckx@hubbrussel.be

Jef L. Teugels
Katholieke Universiteit Leuven, Belgium &
EURANDOM, Technical University Eindhoven, the Netherlands
e-mail: jef.teugels@wis.kuleuven.be
Overview

1. Introduction

2. Test statistic
   (a) Construction
   (b) Extreme value situation
   (c) Asymptotics
   (d) Practical procedure

3. Examples
   (a) Simulation
   (b) Malaysian Stock Index.
       • Classical Approach
       • Improved Approach
   (c) Nile Data
   (d) Swiss-Re Catastrophic Data

4. Conclusions

5. References
1. INTRODUCTION


Changes in
- distribution?
- in parameters of a distribution?
  - central behavior?
  - tail behavior?
2. TEST STATISTIC

2.a. Construction of Test Statistic

Start with a sample \( X_1, \ldots, X_{m^*}, X_{m^*+1}, \ldots X_n \), from a density function \( f(x; \theta_i, \eta) \).

Csörgő and Horváth (1997) test whether \( \theta_i \) changes at some point \( m^* \)

\[
H_0 : \theta_1 = \theta_2 = \ldots = \theta_n \quad \text{versus} \\
H_1 : \theta_1 = \ldots = \theta_{m^*} \neq \theta_{m^*+1} = \ldots = \theta_n \quad \text{for some } m^*.
\]

using the test statistic

\[
Z_n = \sqrt{\max_{1 \leq m < n} (-2 \log \Lambda_m)},
\]

where

\[
\Lambda_m = \frac{\sup_{\theta, \eta} \prod_{i=1}^{n} f(X_i; \theta, \eta)}{\sup_{\theta, \tau, \eta} \prod_{i=1}^{m} f(X_i; \theta, \eta) \prod_{i=m+1}^{n} f(X_i; \tau, \eta)}.
\]
2. TEST STATISTIC

Example

For the exponential distribution where $X_i$ has mean $\theta_i$

$$-2 \log \Lambda_m = 2 \left[ -m \log \frac{1}{m} \sum_{i=1}^{m} X_i - (n - m) \log \frac{1}{n - m} \sum_{i=m+1}^{n} X_i + n \log \frac{1}{n} \sum_{i=1}^{n} X_i \right]$$

For large $n$, $m$ and $n - m$ one can expect 'normal' behaviour expressed in terms of Brownian motions.
2.b. Extreme Value Situation

Assume that $X_{n,n}$ is the maximum in a sample of independent random variables with a common distribution. Maximum domain of attraction condition

$$\lim_{n \to \infty} P \left( \frac{X_{n,n} - b_n}{a_n} \leq x \right) = G_\gamma(x).$$

Under very weak conditions we get the approximation

$$P(X_{n,n} \leq y) \approx G_\gamma(b_n + a_n x)$$

where $\gamma$ is a real-valued extreme value index and

$$G_\gamma(x) = \exp \left\{ -\left[ 1 + \gamma x \right]^{1/\gamma} \right\}$$

an extremal law.

When $\gamma > 0$ we end up with heavy right-tailed distributions, the Pareto-Fréchet Case.
2. TEST STATISTIC

We concentrate on changes of parameters that describe the tail of distributions appearing in extreme value analysis.

• $X$ has a **Pareto-type distribution** with parameter $\theta = \gamma$, when the relative excesses of $X$ over a high threshold $u$, given that $X$ exceeds $u$ satisfy the condition

$$P \left( \frac{X}{u} > x \mid X > u \right) \to x^{-\frac{1}{\gamma}} \cdot u \to \infty,$$

• More generally $X$ follows a **Generalized Pareto distribution (GPD)** with parameter $\theta = (\gamma, \sigma)$ if the behavior of the absolute excesses over a high threshold $u$ satisfies the condition

$$P \left( X - u > x \mid X > u \right) \to \left( 1 + \frac{\gamma x}{\sigma} \right)^{-\frac{1}{\gamma}}, \quad u \to \infty.$$
2. TEST STATISTIC

For large values, \( \log \) of Pareto-type with extreme value index \( \gamma_i \) is close to be exponential with mean \( \gamma_i \).

- The most classical approach for the *estimation* of the extreme value index \( \gamma > 0 \) is to use the **Hill estimator**:

\[
H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}.
\]

Hence, only a segment of the available data is used.

- The determination of the quantity \( k \) is important. Alternatively, we look at extremes above a *threshold* \( u = X_{n-k,n} \). The Hill estimator has
  - small bias but large variance for small \( k \)
  - large bias but small variance for large \( k \).

As a compromise we select \( k \) such that the **empirical mean squared error** is minimal.
2. TEST STATISTIC

1. Pareto-type density
Suppose \( X_1, \ldots, X_m, X_{m+1}, \ldots X_n \) are independent and Pareto-type distributed. We denote the extreme value index for \( X_i \) by \( \gamma_i \). In order to determine whether the index \( \gamma \) changes or not, we perform the following test

\[
H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_n = \gamma \quad \text{versus} \quad H_1 : \gamma_1 = \gamma_{m^*} \neq \gamma_{m^*+1} = \gamma_n \text{ for some } m^*
\]

Hence \( Z_n = \sqrt{\max_{1 \leq m < n} (-2 \log \Lambda_m)} \)

where in turn

\[
\log \Lambda_m = [k_1 \log H_{k_1,m} + (k - k_1) \log H_{k-k_1,n-m} - k \log H_{k,n}]
\]

\[+ \left[\frac{1}{H_{k,n}} (k_1 H_{k_1,m} + (k - k_1) H_{k-k_1,n-m} - k H_{k,n})\right].\]
2. TEST STATISTIC

2. GPD. Suppose now that $X_i$ is GPD with parameters $\theta_i = (\gamma_i, \sigma_i)$. To perform the test

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_n \quad \text{versus} \quad H_1 : \theta_1 = \ldots = \theta_{m^*} \neq \theta_{m^*+1} = \ldots = \theta_n \quad \text{for some } m^*$$

we use as test statistic $Z_n = \sqrt{\max_{1 \leq m < n} (-2 \log \Lambda_m)}$, where

$$-2 \log \Lambda_m = 2 \left[ L_{k_1}(\hat{\theta}_{k_1}) + L_{k_1}^+(\hat{\theta}_{k_1}^+) - L_k(\hat{\theta}_k) \right]$$

$$L_m(\hat{\theta}_m) = -m \log \hat{\sigma}_m - \left( \frac{1}{\hat{\gamma}_m} + 1 \right) \sum_{i=1}^{m} \log \left( 1 + \hat{\gamma}_m \frac{x}{\hat{\sigma}_m} \right)$$

$$L_m^+(\hat{\theta}_m^+) = -(n - m) \log \hat{\sigma}_m^+ - \left( \frac{1}{\hat{\gamma}_m^+} + 1 \right) \sum_{i=m+1}^{n} \log \left( 1 + \hat{\gamma}_m^+ \frac{x}{\hat{\sigma}_m^+} \right)$$

and likelihood estimators $(\hat{\gamma}_m, \hat{\sigma}_m)$ resp. $(\hat{\gamma}_m^+, \hat{\sigma}_m^+) \quad \text{based on} \quad X_1, X_2, \ldots, X_m \quad \text{and} \quad X_{m+1}, \ldots X_n \quad \text{are obtained by numerical procedures.}$
2. TEST STATISTIC

2.c. Asymptotics

Using the procedure suggested by Csörgő and Horváth we have

Theorem Suppose $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are independent and identically distributed. We set the threshold at $u = X_{n-k,n}$. Define

$$Z_n = \sqrt{\max_{c_n \leq m < n-d_n} (-2 \log \Lambda_m)},$$

with $-2 \log \Lambda_m$ as before. Let $n, k \to \infty$ such that $k/n \to 0$. Let further $c_n$ and $d_n$ be intermediate sequences for which $c_n/n \to 0$ and $d_n/n \to 0$. Then, under $H_0$ of our test,

$$Z_n \xrightarrow{d} \begin{cases} \sqrt{\sup_{0 \leq t < 1} \frac{B^2(t)}{t(1-t)}} & \text{if Pareto-type ,} \\ \sqrt{\sup_{0 \leq t < 1} \frac{B_2^2(t)}{t(1-t)}} & \text{if GPD .} \end{cases}$$

$B(t)$ is a Brownian bridge, $B_2(t)$ is a sum of two independent Brownian bridges.
2. TEST STATISTIC

2.d. Practical Procedure

Consecutive steps

1. Check on Pareto-type behavior of the data by $Q - Q$-plots.

2. Select a threshold $u$ or the value of $k = k_{opt,n}$ that minimizes the asymptotic mean square error of the Hill estimator. We choose the optimal threshold $u = X_{n-k_{opt,n}}$.

3. (a) Define $c_n$ as the smallest number such that at least $k_{min} = (\log k_{opt,n})^{3/2}$ of the data points $X_1, \ldots, X_{c_n}$ are larger than $u$.

   (b) Define $d_n$ as the smallest number such that at least $k_{min}$ of the data points $X_{n-d_n+1}, \ldots, X_n$ are larger than $u$.

4. Repeat the next step for all $m$ from $c_n$ up to $n - d_n$.

   (a) Split the data up in two groups $X_1, X_2, \ldots, X_m$ and $X_{m+1}, \ldots, X_n$.

   (b) Calculate $-2 \log \Lambda_m$.

5. Calculate $Z_n = \sqrt{\max_{c_n \leq m < n-d_n} (-2 \log \Lambda_m)}$ and compare $Z_n$ with the critical values for sample size $k$. 


3. EXAMPLES

3.a. Simulation

We simulate 1000 data sets of size $n$ (with $n = 100$, $n = 500$) from the **Burr distribution** $Burr(\beta, \tau, \lambda)$ with parameters as given by

$$P(X > x) = \left( \frac{\beta}{\beta + x^\tau} \right)^\lambda,$$

an example of a GPD with $\gamma = (\lambda \tau)^{-1}$. The rejection probabilities are given below.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$m^*$</th>
<th>$H_0$ true</th>
<th>$H_0$ false</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\gamma = 1$</td>
<td>$\gamma_1 = 1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_2 = 2$</td>
<td>$\gamma_2 = 1$</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.096</td>
<td>.191</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>.075</td>
<td>.517</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>.029</td>
<td>.181</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>.044</td>
<td>.378</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>.019</td>
<td>.894</td>
</tr>
</tbody>
</table>
3. EXAMPLES

The corresponding median of $\hat{m}$ is given in the table below.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$m^*$</th>
<th>$H_0$ false</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\gamma_1 = 1$</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>252</td>
</tr>
</tbody>
</table>
3. EXAMPLES

Figure shows Boxplot of \( \hat{m} \) for the Burr cases for \( n = 500 \) and \( m^* = 100 \).
3. EXAMPLES

3.b. Malaysian Stock Index: Classical approach

Figure below indicates that the data are Pareto-type distributed. If we accept that July 1997 was a change point, then the data before that date give an extreme value index $\gamma_1$ between 0.1 and 0.2 while those after that date give $\gamma_2$ around 0.5. The mean squared error of the Hill estimator based on the whole data set attains a local minimum for the threshold $u$ given by $X_{987-224,987} = 0.0099$ so that $k = k_{\text{opt}} = 224$. 
3. EXAMPLES

1. Pareto-type distribution
First $\sqrt{-2 \log \Lambda_m}, 1 \leq m \leq n - 1$ is plotted below.

Graph of $(m, \sqrt{-2 \log \Lambda_m})$ with critical value indicated with a horizontal line.
We see that $Z_n = \sqrt{\max(-2 \log \Lambda_m)} = 5.8$ falls above the critical value 3.14 and we reject $H_0$. The maximum is attained at $m = 635$, which corresponds to 1/08/1997, shortly after the beginning of the Asian crisis.
3. EXAMPLES

2. GPD

Now $\sqrt{-2\log \Lambda_m}, 1 \leq m \leq n - 1$ is plotted below.

\[
\sqrt{-2\log \Lambda_m}, 1 \leq m \leq n - 1
\]

Since $Z_n = \sqrt{\max(-2\log \Lambda_m)} = 5.93$ is above the critical value 3.18 we again reject $H_0$. Also the instant of change $\hat{m} = 636$ is again very close to the value before.
3. EXAMPLES

3.b. Malaysian Stock Index: Improved approach

In the above analysis, we assumed that the data were independent. But market data are hardly ever independent. However, it is known that the Hill estimator withstands many forms of dependence. Alternatively, one can proceed as follows. The time series and an estimate of the extremal index are given below.

A declustering scheme cuts the data into clusters that can safely be taken as independent. Apply the previous procedure to the 76 cluster maxima.
There is a local maximum for cluster maximum 48 which corresponds to \( m = 631 \) on (28/07/1997). However this local maximum is not larger than the critical. The actual maximum \( Z_n \) is attained for cluster maximum 66 which corresponds to \( m = 854 \) (22/6/98). We cannot reject the hypothesis.
Now $\sqrt{-2 \log \Lambda_m}, 1 \leq m \leq n - 1$ is plotted in the figure. The maximum $Z_n$ is attained for cluster maximum 48 which corresponds to $m = 631(28/07/1997)$. The critical value 2.95 for the test is indicated with a horizontal line. On the basis of this test, we reject the hypothesis of no change.
3. EXAMPLES

3.c. Nile Data

Annual flow volume of the Nile River at Aswan from 1871 to 1970.

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1120</td>
<td>1160</td>
<td>963</td>
<td>1210</td>
<td>1160</td>
<td>1160</td>
<td>813</td>
<td>1230</td>
<td>1370</td>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>995</td>
<td>935</td>
<td>1110</td>
<td>994</td>
<td>1020</td>
<td>960</td>
<td>1180</td>
<td>799</td>
<td>958</td>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>1210</td>
<td>1150</td>
<td>1250</td>
<td>1260</td>
<td>1220</td>
<td>1030</td>
<td>1100</td>
<td>774</td>
<td>840</td>
<td></td>
</tr>
<tr>
<td>874</td>
<td>694</td>
<td>940</td>
<td>833</td>
<td>701</td>
<td>916</td>
<td>692</td>
<td>1020</td>
<td>1050</td>
<td>969</td>
<td></td>
</tr>
<tr>
<td>831</td>
<td>726</td>
<td>456</td>
<td>824</td>
<td>702</td>
<td>1120</td>
<td>1100</td>
<td>832</td>
<td>764</td>
<td>821</td>
<td></td>
</tr>
<tr>
<td>768</td>
<td>845</td>
<td>864</td>
<td>862</td>
<td>698</td>
<td>845</td>
<td>744</td>
<td>796</td>
<td>1040</td>
<td>759</td>
<td></td>
</tr>
<tr>
<td>781</td>
<td>865</td>
<td>845</td>
<td>944</td>
<td>984</td>
<td>897</td>
<td>822</td>
<td>1010</td>
<td>771</td>
<td>676</td>
<td></td>
</tr>
<tr>
<td>649</td>
<td>846</td>
<td>812</td>
<td>742</td>
<td>801</td>
<td>1040</td>
<td>860</td>
<td>874</td>
<td>848</td>
<td>890</td>
<td></td>
</tr>
<tr>
<td>744</td>
<td>749</td>
<td>838</td>
<td>1050</td>
<td>918</td>
<td>986</td>
<td>797</td>
<td>923</td>
<td>975</td>
<td>815</td>
<td></td>
</tr>
<tr>
<td>1020</td>
<td>906</td>
<td>901</td>
<td>1170</td>
<td>912</td>
<td>746</td>
<td>919</td>
<td>718</td>
<td>714</td>
<td>740</td>
<td></td>
</tr>
</tbody>
</table>
3. EXAMPLES

Prior studies indicate

- 1877 (measurement 813) candidate for additive outlier,
- 1913 (measurements 456) candidate for additive outlier,
- 1899 (measurement 774) indicates start of construction of Aswan dam.
3. EXAMPLES

Group 1: first 28 points – Group 2: remaining 71 points with Pareto QQ plots for both groups. Optimal values $k = 17$, resp. $k = 13$ lead to the estimators 0.07 and 0.13.
3. EXAMPLES

The change point detection based on the Pareto and the GPD model are given in the figure, both leading to a significant change point at $\hat{m} = 28$ at the beginning of construction of the Aswan dam.
### 3. EXAMPLES

#### 3.d. Swiss-Re Catastrophic Data

<table>
<thead>
<tr>
<th>PLACE</th>
<th>DATE</th>
<th>VICTIMS</th>
<th>PLACE</th>
<th>DATE</th>
<th>VICTIMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>0.87</td>
<td>300000</td>
<td>Indonesia</td>
<td>34.98</td>
<td>280000</td>
</tr>
<tr>
<td>China</td>
<td>6.57</td>
<td>255000</td>
<td>Bangladesh</td>
<td>21.33</td>
<td>138000</td>
</tr>
<tr>
<td>Peru</td>
<td>0.51</td>
<td>66000</td>
<td>Gilan (Iran)</td>
<td>20.47</td>
<td>50000</td>
</tr>
<tr>
<td>Bam (Iran)</td>
<td>33.98</td>
<td>26271</td>
<td>Tabas (Iran)</td>
<td>8.71</td>
<td>25000</td>
</tr>
<tr>
<td>Armenia</td>
<td>18.93</td>
<td>25000</td>
<td>Colombia</td>
<td>15.87</td>
<td>23000</td>
</tr>
<tr>
<td>Guatemala</td>
<td>6.10</td>
<td>22084</td>
<td>Izmit (Turkey)</td>
<td>29.63</td>
<td>19118</td>
</tr>
<tr>
<td>Gujarat (India)</td>
<td>31.07</td>
<td>15000</td>
<td>India</td>
<td>8.67</td>
<td>15000</td>
</tr>
<tr>
<td>India</td>
<td>29.83</td>
<td>15000</td>
<td>India</td>
<td>9.61</td>
<td>15000</td>
</tr>
<tr>
<td>India</td>
<td>1.83</td>
<td>10800</td>
<td>Venezuela</td>
<td>29.95</td>
<td>10000</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>7.89</td>
<td>10000</td>
<td>Mexico</td>
<td>15.72</td>
<td>9500</td>
</tr>
<tr>
<td>India</td>
<td>23.75</td>
<td>9500</td>
<td>Honduras</td>
<td>28.81</td>
<td>9000</td>
</tr>
<tr>
<td>Kobe (Japan)</td>
<td>25.05</td>
<td>6425</td>
<td>Philippines</td>
<td>21.85</td>
<td>6304</td>
</tr>
<tr>
<td>Pakistan</td>
<td>4.99</td>
<td>5300</td>
<td>Brazil</td>
<td>31.87</td>
<td>5112</td>
</tr>
</tbody>
</table>
3. EXAMPLES

The Pareto QQ plots with the corresponding Hill estimators. The mean squared error is minimal at $k = 39$ for Pareto and $k = 22$ for GPD, both leading to $\hat{\gamma} = 1.3$. 
3. EXAMPLES

The likelihood expression $\sqrt{-2 \log \Lambda_m}$ based on the Pareto model and the GPD model as a function of $m$ where $m$ is indicating where the group is split up in two.

Pictures for Pareto model with critical value 1.6 and GPD model with critical value 1.4.
4. CONCLUSIONS

• What has been shown are just first attempts
• Assumption on positive $\gamma$
• Rounded figures make accurate conclusions harder
• There is a need for sufficiently large data sets
• Need for studies under specific dependence structures
• Multivariate extensions should be possible

5. REFERENCES