Clustering of extreme values: a contribution on inference

Marta Ferreira (CMAT-University of Minho)

Abbreviated abstract: The propensity of data for the occurrence of clusters of extreme values is measured through the extremal index. This is an important parameter to infer the risk associated to extreme phenomena, whose duration in time may cause a great impact on our days life. The estimation is very sensitive and affected by the choice of a high threshold and the definition of the clusters. We consider an estimator that depends on the indication of the cluster size, not requiring the definition of the high level to be considered. We analyze the application of an heuristic method for the cluster size selection through simulation. We illustrate the procedure with real data.

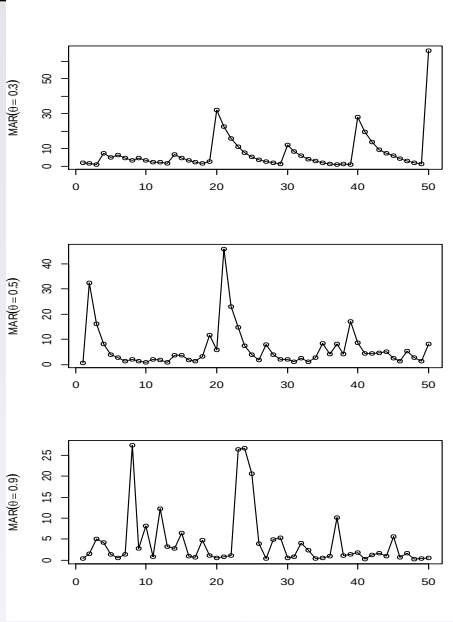
Related publications:

– P. Northrop, Extremes18, 2-603 (2015)
– D.Prata Gomes et al, Journal of Applied Statistics 47 (13-15), 2846-2861 (2020)

Extremal Index θ – clustering degree at extreme values

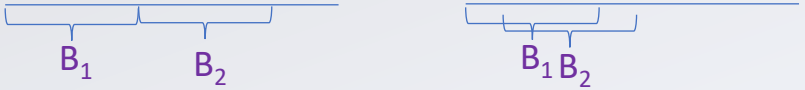
- $\theta \in [0,1]$; IID case $\Rightarrow \theta=1$; smaller $\theta \Rightarrow$ larger dependence and clustering

Under mixing conditions, θ is the reciprocal of the mean cluster size in the point process of exceedance times over a high threshold (Hsing et al, 1988)



Blocks estimator: cut the n obs. into blocks of length b, count in each block the n.º of exceedances over a high threshold, estimate θ^{-1} by the average of exceedances per block among blocks with at least one exceedance (these are the clusters)

- **disjoint** blocks & **sliding** blocks



– specification of 2 parameters: threshold and block length (cluster identif)
– θ estimation highly influenced by both choices -> Alternative approaches:

Threshold methods (threshold choice) & Maxima methods (block size)

Maxima method of P. Northrop (2015): $V = -b\ log\ F(Y) \sim Exp(\theta)$, $E(V) = \theta^{-1}$, where Y is the maximum of any block of length b (i.e., b consecutive obs. from $df\ F$) ⇝ MLE estim

– How to choose the block size? ↣ heuristic path stability algorithm in D. Prata Gomes et al (2020)

Disjoint and Sliding estimators | Algorithm to find block size

Disjoint and sliding estimators: $\tilde{\theta}_{dj}$ and $\tilde{\theta}_{sl}$ depending on block size b (Northrop, 2015) | Simulation study: 1000 replicas of size 1000

Algorithm: (the same for $\tilde{\theta}_{sl}$)

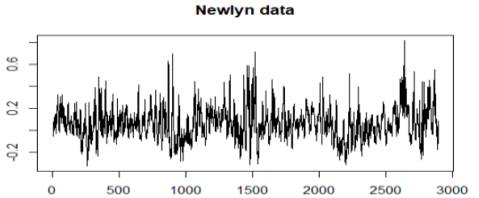
1. Compute estimates $\tilde{\theta}_{dj}$(b) for [n/100]≤b≤[n/4]
2. Round the estimates to 0 decimal places
3. Consider the sets of values associated to equal consecutive values and take b_{min} and b_{max} the minimum and the maximum of largest range set
4. Consider all estimates $\tilde{\theta}_{dj}$(b) for b_{min} ≤b≤ b_{max} with 2 decimal places, obtain the mode of $\tilde{\theta}_{dj}$(b) and take $R_{\tilde{\theta}_{dj}}$(b) the set of b-values associated to the mode
5. Get $b_0 = \max(R_{\tilde{\theta}_{dj}}(\text{b}))$ and estimate $\tilde{\theta}_{dj}(b_0)$

		sliding			disjoint		
model	θ	$\overline{b_0}$	abias	rmse	$\overline{b_0}$	abias	rmse
MCBEV [1]	0.328	194(51)	0.019	0.167	193(51)	0.042	0.222
AR [1]	1	203(49)	0.178	0.259	224(52)	0.176	0.263
MAR [1]	0.5	194(48)	0.007	0.196	199(44)	0.040	0.248
MMFrec [1]	0.5	192(49)	0.000	0.167	199(42)	0.026	0.215
MMUnif [2]	1/3	195(52)	0.004	0.085	185(51)	0.010	0.108
ARUnif [3]	0.75	185(71)	0.079	0.206	188(67)	0.092	0.205
ARCauchy [3]	0.64	194(53)	0.020	0.207	177(66)	0.053	0.233

1(Anc.Navarrete et al, 2000); 2(Suveges, 2007); 3(Chernick et al, 1991)

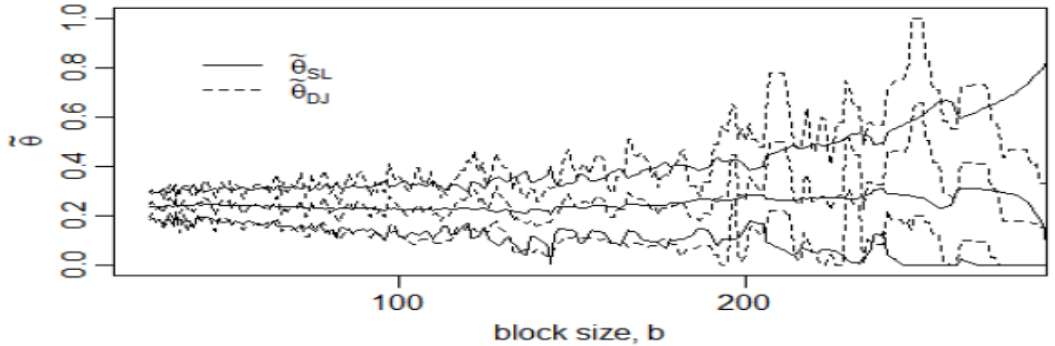
$\overline{b_0}$ block size sample mean and standard-deviation in ()

Application to real data and Conclusions



Series of 2894 maximum hourly (15h) surge heights - coast at Newlyn, Cornwall, UK, 1971–1976.

- Sliding estimator sample path (full line) less oscilating
- The heuristic method leads to the choice of larger block sizes, where confidence bands are larger too but the point estimates are very closed to the ones obtained in P. Northrop, 2015



$\tilde{\theta}_{sl}$	0.224	]0.103,0.345[	b=127
$\tilde{\theta}_{dj}$	0.278	]0.101,0.430[	b=195
$\tilde{\theta}_{sl}$ [1]	0.238	]0.188,0.296[	b=20
$\tilde{\theta}_{dj}$ [1]	0.241	]0.204,0.283[	b=20

[1] P. Northrop, 2015

- ✓ Sliding blocks: lower biases and rmse's (usually performs better than disjoint blocks)
- ✓ Lower biases both associated to maxima models (MCBEV, MAR, MMFrec, MMunif)
- ✓ Large rmse's -> some variability in the method - complement with the estimates path over a wide range of b values (candidate: smallest b above which these estimates appear to be constant with respect to b, Northrop 2015)
- ✓ Also large sd in the algorithm estimates of b – reinforces the method variability
- ✓ Worst performance in θ=1 (typical; plot sample path …)
- ✓ Future work: improve the heuristic method towards robustness