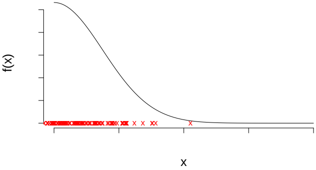
Where does the tail begin? Threshold selection for extremes

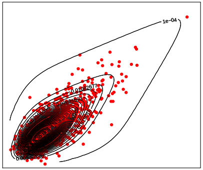
Jenny Wadsworth
Lancaster University

12th February 2016, University of Bristol

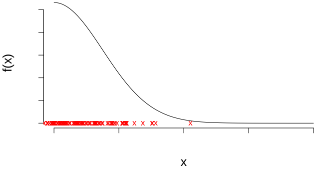
Extreme events

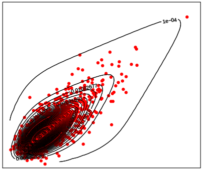
- An extreme event is something which occurs rarely and thus lies in the tail of the distribution (focus here on upper tail)





- An extreme event is something which occurs rarely and thus lies in the tail of the distribution (focus here on upper tail)





Goal of extreme value theory

Estimate probabilities of extreme events by estimating the tails of probability distributions

- Use existing extreme data to fit an asymptotically justified model

Univariate extremes

Distributions of univariate extremes

Let

- $X_i \sim F$
- $u_n \in \mathbb{R}$ s.t. $F(u_n) \to 1$ as $n \to \infty$

If there exists $\sigma_n > 0$ s.t.

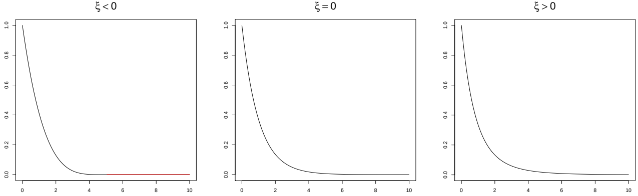
$$\mathrm{P}\left(\frac{X_i - u_n}{\sigma_n} \leq x \,\middle|\, X_i > u_n \right) \to H(x)$$

for non-degenerate H then

$$H(x) = 1 - \left[1 + \xi \left(\frac{x}{\sigma}\right)\right]_+^{-1/\xi}, \quad \sigma > 0,\ \xi \in \mathbb{R}$$

is the generalized Pareto or GP distribution.

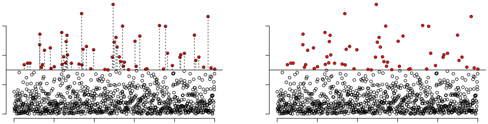
Tail behaviour determined by sign of ξ



$$\xi \begin{cases} < 0 & \text{light tail} \\ \to 0 & \text{exponential tail} \\ > 0 & \text{heavy tail} \end{cases}$$

Alternative characterization

GP distribution gives a model for sizes of excesses conditional upon being an excess.



More "complete" characterization of tail from Pickands (1971) point process representation. Assuming "weak long range dependence",

$$\sum_{i=1}^{n} \delta_{\left(\frac{i}{n+1}, \frac{X_i - u_n}{\sigma_n}\right)} \to \sum_{i \geq 1} \delta_{(T_i, Z_i)},$$

a non-homogeneous Poisson point process on $[0, 1] \times (\lim_{n\to\infty}(x_* - u_n)/\sigma_n, \infty)$ with integrated intensity

$$\Lambda\left((a, b) \times (x, \infty)\right) = (b - a)\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-1/\xi}$$

Statistical models

GP distribution in practice

$$X - u \mid X > u \mathrel{\dot\sim} \mathrm{GP}(\tilde{\sigma}, \xi)$$

Poisson process in practice

$$\left\{\left(\tfrac{i}{n+1}, X_i\right) : X_i > u\right\} \mathrel{\dot\sim} \mathrm{PP}(\mu, \sigma, \xi)$$

Both require specification of a threshold u. Where does the tail begin?

- As high as possible to minimize bias
- As low as possible to minimize variance

Exploiting properties of the limit model

Threshold stability

If $X - u \mid X > u \sim \mathrm{GP}(\tilde{\sigma}, \xi)$, then for $v > 0$

$$X - (u + v) \mid X > u + v \sim \mathrm{GP}(\sigma_v, \xi)$$

with $\sigma_v = \tilde{\sigma} + \xi v$.

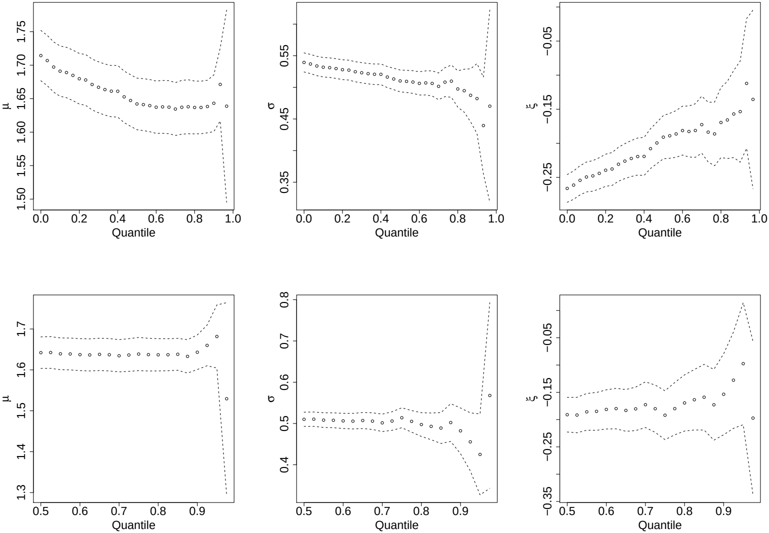
Thus when the GP distribution holds, excesses above a higher threshold also follow a GP distribution with

- the same shape parameter ξ
- modified scale parameter $\sigma_v - \xi v$ invariant to v

For the point process, points above a higher thresholds $u + v$ follow the same Poisson process with parameters (μ, σ, ξ).

$$\left\{\left(\tfrac{i}{n+1}, X_i\right) : X_i > u + v\right\} \sim \mathrm{PP}(\mu, \sigma, \xi)$$

Parameter stability plots



Parameter stability plots?

- Simple, but not sophisticated
- Assumption that we will only take a fixed threshold
  - *The* threshold does not exist
  - What about uncertainty?

Parameter stability plots?

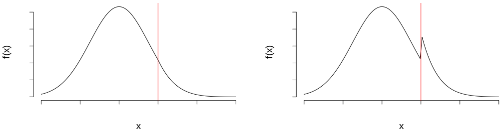
- Simple, but not sophisticated
- Assumption that we will only take a fixed threshold
  - *The* threshold does not exist
  - What about uncertainty?

Alternatives (non-exhaustive) for fixed threshold selection:

- Minimum MSE
  - Of shape parameter (Danielsson et al, 2001)
  - Of specific quantile (Ferreira et al, 2003)
- Second order decay assumptions
  - Peng (1998); Feuerverger and Hall (1999); Beirlant et al. (1999); Guillou and Hall (2001)

Alternatives for threshold uncertainty

Turning the threshold into a parameter necessitates some modelling below u:



- Parametric model
  - Gaussian, gamma, . . . (Frigessi et al., 2002; Behrens et al., 2004; Mendes and Lopes, 2004; Carreau and Bengio, 2009)
  - Extended Poisson process (Wadsworth and Tawn, 2012)
- Semi/Non-parametric model
  - Mixture of uniforms (Tancredi et al, 2006)
  - Kernel density estimation (MacDonald et al, 2011)

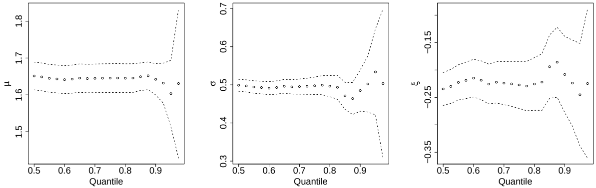
- ✗ Virtually all methods require specification of a tuning parameter: shifts the problem elsewhere
- ✓ But: sensitivity to the tuning parameter may be reduced compared to threshold sensitivity
- ✗ Bespoke coding and idea that this is "just one method" offputting

Simplicity of parameter stability plots $\Rightarrow$ still commonly used in practice

- Only need to fit model and calculate Hessian at a sequence of thresholds
- Can we keep it simple, but do more with the information we have?

More information from the same plot

Difficulty in interpretation stems from dependent estimates / CIs



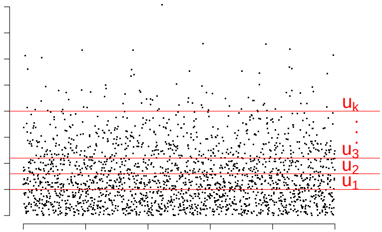
Idea:

- Find the joint (asymptotic) distribution of the MLEs calculated using different thresholds
- Use this distribution to suggest modifications to the plots to aid interpretability

Focus on NHPP representation and the parameter stability plot for the shape parameter ξ.

Set-up and notation

- Consider thresholds $u_1 < u_2 < \cdots < u_k$
- Fit the NHPP model separately above these k thresholds
- Denote the MLEs of $\boldsymbol{\theta} = (\mu, \sigma, \xi)$ from data on $(u_1, \infty), \ldots, (u_k, \infty)$, by $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_k$



Asymptotic distribution of MLEs

Let

- $l_1(\boldsymbol{\theta}), \ldots, l_k(\boldsymbol{\theta})$ log-likelihoods on $(u_1, \infty), \ldots, (u_k, \infty)$
- $\boldsymbol{\theta}_0$ true parameter value
- m grow with length of series s.t. $m \propto \mathrm{E}(\text{number of data points on } (u_j, \infty))$

Asymptotic distribution of MLEs

Let

- $l_1(\boldsymbol{\theta}), \ldots, l_k(\boldsymbol{\theta})$ log-likelihoods on $(u_1, \infty), \ldots, (u_k, \infty)$
- $\boldsymbol{\theta}_0$ true parameter value
- m grow with length of series s.t. $m \propto \mathrm{E}(\text{number of data points on } (u_j, \infty))$

Under the true model + regularity conditions $(\xi > -1/2)$

$$m^{1/2}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0) = \{-\nabla^2 l_j(\hat{\boldsymbol{\theta}}_j)/m\}^{-1} m^{-1/2} \nabla l_j(\boldsymbol{\theta}_0) + o_p(1), m \to \infty$$

Asymptotic normality of $\nabla l_j(\boldsymbol{\theta}_0)$ gives

$$\hat{\boldsymbol{\theta}}_j \dot{\sim} N_3(\boldsymbol{\theta}_0, J_j^{-1}/m)$$

with $J_j = \mathrm{E}\left[-\nabla^2 l_j(\hat{\boldsymbol{\theta}}_j)\right]$ expected / Fisher information.

For joint distribution of $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_k$, require joint distribution of scores

$$\nabla l_1(\boldsymbol{\theta}), \ldots, \nabla l_k(\boldsymbol{\theta})$$

Noting that they are sums of independent or overlapping components gives joint asymptotic distribution of scores as

$$N_{3k}\left(\mathbf{0}, \{J_{\max(i,j)}\}_{1\leq i\leq k, 1\leq j\leq k}\right)$$

and approximate asymptotic joint distribution of MLEs as

$$N_{3k}(\boldsymbol{\theta}_0, \{(J^{-1})_{\min(i,j)}\}_{1\leq i\leq k, 1\leq j\leq k}/m)$$

Consequence of the joint distribution

Consequence: independent increments property

$$\begin{pmatrix} (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2) \\ (\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_3) \\ \vdots \\ (\hat{\boldsymbol{\theta}}_{k-1} - \hat{\boldsymbol{\theta}}_k) \end{pmatrix} \dot{\sim} N_{3(k-1)} \left(\mathbf{0}, \frac{1}{m} \text{BlockDiag} \left(J_{i+1}^{-1} - J_i^{-1} \right)_{1 \le i \le k-1} \right).$$

Focussing on ξ this gives

$$\boldsymbol{\xi}^* = \begin{pmatrix} \xi_1^* \\ \xi_2^* \\ \vdots \\ \xi_{k-1}^* \end{pmatrix} := m^{1/2} \begin{pmatrix} \frac{(\hat{\xi}_1 - \hat{\xi}_2)}{\{(J_2^{-1} - J_1^{-1})_{\xi,\xi}\}^{1/2}} \\ \frac{(\hat{\xi}_2 - \hat{\xi}_3)}{\{(J_3^{-1} - J_2^{-1})_{\xi,\xi}\}^{1/2}} \\ \vdots \\ \frac{(\hat{\xi}_{k-1} - \hat{\xi}_k)}{\{(J_k^{-1} - J_{k-1}^{-1})_{\xi,\xi}\}^{1/2}} \end{pmatrix} \dot{\sim} N_{k-1}\left(\mathbf{0}, I_{k-1}\right).$$

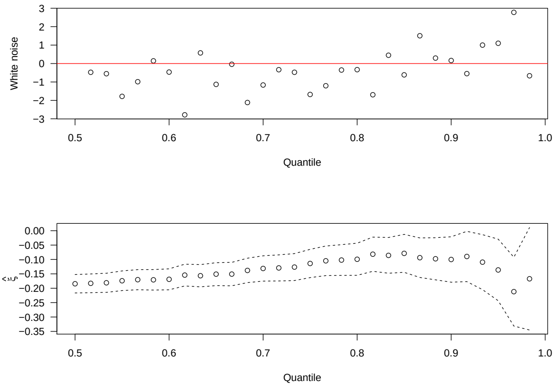
i.e. independent standard normal r.v.s. Call $\boldsymbol{\xi}^*$ the *white noise process.*

Parameter stability and white noise

- Use estimates of the information matrices to get realisations of $\boldsymbol{\xi}^*$
- Numerically-differenced Hessian can be poor, expected info much better

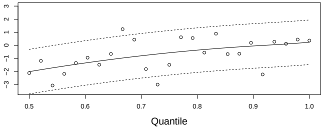
Parameter stability and white noise

- Use estimates of the information matrices to get realisations of ξ^*
- Numerically-differenced Hessian can be poor, expected info much better



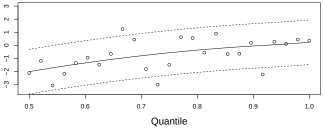
Testing for white noise

- Let $\boldsymbol{\xi}^*_{1:j} = (\xi^*_1, \ldots, \xi^*_j)$ etc.
- Structure of extreme value problems suggests $\boldsymbol{\xi}^*_{1:j}$ is less likely to be white noise than $\boldsymbol{\xi}^*_{j+1:k-1}$



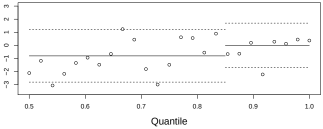
Testing for white noise

- Let $\boldsymbol{\xi}^*_{1:j} = (\xi^*_1, \ldots, \xi^*_j)$ etc.
- Structure of extreme value problems suggests $\boldsymbol{\xi}^*_{1:j}$ is less likely to be white noise than $\boldsymbol{\xi}^*_{j+1:k-1}$



One possibility: assume a simple changepoint model

$$
\begin{aligned}
\xi^*_i &\sim N(\beta, \gamma) \ \text{ iid}, \quad i = 1, \ldots, j, \\
\xi^*_i &\sim N(0, 1) \ \text{ iid}, \quad i = j+1, \ldots, k-1,
\end{aligned}
$$



Likelihood for changepoint model:

$$L(\beta, \gamma, j) = \prod_{i=1}^{k-1} \phi(\xi_i^*; \beta, \gamma)^{\mathbb{1}(i \leq j)} \phi(\xi_i^*; 0, 1)^{\mathbb{1}(i > j)}, \quad \beta \in \mathbb{R}, \gamma > 0, j \in \{2, \ldots, k-1\},$$

- Maximize the profile likelihood $L_p(j) = L(\hat{\beta}_j, \hat{\gamma}_j, j)$
  - $(\hat{\beta}_j, \hat{\gamma}_j)$ the MLEs for a fixed j
- Define $j^* := \arg\max_j L_p(j)$

Likelihood for changepoint model:

$$L(\beta,\gamma,j) = \prod_{i=1}^{k-1} \phi(\xi_i^*;\beta,\gamma)^{\mathbb{1}(i\le j)}\phi(\xi_i^*;0,1)^{\mathbb{1}(i>j)}, \quad \beta\in\mathbb{R}, \gamma>0, j\in\{2,\dots,k-1\},$$

- Maximize the profile likelihood $L_p(j) = L(\hat{\beta}_j, \hat{\gamma}_j, j)$
  - $(\hat{\beta}_j, \hat{\gamma}_j)$ the MLEs for a fixed j
- Define $j^* := \arg\max_j L_p(j)$
- "Does $L(\hat{\beta}_{j^*}, \hat{\gamma}_{j^*}, j^*)$ give a significantly better fit to $\boldsymbol{\xi}^*$ than $L(0,1,0)$?"
  - $L(0,1,0) = \prod_{i=1}^{k-1} \phi(\xi_i^*;0,1)$
- Use likelihood ratio test statistic

$$T = \frac{L(\hat{\beta}_{j^*}, \hat{\gamma}_{j^*}, j^*)}{L(0,1,0)}$$

with null distribution by simulation

Likelihood for changepoint model:

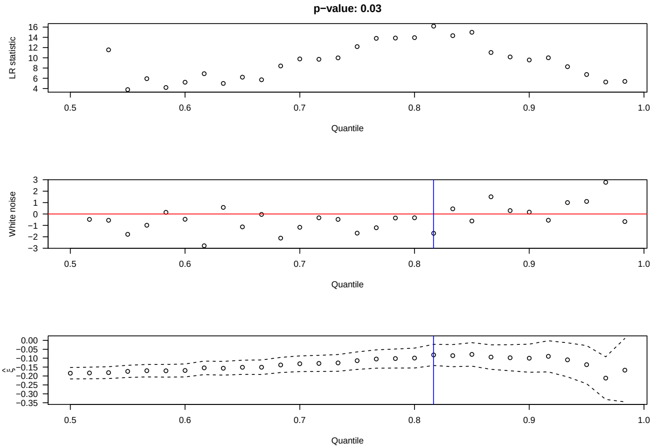
$$L(\beta, \gamma, j) = \prod_{i=1}^{k-1} \phi(\xi_i^*; \beta, \gamma)^{\mathbb{1}(i \le j)} \phi(\xi_i^*; 0, 1)^{\mathbb{1}(i > j)}, \quad \beta \in \mathbb{R}, \gamma > 0, j \in \{2, \ldots, k-1\},$$

- Maximize the profile likelihood $L_p(j) = L(\hat{\beta}_j, \hat{\gamma}_j, j)$
  - $(\hat{\beta}_j, \hat{\gamma}_j)$ the MLEs for a fixed j
- Define $j^* := \arg\max_j L_p(j)$
- "Does $L(\hat{\beta}_{j^*}, \hat{\gamma}_{j^*}, j^*)$ give a significantly better fit to $\boldsymbol{\xi}^*$ than $L(0, 1, 0)$?"
  - $L(0, 1, 0) = \prod_{i=1}^{k-1} \phi(\xi_i^*; 0, 1)$
- Use likelihood ratio test statistic

$$T = \frac{L(\hat{\beta}_{j^*}, \hat{\gamma}_{j^*}, j^*)}{L(0, 1, 0)}$$

  with null distribution by simulation
- If "significant" set $u^* = u_{j^*+1}$; else set $u^* = u_1$ (lowest threshold considered)

Testing for white noise



Caveats

- Enough data needed for joint distribution to be reasonably multivariate normal under the null
  - Number of thresholds k has some effect (tuning parameter?!)
  - Assessed by checking approximate uniformity of p-values under the null
- No theory developed for sequential testing; might be necessary in applications
- Still best combined with "educated interpretation"

Multivariate extremes

Multivariate extremes

Often extreme events are caused by the effect of more than one variable

Example

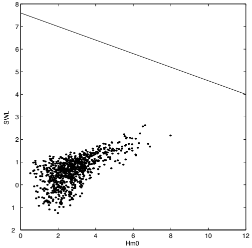


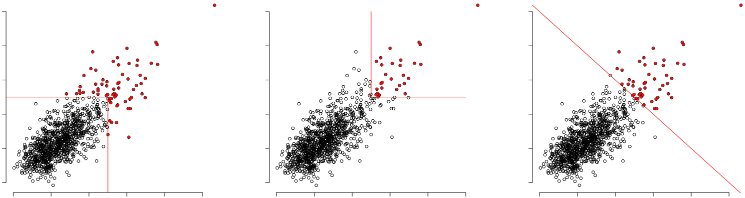
Figure 1. Wave height HmO and sea level SWL recorded during 828 storm events for the Dutch Coast. The area above the solid line represents a possible failure area.

Sea walls breached in storms due to combination of still water level and wave height

RHS plot: de Haan, L. and de Ronde, J. (1998) *Sea and wind: multivariate extremes at work*

Multivariate extremes

- Similar problems exist in defining where the tail begins
- But we also need to define what the tail *is*



- Tail definition linked to type of limit theory we wish to employ (will not focus on this aspect today)

Given a definition of the multivariate tail, how can we select a threshold?

Let

- $\boldsymbol{X}_i \sim F$
- $\boldsymbol{u}_n \in \mathbb{R}^d$ s.t. $F(\boldsymbol{u}_n) \to 1$ as $n \to \infty$

If there exists $\boldsymbol{\sigma}_n > 0$ s.t.

$$\mathrm{P}\left(\frac{\boldsymbol{X}_i - \boldsymbol{u}_n}{\boldsymbol{\sigma}_n} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X}_i \not\leq \boldsymbol{u}_n \right) \to H_\ell(\boldsymbol{x}; \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\tau})$$

for non-degenerate H then this is the multivariate generalized Pareto or MGP distribution (Rootzén and Tajvidi, 2006; Beirlant et al., 2004, Ch. 8).

Models for multivariate extremes

Let

- $\boldsymbol{X}_i \sim F$
- $\boldsymbol{u}_n \in \mathbb{R}^d$ s.t. $F(\boldsymbol{u}_n) \to 1$ as $n \to \infty$

If there exists $\boldsymbol{\sigma}_n > 0$ s.t.

$$\mathrm{P}\left(\frac{\boldsymbol{X}_i - \boldsymbol{u}_n}{\boldsymbol{\sigma}_n} \leq \boldsymbol{x} \,\middle|\, \boldsymbol{X}_i \not\leq \boldsymbol{u}_n \right) \to H_\ell(\boldsymbol{x}; \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\tau})$$

for non-degenerate H then this is the multivariate generalized Pareto or MGP distribution (Rootzén and Tajvidi, 2006; Beirlant et al., 2004, Ch. 8).

$$H_\ell = \frac{\ell\left(\boldsymbol{\tau}\left(\mathbf{1} + \boldsymbol{\xi}\min(\boldsymbol{x}, \mathbf{0})/\boldsymbol{\sigma}\right)_+^{-1/\boldsymbol{\xi}}\right) - \ell\left(\boldsymbol{\tau}\left(\mathbf{1} + \boldsymbol{\xi}\boldsymbol{x}/\boldsymbol{\sigma}\right)_+^{-1/\boldsymbol{\xi}}\right)}{\ell\left(\boldsymbol{\tau}\right)}$$

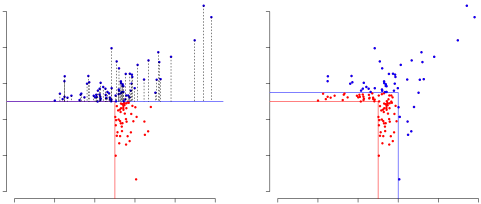
- $\ell : (0, \infty)^d \to (0, \infty)$ stable tail dependence function capturing extremal dependence

Salient properties of MGP distributions

Suppose $\boldsymbol{Z} | \boldsymbol{Z} \not\leq \boldsymbol{0} \sim H_\ell(\boldsymbol{x}; \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\tau})$. Then

- $Z_j | Z_j > 0 \sim \mathrm{GP}(\sigma_j, \xi_j)$
- For $\boldsymbol{v} > \boldsymbol{0}$

$$\boldsymbol{Z} - \boldsymbol{v} | \boldsymbol{Z} \not\leq \boldsymbol{v} \sim H_\ell(\boldsymbol{x}; \boldsymbol{\sigma_v}, \boldsymbol{\xi}, \boldsymbol{\tau}_v)$$



Statistical model

Analogous to the univariate case assume

$$\boldsymbol{X} - \boldsymbol{u} | \boldsymbol{X} \not\leq \boldsymbol{u} \dot{\sim} H_\ell(\boldsymbol{x}; \tilde{\boldsymbol{\sigma}}, \boldsymbol{\xi}, \tilde{\boldsymbol{\tau}})$$

Need to pick a threshold $\boldsymbol{u}$ such that:

- $X_j - u_j | X_j > u_j \sim \mathrm{GP}(\tilde{\sigma}_j, \xi_j)$ (See Part 1!)
- The dependence structure is well described by a MGP distribution

Statistical model

Analogous to the univariate case assume

$$\boldsymbol{X} - \boldsymbol{u} | \boldsymbol{X} \not\leq \boldsymbol{u} \,\dot{\sim}\, H_\ell(\boldsymbol{x}; \tilde{\boldsymbol{\sigma}}, \boldsymbol{\xi}, \tilde{\boldsymbol{\tau}})$$

Need to pick a threshold $\boldsymbol{u}$ such that:

- $X_j - u_j | X_j > u_j \sim \mathrm{GP}(\tilde{\sigma}_j, \xi_j)$ (See Part 1!)
- The dependence structure is well described by a MGP distribution
  - ℓ has no finite-dimensional parameterization
  - Any given parametric model may fit the data badly... doesn't mean not MGP

Dependence in MGP distributions

Key summary parameter for multivariate extremal dependence is

$$\chi_{1:d} = \lim_{q \to 1} \frac{\mathrm{P}(F_1(X_1) > q, \ldots, F_d(X_d) > q)}{1-q}$$

Often studied as a function of q for q near 1:

$$\chi_{1:d}(q) = \frac{\mathrm{P}(F_1(X_1) > q, \ldots, F_d(X_d) > q)}{1-q}$$

Dependence in MGP distributions

Key summary parameter for multivariate extremal dependence is

$$\chi_{1:d} = \lim_{q \to 1} \frac{\mathrm{P}(F_1(X_1) > q, \ldots, F_d(X_d) > q)}{1-q}$$

Often studied as a function of q for q near 1:

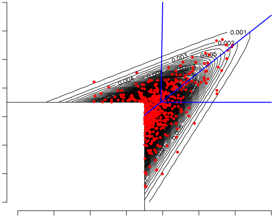
$$\chi_{1:d}(q) = \frac{\mathrm{P}(F_1(X_1) > q, \ldots, F_d(X_d) > q)}{1-q}$$

If

$$\boldsymbol{X} - \boldsymbol{u} | \boldsymbol{X} \not\leq \boldsymbol{u} \sim H_\ell(\boldsymbol{x}; \tilde{\boldsymbol{\sigma}}, \boldsymbol{\xi}, \tilde{\boldsymbol{\tau}})$$

then $\chi_{1:d}(q)$ is constant when $\boldsymbol{X} > \boldsymbol{u}$

Dependence in MGP distributions



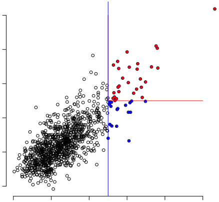
- $\chi_{1:d}(q)$ constant when $\boldsymbol{X} > \boldsymbol{u}$... suggests $\boldsymbol{u} = (F_1^{-1}(q), \ldots, F_d^{-1}(q))$
- But $\boldsymbol{u}$ need not correspond to equal quantiles
- Identifying q above which $\chi_{1:d}(q)$ constant gives maximum marginal quantile above which dependence assumption should hold
- Common in practice to make dependence assumption above equal quantiles

Parameter stability for χ

Empirical estimate for χ:

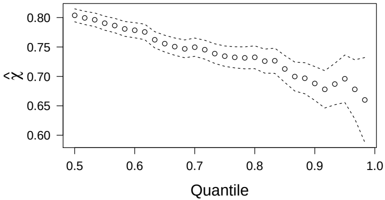
$$\hat{\chi}_{1:d}(q) = \frac{1}{n}\sum_{k=1}^{n} \frac{\mathbb{1}(\min\{\tilde{F}_1(X_{k,1}), \dots, \tilde{F}_d(X_{k,d})\} > q)}{1-q}$$

- $\tilde{F}_j$ empirical cdfs
- MLE based on binomial assumption

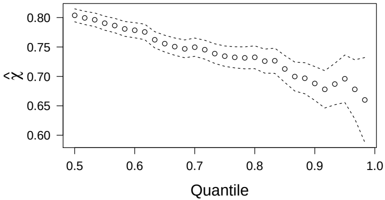


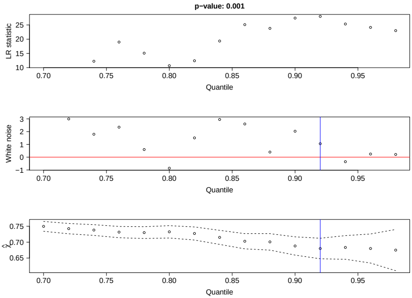
Use parameter stability plots to identify where $\hat{\chi}_{1:d}(q)$ becomes constant

Parameter stability for χ

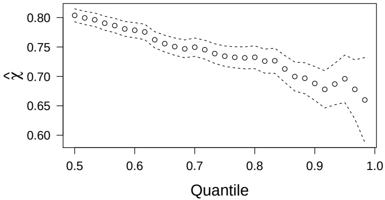


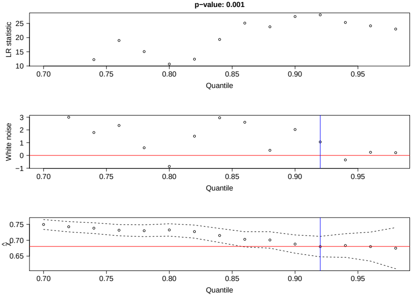
Parameter stability for χ





Parameter stability for χ

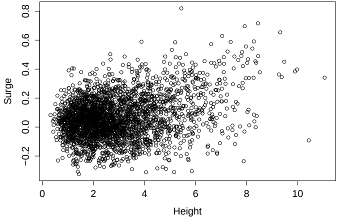




Wave Height Example

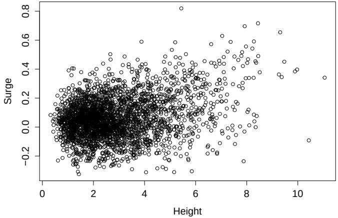
Data

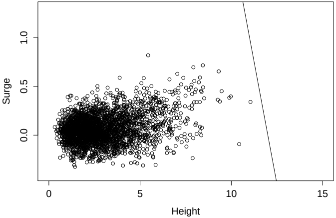
- 2894 measurements of wave height and surge from Newlyn, UK
- Filtered for “approximate temporal independence”



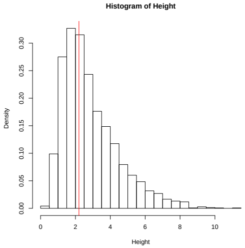
Data

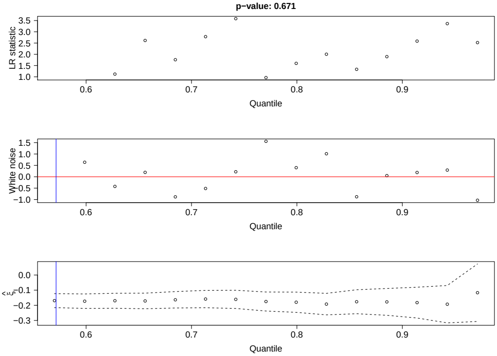
- 2894 measurements of wave height and surge from Newlyn, UK
- Filtered for “approximate temporal independence”



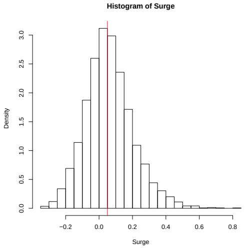


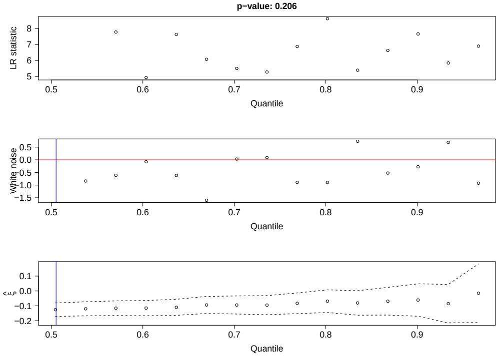
Margins: Height



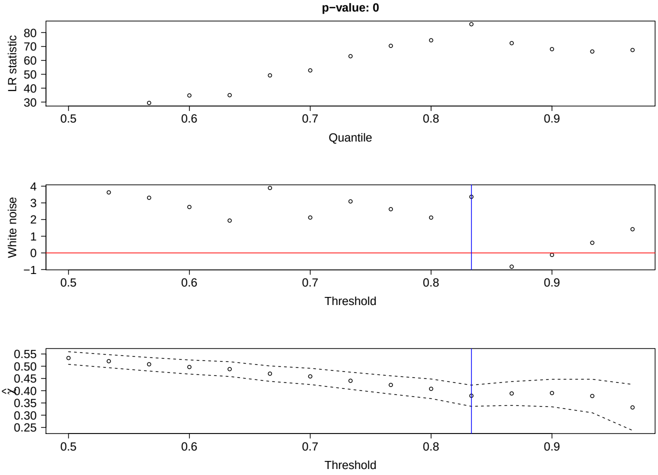


Margins: Surge





Dependence

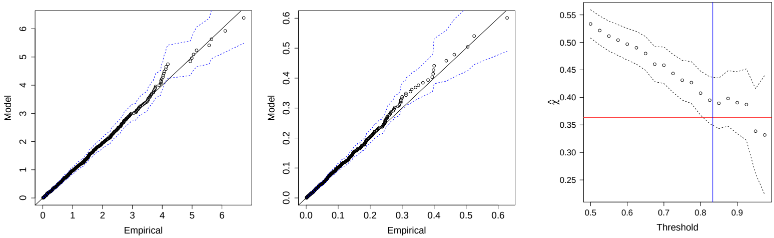

p−value: 0
LR statistic
80
70
60
50
40
30
0.5
0.6
0.7
0.8
0.9
Quantile
White noise
4
3
2
1
0
−1
0.5
0.6
0.7
0.8
0.9
Threshold
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.5
0.6
0.7
0.8
0.9
Threshold

Putting it together

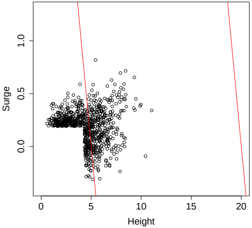
Quantiles implicated:

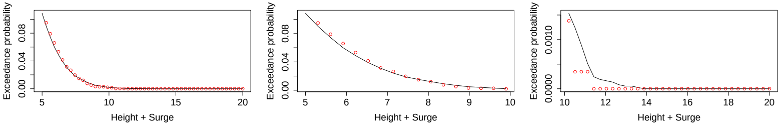
- Height marginal: 0.57 quantile
- Surge marginal: 0.505 quantile
- Dependence: maximum marginal quantile 0.83

Use 0.83 quantile for both margins



Putting it together





Summary

- Threshold selection is challenging!
  - Practitioners will use simple methods unless something else convincingly better
  - Idea in this talk: make "cheap and dirty" methods slightly less dirty
- Univariate threshold selection has received a lot of attention
  - Parameter stability plots can be used in MV contexts too; as can joint distribution of MLEs
- Multivariate extremal modelling can involve *lots* of threshold selection — can we simplify?

Main reference:

Wadsworth, J. L. (2016) *Exploiting structure of maximum likelihood estimators for extreme value threshold selection*, to appear in Technometrics

Some code available at:

`http://www.lancaster.ac.uk/~wadswojl/RCode.html`

References

Behrens, C., Lopes, H. and Gamerman, D. (2004)
Bayesian analysis of extreme events with threshold estimation
Statist. Mod. 4, 227-244

Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G. (1999)
Tail Index Estimation and an Exponential Regression Model
Extremes 2(2), 177-200

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004)
Statistics of Extremes
Wiley

Carreau, J. and Bengio, Y. (2009)
A hybrid Pareto model for asymmetric fat-tailed data: the univariate case
Extremes 12, 53-76

Danielsson, J., de Haan, L., Peng, L. and de Vries C. (2001)
Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation
J. Mult. Analysis 76(2), 226-248

Ferreira, A., de Haan, L. and Peng, L. (2003)
On optimising the estimation of high quantiles of a probability distribution
Statistics 37(5), 401-434

Feuerverger, A. and Hall, P. (1999)
Estimating a Tail Exponent by Modelling Departure from a Pareto Distribution
Ann. Stat. 27(2), 760–781

Frigessi, A. and Haug, O. and Rue, H. (2003)
A dynamic mixture model for unsupervised tail estimation without threshold selection
Extremes 5 219-235

References

Guillou, A. and Hall, P. (2001)
A diagnostic for selecting the threshold in extreme value analysis
J. Roy. Stat. Soc. B 63(2), 293-305

Lee, J. and Fan, Y. and Sisson, S. (2014)
Bayesian threshold selection for extremal models using measures of surprise
http://arxiv.org/pdf/1311.2994

MacDonald, A., Scarrott, C.J., Lee, D., Darlow, B., Reale, M. and Russell, G.
A flexible extreme value mixture model
Comp. Statist. Data Anal. 55, 2137-2157

Mendes, B. and Lopes, H. F. (2004)
Data driven estimates for mixtures
Comp. Statist. Data Anal. 47, 583-598

Rootzén, H. and Tajvidi, N. (2006)
Multivariate generalized Pareto distributions
Bernoulli 5, 917-930

Pickands, J. III (1971)
The two-dimensional Poisson process and extremal processes
J. App. Prob 8(4), 745-756

Tancredi, A., Anderson, C. W. and O'Hagan, A. (2006)
Accouting for threshold uncertainty in extreme value estimation
Extremes 9, 87-106

Wadsworth and Tawn (2012)
Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling
J. Roy. Stat. Soc. B 74(3), 543-567