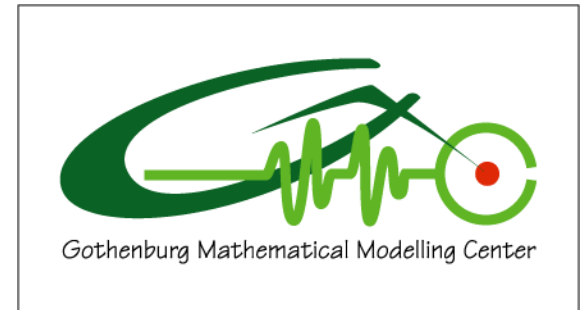


Limit Theorems for Empirical Processes of Cluster Functionals

Holger Drees
University of Hamburg

and

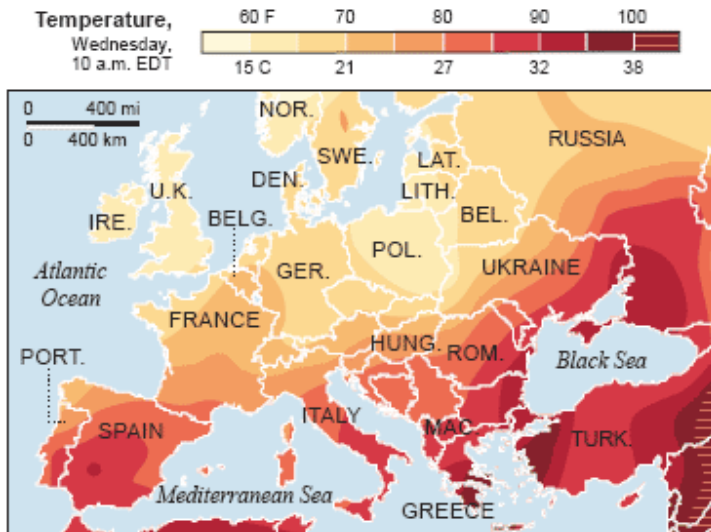
Holger Rootzén
GMMC & Stochastic Centre
Chalmers & Gothenburg University



www.math.chalmers.se/~rootzen/

Why models for clusters of extremes?

- Heat waves may kill many 1000. Not just one high temperature which kills, but entire sequence of temperatures – cf 2003 heat wave in Europe.
- Floods may be caused by extreme precipitation when the ground is saturated by water from a previous rain – cf 2000 flood in northern Sweden



Why empirical process theory?

- Omnibus method to prove asymptotic normality of estimators and goodness of fit tests
- Proofs are important because
 - Sometimes heuristics are wrong
 - Sometimes exact limits of validity of heuristics are unclear
 - Sometimes they provide more detailed information
 - And, they lead to clearer and more general thinking which gives base for the next heuristic
- May stimulate development of new methods

Empirical limit theory – uniform central limit theorems

X_1, X_2, \dots i.i.d \mathcal{F} class of functions, e.g. $\mathcal{F} = \{\mathbf{1}_{(-\infty, t]} \mid t > 0\}$

$Z_n : \mathcal{F} \rightarrow \mathbf{R}$

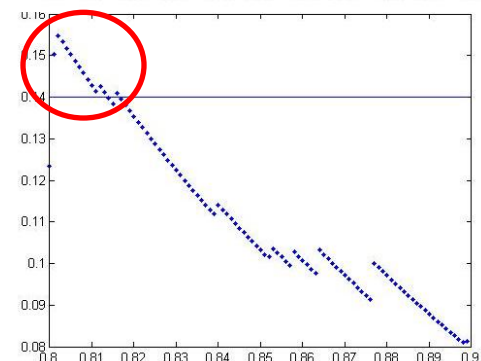
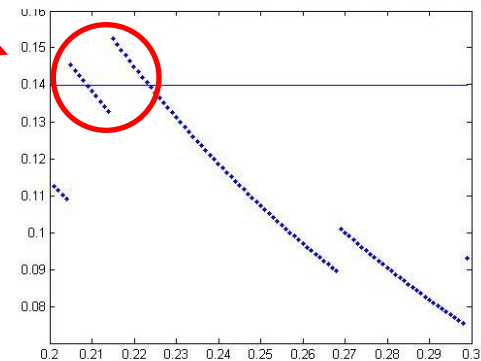
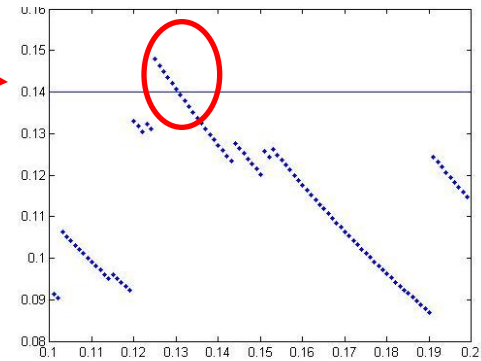
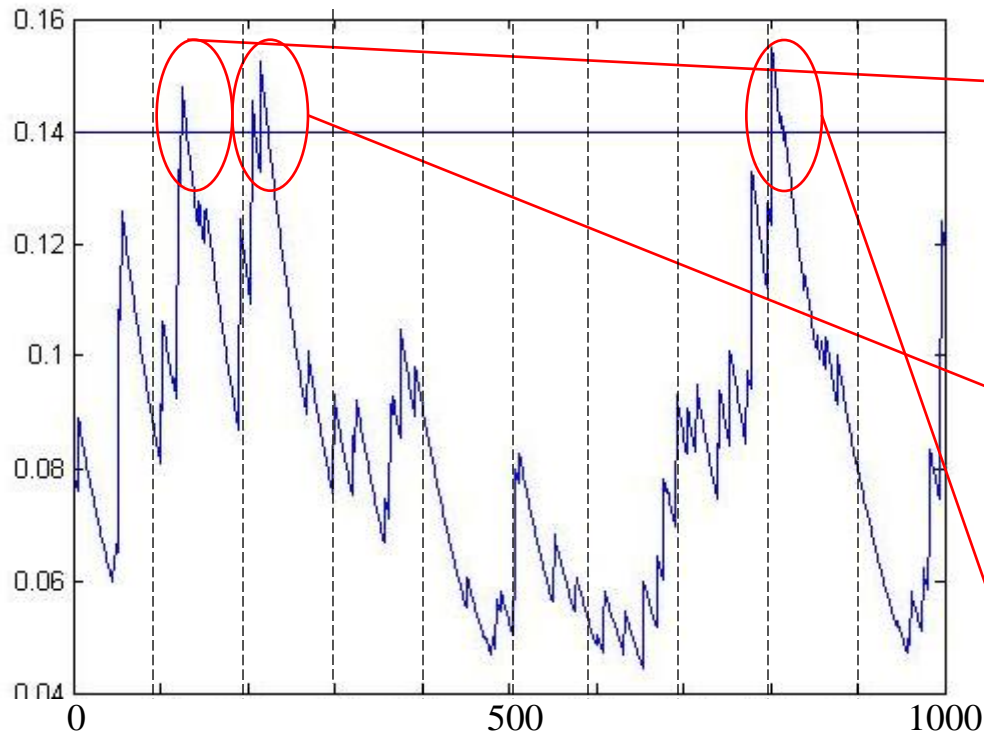
empirical function

$$f \rightarrow Z_n(f) := \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - E(f(X_i)))$$

often $\sqrt{n} (\hat{\theta}_n - \theta) \approx h(Z_n)$ and one can hope that $Z_n \xrightarrow{d} Z$
implies that $\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{d} h(Z)$

“True” if $h : \ell^\infty(\mathcal{F}) \rightarrow R$ is continuous (and this in fact is the
“definition” of \xrightarrow{d})

Cores --- clusters of extremes



Blocks – separated by dashed lines:
increase as sample size increases

Core -- the large values in a block:
starts with first large value in block,
ends with the last

$(X_{n,i})$ triangular array of real variables, stationary in rows

typical example: $X_{n,i} = \frac{X_i - u_n}{a_n}$, with X_i stationary

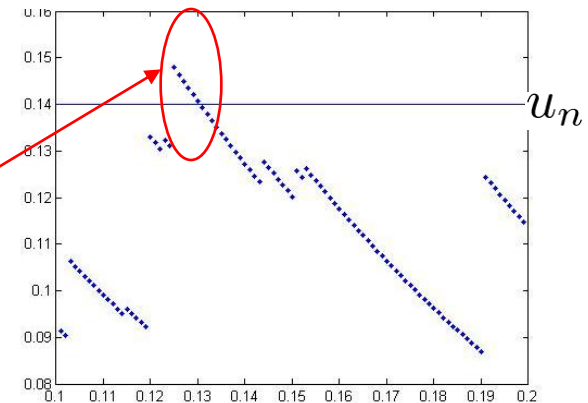
$Y_{n,i} = (X_{n,j})_{(i-1)r_n < j \leq ir_n}$ the i -th block:

↑
blocklength

core: shortest consecutive vector containing all positive values in a block

$Y_{n,i}^c = (Y_{n,i,j})_{M_i \leq j \leq N_i}$ (or 0 if no pos value)

$Y_{n,i}, Y_{n,i}^c \in \mathbf{R}_U := \cup_{d=1}^{\infty} \mathbf{R}^d$



goal: "limit distribution of cores" in general setting

$f : \mathbf{R}_U \rightarrow \mathbf{R}$ **cluster functional** if f only depends on the core, i.e. if

$$f(Y) = f(Y^c), \quad Y \in \mathbf{R}_U, \quad \text{and} \quad f(0) = 0$$

Yun (2000)
Segers (2003)

\mathcal{F} class of cluster functionals

the number of blocks

$$Z_n(f) := \frac{1}{\sqrt{nv_n}} \sum_{i=1}^{m_n} (f(Y_{n,i}) - E(f(Y_{n,i}))), \quad f \in \mathcal{F} \quad \text{empirical process}$$

\uparrow
 $= P(X_{n,1} > 0) \rightarrow 0 \quad \text{and} \quad nv_n \rightarrow \infty$

Prove uniform central limit theorems for Z_n

for interesting classes \mathcal{F} under suitable conditions

“big blocks – small blocks” proofs

- beta-mixing:

$$\beta_{n,\ell_n} \frac{n}{r_n} \rightarrow 0 \quad \text{for} \quad \beta_{n,\ell} = \sup_{k \geq 1} E_n \sup_{A \in \mathcal{B}_{n,\ell+k+1}^k} |P_n(A | \mathcal{B}_{n,1}^k) - P_n(A)|$$

length of short blocks = $o(r_n)$

- sums over small blocks tend to zero (i.i.d. calculation)

makes it possible to assume blocks are independent in proofs
(assumed without further comment in sequel)

Finite-dimensional convergence:

$$(Z_n(f_1), \dots, Z_n(f_d)) \xrightarrow{d} (Z(f_1), \dots, Z(f_d)), \quad \forall d, f_1, \dots, f_d \in \mathcal{F}$$

- Lindeberg (typically consequence of tightness condition)
- covariances converge: Covariance function for limit Z

$$\frac{1}{r_n v_n} \text{Cov}(f(Y_{n,1}), g(Y_{n,1})) \rightarrow r(f, g), \quad \text{all } f, g \in \mathcal{F}$$

(can use results of Segers (2003) to prove convergence of covariances)

Asymptotic equicontinuity / tightness

Ensures:

- Distribution of entire process can be approximated by finite-dimensional distributions
- Continuity of limiting Gaussian process

Tail empirical function (uniformly distributed variables)

If $f_t : \mathbf{R}_U \rightarrow \mathbf{R}$, $f_t(y_1, \dots, y_k) = \sum_{i=0}^k \mathbf{1}_{(t, \infty)}(y_i)$

then $Z_n(f_t) \approx \frac{1}{\sqrt{r_n v_n}} \sum_{i=1}^n (\mathbf{1}_{\{(X_i) - u_n\}/a_n > t\}} - \bar{F}(a_n t + u_n))$

Uniform central limit theorem if

$$\frac{1}{r_n v_n} E \left(\left(\sum_{i=1}^{r_n} \mathbf{1}_{\{x < X_{n,i} \leq y\}} \right)^2 \right) \leq |\log(y - x)|^{-(1+\epsilon)}$$

$$\frac{1}{r_n v_n} \text{Cov}(f_s(Y_{n,1}), f_t(Y_{n,1})) \rightarrow r(s, t), \quad \text{all } f, g \in \mathcal{F}$$

Special case of tail array sums (Leadbetter)

Φ class of functions $\phi : \mathbf{R}^k \rightarrow \mathbf{R}$ with $\phi(\mathbf{0}) = 0$

Define $\mathcal{F} = \{f_\phi \mid \phi \in \Phi\}$ by

$$f_\phi : \mathbf{R}_\cup \rightarrow \mathbf{R}, \quad f_\phi(y_1, \dots, y_k) = \sum_{i=0}^{k-d} \phi(y_{i+1}^+, \dots, y_{i+d}^+)$$

- $\phi_t(y) = \mathbf{1}_{(t, \infty)}(y)$ tail empirical process
- $\phi_t(y) = (y - t)^+$ total mass above t
- $\phi_{s,t}(y_1, y_2) = \mathbf{1}_{(-\infty, s] \times (t, \infty)}(y_1, y_2)$ number of upcrossings of $(s, t]$

Joint distribution function of cluster values

$\ell(y)$ = length of core of y

\mathcal{F} = indicator functions of sets of the form

$$\{y_1^c \in [0, t_1], \dots, y_j^c \in [0, t_j], \ell(x) = j\}, \quad j = 1, 2, \dots$$

Uniform central limit theorem if

finite-dimensional marginal distributions belong to the domain of attraction of multivariate extreme value distribution

$$E \left(\ell(Y_{n,1})^{1/2+\epsilon} \mid Y_{n,1} \neq 0 \right) = O(1), \quad \text{some } \epsilon > 0$$

Extensions/problems

- Other cluster definitions than the blocks definition, e.g. clusters start when there is an " r_n -upcrossing", i.e. a positive $X_{n,i}$ preceded by r_n values which are negative or zero.
- Random level, say the c_n -th largest order statistic instead of the non-random u_n : gives different limit, and new difficulties of proof

