Extreme Value Theory in Time Series Analysis

Richard A. Davis
Columbia University

http://www.stat.columbia.edu/~rdavis

Strasbourg 6/08

1

Game Plan

➢ Quick overview of classical extreme value theory
- Extremal types theorem
- Generalized extreme value distributions
- Examples
- Domain of attraction
- Generalized Pareto distribution

➢ Point Processes
- Basic results
- Application to extremes

➢ Extensions to stationary time series
- Extremal index
- Mixing conditions
- Point process examples

➢ Regular variation
- Univariate
- Multivariate
- Estimation
- Examples

➢ Application to Crystal River Flows
➢ Bonus example!!

Strasbourg 6/08

2

References

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*. Cambridge University Press, Cambridge

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

Galambos, J. (1978) *The Asymptotic Theory of Order Statistics*. Wiley, New York.

Haan, L. de, Fereira, A. (2006) *Extreme Value Theory: An Introduction.* Springer, Berlin, New York.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.

Resnick, S.I. (1986) Point processes, regular variation and weak convergence. *Adv. Appl. Prob.* 18, 66–138.

Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

Resnick, S.I. (20067 *Heavy Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance*. Chapman and Hall, London.

Classical Extreme Value Theory

Setup:

- $\{X_t\} \sim \text{IID}(F)$

- $M_n = \max\{X_1, \ldots, X_n\}$

- $P(M_n \leq x) = F^n(x)$

Now in order for the right hand side to converge to a nonzero value, must let $x \to \infty$ with n. Replacing x by $u_n \to \infty$,

$$P(M_n \leq u_n) = F^n(u_n)$$

$$= (1-(1-F(u_n)))^n$$

$$= (1-n(1-F(u_n))/n)^n \to e^{-\tau}$$

if and only if

$$n(1-F(u_n)) \to \tau$$

Classical EVT— Extremal Types Theorem

Convergence of types: Now taking $u_n = a_n x + b_n$, $a_n > 0$,

$$P\,(a_n^{-1}(M_n - b_n) \le x) = F^n(a_n x + b_n)$$

$$\rightarrow G(x)$$

if and only if

$$n(1 - F(a_n x + b_n)) \rightarrow -\log G(x)$$

Theorem. If G is a nondegenerate distribution, then G has to be one of the three types,

1. $G(x) = \exp(-e^{-x})$ (Gumbel)

2. $G(x) = \exp(-x^{-\alpha})$, $x \ge 0$ (Fréchet)

3. $G(x) = \exp(-(-x)^{\alpha})$, $x \le 0$ (Weibull)

Strasbourg 6/08

Classical EVT— GEV

The three types of extreme value distributions can be parameterized as one family (**GEV family**) via,

$$G(x) = \exp(-(1 + \xi x)^{-1/\xi}), \quad 1 + \xi x > 0$$

where

$\xi = 0$ implies $G(x) = \exp(-e^{-x})$ (Gumbel)

$\xi > 0$ implies $G(x)$ is Fréchet ($\alpha = 1/\xi$)

$\xi < 0$ implies $G(x)$ is Weibull ($x < -1/\xi$, $\alpha = -1/\xi$)

Summary

- Gumbel—light tailed

- Fréchet—heavy tailed

- Weibull—finite endpoint

Strasbourg 6/08

Classical EVT— Examples

In looking at maxima, usually one considers only the Gumbel and Fréchet distributions for modeling.

Examples

1. uniform. F(x)=x, 0 < x < 1. With a_n=1/n, b_n=1, we have for n large

$$n(1\text{-}F(a_nx+ b_n)) = -x = -\log G(x) = -\log(\ \exp(-(-x)^1))$$

$$P(n(M_n - 1) \le x) \to e^x \quad (x \le 0). \text{ Weibull limit}$$

2. exponential. F(x)=1-e^{-x}, x > 0. With a_n=1, b_n=log n, we have

$$n(1\text{-}F(a_nx+ b_n)) = n\ e^{-x\text{-log }n} = e^{-x} = -\log(\exp(-e^{-x}))$$

$$P(\ M_n - \log n \le x) \to \exp(-e^{-x}) \quad \text{(Gumbel limit)}$$

Application to Fisher's test for hidden periodicity in time series (see B&D (1991) and Davis and Mikosch (1999)).
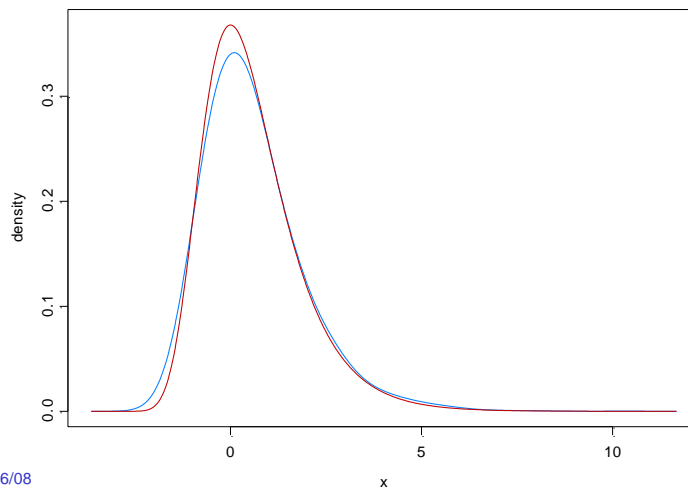
7

Classical EVT— Examples

Exponential. $P(\ M_n - \log n \le x) \to \exp(-e^{-x})$

Empirical distribution of $M_n - \log n$, n=100   Limit density
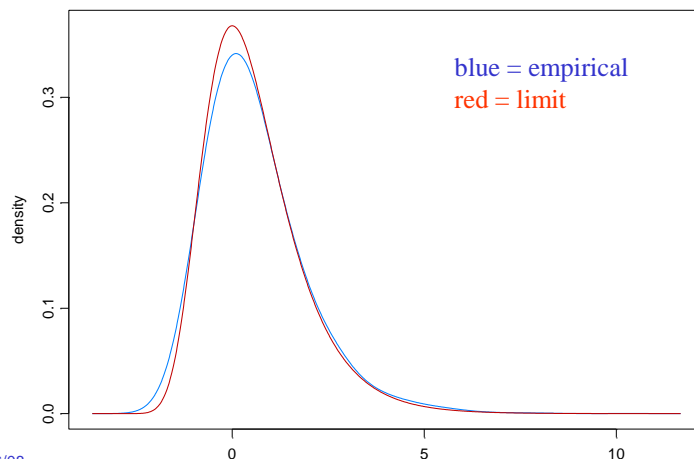

8

4

Classical EVT— Examples

3. Pareto. $F(x)=1-x^{-\alpha}$, $1 < x$. With $a_n=n^{\alpha}$, $b_n=0$, we have for n large

$n(1-F(a_nx)) = x^{-\alpha} = -\log G(x) = -\log(\ \exp(-x^{-\alpha})\ )$

$P(n^{-\alpha}M_n \le x) \to \exp(-x^{-\alpha})\ \ (x \ge 0)$. Fréchet limit

blue = empirical
red = limit

9

Classical EVT— Examples
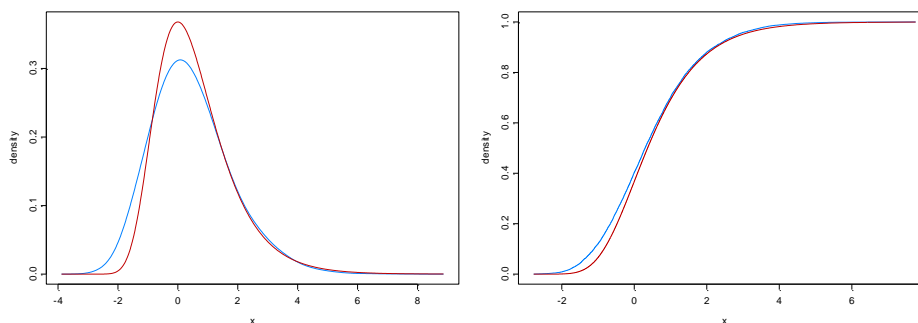
4. Gaussian.

$a_n= (2\log(n))^{-1/2}$, $b_n=(2\log(n))^{1/2} - .5\ (2\log(n))^{-1/2}(\log\log(n)+\log(4pi))$,

$n(1-F(a_nx+ b_n)) \to e^{-x} = -\log(\exp(-e^{-x}))$

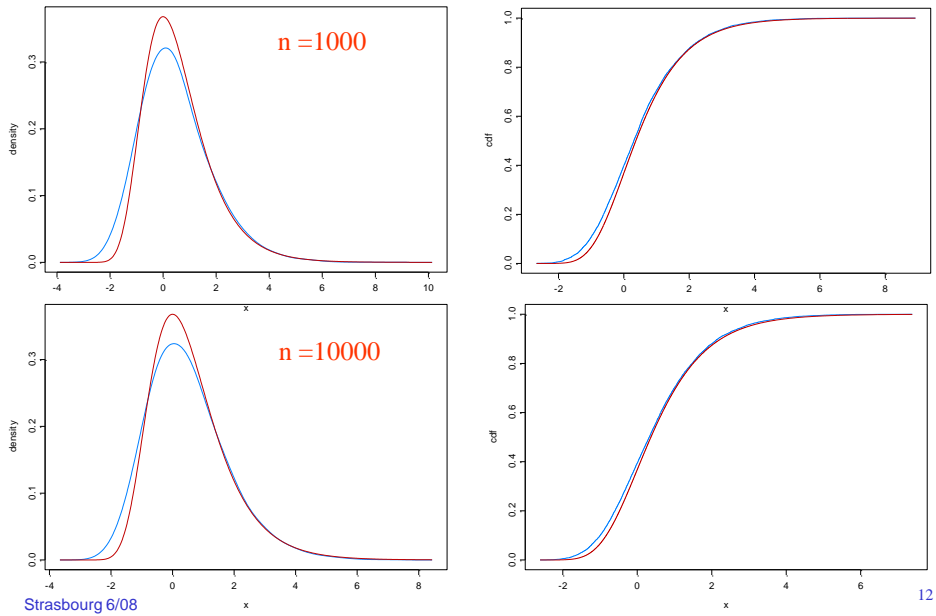$P((2\log(n))^{1/2}(M_n - b_n)\le x) \to \exp(-e^{-x})$ (Gumbel limit)

n=100

10

5

Classical EVT— Examples

n =1000

n =10000

Strasbourg 6/08

12

Classical EVT— Examples

n =1000

n =100000

Strasbourg 6/08
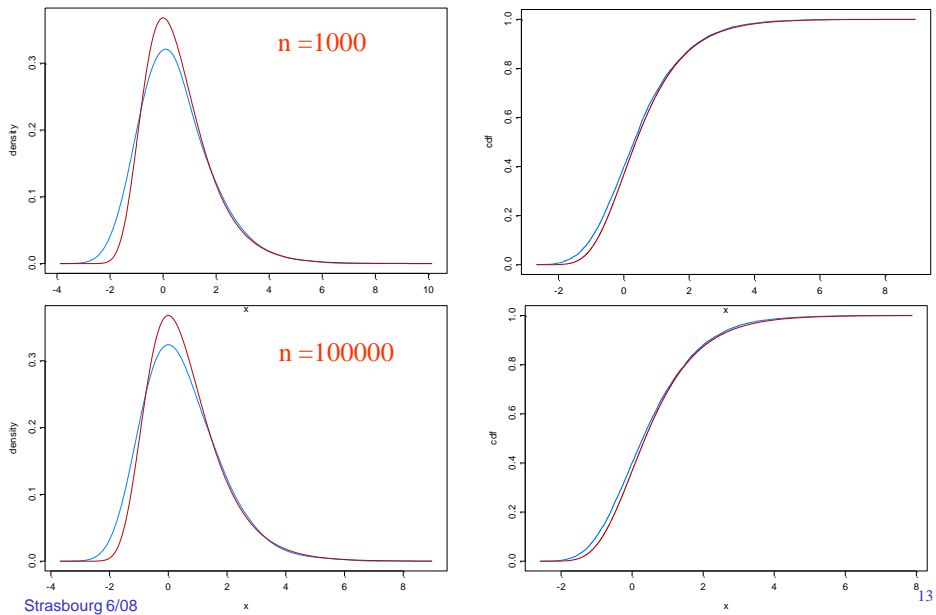
13

Classical EVT— Domain of Attraction

If

$$P((M_n - b_n)/ a_n \leq x) = F^n(a_n x + b_n) \rightarrow G(x)$$

which is equivalent to

$$n(1 - F(a_n x + b_n)) \rightarrow -\log G(x) = (1 + \xi x)^{-1/\xi},$$

then we say that F (or X) is in the **domain of attraction of G** (write F or X ϵ D(G)).

Note that if x=0, then

$$n(1 - F(b_n)) \rightarrow 1,$$

or

$$P(X_1 > b_n) \sim 1/n.$$

It follows that

$$P(X_1 - b_n > a_n x \mid X_1 > b_n) \rightarrow (1 + \xi x)^{-1/\xi}$$

Strasbourg 6/08

14

Classical EVT— GPD

$$P(X_1 - b_n > a_n x \mid X_1 > b_n) \rightarrow (1 + \xi x)^{-1/\xi}$$

Replacing b_n with u, it follows that conditional distribution of $X_1 - u$ given $X_1 > u$ has an approximate generalized Pareto distribution,

$$P(X_1 - u \leq a_n x \mid X_1 > u) \sim H(x) = 1 - (1 + \xi x)^{-1/\xi}$$

for u large.

Generalized Pareto Distribution (GPD)

$$H(x) = 1 - (1 + \xi(x - \mu)/\sigma)^{-1/\xi}$$

The limit above is the basis for fitting the tail of the distribution of a random variable using a GPD. There are myriad of procedures for estimating the parameters in this model.

Strasbourg 6/08

15

7

Classical EVT— Domains of Attraction

Domains of attraction: There are necessary and sufficient conditions for F ϵ D(G) for the three extreme value distributions. The heavy-tailed Fréchet, which is perhaps the most commonly used extreme value distribution, has the easiest n.a.s. to state (and check!). In this case,

F ϵ D(exp(-$x^{-\alpha}$))   if and only if   F is RV(α) for some $\alpha > 0$.

Regular variation: F is RV(α)   if and only if

$$\frac{\overline{F}(tx)}{\overline{F}(t)} = \frac{P(X > tx)}{P(X > t)} \to x^{-\alpha} \quad \text{as} \quad t \to \infty,$$

for every x > 0.

16

Classical EVT — Domains of Attraction

Examples:

      Pareto,
      log-gamma,
      infinite variance stable,
      Cauchy,
      student,
      Fréchet.

Remarks:

• P(X > x) = L(x)$x^{-\alpha}$ where L is slowly varying, L(tx)/L(t)$\to$1 (think log function for L)

• Can take b_n = 0 and a_n as the 1-1/n quantile of F, i.e., n(1-F(a_n)) $\to$1. It turns out that $a_n = L_1(n)n^{1/\alpha}$

• P(a_n^{-1} M_n $\leq$ x) $\to$ exp(-$x^{-\alpha}$).

17

Point Processes

The theory of point processes plays a central role in extreme value theory. Applications include:

- Derivation of joint limiting distribution of order statistics, i.e., kth largest order statistic, limiting distribution of maximum and minimum, etc.

- Calculation of limit distribution of exceedances of a high level.

- Extensions to stationary processes.

- Provides a useful tool in heavy-tailed case for deriving limiting behavior of various statistics, e.g., sample mean, sample autocovariances, etc, which are often determined by the behavior of the extreme order statistics.

18

Point Processes — Definition and Basic Results.

Setup:

- Suppose (X_t) is an iid sequence with common distribution F.

- $F \in D(\exp(-x^{-\alpha}))$ (Can be formulated for general extreme value distribution, but for simplicity we will restrict attention to Fréchet.)

We have

$$n(1-F(a_n x)) = nP(a_n^{-1}X > x)) \to x^{-\alpha}$$

from which it follows that

$$nP(a_n^{-1}X \in (a,b]) \to a^{-\alpha} - b^{-\alpha} =: \nu(a,b],$$

where ν is the measure on $(0,\infty]$ given by $\nu(dx) = \alpha x^{-\alpha-1}(dx)$.

More generally, we have

$$nP(a_n^{-1}X \in B) \to \nu(B)$$

for all nice sets B, those bounded away from 0.

19

9

Point Processes — Definition and Basic Results.

Since

$$nP(a_n^{-1} X \in B) \to \nu(B)$$

for nice sets B, it follows that

$$nP(a_n^{-1} X \in \cdot) \to_v \nu(\cdot),$$

where $\to_v$ denotes vague convergence of measures.

Now for a set B bounded away from 0, define the sequence of point processes by

$$N_n(B) = \#\{t=1,\dots,n: \ a_n^{-1} X_t \in B\} \ = \ \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t}(B)$$

where ε_x is the Dirac measure at x.

20

Point Processes — Definition and Basic Results.

$$N_n(B) = \#\{t=1,\dots,n: \ a_n^{-1} X_t \in B\} \ = \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t}(B)$$

Properties:

• $N_n(B) \sim \text{Bin}(n, p_n)$, where $p_n = P(a_n^{-1} X \in B)$ and since

$$nP(a_n^{-1} X \in B) \to \nu(B),$$

it follows from convergence of binomial to Poisson that

$$N_n(B) \to_d N(B),$$

where N(B) is a Poisson random variable with mean ν(B).

• In fact, we have the stronger point process convergence,

$$N_n \to_d N$$

where N is a Poisson process on $(0,\infty]$ with mean measure ν (dx)

and $\to_d$ denotes convergence in distribution of point processes.

21

Point Processes — Definition and Basic Results.

Convergence for point processes.

For our purposes, $\rightarrow_d$ for point processes means that for any collection of Borel sets $B_1, \ldots, B_k$ that are bounded away from 0 and $P(N(\partial B_j) > 0) = 0$, $j = 1, \ldots, k$, we have

$$(N_n(B_1), \ldots, N_n(B_k)) \rightarrow_d (N(B_1), \ldots, N(B_k))$$

Point Processes — Application to Extremes

Application

Define $M_{n,2}$ to be the second largest among $X_1, \ldots, X_n$. Since the event

$$\{a_n^{-1} M_{n,2} \leq y\} \text{ is the same as } \{N_n (y,\infty) \leq 1\},$$

we conclude from the point process convergence that

$$P(a_n^{-1} M_{n,2} \leq y) = P(N_n(y,\infty) \leq 1)$$

$$\rightarrow P(N(y,\infty)) \leq 1)$$

$$= \exp\{-y^{-\alpha}\}(1 - \log(\exp\{-y^{-\alpha}\})).$$

Point Processes — Application to Extremes

Application

Similarly, the joint limiting distribution of $(M_n, M_{n,2})$ can be calculated by noting that for $y \leq x$,

$\{a_n^{-1} M_n \leq x, a_n^{-1} M_{n,2} \leq y\}$ is the same as $\{N_n (x, \infty) \leq 0, N_n (y,x] \leq 1\}$.

Hence,

$$P(a_n^{-1} M_n \leq x, a_n^{-1} M_{n,2} \leq y) = P(N_n (x, \infty) \leq 0, N_n(y,x] \leq 1)$$

$$\rightarrow P(N(x, \infty) \leq 0, N(y,x] \leq 1)$$

$$= \exp\{-y^{-\alpha}\}(1 + y^{-\alpha} - x^{-\alpha}).$$

Point Processes — Representation of Limit.

Representation of limit Poisson process

The points of the limit Poisson process can be displayed in an explicit fashion. Set

$$\Gamma_k = E_1 + \cdots + E_k,$$

where $E_1, E_2, \ldots$ are iid unit exponentials. Then the limit Poisson process N has the representation

$$N = \sum_{k=1}^{\infty} \varepsilon_{\Gamma_k^{-1/\alpha}}$$

and

$$N_n = \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t} \rightarrow_d N = \sum_{t=1}^{n} \varepsilon_{\Gamma_k^{-1/\alpha}}$$

If we order the data, then we can read off the weak convergence for the kth-largest $M_{n,k}$, i.e.,

$$a_n^{-1} M_{n,k} \rightarrow_d \Gamma_k^{-1/\alpha} \quad \text{(joint in k)}$$

Point Processes — Exceedances

Point Process of Exceedances.

Sometimes, it is convenient to consider the two-dimensional point process defined by

$$N_n = \sum_{t=1}^{\infty} \varepsilon_{(t/n, a_n^{-1} X_t)}.$$

For x > 0 fixed, the point process

$$N_n(\bullet \times (x, \infty)) = \sum_{t=1}^{\infty} \varepsilon_{(t/n, a_n^{-1} X_t)}(\bullet \times (x, \infty))$$

Is the point process of exceedances of the level $a_n x$. This point process converges in distribution to a homogeneous Poisson process rate with intensity $x^{-\alpha}$.

Bottom line: point process of exceedances is approximately Poisson.

Point Processes — An Application

Cool application

If $\alpha < 1$, then

$$N_n = \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t} \to_d N = \sum_{t=1}^{n} \varepsilon_{\Gamma_k^{-1/\alpha}}$$

implies, by an application of the continuous mapping theorem, that

$$a_n^{-1} \sum_{t=1}^{n} X_t \to_d \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha}.$$

The limit random variable is positive stable with index α.

Extension to Stationary Time Series

Let (X_t) is a strictly stationary sequence with common df $F \in D(G)$, i.e.,

$$F^n(a_n x + b_n) \to G(x).$$

Theorem If (X_t) satisfies a mixing condition (like strong mixing) and

$$P(\, a_n^{-1}(M_n - b_n) \le x) \to H(x),$$

H nondegenerate, then there exists a $\theta \in (0,1]$ such that

$$H(x) = G^\theta(x).$$

The parameter θ is called the **extremal index** and is a measure of extremal clustering.

Extension to Stationary Time Series—Extremal Index

$$F^n(a_n x + b_n) \to G(x) \quad P(\, a_n^{-1}(M_n - b_n) \le x) \to G^\theta(x).$$

Properties

- $\theta < 1$ implies clustering of exceedances
- Suppose c is a threshold such that $F^n(c) \sim .95$ and $\theta = .5$. Then
$$P(M_n \le c) \sim .95^{1/2} = .975$$
- $1/\theta$ is the mean cluster size of exceedances.
- In a certain sense, one can view θ as a measure of statistical efficiency relative to the iid case. That is, one needs $1/\theta$ more observations to match the behavior of the iid case. Specifically,
$$P(M_{n/\theta} \le x) \sim F^n(x)$$

Extension to Stationary Time Series—Example

Example (max-moving average) Let (Z_t) be iid with a Pareto distribution, i.e., $P(Z_1 > x) = x^{-\alpha}$ for $x \geq 1$, and set

$$X_t = \max(Z_t, \phi Z_{t-1}), \quad \phi \in [0,1].$$

Then

$$nP(X_1 > xn^{1/\alpha}) \to (1+\phi^\alpha)x^{-\alpha} \text{ and } F^n(a_n x) \to \exp(-(1+\phi^\alpha)x^{-\alpha}).$$

On the other hand

$$P(n^{-1/\alpha} M_n \leq x) = P(n^{-1/\alpha} \max(Z_0, \dots, Z_n) \leq x) \to \exp(-x^{-\alpha}).$$

Thus $\theta = 1/(1+\phi^\alpha)$.

Extension to Stationary Time Series—Example

iid (pareto $\alpha = 3$) $\qquad\qquad$ max-moving average ($\phi = 1$)

$\theta = 1$ $\qquad\qquad\qquad\qquad\qquad$ $\theta = 1/2$
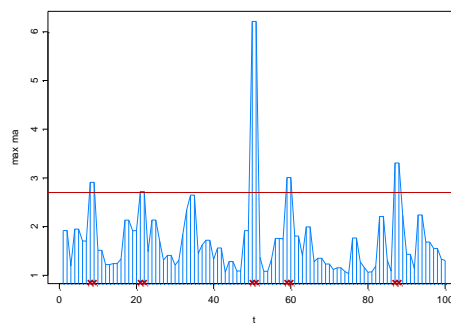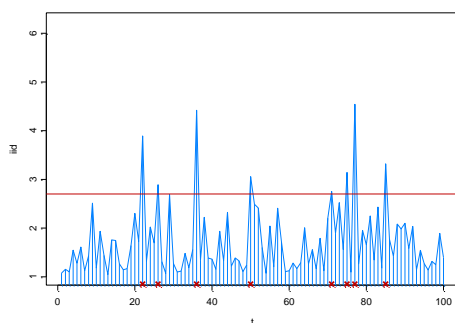
Note that cluster size is exactly 2 in this case.

Extension to Stationary Time Series—Mixing Conditions

Strong Mixing:

$$\sup_{A\in\sigma(X_s,s\leq 0),\,B\in\sigma(X_s,s\geq k)} |P(A\bigcap B)-P(A)P(B)|=\alpha_k \downarrow 0 \ \ \text{as}\ \ k\to\infty.$$

Remarks:
• Since mixing is defined via σ-fields, measurable functions of (X_t) inherit the same mixing property. For example, if the stationary sequence (X_t) is strongly mixing, so are $(|X_t|)$ and (X_t^2) with a rate function of similar order.
• If (α_k) decays to zero at an exponential rate, (X_t) is strongly mixing with geometric rate, i.e., the memory between past and future dies out exponentially fast.
• Strong mixing is much stronger than Leadbetter's dependence condition $D(u_n)$.

32

Extension to Stationary Time Series—Mixing Conditions

• If (α_k) decays to zero at an exponential rate, (X_t) is strongly mixing with geometric rate, i.e., the memory between past and future dies out exponentially fast.
• The rate function is closely related to the ACVF of the process:
if $E|X_1|^{2+\delta} < \infty$ for some $\delta > 0$, then

$$|Cov(X_0,X_h)|\leq c\ \alpha_k^{\delta/(2+\delta)}$$

• Many commonly used time series models, such as ARMA, GARCH, stochastic volatility processes, and Markov processes are strong mixing with a geometric rate.

33

16

Extension to Stationary Time Series—D'

Anti-clustering condition D'(u_n): Think of u_n as $a_n x + b_n$.

$$\limsup_{n\to\infty} n\sum_{t=2}^{n/k} P(X_1 > u_n, X_t > u_n) = O(1/k)$$

as $k \to \infty.$

Theorem: If (X_t) satisfies D and D', FϵD(G), then $\theta = 1$ (i.e., **no clustering)**.

Remarks:

• If (X_t) is iid, then the lim sup of the sum is

$\limsup_n n^2/k\ P^2(X_1 > u_n) = O(1/k).$

• If (X_t) is a stationary Gaussian process with ACF $\rho(h)=o(1/\log h)$, then D and D' hold and there is **no clustering** for Gaussian processes.

34

Extension to Stationary Time Series—Example

IID N(0,1/(1-.9²))   AR(1): $X_t = .9\ X_{t-1} + Z_t$, (Z_t)~IID N(0,1)
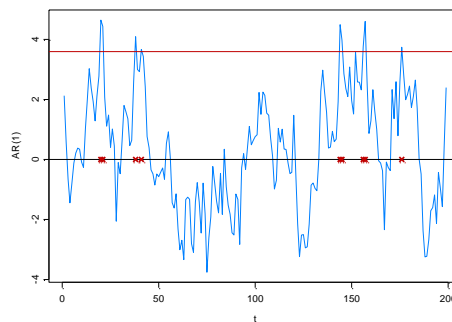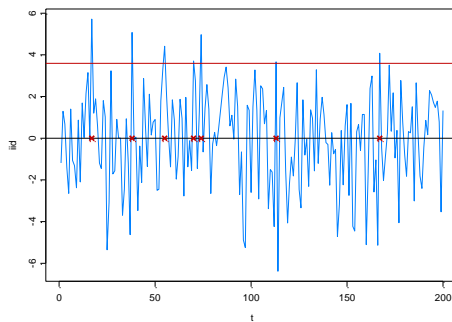
• Even though $\theta = 1$, there appears to be some clustering for small n.
• Hsing, Hüsler, Reiss (1996) overcome this problem for Gaussian processes by considering a triangular array or rvs.

35

17

Extension to Stationary Time Series—Example

Max-moving average: Let (Z_t) be iid with a Pareto distribution, i.e.,
$P(Z_1 > x) = x^{-\alpha}$ for $x \geq 1$, and set
$$X_t = \max(Z_\tau, \phi Z_{\tau-1}), \ \phi \in [0,1]$$

Then, since (X_t) is 1-dependent, the mixing condition D is automatically satisfied. As for D',

$$\limsup_{n\to\infty} n\sum_{t=2}^{n/k} P(X_1 > u_n, X_t > u_n)$$
$$= \lim_n nP(X_1 > u_n, X_2 > u_n) + O(1/k)$$
$$= \lim_n nP(\max(Z_1, \phi Z_0) > u_n, \max(Z_2, \phi Z_1) > u_n) + O(1/k)$$
$$= \lim_n nP(\phi Z_1 > u_n) + O(1/k)$$
$$= \phi^\alpha x^{-\alpha} + O(1/k)$$
$$= \left(\phi^\alpha/(1+\phi^\alpha)\right)(1+\phi^\alpha)x^{-\alpha} + O(1/k)$$

Strasbourg 6/08   36

Point Process Example—baby steps

In particular, for one-dependent sequences,
$$P(X_2 > x | X_1 > x) \to 1-\theta = \phi^\alpha/(1+\phi^\alpha).$$

Point process convergence (max-moving average): With $a_n = n^{1/\alpha}$
$$nP(Z_1 > a_n x) \to x^{-\alpha} \text{ and } nP(X_1 > a_n x) \to (1+\phi^\alpha)x^{-\alpha}$$

Define the sequence of point processes by
$$N_n^* = \sum_{t=1}^n \varepsilon_{a_n^{-1}(Z_t, Z_{t-1})}$$

From the convergence
$$\sum_{t=1}^n \varepsilon_{a_n^{-1}Z_t} \to_d \sum_{t=1}^n \varepsilon_{\Gamma_k^{-1/\alpha}}, \quad \Gamma_k = E_1 + \cdots + E_k,$$

one can show
$$N_n^* = \sum_{t=1}^n \varepsilon_{a_n^{-1}(Z_t, Z_{t-1})} \to_d \sum_{k=1}^\infty \left(\varepsilon_{(\Gamma_k^{-1/\alpha}, 0)} + \varepsilon_{(0, \Gamma_k^{-1/\alpha})}\right)$$

Strasbourg 6/08   37
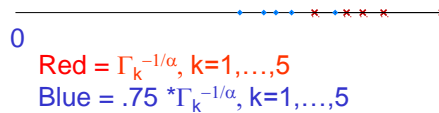
18

Point Process Example—baby steps

Applying the continuous mapping theorem (need to be careful),

$$N_n^* = \sum_{t=1}^{n} \varepsilon_{a_n^{-1}(Z_t, Z_{t-1})} \to_d \sum_{k=1}^{\infty} (\varepsilon_{(\Gamma_k^{-1/\alpha}, 0)} + \varepsilon_{(0, \Gamma_k^{-1/\alpha})})$$

we have

$$N_n = \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t} = \sum_{t=1}^{n} \varepsilon_{a_n^{-1} \max(Z_t, \phi Z_{t-1})}$$

$$\to_d \sum_{k=1}^{\infty} (\varepsilon_{\max(\Gamma_k^{-1\alpha}, 0)} + \varepsilon_{\max(0, \phi \Gamma_k^{-1\alpha})})$$

$$= \sum_{k=1}^{\infty} (\varepsilon_{\Gamma_k^{-1\alpha}} + \varepsilon_{\phi \Gamma_k^{-1\alpha}}) =: N$$

0

Red = $\Gamma_k^{-1/\alpha}$, k=1,…,5
Blue = .75 *$\Gamma_k^{-1/\alpha}$, k=1,…,5

38

Regular Variation — univariate case

<u>Def</u>: The random variable X is *regularly varying with index* α if

$$P(|X|> t\, x)/P(|X|>t) \to x^{-\alpha} \text{ and } P(X> t)/P(|X|>t) \to p,$$

or, equivalently, if

$$P(X> t\, x)/P(|X|>t) \to p x^{-\alpha} \text{ and } P(X< -t\, x)/P(|X|>t) \to q x^{-\alpha},$$

where $0 \le p \le 1$ and p+q=1.

<u>Equivalence:</u>

X is RV(-α) *if and only if* $P(X \in t \bullet) /P(|X|>t) \to_v \mu(\bullet)$

($\to_v$ vague convergence of measures on R\{0}). In this case,

$$\mu(dx) = \big(p\alpha\, x^{-\alpha-1}\, I(x>0) + q\alpha\, (-x)^{-\alpha-1}\, I(x<0)\big)\, dx$$

Note: $\mu(tA) = t^{\alpha}\, \mu(A)$ for every t and A bounded away from 0.

39

19

Regular Variation — univariate case

Another formulation (polar coordinates):

Define the ± 1 valued rv θ, $P(\theta = 1) = p$, $P(\theta = -1) = 1 - p = q$.

Then

X is RV(α) *if and only if*

$$\frac{P(|X| > t\,x, X/|X| \in S)}{P(|X| > t)} \to x^{-\alpha} P(\theta \in S)$$

or

$$\frac{P(|X| > t\,x, X/|X| \in \bullet)}{P(|X| > t)} \to_v x^{-\alpha} P(\theta \in \bullet)$$

($\to_v$ vague convergence of measures on $S^0 = \{-1,1\}$).

40

Regular Variation — multivariate case

Multivariate regular variation of $\mathbf{X} = (X_1, \ldots, X_m)$: There exists a random vector $\theta \in S^{m-1}$ such that

$$P(|\mathbf{X}| > t\,x, \mathbf{X}/|\mathbf{X}| \in \bullet)/P(|\mathbf{X}| > t) \to_v x^{-\alpha} P(\theta \in \bullet)$$

($\to_v$ vague convergence on S^{m-1}, unit sphere in R^m) .

• $P(\theta \in \bullet)$ is called the spectral measure

• α is the index of $\mathbf{X}$.

Equivalence:

$$\frac{P(\mathbf{X} \in t\bullet)}{P(|\mathbf{X}| > t)} \to_v \mu(\bullet)$$

μ is a measure on R^m which satisfies for $x > 0$ and A bounded away from 0,
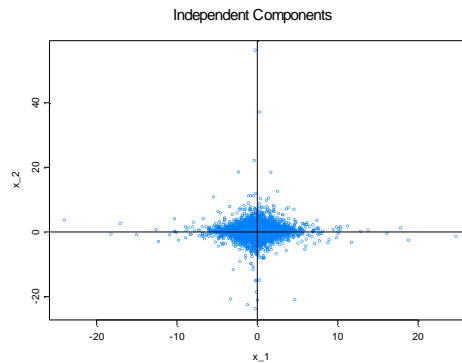
$$\mu(xB) = x^{-\alpha} \mu(xA).$$

41

20

Regular Variation — multivariate case

Examples: 1. If X_1 and X_2 are iid $RV(\alpha)$, then $\mathbf{X} = (X_1, X_2)$ is multivariate regularly varying with index α and *spectral distribution* (assuming symmetry)

$$P(\theta = \pi k/2) = \tfrac{1}{4} \quad k=1,2,3,4 \text{ (mass on axes).}$$

Interpretation: Unlikely that X_1 and X_2 are very large at the same time.

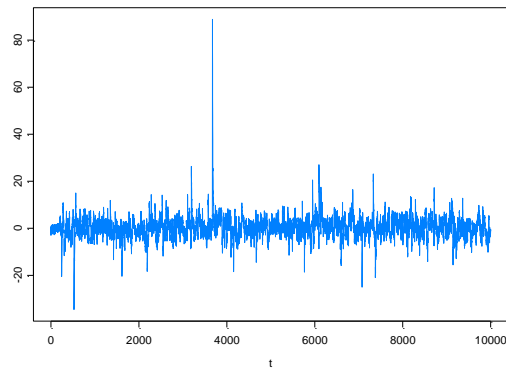Figure: plot of (X_{t1}, X_{t2}) for realization of 10,000.

Independent Components

42

2. If $X_1 = X_2 > 0$, then $\mathbf{X} = (X_1, X_2)$ is multivariate regularly varying with index α and *spectral distribution*

$$P(\theta = \pi/2) = 1.$$

3. AR(1): $X_t = .9 \, X_{t-1} + Z_t$, $\{Z_t\} \sim$ IID t(3)

$$P(\theta = \pm \arctan(.9)) = .9898 \quad P(\theta = \pm \pi/2) = .0102$$

43

21

Figure: plot of (X_t, X_{t+1}) for realization of 10,000.

$$X_t = .9\ X_{t-1} + Z_t$$

AR(1), X_{t+1} vs. X_t

44
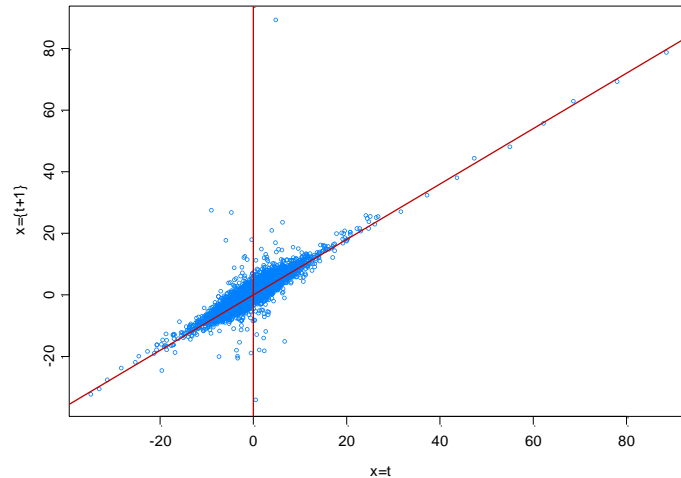
Estimation of α and θ

The marginal distribution *F* for heavy-tailed data is often modeled using *Pareto-like tails*,

$$1 - F(x) = x^{-\alpha} L(x),$$

for *x* large, where $L(x)$ is a slowly varying function ($L(xt)/L(x) \to 1$, as $x \to 1$). Now if

$$X \sim F, \text{ then } P(\log X > x) = P(X > \exp(x)) \sim \exp(-\alpha x),$$

and hence $\log X$ has an approximate exponential distribution for large *x*. The spacings,

$$\log(X_{(n-j)}) - \log(X_{(n-j-1)}),\ j=0,1,2,\ldots,m,$$

from a sample of size *n* from an exponential distr are approximately independent and *Exp*(α(j+1)) distributed. This suggests estimating α^{-1} by

$$\hat{\alpha}^{-1} = \frac{1}{m}\sum_{j=0}^{m-1}\left(\log X_{(n-j)} - \log X_{(n-j-1)}\right)(j+1)$$

$$= \frac{1}{m}\sum_{j=0}^{m-1}\left(\log X_{(n-j)} - \log X_{(n-m)}\right)$$

46

22

Hill's estimate of α

<u>Def:</u> The *Hill estimate* of α for heavy-tailed data with distribution given by

$$1 - F(x) = x^{-\alpha}L(x),$$

is

$$\hat{\alpha}^{-1} = \frac{1}{m}\sum_{j=0}^{m-1}\left(\log X_{(n-j)} - \log X_{(n-j-1)}\right)(j+1)$$

$$= \frac{1}{m}\sum_{j=0}^{m-1}\left(\log X_{(n-j)} - \log X_{(n-m)}\right)$$

The *asymptotic variance* of this estimate for α is
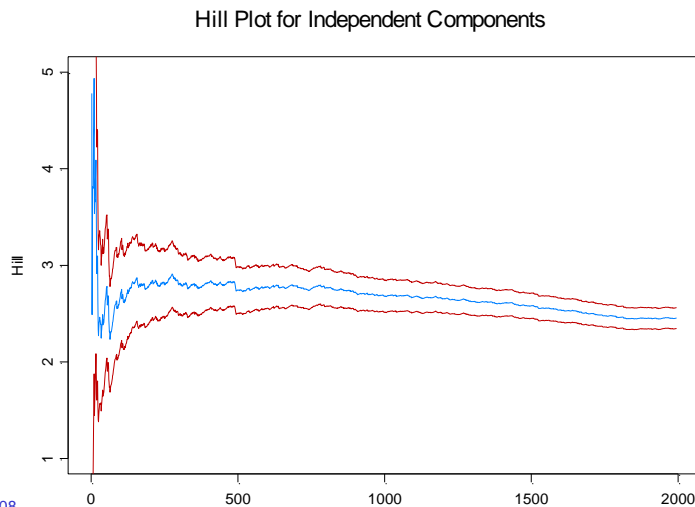
$$\alpha^2 / m$$ and estimated by $\hat{\alpha}^2 / m.$

(See also GPD=generalized Pareto distribution.)

Strasbourg 6/08
47

Hill's estimate of α

For a bivariate series, we will estimate α for the univariate series using the Euclidean norm of the two components.

Hill Plot for Independent Components

Strasbourg 6/08
48

23

Hill's estimate of α

Hill Plot for AR(1)
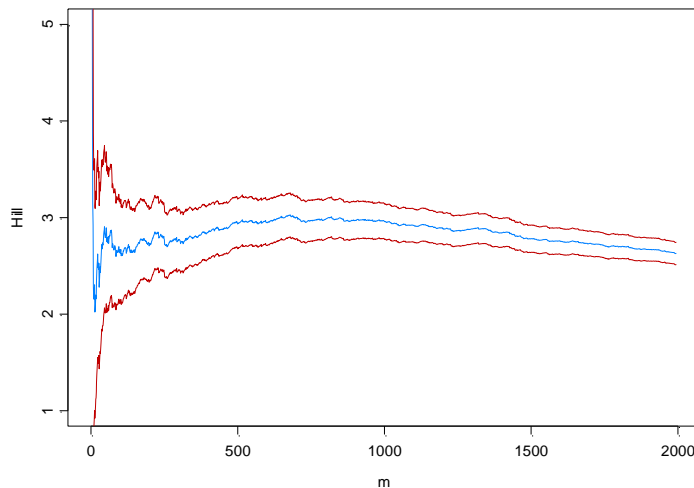
Estimation of the spectral distribution of θ

Based on the relation

$$P(|\mathbf{X}|> t\, x,\ \mathbf{X}/|\mathbf{X}| \in \bullet\ )/P(|\mathbf{X}|>t) \to_v x^{-\alpha} P(\ \theta \in \bullet\ )$$

a naïve estimate of the distribution of θ is based on the angular components $\mathbf{X}_t/|\mathbf{X}_t|$ in the sample. One simply uses the empirical distribution of these angular pieces for which the modulus $|\mathbf{X}_t|$ exceeds some large threshold. In the examples given below, we use a kernel density estimate of these angular components for those observations whose moduli exceed some large threshold. Here we only consider two components, i.e., θ is one dimensional.

Estimation of the spectral distribution of θ

Independent Components

Independent Components

Estimation of θ

AR(1), X_{t+1} vs. X_t

AR(1)
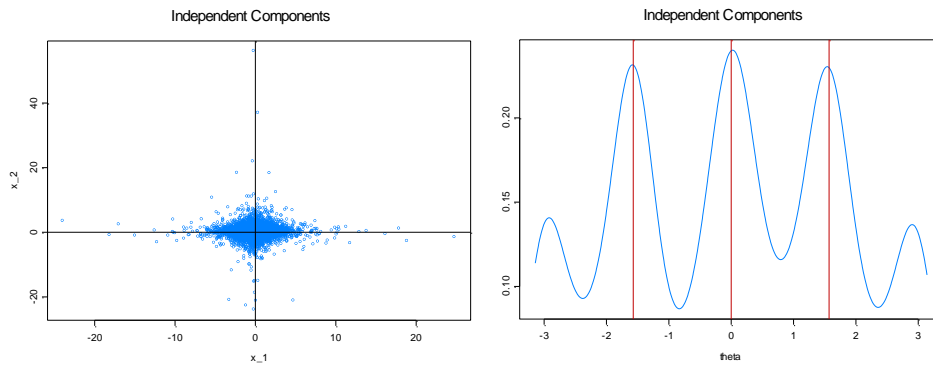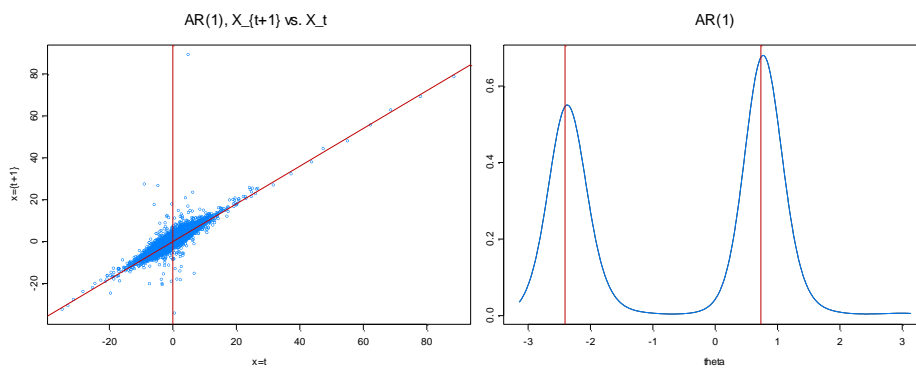
Vertical lines on right are at arctan(.9) and arctan(.9) -π

Examples of Processes that are Regular Varying

ARCH(1): $X_t = (\alpha_0 + \alpha_1 X^2_{t-1})^{1/2} Z_t$, $\{Z_t\} \sim$ IID.

α found by solving $E|\alpha_1 Z^2|^{\alpha/2} = 1$.

α_1	.312	.577	1.00	1.57
α	8.00	4.00	2.00	1.00

Distr of θ:

$$P(\theta \in \bullet) = E\{||(B,Z)||^\alpha \, I(\arg((B,Z)) \in \bullet)\} / E||(B,Z)||^\alpha$$

where

$$P(B = 1) = P(B = -1) = .5$$

53

Examples of Processes that are Regular Varying

Example of ARCH(1): $\alpha_0 = 1$, $\alpha_1 = 1$, $\alpha = 2$, $X_t = (\alpha_0 + \alpha_1 X^2_{t-1})^{1/2} Z_t$, $\{Z_t\} \sim$ IID

<u>Figures:</u> plots of (X_t, X_{t+1}) and estimated distribution of θ for realization of 10,000.

54

26
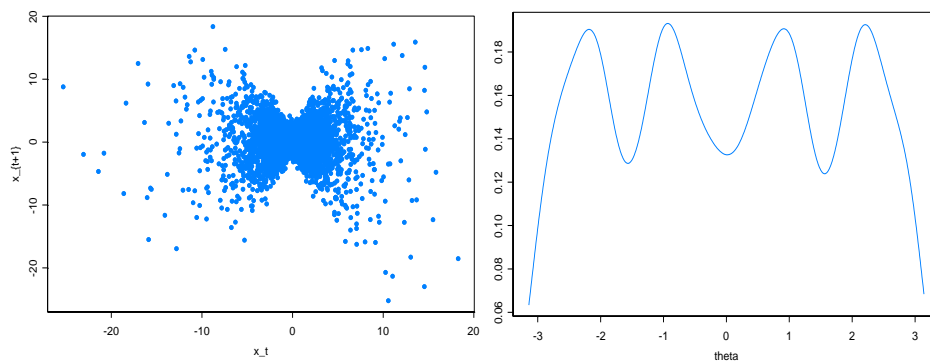
Examples of Processes that are Regular Varying

Example: SV model $X_t = \sigma_t Z_t$

Suppose $Z_t \sim RV(\alpha)$ and
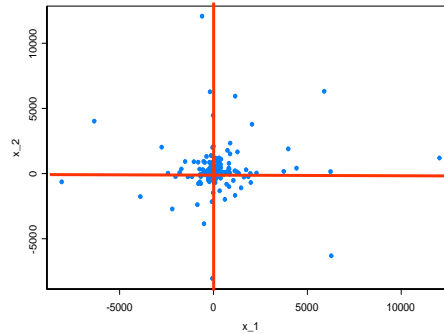
$$\log \sigma_t^2 = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \ \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty, \{\varepsilon_t\} \sim \text{IID N}(0, \sigma^2).$$

Then $\mathbf{Z}_n = (Z_1,\ldots,Z_n)'$ is regulary varying with index α and so is

$$\mathbf{X}_n = (X_1,\ldots,X_n)' = \text{diag}(\sigma_1,\ldots,\sigma_n)\, \mathbf{Z}_n$$

with spectral distribution concentrated on $(\pm 1, 0)$, $(0, \pm 1)$.

Figure: plot of (X_t, X_{t+1}) for realization of 10,000.

55

Point process Convergence

Theorem (Davis & Hsing `95, Davis & Mikosch `97). Let $\{\mathbf{X}_t\}$ be a stationary sequence of random m-vectors. Suppose

(i) finite dimensional distributions are jointly regularly varying (let $(\boldsymbol{\theta}_{-k}, \ldots, \boldsymbol{\theta}_k)$ be the vector in $S^{(2k+1)m-1}$ in the definition).

(ii) mixing condition $\mathcal{A}(a_n)$ or strong mixing.

(iii) $\lim_{k\to\infty} \limsup_{n\to\infty} P(\bigvee_{k \le |t| \le r_n} |\mathbf{X}_t| > a_n y \mid |\mathbf{X}_0| > a_n y) = 0.$

Then

$$\theta = \lim_{k\to\infty} E(|\theta_0^{(k)}|^{\alpha} - \bigvee_{j=1}^{k} |\theta_j^{(k)}|)_+ / E|\theta_0^{(k)}|^{\alpha} \qquad \text{(extremal index)}$$

exists. If $\theta > 0$, then

$$N_n := \sum_{t=1}^{n} \varepsilon_{\mathbf{X}_t / a_n} \xrightarrow{d} N := \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varepsilon_{P_i \mathbf{Q}_{ij}},$$

56

27

Point process convergence(cont)

- (P_i) are points of a Poisson process on $(0,\infty)$ with intensity function
$$\nu(dy)=\theta\alpha y^{-\alpha-1}dy.$$

- $\sum_{j=1}^{\infty}\varepsilon_{\mathbf{Q}_{ij}}$, $i \geq 1$, are iid point process with distribution Q, and Q is the weak limit of

$$\lim_{k\to\infty} E(|\theta_0^{(k)}|^\alpha - \bigvee_{j=1}^{k}|\theta_j^{(k)}|)_+ \, I_{\bullet}(\sum_{|t|\leq k}\varepsilon_{\theta_t^{(k)}})/E(|\theta_0^{(k)}|^\alpha - \bigvee_{j=1}^{k}|\theta_j^{(k)}|)_+$$

Remarks:

1. GARCH and SV processes satisfy the conditions of the theorem.

2. Limit distribution for sample extremes and sample ACF follows from this theorem.

57

Application to GARCH and SV Models

Setup

- $X_t = \sigma_t Z_t$, $\{Z_t\} \sim$ IID $(0,1)$
- X_t is RV (α)
- Choose $\{a_n\}$ s.t. $nP(X_t > a_n) \to 1$

Then
$$P^n(a_n^{-1}X_1 \leq x) \to \exp\{-x^{-\alpha}\}.$$

Then, with $M_n= \max\{X_1, \dots, X_n\}$,

(i) GARCH:
$$P(a_n^{-1}M_n \leq x) \to \exp\{-\theta x^{-\alpha}\},$$
θ is extremal index ($0 < \theta < 1$).

(ii) SV model:
$$P(a_n^{-1}M_n \leq x) \to \exp\{-x^{-\alpha}\},$$
extremal index $\theta = 1$ no clustering.

58

Application to Linear Processes

Suppose (X_t) is the linear process

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \qquad (Z_t) \sim \text{IID RV}(\alpha)$$

If a_n is chosen such that $nP(|Z_t| > a_n) \to 1$, then

$$nP(|X_t| > a_n x) = \sum_{j=-\infty}^{\infty} |\psi_j|^\alpha x^{-\alpha}$$

and

$$N_n := \sum_{t=1}^{n} \varepsilon_{X_t/a_n} \xrightarrow{\ d\ } N := \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} \varepsilon_{\Gamma_k^{-\alpha} \psi_j},$$

That is, each Poisson point $\Gamma_k^{-1/\alpha}$ gives rise to a cluster of points (deterministic) given by the coefficients of the filter, i.e., $\psi_j \Gamma_k^{-1/\alpha}$.

Extremal index and other extremal properties can be determined from this limit result.

Strasbourg 6/08

59

Application to Crystal River

River flow rate for Crystal River located in the mountain of Western Colorado (see Cooley et al. (2007)). After deasonalizing the data, we obtain 728 weekly observations from Oct 1, 1990 to Oct 1, 2005.
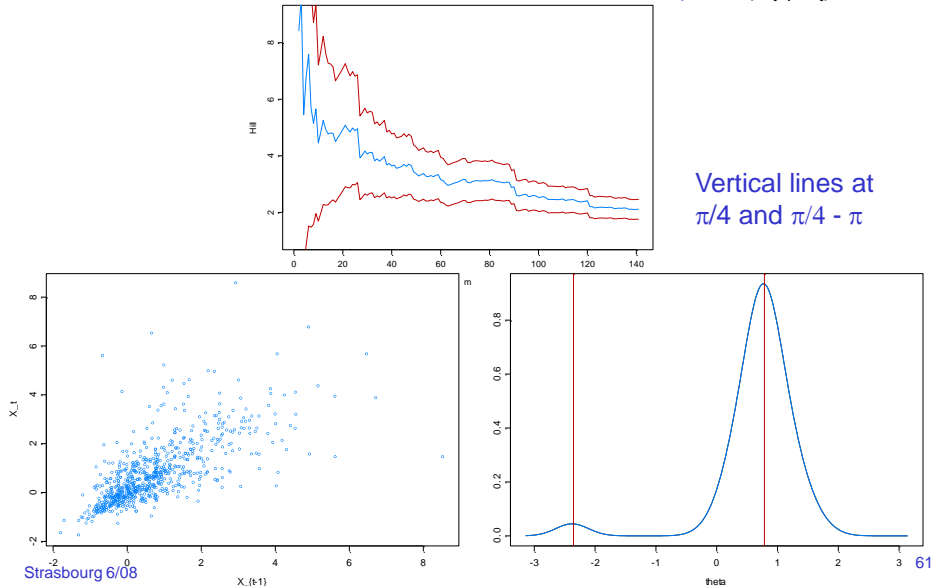

Strasbourg 6/08

60

29
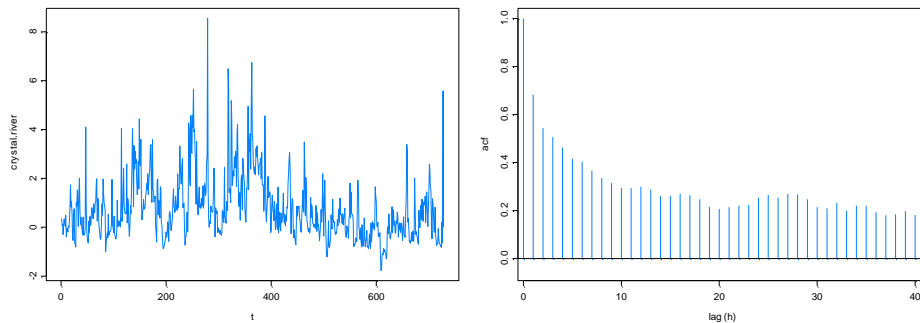
Application to Crystal River

Estimates of α and the distribution of θ for bivariate pairs (X_{t-1}, X_t)

Vertical lines at
$\pi/4$ and $\pi/4 - \pi$

Strasbourg 6/08

61

The Extremogram

The extremogram of a stationary time series (X_t) can be viewed as the analogue of the correlogram for measuring dependence in extremes (see Davis and Mikosch (2008)).

Definition: For two sets A & B bounded away from 0, the *extremogram* is defined as

$$\rho_{A,B}(h) = \lim_{n\to\infty} P(a_n^{-1}X_0 \in A, a_n^{-1}X_h \in B)/ P(a_n^{-1}X_0 \in A)$$

In many examples, this can be computed explicitly. If one takes A=B=(1,∞), then

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(X_h >x, \mid X_0 >x) = \lambda(X_0,X_h)$$

often called the *extremal dependence coefficient.*

The Extremogram

The extremogram of a stationary time series (X_t) can be viewed as the analogue of the correlogram for measuring dependence in extremes (see Davis and Mikosch (2008)).

Definition: For two sets A & B bounded away from 0, the *extremogram* is defined as

$$\rho_{A,B}(h) = \lim_{n\to\infty} P(a_n^{-1}X_0 \in A, a_n^{-1}X_h \in B)/ P(a_n^{-1}X_0 \in A)$$

In many examples, this can be computed explicitly. If one takes A=B=(1,∞), then

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(X_h > x, | X_0 > x) = \lambda(X_0, X_h)$$

often called the *extremal dependence coefficient* ($\lambda = 0$ means independence or asymptotic independence).

63

The Extremogram

The extremogram is estimated via the empirical extremogram defined by

$$\hat{\rho}_{A,B}(h) = \frac{\dfrac{m}{n}\sum_{t=1}^{n-h} I_{\{a_m^{-1}X_t \in A, a_m^{-1}X_{t+h} \in B\}}}{\dfrac{m}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}X_t \in A\}}}$$

where m→∞ with m/n →0. Note that the limit of the expectation of the numerator is

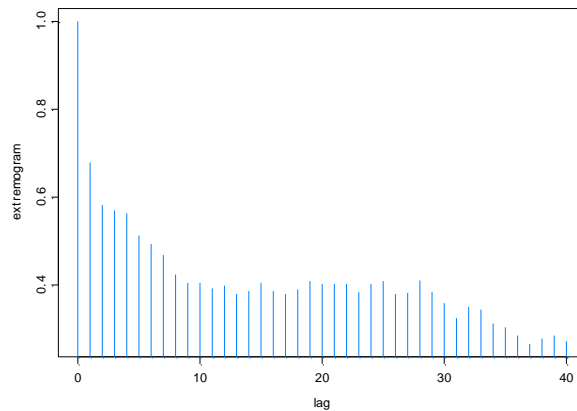$$mP(a_m^{-1}X_0 \in A, a_m^{-1}X_h \in B) \to \mu(A\times B),$$

where μ is the measure defined in the statement of regular variation. Hence the empirical estimate is asymptotically unbiased. Under suitable mixing conditions, a CLT for the empirical estimate is established in &M (2008).

64

Application to Crystal River
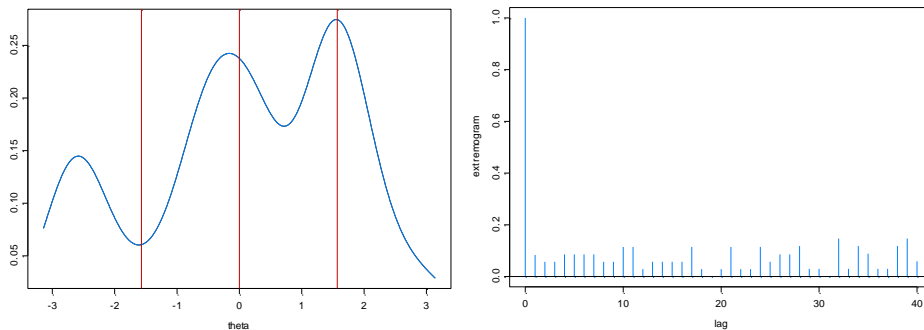
Extremogram for Crystal River $A = B = (1, \infty)$

Application to Crystal River

Fit an AR(6) model to the data (remove all appreciable autocorrelation in the data).   Now we estimate the distribution of θ and the extremogram based on the residuals.
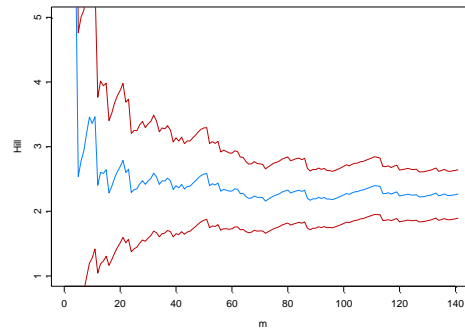

Vertical lines at $-\pi/2$, 0, and $\pi/2$

Application to Crystal River

There is still a touch of autocorrelation in the absolute values and squares of the residuals. We remove these by fitting a GARCH model to these residuals. The degrees of freedom for the noise was 3.43

Bonus Example

Wind Speed at Kilkenny 1/1/61-1/17/78
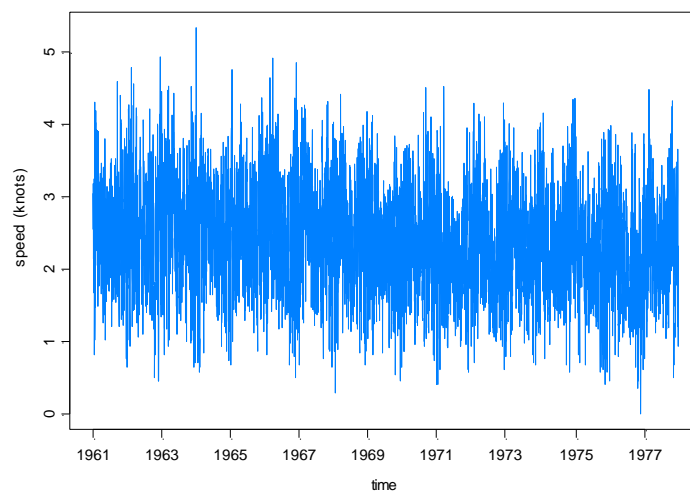
Bonus Example
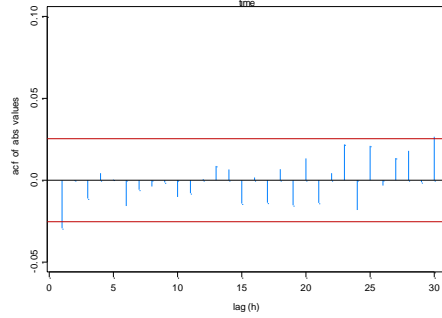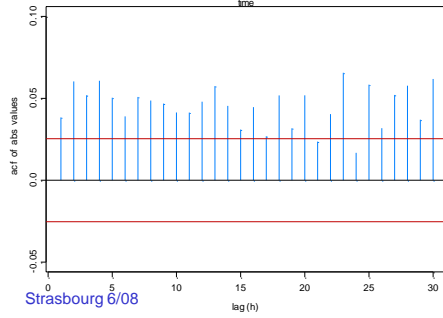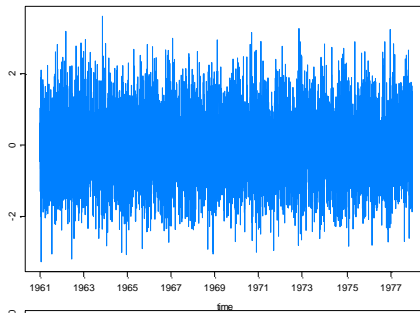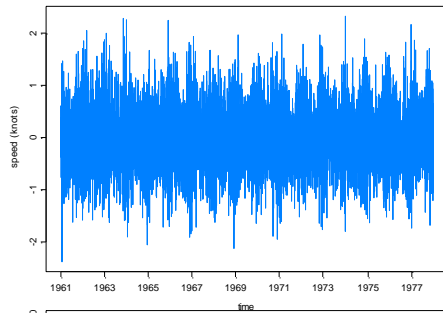
Wind speed at Kilkenny adjusted

Wind speed at Kilkenny Adjusted for Volatility

Strasbourg 6/08

69

34