A new shape of extremal clusters for certain stationary semi-exponential processes with moderate long range dependence

with Z. Chen

Extremal clustering and extremal index

Extremal clustering and extremal index

- Extremes of an i.i.d. sequence do not cluster!

Extremal clustering and extremal index

- Extremes of an i.i.d. sequence do not cluster!
- $Y_1, Y_2, \dots$ i.i.d., common distribution H,
  $M_n^{(0)} = \max(Y_1, \dots, Y_n)$.

Extremal clustering and extremal index

- Extremes of an i.i.d. sequence do not cluster!
- $Y_1, Y_2, \dots$ i.i.d., common distribution H,
  $M_n^{(0)} = \max(Y_1, \dots, Y_n)$.
- H is in a maximum domain of attraction if there are (a_n), (b_n) such that $(M_n^{(0)} - b_n)/a_n \Rightarrow G$, some nondegenerate G.

- Automatically $G(x) = G_\gamma(Ax + B)$, $x \in \mathbb{R}$ for some $\gamma \in \mathbb{R}$ and $A > 0, B \in \mathbb{R}$.

- Automatically $G(x) = G_\gamma(Ax + B)$, $x \in \mathbb{R}$ for some $\gamma \in \mathbb{R}$ and $A > 0, B \in \mathbb{R}$.
- The "standard" distributions G_γ:
  1. Fréchet $G_\gamma(x) = \exp\{-x^{-1/\gamma}\}$, $x \geq 0$ if $\gamma > 0$,
  2. Gumbel $G_0(x) = \exp\{-e^{-x}\}$, $x \in \mathbb{R}$,
  3. Weibull $G_\gamma(x) = \exp\{-(-x)^{-1/\gamma}\}$, $x \leq 0$ if $\gamma < 0$.

- If $X_1, X_2, \ldots$ is a stationary sequence with a marginal distribution H, its extreme values may cluster.

- If $X_1, X_2, \ldots$ is a stationary sequence with a marginal distribution H, its extreme values may cluster.
- A common numerical measure of clustering: the extremal index (Davis (1979), Leadbetter (1983)).

- If $X_1, X_2, \ldots$ is a stationary sequence with a marginal distribution H, its extreme values may cluster.
- A common numerical measure of clustering: the extremal index (Davis (1979), Leadbetter (1983)).
- $M_n = \max(X_1, \ldots, X_n)$, $M_n^{(0)} = \max(Y_1, \ldots, Y_n)$.

- If $X_1, X_2, \ldots$ is a stationary sequence with a marginal distribution H, its extreme values may cluster.

- A common numerical measure of clustering: the extremal index (Davis (1979), Leadbetter (1983)).

- $M_n = \max(X_1, \ldots, X_n)$, $M_n^{(0)} = \max(Y_1, \ldots, Y_n)$.

  Y_1, $Y_2, \ldots$ i.i.d. with the same marginal distribution H.

The stationary sequence $X_1, X_2, \ldots$ has extremal index θ if for some nondegenerate G both

$$(M^{(0)}_{[n\theta]} - b_n)/a_n \Rightarrow G^\theta \quad \text{and} \quad (M_n - b_n)/a_n \Rightarrow G^\theta,$$

The stationary sequence $X_1, X_2, \ldots$ has extremal index θ if for some nondegenerate G both

$$(M^{(0)}_{[n\theta]} - b_n)/a_n \Rightarrow G^\theta \quad \text{and} \quad (M_n - b_n)/a_n \Rightarrow G^\theta,$$

- M_n is "similar" to $M^{(0)}_{[n\theta]}$.

The stationary sequence $X_1, X_2, \ldots$ has extremal index θ if for some nondegenerate G both

$$(M^{(0)}_{[n\theta]} - b_n)/a_n \Rightarrow G^\theta \text{ and } (M_n - b_n)/a_n \Rightarrow G^\theta,$$

- M_n is "similar" to $M^{(0)}_{[n\theta]}$.
- An extremal index, if it exists, is in the range $0 < \theta \leq 1$.

The stationary sequence $X_1, X_2, \ldots$ has extremal index θ if for some nondegenerate G both

$$(M^{(0)}_{[n\theta]} - b_n)/a_n \Rightarrow G^\theta \;\text{ and }\; (M_n - b_n)/a_n \Rightarrow G^\theta,$$

- M_n is "similar" to $M^{(0)}_{[n\theta]}$.
- An extremal index, if it exists, is in the range $0 < \theta \leq 1$.
- The extremes cluster together, the average cluster size $1/\theta$.

- The situation becomes complicated when the stationary sequence has long memory with respect to its extremes.

- The situation becomes complicated when the stationary sequence has long memory with respect to its extremes.
- The average cluster size becomes infinite!

- The situation becomes complicated when the stationary sequence has long memory with respect to its extremes.
- The average cluster size becomes infinite!
- New normalization and centering are needed.

- The situation becomes complicated when the stationary sequence has long memory with respect to its extremes.
- The average cluster size becomes infinite!
- New normalization and centering are needed.
- What do extremal clusters look like?

Two ways to look at the shape of extremal clusters

Two ways to look at the shape of extremal clusters

1 *Point process convergence*

Two ways to look at the shape of extremal clusters

1 *Point process convergence* Let

$$M_n^{\mathrm{pp}} = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}.$$

Two ways to look at the shape of extremal clusters

1 *Point process convergence* Let

$$M_n^{\mathrm{pp}} = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}.$$

Under some conditions (if the memory is not too long)

$$M_n^{\mathrm{pp}} \Rightarrow M_\infty^{\mathrm{pp}} \ \text{ weakly, in the vague topology,}$$

Two ways to look at the shape of extremal clusters

1 *Point process convergence* Let

$$M_n^{\mathrm{pp}} = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}.$$

Under some conditions (if the memory is not too long)

$$M_n^{\mathrm{pp}} \Rightarrow M_\infty^{\mathrm{pp}} \text{ weakly, in the vague topology,}$$

$$M_\infty^{\mathrm{pp}} = \sum_{j=1}^{\infty} \sum_{x \in S_j(\Gamma_j)} \delta_x$$

Two ways to look at the shape of extremal clusters

1 *Point process convergence* Let

$$M_n^{\mathrm{pp}} = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}.$$

Under some conditions (if the memory is not too long)

$$M_n^{\mathrm{pp}} \Rightarrow M_\infty^{\mathrm{pp}} \text{ weakly, in the vague topology,}$$

$$M_\infty^{\mathrm{pp}} = \sum_{j=1}^{\infty} \sum_{x \in S_j(\Gamma_j)} \delta_x$$

(Γ_j): standard Poisson arrivals, $(S_j)(\cdot)$: (random) extremal clusters.

2 *Sup measures convergence*

2 *Sup measures convergence* For open set $G \in [0,1]$ let

$$M_n^{\mathrm{sm}}(G) = \sup_{i/n \in G} X_i.$$

2 *Sup measures convergence* For open set $G \in [0,1]$ let

$$M_n^{\mathrm{sm}}(G) = \sup_{i/n \in G} X_i.$$

Under some conditions

$$\left(M_n^{\mathrm{sm}} - b_n^{\mathrm{sm}}\right)/a_n^{\mathrm{sm}} \Rightarrow M_\infty^{\mathrm{sm}} \text{ weakly, in the sup vague topology,}$$

2 *Sup measures convergence* For open set $G \in [0,1]$ let

$$M_n^{\mathrm{sm}}(G) = \sup_{i/n \in G} X_i.$$

Under some conditions

$$\big(M_n^{\mathrm{sm}} - b_n^{\mathrm{sm}}\big)/a_n^{\mathrm{sm}} \Rightarrow M_\infty^{\mathrm{sm}} \ \text{ weakly, in the sup vague topology,}$$

$$M_\infty^{\mathrm{sm}}(G) = \sup_{t \in G} \sum_j D(t, \Gamma_j) \mathbf{1}\big(t \in S_j\big).$$

$D(t,\cdot)$: random function, (Γ_j): Poisson arrivals, (S_j): random sets.

Stationary infinitely divisible processes with long memory

$$X_n = \int_E f \circ \theta^n(x)\, M(dx),\ \ n \in \mathbb{Z}\,.$$

Stationary infinitely divisible processes with long memory

$$X_n = \int_E f \circ \theta^n(x)\, M(dx),\ \ n \in \mathbb{Z}\,.$$

- θ is the left shift operator on $E = \mathbb{Z}^{\mathbb{Z}}$;

Stationary infinitely divisible processes with long memory

$$X_n = \int_E f \circ \theta^n(x)\, M(dx),\ \ n \in \mathbb{Z}\,.$$

- θ is the left shift operator on $E = \mathbb{Z}^{\mathbb{Z}}$;
- M is an infinitely divisible random measure on $(E, \mathcal{E})$;

Stationary infinitely divisible processes with long memory

$$X_n = \int_E f \circ \theta^n(x)\, M(dx), \;\; n \in \mathbb{Z}\,.$$

- θ is the left shift operator on $E = \mathbb{Z}^{\mathbb{Z}}$;
- M is an infinitely divisible random measure on $(E, \mathcal{E})$;
- The marginal Lévy measure ν has a subexponential right tail, and so is marginal tail of X.

- The control measure of M has the form

$$\mu(\cdot) := \sum_{i \in \mathbb{Z}} \pi_i P_i(\cdot)\,,$$

- The control measure of M has the form

$$\mu(\cdot) := \sum_{i \in \mathbb{Z}} \pi_i P_i(\cdot)\,,$$

- (π_i): the invariant measure of an irreducible, aperiodic, and null recurrent Markov chain $\{Y_t\}_{t\in\mathbb{Z}}$ on $\mathbb{Z}$;

- The control measure of M has the form

$$\mu(\cdot) := \sum_{i \in \mathbb{Z}} \pi_i P_i(\cdot)\,,$$

- (π_i): the invariant measure of an irreducible, aperiodic, and null recurrent Markov chain $\{Y_t\}_{t\in\mathbb{Z}}$ on $\mathbb{Z}$;
- P_i is the probability law of $\{Y_t\}_{t\in\mathbb{Z}}$ on $(E, \mathcal{E})$ given $Y_0 = i$.

- The control measure of M has the form

$$\mu(\cdot) := \sum_{i \in \mathbb{Z}} \pi_i P_i(\cdot)\,,$$

- (π_i): the invariant measure of an irreducible, aperiodic, and null recurrent Markov chain $\{Y_t\}_{t \in \mathbb{Z}}$ on $\mathbb{Z}$;

- P_i is the probability law of $\{Y_t\}_{t \in \mathbb{Z}}$ on $(E, \mathcal{E})$ given $Y_0 = i$.

- The key assumption:

$$P_0\Big(\inf\{n \geq 1 : Y_n = 0\} > m\Big) \in \mathrm{RV}_{-\beta},\ 0 < \beta < 1.$$

- The normalization and centering in the extremal limit theorems depend on the marginal tails and on the memory encoded in β.

- The normalization and centering in the extremal limit theorems depend on the marginal tails and on the memory encoded in β.
- So do the shape of the extremal clusters S_j and the mass distribtion $D(t, \Gamma_j)$ over S_j.

- The normalization and centering in the extremal limit theorems depend on the marginal tails and on the memory encoded in β.
- So do the shape of the extremal clusters S_j and the mass distribtion $D(t, \Gamma_j)$ over S_j.
- The lighter is the tails and the longer is the memory, the more intricate is the picture.

- The normalization and centering in the extremal limit theorems depend on the marginal tails and on the memory encoded in β.
- So do the shape of the extremal clusters S_j and the mass distribtion $D(t, \Gamma_j)$ over S_j.
- The lighter is the tails and the longer is the memory, the more intricate is the picture.
- Assume **moderately long memory**: $0 < \beta < 1/2$.

We assume that the marginal tails are *subexponential*:

We assume that the marginal tails are *subexponential*:
A distribution H is subexponential if

$$\lim_{x\to\infty} \frac{\overline{H * H}(x)}{\overline{H}(x)} = 2\,.$$

We assume that the marginal tails are *subexponential*:
A distribution H is subexponential if

$$\lim_{x \to \infty} \frac{\overline{H * H}(x)}{\overline{H}(x)} = 2\,.$$

Case 1: regularly varying tails $\nu\big((x,\infty)\big) \in \mathrm{RV}_{-\alpha}$, $\alpha > 0$.

We assume that the marginal tails are *subexponential*:
A distribution H is subexponential if

$$\lim_{x\to\infty} \frac{\overline{H * H}(x)}{\overline{H}(x)} = 2\,.$$

Case 1: regularly varying tails $\nu\big((x,\infty)\big) \in \mathrm{RV}_{-\alpha}$, $\alpha > 0$.

Case 2: lognormal-type tails

$$\nu\big((x,\infty)\big) \sim cx^{\beta}(\log x)^{\xi}\exp\left(-\lambda(\log x)^{\gamma}\right)$$

for some $\gamma > 1$, $\lambda, c > 0$ and $\beta, \xi \in \mathbb{R}$.

Case 3: super-lognormal-type tails

$$\nu\big((x,\infty)\big) \sim c x^{\beta} (\log x)^{\xi} \exp\left(\lambda (\log x)^{\gamma}\right) \exp\left(-\rho \exp\left(\mu (\log x)^{\alpha}\right)\right)$$

for some $\alpha \in (0,1)$, $c, \mu, \rho > 0$ and $\beta, \xi, \lambda, \gamma \in \mathbb{R}$.

Case 3: super-lognormal-type tails

$$\nu\big((x,\infty)\big) \sim cx^{\beta}(\log x)^{\xi}\exp\left(\lambda(\log x)^{\gamma}\right)\exp\left(-\rho\exp\left(\mu(\log x)^{\alpha}\right)\right)$$

for some $\alpha \in (0,1)$, $c, \mu, \rho > 0$ and $\beta, \xi, \lambda, \gamma \in \mathbb{R}$.

Case 4: semiexponential tails

$$\nu\big((x,\infty)\big) \sim \exp\left(-x^{-\alpha}L(x)\right)$$

for some $\alpha \in (0,1)$ and a slowly varying L.

Stable regenerative sets

Stable regenerative sets

- $\{L_\beta(t)\}_{t\geqslant 0}$: β-stable subordinator, $0 < \beta < 1$.

Stable regenerative sets

- $\{L_\beta(t)\}_{t \geqslant 0}$: β-stable subordinator, $0 < \beta < 1$.
- β- stable regenerative set: $\overline{\{L_\beta(t) : t \in \mathbb{R}_+\}}$.

Stable regenerative sets

- $\{L_\beta(t)\}_{t \geqslant 0}$: β-stable subordinator, $0 < \beta < 1$.
- β- stable regenerative set: $\overline{\{L_\beta(t) : t \in \mathbb{R}_+\}}$.
- Two independent β- stable regenerative sets intersect with probability 0 (if $0 < \beta \leq 1/2$) or 1 (if $1/2 < \beta < 1$).

In the case of regularly varying tails:

In the case of regularly varying tails:

- marginally, the Fréchet maximum domain of attraction;

In the case of regularly varying tails:

- marginally, the Fréchet maximum domain of attraction;
- Extremal clusters: iid, $S_j = R_j + U_j$;

In the case of regularly varying tails:

- marginally, the Fréchet maximum domain of attraction;
- Extremal clusters: iid, $S_j = R_j + U_j$;
  (R_j) β- stable regenerative sets,
  $P(U_j \leq x) = x^{1-\beta}$, $0 \leq x \leq 1$.

In the case of regularly varying tails:

- marginally, the Fréchet maximum domain of attraction;
- Extremal clusters: iid, $S_j = R_j + U_j$;
  (R_j) β- stable regenerative sets,
  $P(U_j \le x) = x^{1-\beta}$, $0 \le x \le 1$.
- Spread of the mass: $D(t, \Gamma_j) = \Gamma_j^{-1/\alpha} \mathbf{1}(t \in S_j)$.

In the case of lognormal-type and super-lognormal-type tails:

In the case of lognormal-type and super-lognormal-type tails:

- marginally, the Gumbel maximum domain of attraction;

In the case of lognormal-type and super-lognormal-type tails:

- marginally, the Gumbel maximum domain of attraction;
- Extremal clusters: the shifted stable regenerative sets.

In the case of lognormal-type and super-lognormal-type tails:

- marginally, the Gumbel maximum domain of attraction;
- Extremal clusters: the shifted stable regenerative sets.
- Spread of the mass: $D(t, \Gamma_j) = -\log \Gamma_j \, \mathbf{1}(t \in S_j)$.

In the case of semiexponential tails:

In the case of semiexponential tails:

- marginally, the Gumbel maximum domain of attraction;

In the case of semiexponential tails:

- marginally, the Gumbel maximum domain of attraction;
- Extremal clusters: the shifted stable regenerative sets (S_j).

In the case of semiexponential tails:

- marginally, the Gumbel maximum domain of attraction;
- Extremal clusters: the shifted stable regenerative sets (S_j).
- Spread of the mass: on each S_j take an iid sample $(W_{k,j})$ from the normalized Hausdorff measure;

In the case of semiexponential tails:

- marginally, the Gumbel maximum domain of attraction;
- Extremal clusters: the shifted stable regenerative sets (S_j).
- Spread of the mass: on each S_j take an iid sample $(W_{k,j})$ from the normalized Hausdorff measure; $(\Gamma_{k,j})$: Poisson arrivals.

In the case of semiexponential tails:

- marginally, the Gumbel maximum domain of attraction;
- Extremal clusters: the shifted stable regenerative sets (S_j).
- Spread of the mass: on each S_j take an iid sample $(W_{k,j})$ from the normalized Hausdorff measure; $(\Gamma_{k,j})$: Poisson arrivals.

$$D(W_{k,j}, \Gamma_j) = -\log \Gamma_j - \frac{1}{C_{\alpha,\beta}} \log \left(\frac{c_\beta \Gamma_{k,j}}{m_\beta(S_j)} \right)$$

In the case of semiexponential tails:

- marginally, the Gumbel maximum domain of attraction;
- Extremal clusters: the shifted stable regenerative sets (S_j).
- Spread of the mass: on each S_j take an iid sample $(W_{k,j})$ from the normalized Hausdorff measure; $(\Gamma_{k,j})$: Poisson arrivals.

$$D(W_{k,j},\Gamma_j) = -\log\Gamma_j - \frac{1}{C_{\alpha,\beta}}\log\left(\frac{c_\beta\Gamma_{k,j}}{m_\beta(S_j)}\right)$$

$C_{\alpha,\beta}, c_\beta$ constants, $m_\beta(S_j)$ the Hausdorff measure of S_j.