# Inference for the limiting cluster size distribution of extreme values

Christian Y. Robert

4th Conference on Extreme Value Analysis
Gothenburg, August 15-19, 2005

# The point process of exceedances - independent case

Let $(X_i)$ be an iid sequence of rvs with distribution $F$. The point process of time normalized exceedances $N_n^{(\tau)}(\cdot)$ is defined by

$$N_n^{(\tau)}(B) = \sum_{i=1}^{n} \mathbb{I}_{\{i/n \in B, X_i > u_n(\tau)\}},$$

for any Borel set $B \subset E := (0,1]$, where $(u_n(\tau))$ is a sequence of deterministic thresholds.

**Theorem**

*Let $(u_n(\tau))$ be such that $\lim_{n\to\infty} n\bar{F}(u_n(\tau)) = \tau$ where $\bar{F} := 1 - F$. Then $N_n^{(\tau)}$ converges weakly to a homogeneous Poisson point process $N$ on $(0,1]$ with intensity $\tau\,|\cdot|$, where $|\cdot|$ denotes the Lebesgue measure.*

If $X_{(k)}$ is the $k$-th largest of $X_1, ..., X_n$, then

$$\lim_{n\to\infty} \mathbb{P}\left(X_{(k)} < u_n(\tau)\right) = \lim_{n\to\infty} \mathbb{P}\left(N_n^{(\tau)}(E) < k\right) = e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!}.$$

# The point process of exceedances - independent case

Let $(X_i)$ be an iid sequence of rvs with distribution $F$. The point process of time normalized exceedances $N_n^{(\tau)}(\cdot)$ is defined by

$$N_n^{(\tau)}(B) = \sum_{i=1}^{n} \mathbb{I}_{\{i/n \in B, X_i > u_n(\tau)\}},$$

for any Borel set $B \subset E := (0, 1]$, where $(u_n(\tau))$ is a sequence of deterministic thresholds.

**Theorem**

*Let $(u_n(\tau))$ be such that $\lim_{n\to\infty} n\bar{F}(u_n(\tau)) = \tau$ where $\bar{F} := 1 - F$. Then $N_n^{(\tau)}$ converges weakly to a homogeneous Poisson point process $N$ on $(0, 1]$ with intensity $\tau\,|\cdot|$, where $|\cdot|$ denotes the Lebesgue measure.*

If $X_{(k)}$ is the $k$-th largest of $X_1, \ldots, X_n$, then

$$\lim_{n\to\infty} \mathbb{P}\left(X_{(k)} < u_n(\tau)\right) = \lim_{n\to\infty} \mathbb{P}\left(N_n^{(\tau)}(E) < k\right) = e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!}.$$

# The point process of exceedances - independent case

Let $(X_i)$ be an iid sequence of rvs with distribution $F$. The point process of time normalized exceedances $N_n^{(\tau)}(\cdot)$ is defined by

$$N_n^{(\tau)}(B) = \sum_{i=1}^{n} \mathbb{I}_{\{i/n \in B, X_i > u_n(\tau)\}},$$

for any Borel set $B \subset E := (0, 1]$, where $(u_n(\tau))$ is a sequence of deterministic thresholds.

## Theorem

*Let $(u_n(\tau))$ be such that $\lim_{n\to\infty} n\bar{F}(u_n(\tau)) = \tau$ where $\bar{F} := 1 - F$. Then $N_n^{(\tau)}$ converges weakly to a homogeneous Poisson point process $N$ on $(0, 1]$ with intensity $\tau\,|\cdot|$, where $|\cdot|$ denotes the Lebesgue measure.*

If $X_{(k)}$ is the $k$-th largest of $X_1, ..., X_n$, then

$$\lim_{n\to\infty} \mathbb{P}\left(X_{(k)} < u_n(\tau)\right) = \lim_{n\to\infty} \mathbb{P}\left(N_n^{(\tau)}(E) < k\right) = e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!}.$$

# The point process of exceedances - dependent case

Let $(X_n)$ be a strictly stationary sequence with distribution $F$.

$$\begin{aligned}\beta_{n,l}(\tau) \equiv & \sup \left|\mathbb{P}(X_i \le u_n(\tau), i \in A \cup B)\right. \\ & \left.-\mathbb{P}(X_i \le u_n(\tau), i \in A)\,\mathbb{P}(X_i \le u_n(\tau), i \in B)\right|,\end{aligned}$$

where $A \subset \{1, ..., k\}$, $B \subset \{k+l, ..., n\}$, and $1 \le k \le n-l$.

**Condition**

*$D(u_n(\tau))$ is satisfied if there exists a sequence $l_n = o(n)$ such that $l_n \to \infty$ and $\beta_{n,l_n}(\tau) \to 0$ when $n \to \infty$.*

**Definition**

*Suppose that $D(u_n(\tau))$ is satisfied, $\theta$ $(0 < \theta \le 1)$ is called the extremal index of the process $(X_n)$ if for each $\tau > 0$ :*
*(i) there exists $u_n(\tau)$ such that $n\bar{F}(u_n(\tau)) \to \tau$,*
*(ii) $P\left(N_n^{(\tau)}(E) = 0\right) \to \exp(-\theta\tau)$.*

# The point process of exceedances - dependent case

Let $(X_n)$ be a strictly stationary sequence with distribution $F$.

$$\begin{aligned}\beta_{n,l}(\tau) \equiv & \sup|\mathbb{P}(X_i \leq u_n(\tau), i \in A \cup B) \\ & -\mathbb{P}(X_i \leq u_n(\tau), i \in A)\mathbb{P}(X_i \leq u_n(\tau), i \in B)|,\end{aligned}$$

where $A \subset \{1, ..., k\}$, $B \subset \{k+l, ..., n\}$, and $1 \leq k \leq n-l$.

## Condition

$D(u_n(\tau))$ *is satisfied if there exists a sequence* $l_n = o(n)$ *such that* $l_n \to \infty$ *and* $\beta_{n,l_n}(\tau) \to 0$ *when* $n \to \infty$.

## Definition

*Suppose that* $D(u_n(\tau))$ *is satisfied,* $\theta$ $(0 < \theta \leq 1)$ *is called the extremal index of the process* $(X_n)$ *if for each* $\tau > 0$ *:*
*(i) there exists* $u_n(\tau)$ *such that* $n\bar{F}(u_n(\tau)) \to \tau$,
*(ii)* $P\left(N_n^{(\tau)}(E) = 0\right) \to \exp(-\theta\tau)$.

# The point process of exceedances - dependent case

Let $(X_n)$ be a strictly stationary sequence with distribution $F$.

$$\begin{aligned}\beta_{n,l}(\tau) \equiv & \sup\left|\mathbb{P}\left(X_i \leq u_n(\tau), i \in A \cup B\right)\right. \\ & \left.-\mathbb{P}\left(X_i \leq u_n(\tau), i \in A\right)\mathbb{P}\left(X_i \leq u_n(\tau), i \in B\right)\right|,\end{aligned}$$

where $A \subset \{1, ..., k\}$, $B \subset \{k+l, ..., n\}$, and $1 \leq k \leq n-l$.

## Condition

*$D(u_n(\tau))$ is satisfied if there exists a sequence $l_n = o(n)$ such that $l_n \to \infty$ and $\beta_{n,l_n}(\tau) \to 0$ when $n \to \infty$.*

## Definition

*Suppose that $D(u_n(\tau))$ is satisfied, $\theta$ $(0 < \theta \leq 1)$ is called the extremal index of the process $(X_n)$ if for each $\tau > 0$ :*
*(i) there exists $u_n(\tau)$ such that $n\bar{F}(u_n(\tau)) \to \tau$,*
*(ii) $P\left(N_n^{(\tau)}(E) = 0\right) \to \exp(-\theta\tau)$.*

# The point process of exceedances - dependent case

Let $\mathcal{F}_{p,q} = \mathcal{F}_{p,q}(\tau)$ be the $\sigma$-algebra generated by the events $\{X_i > u_n(\tau)\}$, $p \leq i \leq q$, and

$$\alpha_{n,l}(\tau) \equiv \sup |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) : A \in \mathcal{F}_{1,t}, B \in \mathcal{F}_{t+l,\infty}, t \geq 1| .$$

**Condition**

*$\Delta(u_n(\tau))$ is satisfied if there exists a sequence $l_n = o(n)$ such that $l_n \to \infty$ and $\alpha_{n,l_n}(\tau) \to 0$ when $n \to \infty$.*

**Theorem**

*Assume that $\Delta(u_n(\tau))$ is satisfied and $\lim_{n\to\infty} n\bar{F}(u_n(\tau)) = \tau$. If the limiting point process of $N_n^{(\tau)}$ exists, it is necessarily a homogeneous compound Poisson point process with intensity $\tau\theta|\cdot|$ and limiting cluster size distribution $\pi$ (Hsing et al. (1988)).*

# The point process of exceedances - dependent case

Let $\mathcal{F}_{p,q} = \mathcal{F}_{p,q}(\tau)$ be the $\sigma$-algebra generated by the events $\{X_i > u_n(\tau)\}$, $p \leq i \leq q$, and

$$\alpha_{n,l}(\tau) \equiv \sup |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) : A \in \mathcal{F}_{1,t}, B \in \mathcal{F}_{t+l,\infty}, t \geq 1| .$$

**Condition**

*$\Delta(u_n(\tau))$ is satisfied if there exists a sequence $l_n = o(n)$ such that $l_n \to \infty$ and $\alpha_{n,l_n}(\tau) \to 0$ when $n \to \infty$.*

**Theorem**

*Assume that $\Delta(u_n(\tau))$ is satisfied and $\lim_{n\to\infty} n\bar{F}(u_n(\tau)) = \tau$. If the limiting point process of $N_n^{(\tau)}$ exists, it is necessarily a homogeneous compound Poisson point process with intensity $\tau\theta|\cdot|$ and limiting cluster size distribution $\pi$ (Hsing et al. (1988)).*

# The point process of exceedances - dependent case

Let $\mathcal{F}_{p,q} = \mathcal{F}_{p,q}(\tau)$ be the $\sigma$-algebra generated by the events $\{X_i > u_n(\tau)\}$, $p \leq i \leq q$, and

$$\alpha_{n,l}(\tau) \equiv \sup \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) : A \in \mathcal{F}_{1,t}, B \in \mathcal{F}_{t+l,\infty}, t \geq 1 \right|.$$

## Condition

*$\Delta(u_n(\tau))$ is satisfied if there exists a sequence $l_n = o(n)$ such that $l_n \to \infty$ and $\alpha_{n,l_n}(\tau) \to 0$ when $n \to \infty$.*

## Theorem

*Assume that $\Delta(u_n(\tau))$ is satisfied and $\lim_{n\to\infty} n\bar{F}(u_n(\tau)) = \tau$. If the limiting point process of $N_n^{(\tau)}$ exists, it is necessarily a homogeneous compound Poisson point process with intensity $\tau\theta\,|\cdot|$ and limiting cluster size distribution $\pi$ (Hsing et al. (1988)).*

# The point process of exceedances - dependent case

Let $\pi_n(m; q_n, u_n(\tau)) = \mathbb{P}\left( N_n^{(\tau)}((0; q_n/n]) = m \middle| N_n^{(\tau)}((0; q_n/n]) > 0 \right)$.

**Proposition**

*Suppose that extremal index $\theta$ exists, then a necessary and sufficient condition for the convergence of $N_n^{(\tau)}$ is*

$$\lim_{n\to\infty} \pi_n(m; q_n, u_n(\tau)) = \pi(m),$$

*where $(q_n)$ is a sequence of positive integers such that there exists a sequence $(l_n)$ satisfying $l_n = o(q_n)$, $q_n = o(n)$ and $nq_n^{-1}\alpha_{n,l_n}(\tau) \to 0$.*

If $\Delta(u_n(\tau))$ holds for each $\tau > 0$, then $\theta$ and $\pi$ do not depend on $\tau$.

# The point process of exceedances - dependent case

Let $\pi_n(m; q_n, u_n(\tau)) = \mathbb{P}\left( N_n^{(\tau)}((0; q_n/n]) = m \middle| N_n^{(\tau)}((0; q_n/n]) > 0 \right)$.

## Proposition

*Suppose that extremal index $\theta$ exists, then a necessary and sufficient condition for the convergence of $N_n^{(\tau)}$ is*

$$\lim_{n\to\infty} \pi_n(m; q_n, u_n(\tau)) = \pi(m),$$

*where $(q_n)$ is a sequence of positive integers such that there exists a sequence $(l_n)$ satisfying $l_n = o(q_n)$, $q_n = o(n)$ and $nq_n^{-1}\alpha_{n,l_n}(\tau) \to 0$.*

If $\Delta(u_n(\tau))$ holds for each $\tau > 0$, then $\theta$ and $\pi$ do not depend on $\tau$.

# The point process of exceedances - dependent case

Let $\pi_n(m; q_n, u_n(\tau)) = \mathbb{P}\left( N_n^{(\tau)}((0; q_n/n]) = m \,\middle|\, N_n^{(\tau)}((0; q_n/n]) > 0 \right)$.

## Proposition

*Suppose that extremal index $\theta$ exists, then a necessary and sufficient condition for the convergence of $N_n^{(\tau)}$ is*

$$\lim_{n\to\infty} \pi_n(m; q_n, u_n(\tau)) = \pi(m),$$

*where $(q_n)$ is a sequence of positive integers such that there exists a sequence $(l_n)$ satisfying $l_n = o(q_n)$, $q_n = o(n)$ and $nq_n^{-1}\alpha_{n,l_n}(\tau) \to 0$.*

If $\Delta(u_n(\tau))$ holds for each $\tau > 0$, then $\theta$ and $\pi$ do not depend on $\tau$.

# Estimation of the extremal index and the limiting cluster size distribution

The natural approach to estimating $\theta$ and $\pi$ is to identify the clusters of exceedances above a high threshold, then evaluate for each cluster the characteristic of interest and construct estimates from these values.

- The blocks declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a block length $r_n = o(s_n)$, and partitionning the $n$ observations into $k_n = \lceil n/r_n \rceil$ blocks.
- The runs declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a run length $p_n = o(s_n)$, and stipulating that any extreme observations separated by fewer than $p_n$ non-extreme observations belong to the same cluster.
- The automatic declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and working with inter-exceedance times (Ferro and Segers (2003)).

# Estimation of the extremal index and the limiting cluster size distribution

The natural approach to estimating $\theta$ and $\pi$ is to identify the clusters of exceedances above a high threshold, then evaluate for each cluster the characteristic of interest and construct estimates from these values.

- The blocks declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a block length $r_n = o(s_n)$, and partitionning the $n$ observations into $k_n = \lceil n/r_n \rceil$ blocks.
- The runs declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a run length $p_n = o(s_n)$, and stipulating that any extreme observations separated by fewer than $p_n$ non-extreme observations belong to the same cluster.
- The automatic declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and working with inter-exceedance times (Ferro and Segers (2003)).

# Estimation of the extremal index and the limiting cluster size distribution

The natural approach to estimating $\theta$ and $\pi$ is to identify the clusters of exceedances above a high threshold, then evaluate for each cluster the characteristic of interest and construct estimates from these values.

- The blocks declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a block length $r_n = o(s_n)$, and partitionning the $n$ observations into $k_n = \lceil n/r_n \rceil$ blocks.
- The runs declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a run length $p_n = o(s_n)$, and stipulating that any extreme observations separated by fewer than $p_n$ non-extreme observations belong to the same cluster.
- The automatic declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and working with inter-exceedance times (Ferro and Segers (2003)).

# Estimation of the extremal index and the limiting cluster size distribution

The natural approach to estimating $\theta$ and $\pi$ is to identify the clusters of exceedances above a high threshold, then evaluate for each cluster the characteristic of interest and construct estimates from these values.

- The blocks declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a block length $r_n = o(s_n)$, and partitionning the $n$ observations into $k_n = \lceil n/r_n \rceil$ blocks.
- The runs declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and a run length $p_n = o(s_n)$, and stipulating that any extreme observations separated by fewer than $p_n$ non-extreme observations belong to the same cluster.
- The automatic declustering scheme consists in choosing a threshold $u_{s_n}(\tau)$ where $s_n = o(n)$ and working with inter-exceedance times (Ferro and Segers (2003)).

# Extremal index estimators - The blocks method

Leadbetter (1983) showed that

$$\theta = \lim_{n\to\infty} s_n \mathbb{P}\left(\max_{1\le i\le r_n} X_i > u_{s_n}(\tau)\right) / (r_n \tau),$$

where $s_n = o(n)$ and $r_n = o(s_n)$.

This relation motivates the following estimator

$$\hat{\theta}_n = \frac{s_n K_{k_n}(\hat{u}_{s_n}(\tau))}{r_n \tau},$$

where $K_{k_n}(u) = k_n^{-1} \sum_{i=1}^{k_n} \mathbb{I}_{\{M_{r_n}^i > u\}}$ is the mean number of blocks with one or more exceedances of $u$ and $\hat{u}_{s_n}(\tau) = X_{(\lceil n\tau/s_n \rceil)}$.

# Extremal index estimators - The blocks method

Leadbetter (1983) showed that

$$\theta = \lim_{n\to\infty} s_n \mathbb{P}\left(\max_{1\le i\le r_n} X_i > u_{s_n}(\tau)\right) / (r_n \tau),$$

where $s_n = o(n)$ and $r_n = o(s_n)$.

This relation motivates the following estimator

$$\hat{\theta}_n = \frac{s_n K_{k_n}(\hat{u}_{s_n}(\tau))}{r_n \tau},$$

where $K_{k_n}(u) = k_n^{-1} \sum_{i=1}^{k_n} \mathbb{I}_{\left\{M_{r_n}^i > u\right\}}$ is the mean number of blocks with one or more exceedances of $u$ and $\hat{u}_{s_n}(\tau) = X_{(\lceil n\tau/s_n \rceil)}$.

# Extremal index estimators - The runs method

O'Brien (1987) showed that

$$\theta = \lim_{n\to\infty} \mathbb{P}\left( \max_{2\leq i\leq p_n} X_i \leq u_{s_n}(\tau) \middle| X_1 > u_{s_n}(\tau) \right),$$

where $s_n = o(n)$ and $p_n = o(s_n)$.

This relation motivates the following estimator

$$\hat{\theta}_n = \frac{\sum_{i=1}^{n-p_n} \mathbb{I}_{\{A_{i,p_n}(\hat{u}_{s_n}(\tau))\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{X_i > \hat{u}_{s_n}(\tau)\}}},$$

where $A_{i,p_n}(u) = \{X_i > u, X_{i+1} \leq u, ..., X_{i+p_n} \leq u\}$ and $\hat{u}_{s_n}(\tau) = X_{(\lceil n\tau/s_n \rceil)}$.

# Extremal index estimators - The runs method

O'Brien (1987) showed that

$$\theta = \lim_{n\to\infty} \mathbb{P}\left( \max_{2\le i\le p_n} X_i \le u_{s_n}(\tau) \middle| X_1 > u_{s_n}(\tau) \right),$$

where $s_n = o(n)$ and $p_n = o(s_n)$.

This relation motivates the following estimator

$$\hat{\theta}_n = \frac{\sum_{i=1}^{n-p_n} \mathbb{I}_{\{A_{i,p_n}(\hat{u}_{s_n}(\tau))\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{X_i > \hat{u}_{s_n}(\tau)\}}},$$

where $A_{i,p_n}(u) = \{X_i > u, X_{i+1} \le u, ..., X_{i+p_n} \le u\}$ and $\hat{u}_{s_n}(\tau) = X_{(\lceil n\tau/s_n \rceil)}$.

# Extremal index estimators - The automatic method

Let $T(u)$ be the inter-exceedance time, i.e. $\min\{i \geq 1, X_{i+1} > u\}$ given that $X_1 > u$. Ferro and Segers (2003) showed that

$$\lim_{n\to\infty} \mathbb{P}\left(\bar{F}\left(u_{s_n}(\tau)\right) T\left(u_{s_n}(\tau)\right) > t\right) = \theta e^{-\theta t}, \qquad t > 0,$$

where $s_n = o(n)$.

This relation motivates the following moment estimator

$$\hat{\theta}_n = \frac{2\left(\sum_{i=1}^{N_n(\hat{u}_{s_n}(\tau))}\left(T_i\left(\hat{u}_{s_n}(\tau)\right) - 1\right)\right)^2}{\left(N_n(\hat{u}_{s_n}(\tau)) - 1\right)\sum_{i=1}^{N_n(\hat{u}_{s_n}(\tau))}\left(T_i\left(\hat{u}_{s_n}(\tau)\right) - 1\right)\left(T_i\left(\hat{u}_{s_n}(\tau)\right) - 2\right)},$$

where $N_n(u) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i > u\}}$ is the number of exceedances of $u$, $T_i(u)$ is the $i^{th}$ inter-exceedance time of $u$ and $\hat{u}_{s_n}(\tau) = X_{(\lceil n\tau/s_n \rceil)}$.

# Extremal index estimators - The automatic method

Let $T(u)$ be the inter-exceedance time, i.e. $\min\{i \geq 1, X_{i+1} > u\}$ given that $X_1 > u$. Ferro and Segers (2003) showed that

$$\lim_{n\to\infty} \mathbb{P}\left(\bar{F}\left(u_{s_n}(\tau)\right) T\left(u_{s_n}(\tau)\right) > t\right) = \theta e^{-\theta t}, \qquad t > 0,$$

where $s_n = o(n)$.

This relation motivates the following moment estimator

$$\hat{\theta}_n = \frac{2\left(\sum_{i=1}^{N_n(\hat{u}_{s_n}(\tau))}\left(T_i\left(\hat{u}_{s_n}(\tau)\right)-1\right)\right)^2}{\left(N_n(\hat{u}_{s_n}(\tau))-1\right)\sum_{i=1}^{N_n(\hat{u}_{s_n}(\tau))}\left(T_i\left(\hat{u}_{s_n}(\tau)\right)-1\right)\left(T_i\left(\hat{u}_{s_n}(\tau)\right)-2\right)},$$

where $N_n(u) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i > u\}}$ is the number of exceedances of $u$, $T_i(u)$ is the $i^{th}$ inter-exceedance time of $u$ and $\hat{u}_{s_n}(\tau) = X_{(\lceil n\tau/s_n\rceil)}$.

# Limiting cluster size distribution estimators - The blocks method

Let us recall that

$$\lim_{n\to\infty} \mathbb{P}\left( N_n^{(\tau)}\left((0;q_n/n]\right) = m \middle| N_n^{(\tau)}\left((0;q_n/n]\right) > 0 \right) = \pi\left(m\right),$$

where $q_n = o\left(n\right)$.

This relation motivates the following estimators (Hsing (1991))

$$\hat{\pi}_n\left(m; r_n, \hat{u}_{s_n}\left(\tau\right)\right) = \frac{\sum_{j=1}^{k_n} \mathbb{I}_{\{Y_{n,j}(\hat{u}_{s_n}(\tau))=m\}}}{\sum_{j=1}^{k_n} \mathbb{I}_{\{Y_{n,j}(\hat{u}_{s_n}(\tau))>0\}}},$$

where $Y_{n,j}\left(u\right) = \sum_{i=(j-1)r_n+1}^{jr_n} \mathbb{I}_{\{X_i>u\}}$ is the number of exceedances of $u$ for the $j^{th}$ block and $\hat{u}_{s_n}\left(\tau\right) = X_{(\lceil n\tau/s_n \rceil)}$.

# Limiting cluster size distribution estimators - The blocks method

Let us recall that

$$\lim_{n\to\infty} \mathbb{P}\left( N_n^{(\tau)}\left((0; q_n/n]\right) = m \middle| N_n^{(\tau)}\left((0; q_n/n]\right) > 0 \right) = \pi\left(m\right),$$

where $q_n = o\left(n\right)$.

This relation motivates the following estimators (Hsing (1991))

$$\hat{\pi}_n\left(m; r_n, \hat{u}_{s_n}\left(\tau\right)\right) = \frac{\sum_{j=1}^{k_n} \mathbb{I}_{\{Y_{n,j}(\hat{u}_{s_n}(\tau))=m\}}}{\sum_{j=1}^{k_n} \mathbb{I}_{\{Y_{n,j}(\hat{u}_{s_n}(\tau))>0\}}},$$

where $Y_{n,j}\left(u\right) = \sum_{i=(j-1)r_n+1}^{jr_n} \mathbb{I}_{\{X_i>u\}}$ is the number of exceedances of $u$ for the $j^{th}$ block and $\hat{u}_{s_n}\left(\tau\right) = X_{(\lceil n\tau/s_n\rceil)}$.

# Limiting cluster size distribution estimators - The automatic method (Ferro (2003))

Let $T_1$ and $T_{1+j}$ be the inter-exceedance times separeted by $j-1$ other inter-exceedance times. Then

$$\begin{aligned} & \lim_{n\to\infty} \mathbb{P}\left(\bar{F}\left(u_{s_n}\left(\tau\right)\right)T_1\left(u_{s_n}\left(\tau\right)\right) > t, \bar{F}\left(u_{s_n}\left(\tau\right)\right)T_{1+j}\left(u_{s_n}\left(\tau\right)\right) > s\right) \\ = \quad & \theta e_j e^{-\theta(t+s)}, \end{aligned}$$

where $s_n = o\left(n\right)$ and $e_j$ is defined recursively by $e_1 = 1$ and

$$e_j \quad = \quad \pi\left(1\right)e_{j-1} + ... + \pi\left(j-1\right)e_1 + \pi\left(j\right).$$

Ferro (2003) introduced moments estimators.

In our paper, we introduce new estimators of the limiting cluster size probabilities. They are constructed from the compound probabilities of the limiting point process through a recursive algorithm.

# Limiting cluster size distribution estimators - The automatic method (Ferro (2003))

Let $T_1$ and $T_{1+j}$ be the inter-exceedance times separeted by $j-1$ other inter-exceedance times. Then

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{P}\left(\bar{F}\left(u_{s_n}\left(\tau\right)\right)T_1\left(u_{s_n}\left(\tau\right)\right)>t,\bar{F}\left(u_{s_n}\left(\tau\right)\right)T_{1+j}\left(u_{s_n}\left(\tau\right)\right)>s\right)\\ =\;&\theta e_j e^{-\theta(t+s)},\end{aligned}$$

where $s_n = o(n)$ and $e_j$ is defined recursively by $e_1 = 1$ and

$$e_j = \pi(1)\,e_{j-1} + ... + \pi(j-1)\,e_1 + \pi(j).$$

Ferro (2003) introduced moments estimators.

In our paper, we introduce new estimators of the limiting cluster size probabilities. They are constructed from the compound probabilities of the limiting point process through a recursive algorithm.

# Panjer's algorithm

Let us denote by $N_E^{(\tau)}$ the weak limit of $N_n^{(\tau)}(E)$ as $n \to \infty$ when it exists and by $p^{(\tau)} = \left(p^{(\tau)}(m)\right)_{m \geq 0}$ its distribution. Then

$$N_E^{(\tau)} \overset{d}{=} \sum_{i=1}^{\eta(\theta\tau)} \zeta_i,$$

where $(\zeta_n)$ is a sequence of iid integer rvs with distribution $\pi$ and $\eta(\theta\tau)$ is an independent Poisson rv with parameter $\theta\tau$.

We have

$$p^{(\tau)}(0) = e^{-\theta\tau}, \qquad p^{(\tau)}(m) = e^{-\theta\tau} \sum_{j=1}^{m} \frac{(\theta\tau)^j}{j!} \pi^{*j}(m), \qquad m \geq 1,$$

where $\pi^{*j}$ is the $j^{th}$ convolution of $\pi$. Panjer's algorithm is a recursive algorithm which can be used to compute $p^{(\tau)}$

$$p^{(\tau)}(0) = e^{-\theta\tau}, \; p^{(\tau)}(m) = -\frac{\ln\left(p^{(\tau)}(0)\right)}{m} \sum_{j=1}^{m} j\pi(j)\, p^{(\tau)}(m-j), \; m \geq 1.$$

# Panjer's algorithm

Let us denote by $N_E^{(\tau)}$ the weak limit of $N_n^{(\tau)}(E)$ as $n \to \infty$ when it exists and by $p^{(\tau)} = \left(p^{(\tau)}(m)\right)_{m \geq 0}$ its distribution. Then

$$N_E^{(\tau)} \stackrel{d}{=} \sum_{i=1}^{\eta(\theta\tau)} \zeta_i,$$

where $(\zeta_n)$ is a sequence of iid integer rvs with distribution $\pi$ and $\eta(\theta\tau)$ is an independent Poisson rv with parameter $\theta\tau$.
We have

$$p^{(\tau)}(0) = e^{-\theta\tau}, \qquad p^{(\tau)}(m) = e^{-\theta\tau} \sum_{j=1}^{m} \frac{(\theta\tau)^j}{j!} \pi^{*j}(m), \qquad m \geq 1,$$

where $\pi^{*j}$ is the $j^{th}$ convolution of $\pi$. Panjer's algorithm is a recursive algorithm which can be used to compute $p^{(\tau)}$

$$p^{(\tau)}(0) = e^{-\theta\tau}, \; p^{(\tau)}(m) = -\frac{\ln\left(p^{(\tau)}(0)\right)}{m} \sum_{j=1}^{m} j\pi(j)\, p^{(\tau)}(m-j), \; m \geq 1.$$

Note that $p^{(\tau)}(m)$ can be expressed as a function of the $\pi(j)$, $j = 1, ..., m$. But it is also possible to reverse the algorithm and to evaluate recursively $\pi(m)$ from the $p^{(\tau)}(j)$, $j = 0, ..., m$, in the following way

$$\begin{aligned} \pi(1) &= -\frac{p^{(\tau)}(1)}{\ln\left(p^{(\tau)}(0)\right) p^{(\tau)}(0)}, \\ \pi(m) &= \frac{\pi(1)}{p^{(\tau)}(1)} \left( p^{(\tau)}(m) + \frac{\ln\left(p^{(\tau)}(0)\right)}{m} \sum_{j=1}^{m-1} j\pi(j) p^{(\tau)}(m-j) \right). \end{aligned}$$

We deduce that there exist differentiable functions $f_m : \mathbb{R}_+^{m+1} \setminus \{0\} \to \mathbb{R}$, such that

$$\pi(m) = f_m\left(p^{(\tau)}(0), p^{(\tau)}(1), ..., p^{(\tau)}(m)\right), \qquad m \geq 1.$$

Corollary: it suffices to construct an estimate of $p^{(\tau)}$ to have an estimate of $\pi$.

Note that $p^{(\tau)}(m)$ can be expressed as a function of the $\pi(j)$, $j = 1, ..., m$. But it is also possible to reverse the algorithm and to evaluate recursively $\pi(m)$ from the $p^{(\tau)}(j)$, $j = 0, ..., m$, in the following way

$$\begin{aligned} \pi(1) &= -\frac{p^{(\tau)}(1)}{\ln\left(p^{(\tau)}(0)\right) p^{(\tau)}(0)}, \\ \pi(m) &= \frac{\pi(1)}{p^{(\tau)}(1)} \left( p^{(\tau)}(m) + \frac{\ln\left(p^{(\tau)}(0)\right)}{m} \sum_{j=1}^{m-1} j\pi(j) p^{(\tau)}(m-j) \right). \end{aligned}$$

We deduce that there exist differentiable functions $f_m : \mathbb{R}_+^{m+1} \backslash \{0\} \rightarrow \mathbb{R}$, such that

$$\pi(m) = f_m\left(p^{(\tau)}(0), p^{(\tau)}(1), ..., p^{(\tau)}(m)\right), \qquad m \geq 1.$$

Corollary: it suffices to construct an estimate of $p^{(\tau)}$ to have an estimate of $\pi$.

# Defining the estimators

We use the blocks declustering scheme: we divide $[1,...,n]$ into $k_n$ blocks of length $r_n$, $I_j = [(j-1)r_n + 1, ..., jr_n]$ for $j = 1, ..., k_n$, and a last block $I_{k_n+1} = [r_n k_n + 1, ..., n]$.

We define

- the number of observations above the threshold $u_{r_n}(\tau)$ within the $j$-th block $N_{r_n,j}^{(\tau)} = \sum_{i \in I_j} \mathbb{I}_{\{X_i > u_{r_n}(\tau)\}}$;
- the empirical distribution of the number of exceedances within a block $p_n^{(\tau)}(i) = k_n^{-1} \sum_{j=1}^{k_n} \mathbb{I}_{\left\{N_{r_n,j}^{(\tau)} = i\right\}}$;

We assume that $F$ belongs to the domain of attraction of the generalized extreme value (GEV) distribution with index $\gamma \in \mathbb{R}$, i.e. there exist two functions $a$ and $b$ such that $F$ satisfies the relation

$$\lim_{n\to\infty} n\bar{F}(a(n)x + b(n)) = (1 + \gamma x)^{-1/\gamma}.$$

# Defining the estimators

We use the blocks declustering scheme: we divide $[1, ..., n]$ into $k_n$ blocks of length $r_n$, $I_j = [(j-1)r_n + 1, ..., jr_n]$ for $j = 1, ..., k_n$, and a last block $I_{k_n+1} = [r_n k_n + 1, ..., n]$.

We define

- the number of observations above the threshold $u_{r_n}(\tau)$ within the $j$-th block $N_{r_n,j}^{(\tau)} = \sum_{i \in I_j} \mathbb{I}_{\{X_i > u_{r_n}(\tau)\}}$;
- the empirical distribution of the number of exceedances within a block $p_n^{(\tau)}(i) = k_n^{-1} \sum_{j=1}^{k_n} \mathbb{I}_{\left\{N_{r_n,j}^{(\tau)} = i\right\}}$;

We assume that $F$ belongs to the domain of attraction of the generalized extreme value (GEV) distribution with index $\gamma \in \mathbb{R}$, i.e. there exist two functions $a$ and $b$ such that $F$ satisfies the relation

$$\lim_{n \to \infty} n\bar{F}(a(n)x + b(n)) = (1 + \gamma x)^{-1/\gamma}.$$

The threshold $u_{r_n}(\tau)$ may be chosen as

$$u_{r_n}(\tau) = \gamma^{-1}\left(\tau^{-\gamma} - 1\right) a(r_n) + b(r_n).$$

An estimator of the level $u_{r_n}(\tau)$ is given by

$$\hat{u}_{r_n}(\tau) = \hat{\gamma}_n^{-1}\left(\tau^{-\hat{\gamma}_n} - 1\right) \hat{a}(r_n) + \hat{b}(r_n),$$

where $\hat{\gamma}_n$, $\hat{b}(r_n)$ and $\hat{a}(r_n)$ are suggested in Dekkers, Einmahl and de Haan (1989).
Then we define the counterpart of $N_{r_n,j}^{(\tau)}$, $p_n^{(\tau)}(i)$ where $u_{r_n}(\tau)$ is replaced by $\hat{u}_{r_n}(\tau)$

$$\hat{N}_{r_n,j}^{(\tau)} = \sum_{i \in I_j} \mathbb{I}_{\{X_i > \hat{u}_{r_n}(\tau)\}}, \qquad \hat{p}_n^{(\tau)}(i) = \frac{1}{k_n} \sum_{j=1}^{k_n} \mathbb{I}_{\left\{\hat{N}_{r_n,j}^{(\tau)} = i\right\}}.$$

Finally we introduce the estimators of the limiting cluster size distribution

$$\hat{\pi}_n(j) = f_j\left(\hat{p}_n^{(1)}(0), \hat{p}_n^{(1)}(1), ..., \hat{p}_n^{(1)}(j)\right).$$

Let us derive several estimators of the extremal index. This key parameter appears in different moments of the limiting distributions $N_E^{(\tau)}$ and $\zeta_1$

$$p^{(\tau)}(0) = e^{-\theta\tau}, \qquad \mathbb{E}\zeta_1 = \theta^{-1}, \qquad \mathbb{V}N_E^{(\tau)} = \theta\tau\mathbb{E}(\zeta_1)^2.$$

Estimators of $\theta$ can be constructed by equating approximately theoritical moments to their empirical counterparts

$$\hat{\theta}_{1,n} = -\ln\left(\hat{p}_n^{(1)}(0)\right), \; \hat{\theta}_{2,n} = \frac{1}{\sum_{j=1}^m j\hat{\pi}_n(j)}, \; \hat{\theta}_{3,n} = \frac{\sum_{j=0}^m (j-1)^2\hat{p}_n^{(1)}(j)}{\sum_{j=1}^m j^2\hat{\pi}_n(j)},$$

for some $m > 1$.

Finally we introduce the estimators of the limiting cluster size distribution

$$\hat{\pi}_n(j) = f_j\left(\hat{p}_n^{(1)}(0), \hat{p}_n^{(1)}(1), ..., \hat{p}_n^{(1)}(j)\right).$$

Let us derive several estimators of the extremal index. This key parameter appears in different moments of the limiting distributions $N_E^{(\tau)}$ and $\zeta_1$

$$p^{(\tau)}(0) = e^{-\theta\tau}, \qquad \mathbb{E}\zeta_1 = \theta^{-1}, \qquad \mathbb{V}N_E^{(\tau)} = \theta\tau\mathbb{E}(\zeta_1)^2.$$

Estimators of $\theta$ can be constructed by equating approximately theoritical moments to their empirical counterparts

$$\hat{\theta}_{1,n} = -\ln\left(\hat{p}_n^{(1)}(0)\right), \ \hat{\theta}_{2,n} = \frac{1}{\sum_{j=1}^m j\hat{\pi}_n(j)}, \ \hat{\theta}_{3,n} = \frac{\sum_{j=0}^m (j-1)^2 \hat{p}_n^{(1)}(j)}{\sum_{j=1}^m j^2\hat{\pi}_n(j)},$$

for some $m > 1$.

# Condition C0

## Condition

*The stationary sequence* $(X_n)$ *has extremal index* $\theta > 0$*.* $\Delta(u_n(\tau))$ *holds for each* $\tau > 0$ *and there exists a probability measure* $\pi = (\pi(i))_{i \geq 1}$*, such that for all* $i \geq 1$*,*

$$\pi(i) = \lim_{n \to \infty} P\left( N_n^{(\tau)}((0; q_n/n]) = i \,\middle|\, N_n^{(\tau)}((0; q_n/n]) > 0 \right), \qquad \text{(C0.a)}$$

*for some* $\Delta(u_n(\tau))$*-separating sequence* $(q_n)$*. Moreover there exists a constant* $\rho > 2$ *such that for each* $\tau > 0$

$$\sup_{n \geq 1} E\left( N_n^{(\tau)}(E) \right)^{\rho} < \infty. \qquad \text{(C0.b)}$$

# Condition C1

## Condition

*Condition* $(C0)$ *holds.* $\Delta(u_n(\tau_1), u_n(\tau_2))$ *holds for each* $\tau_1 > \tau_2 > 0$ *and there exists a probability measure* $\pi_2$*, such that for all* $i_1 \geq i_2 \geq 0$, $i_1 \geq 1$,

$$\begin{aligned} & \lim_{n\to\infty} \mathbb{P}\left( N_n^{(\tau_j)}((0;q_n/n]) = i_j; j = 1,2 \middle| N_n^{(\tau_1)}((0;q_n/n]) > 0 \right) \\ = \; & \pi_2^{(\tau_2/\tau_1)}(i_1, i_2), \end{aligned}$$

*for some* $\Delta(u_n(\tau_1), u_n(\tau_2))$*-separating sequence* $(q_n)$.

Let us denote by $\left(N_E^{(\tau_1)}, N_E^{(\tau_2)}\right)$ the weak limit of $\left(N_n^{(\tau_1)}(E), N_n^{(\tau_2)}(E)\right)$ as $n \to \infty$ when it exists and by $p_2^{(\tau_1,\tau_2)}$ its distribution. Then

$$\left(N_E^{(\tau_1)}, N_E^{(\tau_2)}\right) \overset{d}{=} \left(\sum_{i=1}^{\eta(\theta\tau_1)} \zeta_{1,i}^{(\tau_2/\tau_1)}, \sum_{i=1}^{\eta(\theta\tau_1)} \zeta_{2,i}^{(\tau_2/\tau_1)}\right)$$

where $\left(\zeta_{1,i}^{(\tau_2/\tau_1)}, \zeta_{2,i}^{(\tau_2/\tau_1)}\right)$ is a sequence of iid integer vector rvs with distribution $\pi_2^{(\tau_2/\tau_1)}$ and $\eta(\theta\tau_1)$ is a Poisson rv with parameter $\theta\tau_1$ and is independent of the $\left(\zeta_{1,i}^{(\tau_2/\tau_1)}, \zeta_{2,i}^{(\tau_2/\tau_1)}\right)$.

The distribution of $\left(N_E^{(\tau_1)}, N_E^{(\tau_2)}\right)$ is given by

$$\begin{aligned} p_2^{(\tau_1,\tau_2)}(0,0) &= \mathbb{P}\left(\eta(\theta\tau_1) = 0\right) = e^{-\theta\tau_1} \\ p_2^{(\tau_1,\tau_2)}(i,j) &= e^{-\theta\tau_1} \sum_{k=1}^{i} \frac{(\theta\tau_1)^k}{k!} \pi_2^{(\tau_2/\tau_1),*k}(i,j). \end{aligned}$$

# Condition C2

## Condition

*Let* $\varepsilon > 0$*. There exist two constants* $C > 0$ *and* $\delta > 6$ *such that*

$$\alpha_{n,l}(\tau_1, ..., \tau_r) \leq \alpha_l := C l^{-\delta-\varepsilon}, \tag{C2.a}$$

*for every choice of* $\tau_1 > ... > \tau_r > 0$*,* $r \geq 1$*,* $n \geq 1$*.* $(r_n)$ *is sequence such that* $r_n \to \infty$ *and* $r_n = o(n)$*. There exists a sequence* $(l_n)$ *satisfying*

$$l_n = o\left(r_n^{2/v}\right) \text{ and } \lim_{n\to\infty} n r_n^{-1} \alpha_{l_n} = 0, \tag{C2.b}$$

*where* $v = 2\delta/(\delta - 3)$*. There exists a constant* $\gamma > 2v$ *such that for each* $\tau_1 > \tau_2 > 0$

$$\sup_{n\geq 1} E\left(N_n^{(\tau_1)}(E) - N_n^{(\tau_2)}(E)\right)^{\gamma} < \infty. \tag{C2.c}$$

# Condition C2

## Condition

*Let* $\left(\zeta_{1,1}^{(\tau_2/\tau_1)}, \zeta_{2,1}^{(\tau_2/\tau_1)}\right)^{2v}$ *be a vector rv with distribution* $\pi_2^{(\tau_2/\tau_1)}$ *defined in* $(C1.a)$. *There exists a positive constant* $D_{2v}$ *such that*

$$E\left(\zeta_{1,1}^{(\tau_2/\tau_1)} - \zeta_{2,1}^{(\tau_2/\tau_1)}\right)^{2v} < D_{2v}\left(1 - \tau_2/\tau_1\right), \qquad \text{(C2.d)}$$

*for every choice of* $\tau_1 > \tau_2 > 0$.

# Condition C3

## Condition

*There exist a function $A$, with $\lim_{n\to\infty} A(n) = 0$, and a function $L$ such that*

$$d_{TV}\left(N_n^{(\tau)}(E), N_E^{(\tau)}\right) \leq L(\tau) A(n). \tag{C3.a}$$

*There exist some constants $\xi \leq 0$ and a regularly varying positive function of index $\xi$, $\Theta$, with $\lim_{n\to\infty} \Theta(n) = 0$, such that*

$$\lim_{n\to\infty} \frac{n\bar{F}(b(n)+a(n)x) - (1+\gamma x)^{-1/\gamma}}{\Theta(n)} = K\left((1+\gamma x)^{1/\gamma}\right), \tag{C3.b}$$

*locally uniformly for $x \in I_\gamma$. The sequence $(r_n)$ satisfies*

$$\lim_{n\to\infty} \sqrt{k_n} A(r_n) = \lim_{n\to\infty} \sqrt{k_n} \Theta(r_n) = 0.$$

## Proposition

*Suppose that* $(C0)$ *holds. Let* $(r_n)$ *be a sequence such that* $r_n \to \infty$ *and* $r_n = o(n)$. *Then* $\hat{\pi}_n(j) \xrightarrow{P} \pi(j)$, $j = 1, \dots, m$, *and*

$$\hat{\theta}_{1,n} \xrightarrow{P} \theta, \ \hat{\theta}_{2,n} \xrightarrow{P} \left( \sum_{j=1}^{m} j\pi(j) \right)^{-1}, \ \hat{\theta}_{3,n} \xrightarrow{P} \frac{\sum_{j=0}^{m} (j-1)^2 p^{(1)}(j)}{\sum_{j=1}^{m} j^2 \pi(j)}.$$

Let us introduce the multivariate empirical process

$$E_{m,n}(\tau) = (e_{0,n}(\tau), \dots, e_{m,n}(\tau), \bar{e}_n(\tau))', \qquad \tau > 0,$$

where

$$\begin{aligned} e_{i,n}(\tau) &= \sqrt{k_n} \left( p_n^{(\tau)}(i) - P\left( N_{r_n,j}^{(\tau)} = i \right) \right), \\ \bar{e}_n(\tau) &= \sqrt{k_n} \left( \bar{p}_n^{(\tau)} - r_n P(X_i > u_{r_n}(\tau)) \right), \\ \bar{p}_n^{(\tau)} &= \sum_{i=1}^{\infty} i p_n^{(\tau)}(i) = \frac{1}{k_n} \sum_{j=1}^{k_n} N_{r_n,j}^{(\tau)} = \frac{1}{k_n} \sum_{i=1}^{r_n k_n} 1_{\{X_i > u_{r_n}(\tau)\}}. \end{aligned}$$

## Proposition

*Suppose that* $(C0)$ *holds. Let* $(r_n)$ *be a sequence such that* $r_n \to \infty$ *and* $r_n = o(n)$. *Then* $\hat{\pi}_n(j) \xrightarrow{P} \pi(j)$, $j = 1, ..., m$, *and*

$$\hat{\theta}_{1,n} \xrightarrow{P} \theta,\ \hat{\theta}_{2,n} \xrightarrow{P} \left( \sum_{j=1}^{m} j\pi(j) \right)^{-1},\ \hat{\theta}_{3,n} \xrightarrow{P} \frac{\sum_{j=0}^{m} (j-1)^2 p^{(1)}(j)}{\sum_{j=1}^{m} j^2 \pi(j)}.$$

Let us introduce the multivariate empirical process

$$E_{m,n}(\tau) = (e_{0,n}(\tau), ..., e_{m,n}(\tau), \bar{e}_n(\tau))', \qquad \tau > 0,$$

where

$$\begin{aligned} e_{i,n}(\tau) &= \sqrt{k_n}\left( p_n^{(\tau)}(i) - P\left( N_{r_n,j}^{(\tau)} = i \right) \right), \\ \bar{e}_n(\tau) &= \sqrt{k_n}\left( \bar{p}_n^{(\tau)} - r_n P(X_i > u_{r_n}(\tau)) \right), \\ \bar{p}_n^{(\tau)} &= \sum_{i=1}^{\infty} i p_n^{(\tau)}(i) = \frac{1}{k_n} \sum_{j=1}^{k_n} N_{r_n,j}^{(\tau)} = \frac{1}{k_n} \sum_{i=1}^{r_n k_n} 1_{\{X_i > u_{r_n}(\tau)\}}. \end{aligned}$$

Let $B > 1$ and $D^{(m)}(0, B)$ be the space of functions from $(0, B)$ to $\mathbb{R}^m$ which are left continuous and have right limits at each point, equipped with the Skorohod's $J_1$-topology.

**Theorem**

*Suppose that $(C1)$ and $(C2)$ hold. There exists a pathwise continuous centered Gaussian process $E_m$ with covariance function*

$$\mathbb{C}\left(E_m\left(\tau_1\right), E_m\left(\tau_2\right)\right) = V^{(m)}\left(\tau_1, \tau_2\right)$$

*which can be expressed as a function of $\tau_1$, $\tau_2$, $p^{(\tau_1)}$, $p^{(\tau_2)}$, $p_2^{(\tau_1,\tau_2)}$, $\pi$, $\pi_2^{(\tau_2/\tau_1)}$, $\theta$, such that $E_{m,n} \Rightarrow E_m$ weakly in $D^{(m+2)}(0, B)$.*

## Theorem

*Suppose that* $(C1)$, $(C2)$ *and* $(C3)$ *hold. Then*

$$\sqrt{k_n}\left(\hat{p}_n^{(\cdot)}(j) - p^{(\cdot)}(j)\right)$$

$$\Rightarrow \left( \begin{array}{c} E_{j+1,m}(\cdot) - h_j(\cdot)(\cdot)^{1+\gamma} \times \\ \left(\gamma^{-1}\left((\cdot)^{-\gamma} - 1\right)A + B + \gamma^{-2}\left(1 - (\cdot)^{-\gamma}(1 + \gamma \ln(\cdot))\right)\Gamma\right) \end{array} \right)$$

*in* $D^{(m+1)}(0, B)$, *where* $h_j(\tau) = \partial p^{(\tau)}(j)/\partial\tau$, $E_{j,m}$ *is the* $j$*-th component of* $E_m$ *and* $A$, $B$, $\Gamma$ *depend on* $\gamma$ *and* $(E_{m+1,m}(\tau))_{0<\tau\leq 1}$.

## Corollary

*Suppose that* $(C1)$, $(C2)$ *and* $(C3)$ *hold. Then*

$$\sqrt{k_n}\left(\hat{p}_n^{(1)}(j) - p^{(1)}(j)\right)_{j=0,\ldots,m} \xrightarrow{d} N\left(0, M^{(m)}\right),$$

*where* $M^{(m)}$ *can be expressed as a function of* $p^{(1)}$, $\pi$.

## Theorem

*Suppose that* $(C1)$, $(C2)$ *and* $(C3)$ *hold. Then*

$$\sqrt{k_n}\left(\hat{p}_n^{(\cdot)}(j) - p^{(\cdot)}(j)\right)$$

$$\Rightarrow \left(\begin{array}{c} E_{j+1,m}(\cdot) - h_j(\cdot)(\cdot)^{1+\gamma} \times \\ \left(\gamma^{-1}\left((\cdot)^{-\gamma}-1\right)A + B + \gamma^{-2}\left(1-(\cdot)^{-\gamma}(1+\gamma\ln(\cdot))\right)\Gamma\right) \end{array}\right)$$

*in* $D^{(m+1)}(0,B)$, *where* $h_j(\tau) = \partial p^{(\tau)}(j)/\partial\tau$, $E_{j,m}$ *is the* $j$*-th component of* $E_m$ *and* $A$, $B$, $\Gamma$ *depend on* $\gamma$ *and* $\left(E_{m+1,m}(\tau)\right)_{0<\tau\leq 1}$.

## Corollary

*Suppose that* $(C1)$, $(C2)$ *and* $(C3)$ *hold. Then*

$$\sqrt{k_n}\left(\hat{p}_n^{(1)}(j) - p^{(1)}(j)\right)_{j=0,\ldots,m} \xrightarrow{d} N\left(0, M^{(m)}\right),$$

*where* $M^{(m)}$ *can be expressed as a function of* $p^{(1)}$, $\pi$.

Let $\hat{\Pi}_{n,m} = k_n^{1/2} \left( \hat{\pi}_n (j) - \pi (j) \right)_{j=1,\ldots,m}$, $f^{(m)} = (f_1, ..., f_m)$ and $\nabla f^{(m)} = \left( \partial f_j / \partial p_{i-1} \right)_{1 \leq i \leq m+1, 1 \leq j \leq m}$.

## Corollary

*Suppose that* $(C1)$, $(C2)$ *and* $(C3)$ *hold. Then*

$$\hat{\Pi}_{n,m} \xrightarrow{d} N\left(0, P^{(m)}\right),$$

*where* $P^{(m)} = \left( \nabla f^{(m)} \right)' M^{(m)} \nabla f^{(m)}$.

## Corollary

*Suppose that* $(C1)$, $(C2)$ *and* $(C3)$ *hold. Then*

$$\sqrt{k_n}\left(\hat{\theta}_{1,n}-\theta\right)\xrightarrow{d}N\left(0,e^{\theta}-2\theta-1+\theta^3\sum_{j=1}^{\infty}j^2\pi\left(j\right)\right),$$

$$\sqrt{k_n}\left(\hat{\theta}_{2,n}-\frac{1}{\sum_{j=1}^{m}j\pi_n\left(j\right)}\right)\xrightarrow{d}N\left(0,\frac{A_m'P^{(m)}A_m}{\left(\sum_{j=1}^{m}j\pi_n\left(j\right)\right)^4}\right),$$

$$\sqrt{k_n}\left(\hat{\theta}_{3,n}-\frac{\sum_{j=0}^{m}\left(j-1\right)^2p^{(1)}\left(j\right)}{\sum_{j=1}^{m}j^2\pi\left(j\right)}\right)\xrightarrow{d}N\left(0,B_m'M^{(m)}B_m\right),$$

where $A_m=(1,...,m)'$ and

$$B_m=\left(\frac{1}{\sum_{l=1}^{m}l^2\pi\left(l\right)}\left(\left(j-1\right)^2-\frac{\sum_{l=0}^{m}\left(l-1\right)^2p^{(1)}\left(l\right)}{\sum_{l=1}^{m}l^2\pi\left(l\right)}\sum_{l=j}^{m}l^2\frac{\partial f_l}{\partial p_j}\right)\right)_{j=0,\dots,m}.$$

# Simulation study

500 sequences of length 2000 were simulated from the three processes:

- an $ARCH(1)$ process: $X_n = (\eta + \lambda X_{n-1}) Z_n^2$, $n \geq 2$, where $Z_n$ are iid standard Gaussian rvs, $\eta = 2.10^{-5}$, $\lambda = 0.5$ and $X_1 = 0$.

$$\begin{array}{lll} \pi(1) = 0.751 & \pi(2) = 0.168 & \pi(3) = 0.055 \\ \pi(4) = 0.014 & \pi(5) = 0.008 & \theta = 0.727. \end{array}$$

- a max-autoregressive process: $X_n = \max\{(1-\theta) X_{n-1}, W_n\}$, $n \geq 2$, where $W_n$ are independent unit Fréchet rvs, $\theta = 0.5$ and $X_1 = W_1/\theta$.

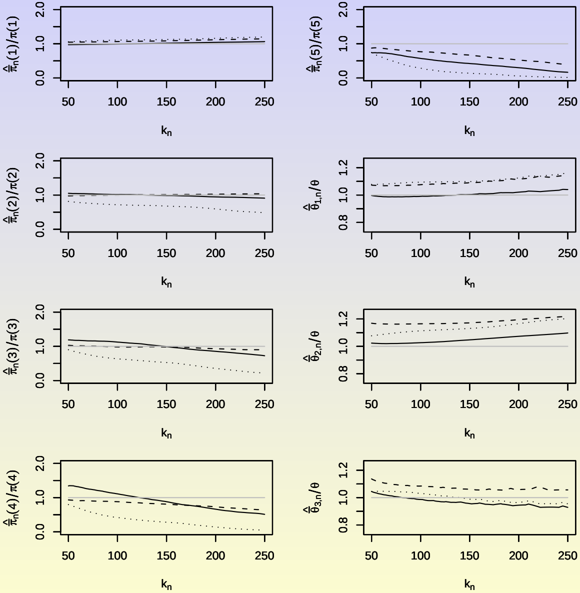
$$\begin{array}{lll} \pi(1) = 0.5 & \pi(2) = 0.25 & \pi(3) = 0.125 \\ \pi(4) = 0.0625 & \pi(5) = 0.031 & \theta = 0.5. \end{array}$$

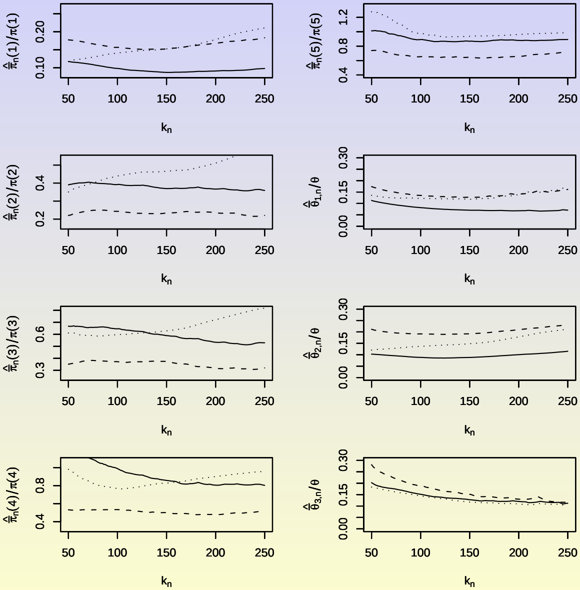
- an $AR(1)$ process with uniform marginal: $X_n = r^{-1}X_{n-1} + \varepsilon_n$, $n \geq 2$, where $(\varepsilon_n)$ are iid and uniformly distributed on $\{0, 1/r, ...., (r-1)/r\}$, $r = 4$ and $X_1$ is uniformly distributed on $(0, 1)$.

$$\begin{array}{lll} \pi(1) = 0.75 & \pi(2) = 0.1875 & \pi(3) = 0.0469 \\ \pi(4) = 0.0117 & \pi(5) = 0.0029 & \theta = 0.75 \end{array}$$

To smooth the discontinuity effect due to the blocks declustering scheme, we computed averages over the estimates corresponding to different block sizes. Morover, we considered the ratios $\widehat{\overline{\pi}}_n(i)/\pi(i)$, $i = 1, ..., 5$ and $\widehat{\overline{\theta}}_{j,n}/\theta$, $j = 1, .., 3$ to compare the performance of the estimators for the three processes.

Legend: $ARCH(1)$ process (——), max-$AR(1)$ process (- - - ), $AR(1)$ process $(\cdot\cdot\cdot\cdot)$.


$\hat{\pi}_n(1)/\pi(1)$
$\hat{\pi}_n(2)/\pi(2)$
$\hat{\pi}_n(3)/\pi(3)$
$\hat{\pi}_n(4)/\pi(4)$
$\hat{\pi}_n(5)/\pi(5)$
$\hat{\theta}_{1,n}/\theta$
$\hat{\theta}_{2,n}/\theta$
$\hat{\theta}_{3,n}/\theta$
$k_n$
0.0 1.0 2.0
0.8 1.0 1.2
50 100 150 200 250


$\hat{\pi}_n(1)/\pi(1)$
$\hat{\pi}_n(2)/\pi(2)$
$\hat{\pi}_n(3)/\pi(3)$
$\hat{\pi}_n(4)/\pi(4)$
$\hat{\pi}_n(5)/\pi(5)$
$\hat{\theta}_{1,n}/\theta$
$\hat{\theta}_{2,n}/\theta$
$\hat{\theta}_{3,n}/\theta$
$k_n$
50 100 150 200 250