Practical conditions on Markov chains for weak convergence of tail empirical processes

Olivier Wintenberger
University of Copenhagen and Paris VI

Joint work with Rafał Kulik and Philippe Soulier

Toronto, May 4 2016

I. Regularly varying time series

II. The tail empirical process

III. Statistical applications

I. Regularly varying time series

Let $\{X_t, t \in \mathbb{Z}\}$ be a real valued stationary time series. It is regularly varying iff for any $k \geq 0$

$$\mathcal{L}\left(\frac{X_0}{x}, \frac{X_1}{x} \ldots, \frac{X_k}{x} \mid |X_0| > x\right) \to_{x\to\infty} (Y_0, \ldots, Y_k).$$

- This defines a process $\{Y_k, k \geq 0\}$ called the tail process, Basrak and Segers (2009).
- Y_0 has a two-sided Pareto distribution on $(-\infty, -1] \cup [1, \infty)$.
- If for some $j \in \{1, \ldots, d\}$, X_j is extremally independent of X_0, then $Y_j \equiv 0$.
- $|Y_0|$ is independent of $Y_j/|Y_0| = \Theta_0$ called the spectral tail process.

Harris recurrent Markov chains

Assume $X_t = g(\Phi_t)$ with $\{\Phi_t\}$ an aperiodic, recurrent and irreducible Markov chain with kernel P and which admits an invariant distribution π such that $X_0 = g(\Phi_0)$ is regularly varying.

- It constitutes a stationary regularly varying time series, Resnick and Zeber (2013), Janssen and Segers (2014).
- Y_t is a multiplicative random walk $Y_t = A_t \cdots A_1 Y_0$, $t \geq 0$ for some (A_t) iid.

Drift condition

Assume the following drift condition holds:

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_{\mathcal{C}}(x) \ ,$$

where $\lambda \in (0, 1)$, $V(x) : E \to [1, \infty)$ and $\mathcal{C}$ is a small set.

- Under these conditions, the chain (Φ_t) is β-mixing with geometric rate and so is (X_t), Meyn and Tweedie (1997).

The DC$_p$, $p > 0$ condition, Mikosch and W. (2014)

Let $\{u_n\}$ be an increasing sequence such that

$$\lim_{n\to\infty} \bar{F}(u_n) = \lim_{n\to\infty} \frac{1}{n\bar{F}(u_n)} = 0 \,.$$

- There exist $p \in (0, \alpha/2)$ and a constant c such that

$$|g|^p \leq cV \,. \tag{1}$$

- For every compact set $[a, b] \subset (0, \infty)$,

$$\limsup_{n\to\infty} \sup_{a\leq s\leq b} \frac{1}{u_n^p \mathbb{P}(X_0 > u_n)} \mathbb{E}\left[V(\Phi_0)\mathbb{1}_{\{su_n < g(\Phi_0)\}}\right] < \infty \,. \tag{2}$$

Roughly (1) and (2) implies the existence of c_1, c_2 and n_0

$$c_1\mathbb{P}(V(\Phi_0) > u_n) \leq \mathbb{P}(|X|^p > u_n) \leq c_2\mathbb{P}(V(\Phi_0) > u_n), \qquad n \geq n_0.$$

Example 1: the GARCH(1,1) process

We consider a GARCH(1,1) process $X_t = \sigma_t Z_t$, where (Z_t) is iid $\mathcal{N}(0,1)$ and

$$\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2(\alpha_1 Z_{t-1}^2 + \beta_1) = \alpha_0 + \sigma_{t-1}^2 A_t.$$

Special SRE then DC$_p$ holds under Kesten's condition $\mathbb{E}[A_0^{\alpha/2}] = 1$ for $p < \alpha$ with $V(x) = |x|^p$ and $g = x$, Mikosch and W. (2014).

Example 2: the AR(p) process

Assume that $\{X_j, j \in \mathbb{Z}\}$ is an AR(p) model

$$X_j = \phi_1 X_{j-1} + \cdots + \phi_p X_{j-p} + \varepsilon_j \ , \quad j \geq 1 \ ,$$

that satisfies the following conditions:

- $\{\varepsilon_j, j \in \mathbb{Z}\}$ are iid, regularly varying with index α,
- the polynomial $1 - \phi_1 z - \cdots - \phi_p z^p$ does not have unit root inside the unit cercle.

As for Random Coefficients AR, DC$_p$ for $p < \alpha$, holds for some V and $g(x_1, \ldots, x_p) = x_1$, Feigin and Tweedie (1985).

Example 3: the Threshold ARCH(1) process

Let $\xi \in \mathbb{R}$. Assume that $\{X_j\}$ is T-ARCH model

$$X_j = (b_{10} + b_{11} X_{j-1}^2)^{1/2} Z_j \mathbb{1}_{\{X_{j-1} < \xi\}} + (b_{20} + b_{21} X_{j-1}^2)^{1/2} Z_j \mathbb{1}_{\{X_{j-1} \geq \xi\}} ,$$

that satisfies the following conditions:

- $b_{ij} > 0$;
- $\{Z_j, j \in \mathbb{Z}\}$ are iid gaussian r.v.,
- the Lyapunov exponent

$$\gamma = p \log b_{11}^{1/2} + (1-p) \log b_{21}^{1/2} + \mathbb{E}[\log(|Z_1|)] ,$$

  where $p = \mathbb{P}(Z_1 < 0)$, is strictly negative;
- $(b_{11} \vee b_{21})^{p/2} \mathbb{E}[|Z_0|^p] < 1$.

Then DC_p holds.

A counterexample

Let $\{Z_j, j \in \mathbb{Z}\}$ be iid positive integer valued r.v. regularly varying with index $\beta > 1$.

Define the Markov chain $\{X_j, j \geq 0\}$ by the following recursion:

$$X_j = \begin{cases} X_{j-1} - 1 & \text{if } X_{j-1} > 1 \, , \\ Z_j & \text{if } X_{j-1} = 1 \, . \end{cases}$$

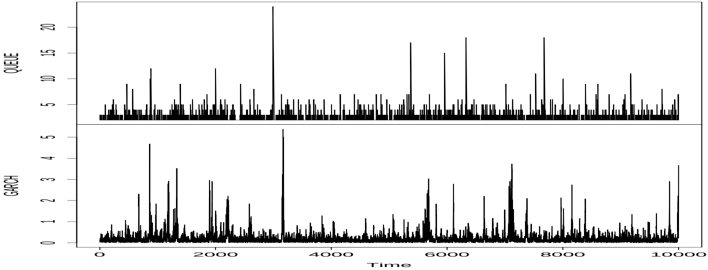
Since $\beta > 1$, the chain admits a stationary distribution π on $\mathbb{N}$ given by

$$\pi(n) = \frac{\mathbb{P}(Z_0 \geq n)}{\mathbb{E}[Z_0]} \, , \quad n \geq 1 \, .$$

Karamata's Theorem implies that π is regularly varying with index $\alpha = \beta - 1$.

- The state $\{1\}$ is an atom and the return time to the atom is distributed as Z_0,
- The chain is not geometrically ergodic and DC$_p$ cannot hold,
- The time series is regularly varying with $Y_j = 1$ for all $j \geq 0$,
- It admits a phantom distribution, the distribution of Z_0, O'Brien (1987), Doukhan *et al.* (2015)

$$\mathbb{P}\Big(\max_{1\leq j\leq n} X_j \leq u_n\Big) - \mathbb{P}(Z_j \leq u_n)^{n/\mathbb{E}[Z]} \to 0$$



II. The tail empirical process.
II.1. Univariate TEP

The univariate tail empirical process is defined by

$$e_n(t) = \frac{1}{n\bar{F}(u_n)} \sum_{i=1}^{n} \{\mathbb{1}_{\{X_i > u_n t\}} - \bar{F}(u_n t)\}$$

relevant to infer the marginal tail.

- In the iid case then

$$\sqrt{n\bar{F}(u_n)}\, e_n(t) \Rightarrow W(t^{-\alpha}) \, .$$

  where W is the standard Brownian motion and $\Rightarrow$ means weak convergence in the space $\mathcal{D}(0,\infty]$ endowed with the J_1 topology, Starica and Resnick (1997).

II.1. Multivariate TEP

For each quantity of interest, an appropriate type of TEP is used.

Joint exceedences, Kulik and Soulier (2015), extremograms, Davis and Mikosch (2009).

$$e_n(t_1, \ldots, t_h) = \frac{1}{n\bar{F}(u_n)} \sum_{i=1}^{n} \{\mathbb{1}_{\{X_{i+1} > u_n t_1, \ldots, X_{i+h} > u_n t_h\}} - \mathbb{P}(X_1 > u_n t_1, \ldots, X_h > u_n t_h)\} \, .$$

Conditional limiting distributions,

$$e_n(t_1, \ldots, t_h) = \frac{1}{n\bar{F}(u_n)} \sum_{i=1}^{n} \mathbb{1}_{\{X_{i+1} \le u_n t_1, \ldots, X_{i+h} \le u_n t_h\}} \mathbb{1}_{\{X_i > u_n\}} - \mathbb{P}(X_1 \le u_n t_1, \ldots, X_h \le u_n t_h \mid X_0 > u_n)$$

Spectral tail process, Drees *et al.* (2015).

The relevant TEP is

$$e_n(t) = \frac{1}{n\bar{F}(u_n)} \sum_{i=1}^{n} \mathbb{1}_{\{X_{i+1}/|X_i| \le t, X_i > u_n\}} - \mathbb{P}(X_1/X_0 \le u_n t \mid X_0 > u_n) .$$

Many variations and more generally, cluster functionals, Drees and Rootzen (2010).

Weighted versions

$$\begin{aligned} e_n(t_1, \ldots, t_h) &= \frac{1}{n\bar{F}(u_n)} \sum_{i=1}^{n} \Psi(X_{i+1}, \ldots, X_{i+h}) \mathbb{1}_{\{X_{i+1} > u_n t_1, \ldots, X_{i+h} > u_n t_h\}} \\ &\quad - \mathbb{E}[\Psi(X_{i+1}, \ldots, X_{i+h}) \mid X_1 > u_n t_1, \ldots, X_h > u_n t_h] . \end{aligned}$$

Extremal quantities of interest

- Univariate:
  - The tail index. Extreme quantiles.
- Multivariate; extremal dependence:
  - various coefficients of extremal dependence; extremograms;
  - distribution of the tail process.
- Multivariate; extremal independence:
  - conditional limiting distributions;
  - conditional scaling exponents.
- Infinite dimensional:
  - extremal index;
  - cluster functionals.

Conditions for time series

To prove the weak convergence of the tail empirical process for time series, three types of conditions are needed.

- Temporal dependence condition;
- Tightness criterion;
- Finite clustering condition;
- Bias conditions.

- We will not discuss the bias issues which are important but of a completely different nature (second order conditions).

Temporal dependence

- β-mixing allowing blocks method, Rootzen (2009)
- For a suitable choice of sequences r_n and u_n, the limiting distribution of e_n is the same as

$$\tilde{e}_n(t) = \sum_{i=1}^{m_n} \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{r_n} \{\mathbb{1}_{\{\tilde{X}_{(i-1)r_n+j}\}} - \bar{F}(u_n t)\}$$

where $m_n = [n/r_n]$ and the coupling blocks, Yu (1994), $\{\tilde{X}_j, j = (i-1)r_n + 1, \ldots, ir_n\}$, $i = 1, \ldots, m_n$ are iid as one original block of X_i.

Given this approximation, we need two more conditions:

- existence of the limiting variance of each block suitably normalized,
- tightness of the TEP.

The variance of one block

By standard computations, we obtain

$$\frac{1}{\bar{F}(u_n)} \mathrm{var}\left(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n s\}}\right)$$

$$\sim \frac{\bar{F}(u_n s)}{\bar{F}(u_n)} + \sum_{j=1}^{r_n} \frac{\mathbb{P}(X_0 > u_n s, X_j > u_n s)}{\bar{F}(u_n)} + O(r_n \bar{F}(u_n))$$

$$\sim s^{-\alpha} + \sum_{j=1}^{r_n} \frac{\mathbb{P}(X_0 > u_n s, X_j > u_n s)}{\bar{F}(u_n)} + O(r_n \bar{F}(u_n)) .$$

If we assume that $r_n \bar{F}(u_n) \to 0$, then the sum right above must a limit.

Note that joint regular variation of (X_0, X_j) implies that for fixed $L > 0$, by definition of the tail process $\{Y_j, j \geq 1\}$,

$$\lim_{n\to\infty} \sum_{j=1}^{L} \frac{\mathbb{P}(X_0 > u_n s, X_j > u_n s)}{\bar{F}(u_n)} = s^{-\alpha} \sum_{j=1}^{L} \mathbb{P}(Y_j > 1) .$$

This implies that for each $L > 1$,

$$\liminf_{n\to\infty} \frac{1}{\bar{F}(u_n)} \mathrm{var}\left(\sum_{j=1}^{r_n} \mathbb{1}_{\{X_j > u_n s\}} \right) \geq s^{-\alpha} \sum_{j=0}^{L} \mathbb{P}(Y_j > 1) .$$

Thus a necessary condition for the quantity on the left hand side to have a finite limit is

$$\sum_{j=1}^{\infty} \mathbb{P}(Y_j > 1) < \infty .$$

Finite Clustering condition

A sufficient condition is

$$\lim_{L\to\infty}\limsup_{n\to\infty}\frac{1}{\bar{F}(u_n)}\sum_{L<|j|\leq r_n}\mathbb{P}(X_0>u_n s, X_j>u_n s)=0\;.$$

(known under the misleading name Anti-Clustering condition).

If Condition FC holds, then for $s \leq t$,

$$\lim_{n\to\infty}\sqrt{n\bar{F}(u_n)}\,\mathrm{cov}(e_n(t),e_n(s))=s^{-\alpha}\sum_{j\in\mathbb{Z}}\mathbb{P}(Y_j>t/s)\;.$$

where $\{Y_j, j\in\mathbb{Z}\}$ is the tail process (extended to negative integers thanks to the forward-backward formula of Basrak and Segers, 2009).

- If the time series is extremally independent, i.e. $Y_j = 0$ for $j \neq 0$, then the limit is the same as in the iid case.
- If Condition FC holds, then the extremal index θ of the chain $\{X_t\}$ is positive and is given by, Basrak and Segers (2009),

$$\theta = \mathbb{P}\Big(\max_{k \geq 1} Y_k \leq 1\Big) > 0\ .$$

- In the counterexample $\theta = 0$, Roberts *et al.* (2006).

Checking Condition FC for Markov chains

Lemma

Under the DC_p and minorization conditions, Condition FC holds.

Proof: The set $\mathcal{C}$ is a regenerative set. Let $\tau_{\mathcal{C}}$ be the first return to $\mathcal{C}$. Fix an integer $L > 0$ and split the sum at $\tau_{\mathcal{C}}$:

$$\frac{1}{\bar{F}(u_n)}\mathbb{E}\Big[\sum_{j=L}^{r_n}\mathbb{1}_{\{su_n<X_0\}}\mathbb{1}_{\{su_n<X_j\}}\Big]$$

$$\leq \frac{1}{\bar{F}(u_n)}\mathbb{E}\left[\sum_{j=L}^{\tau_{\mathcal{C}}}\mathbb{1}_{\{su_n<X_0\}}\mathbb{1}_{\{su_n<X_j\}}\right]$$

$$+ \frac{1}{\bar{F}(u_n)}\mathbb{E}\left[\sum_{j=\tau_{\mathcal{C}}+1}^{r_n}\mathbb{1}_{\{su_n<X_0\}}\mathbb{1}_{\{su_n<X_j\}}\mathbb{1}_{\{\tau_{\mathcal{C}}\leq r_n\}}\right]$$

The last term is dealt with by standard regenerative arguments.

For $r > 1$, the sum up to the first visit is bounded by

$$
\begin{aligned}
\bar{F}^{-1}(u_n)\mathbb{E}\Big[\sum_{j=L}^{\tau_{\mathcal{C}}} \mathbb{1}_{\{su_n<X_0\}}\mathbb{1}_{\{su_n<X_j\}}\Big] \\
&\leq \int_{su_n}^{\infty} \mathbb{E}\Big[\sum_{j=L}^{\tau_{\mathcal{C}}} \mathbb{1}_{\{su_n<X_j\}} \mid X_0 = x\Big]\nu_n(\mathrm{d}x) \\
&\leq C u_n^{-p} r^{-L} \int_{su_n}^{\infty} \mathbb{E}\Big[\sum_{j=L}^{\tau_{\mathcal{C}}} r^j V(X_j) \mid X_0 = x\Big]\nu_n(\mathrm{d}x)
\end{aligned}
$$

The geometric drift condition states precisely that there exists $r > 1$ such that

$$
\mathbb{E}\Big[\sum_{j=L}^{\tau_A} r^k V(\Phi_j) \mid X_0 = x\Big] \leq C|x|^p \; .
$$

Plugging this bound into the integral yields

$$
\bar{F}^{-1}(u_n)\mathbb{E}\Big[\sum_{j=L}^{\tau_A} \mathbb{1}_{\{su_n<X_0\}}\mathbb{1}_{\{su_n<X_j\}}\Big] \leq C r^{-L}.
$$

Tightness

Tightness of the process e_n under the β-mixing condition holds under the following sufficient condition:

- for each $a > 0$, $s, t \geq a$,

$$\limsup_{n\to\infty} \frac{1}{r_n \bar{F}(u_n)} \mathbb{E}\left[\left(\sum_{j=1}^{r_n} \mathbb{1}_{\{u_n s < X_j \leq r_n t\}}\right)^2\right] \leq \left|s^{-\alpha} - t^{-\alpha}\right| .$$

This condition can be proved for Markov chains which satisfy the DC_p and minorization conditions as above.

Moreover, the following restriction on the block size is needed

$$\lim_{n\to\infty} \frac{r_n}{\sqrt{n\bar{F}(u_n)}} = 0 .$$

Theorem

Let DC_p and minorization conditions hold and assume moreover that there exists $\eta > 0$ such that

$$\lim_{n\to\infty} \log^{1+\eta}(n) \left\{ \bar{F}(u_n) + \frac{1}{\sqrt{n\bar{F}(u_n)}} \right\} = 0 \,.$$

Let $s_0 > 0$ be fixed.

- *The TEP converges weakly in $\ell^\infty([s_0\mathbf{1}, \infty))$ to a centered Gaussian process at the rate $\sqrt{n\overline{F}(u_n)}$.*
- *If $\psi : \mathbb{R}^{h+1} \to \mathbb{R}$ is such that*

  $$|\psi(x_0, \dots, x_h)| \le c((|x_0| \vee 1)^{q_0} + \cdots + (|x_h| \vee 1)^{q_h}) \,,$$

  with $q_i + q_{i'} \le p \wedge \alpha/2$ for all $i, i' = 0, \dots, h$, then converges weakly to a centered Gaussian process at the rate $\sqrt{n\overline{F}(u_n)}$.

Examples and a counterexample

Most usual Markovian time series satisfy the DC_p condition

- GARCH(1,1),
- SRE,
- AR(p),
- Functional autoregressive process: $X_{t+1} = g(X_t) + Z_t$ with $\lim_{t\to\infty} |g(x)| \leq \lambda < 1$,
- Threshold models, AR-ARCH models, etc.

The counterexample does not satisfy the DC_p condition and the TEP converges at another rate than the iid case. The limit is not gaussian if $\mathrm{var}(Z) = \infty$.

III. Statistical applications

- In the case of extremal independence, $Y_j = 0$ for all $j \neq 0$, under the DC_p and minorization conditions, the univariate TEP has the same limit as in the iid case.
- Thus most of the statistical inference works similarly as in the iid case.

The Hill estimator

The classical Hill estimator of $\gamma = 1/\alpha$ is defined as

$$\hat{\gamma} = \frac{1}{k} \sum_{j=1}^{n} \log_+ \left(\frac{X_{n:n-j+1}}{X_{n:n-k}} \right) .$$

Corollary

Under the DC_p and minorization conditions and if one can neglect the bias, then

$$\sqrt{k}\,\{\hat{\gamma} - \gamma\} \to \mathcal{N}\left(0, \alpha^{-2}\left\{1 + 2\sum_{j=1}^{\infty} \mathbb{P}(Y_j > 1 \mid Y_0 > 1)\right\}\right) .$$

Notice the limit distribution is the one from the iid case multiplied by the sum of the extremograms.

Estimation of the cluster index

Under Condition DC_p, the cluster index exists, Mikosch and Wintenberger (2014),

$$\lim_{h\to\infty} b_+(h) = b_+, \qquad b_+(h) = \frac{1}{h}\lim_{n\to\infty}\frac{\mathbb{P}(X_0+\cdots+X_h>u_n)}{\overline{F}(u_n)}.$$

Define

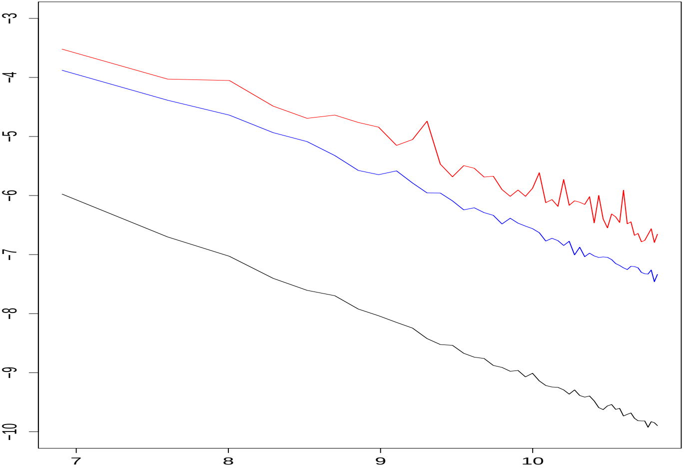
$$\hat{b}_+(h) = \frac{1}{kh}\sum_{j=1}^{n-h}\mathbb{1}_{\{X_j+\cdots+X_{j+h}>X_{n:n-k}\}}\,.$$

Corollary

Under the DC_p and minorization conditions and if one can neglect the bias, $\sqrt{k}(\hat{b}_+(h)-b_+(h))$ converges weakly to a gaussian r.v..

What about the counterexample?

In most statistical applications, we are interested in the behavior of the TEP e_n at $X_{n:n-k}/u_n$. Both quantities behave in a very non usual way but the ultimate estimation seems to work normally.



Thanks!