Applications of weak dependence to Ecology

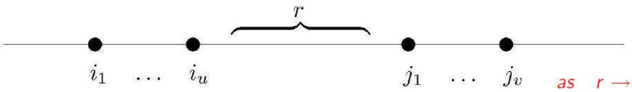
Paul Doukhan
doukhan@cyu.fr
http:www.doukhan.u-cergy.fr

UMR 8088 AGM, CY Cergy Paris University Cergy, & **IUF**
http://doukhan.perso.cyu.fr

February 3, 2023
FiME, IHP

Presentation

Let $(X_t)_{t\in\mathbb{Z}}$ be a time series, a natural question is to quantify the asymptotic independence of this process at the times:



as $r \to \infty$

This problem is considered through elementary ideas and applications adapted to large sample data

Outline:

- From independence to dependence
- Models
- Technique
- Applications, estimation, resampling, Ecology

Independence

We wish to answer the question
How to weaken the independence relation

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \quad ?$$

relating the events $A \in \sigma(P)$ of the past history with those $B \in \sigma(F)$ in a (not so close) future.
This relation is also restated as:

$$\mathrm{Cov}(f(P), g(F)) = 0, \qquad \forall f, g, \quad \|f\|_\infty, \|g\|_\infty \leq 1$$

(Variables P, F denote here Past and Future)

Mixing (Rosenblatt, 1956)

$$\begin{aligned}\alpha(\sigma(P),\sigma(F)) &= \sup_{A,B} |\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)| \\ &= \frac{1}{2}\sup_{\|f\|_\infty,\|g\|_\infty\le 1} |\mathrm{Cov}(f(P),g(F))|\end{aligned}$$

$X=(X_t)_{t\in\mathbb{Z}}$, $P=(X_{i_1},\ldots,X_{i_u})$, $F=(X_{j_1},\ldots,X_{j_v})$,
$i_1\le\cdots\le i_u, j_1\le\cdots\le j_v$ and $r=j_1-i_u$ is large:

$$\alpha(r)=\sup_{P,F}\alpha(\sigma(P),\sigma(F))\to_{r\to\infty} 0$$

See Rio 2000 for sharp technical results, see also Doukhan 1994 and Bradley 2007

Some nonmixing models

$X_t=\frac{1}{2}\left(X_{t-1}+\xi_t\right), \xi_t\sim b\left(\frac{1}{2}\right)$ iid, Andrews-Rosenblatt (1984) ($X_{t-1}=\mathrm{frac}(2X_t)$)
$X_t=\xi_t(1+aX_{t-1})$, $\mathbb{P}(\xi_0=\pm 1)=1/2$, $a\in\left(\frac{3-\sqrt{5}}{2},\frac{1}{2}\right]$, ($X_t=\sum_{j\ge 0}a^j\xi_t\cdots\xi_{t-j}$)

Covariances versus independence

Independence sometimes coincides with orthogonality
$\mathrm{Cov}(X,Y)=0 \Longrightarrow$ independence of a random vector (X,Y) if

$X, Y \in \{0,1\}$	admit Bernoulli distributions
(X,Y)	is a Gaussian vector
(X,Y)	is an associated vector (see below)

$X\in\mathbb{R}^p$ associated $\Leftrightarrow$ $\mathrm{Cov}(f(X),g(X))\geq 0$ for $f,g:\mathbb{R}^p\to\mathbb{R}$ (coordinatewise $\uparrow$)

Then $|\mathrm{Cov}(f(X),g(Y))|\leq \sum_{i,j} a_ib_j|\mathrm{Cov}(X_i,Y_j)|$,

for $(X,Y)\in\mathbb{R}^{p+q}$ associated or Gaussian

$$\begin{aligned}|f(x_1,\ldots,x_p)-f(y_1,\ldots,y_p)| &\leq a_1|x_1-y_1|+\cdots+a_p|x_p-y_p|\\ |g(x_1,\ldots,x_q)-g(y_1,\ldots,y_q)| &\leq b_1|x_1-y_1|+\cdots+b_q|x_q-y_q|\end{aligned}$$

Counterexamples: independent vectors, stability through $\uparrow$ images

A linear process

$$X_t = \sum_{j=-\infty}^{\infty} a_j \xi_{t-j}, \quad \sum_{j=-\infty}^{\infty} |a_j| < \infty, \|\xi_0\|_m < \infty, \quad (\xi_t)_{t\in\mathbb{Z}} \text{ iid}$$

$$X_t^p = \sum_{|j|<p} a_j \xi_{t-j} \Rightarrow \|X_t - X_t^p\|_m \le \|\xi_0\|_m \sum_{|j|\ge p} |a_j|,$$

$t - s > 2p \Rightarrow (X_s^p, X_t^p)$ independent.

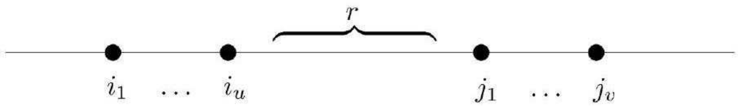
$$\begin{aligned} |\mathrm{Cov}(f(X_s), g(X_t))| &\le |\mathrm{Cov}(f(X_s) - f(X_s^p), g(X_t))| \\ &+ |\mathrm{Cov}(f(X_s^p), g(X_t^p))| + |\mathrm{Cov}(f(X_s^p), g(X_t) - g(X_t^p))| \\ &\le 2\mathrm{Lip}\, g \|f\|_\infty \|X_s - X_s^p\|_1 + 2\mathrm{Lip}\, f \|g\|_\infty \|X_t - X_t^p\|_1 \end{aligned}$$

A definition of weak dependence should be flexible enough to include both this example (which includes ARMA models) and that of associated processes.
It should also yield *reasonable* limit theory in order to work out the consistency of statistical procedures.
Bickel & Bühlmann (1999) also define weak dependence to bootstrap such models: in this case innovations do not admit a density.

General formulation (Doukhan & Louhichi, 1999)

$(X_t)_{t\in\mathbb{Z}}$ ($\in E$), $f: E^u \to \mathbb{R}$ from $\mathcal{F}$, $g: E^v \to \mathbb{R}$ from $\mathcal{G}$:

$$|\mathrm{Cov}\,(f(X_{i_1},\ldots,X_{i_u}),g(X_{j_1},\ldots,X_{j_v}))| \le \Psi(f,g)\epsilon(r), \quad \epsilon(r)\downarrow 0$$



$$\begin{aligned}\Psi(f,g) &= v\mathrm{Lip}\,g, & \epsilon(r)=\theta(r),\\ &= u\mathrm{Lip}\,f + v\mathrm{Lip}\,g + uv\mathrm{Lip}\,f\cdot\mathrm{Lip}\,g, & \epsilon(r)=\lambda(r)\end{aligned}$$

$$\mathrm{Lip}\,f = \sup_{(y_1,\ldots,y_u)\neq(x_1,\ldots,x_u)} \frac{|f(y_1,\ldots,y_u)-f(x_1,\ldots,x_u)|}{\|y_1-x_1\|+\cdots+\|y_u-x_u\|}.$$

Noncausal coefficients correspond to symmetric Ψ's.

Random fields or metric index sets are also considered (think of point processes).

vector LARCH(∞) models

$$X_t = \xi_t\Big(a + \sum_{j=1}^{\infty} a_j X_{t-j}\Big), \quad X_t(n\times 1), \xi_t(n\times p), a(p\times 1), a_j(p\times n)$$

$\phi = \|\xi_0\|_m \sum_j \|a_j\| < 1$, a $\mathbb{L}^m$-solution for (8) writes

$$X_t = \xi_t\Big(a + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k\ge 1} a_{j_1}\xi_{t-j_1}\cdots a_{j_k}\xi_{t-j_1-\cdots-j_k} a\Big)$$

Then	$\theta(t)$	$\le$	$C\,t^{-b}$,	$C\,(q\vee\phi)^{\sqrt{t}}$,		$C\,e^{-bt}$
if respectively	$A(s)$	$\le$	$C' s^{-b}$,	$C' q^s$,	or	$a_j = 0,\ j > C'$
	$A(s) = \|\xi_0\|_m \sum_{j\ge s}\|a_j\|$					

- GARCH(p,q) (Engle, Granger) $r_t = \sigma_t\epsilon_t,\ \ \sigma_t^2 = \sum_{j=1}^p \beta_j\sigma_{t-j}^2 + \gamma_0 + \sum_{j=1}^q \gamma_j r_{t-j}^2$
- ARCH(∞) (Surgailis *et al.* 2001) $r_t = \sigma_t\epsilon_t, \quad \sigma_t^2 = \beta_0 + \sum_{j=1}^{\infty}\beta_j r_{t-j}^2$
- Bilinear (Giraitis, Surgailis, 2003) $X_t = \zeta_t\Big(a + \sum_{j=1}^{\infty} a_j X_{t-j}\Big) + b + \sum_{j=1}^{\infty} b_j X_{t-j}$

General memory models

$X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \dots; \xi_t)$, $(\xi_t)_{t\in\mathbb{Z}}$ iid, $F : (\mathbb{R}^d)^{\mathbb{N}} \times \mathbb{R}^D \to \mathbb{R}^d$

with $\|F(x_1, x_2, x_3, \dots; \xi_t) - F(y_1, y_2, y_3, \dots; \xi_t)\|_m \le \sum_{j=1}^{\infty} a_j \|x_j - y_j\|$, then:

$\|F(0,0,0,\dots;\xi_t)\|_m < \infty$, $a = \sum_{j=1}^{\infty} a_j < 1$ $(m \ge 1)$ imply existence in $\mathbb{L}^m$, stationarity and weak dependence:

$$\theta(r) \le C \inf_{N>0} \left(\sum_{j\ge N} a_j + a^{\frac{r}{N}} \right)$$

- Regression models $X_t = f(X_{t-1}, \dots, X_{t-k}) + \zeta_t g(X_{t-1}, \dots, X_{t-k}) + \xi_t$
- variations on LARCH $X_t = \xi_t\left(a + \sum_{j=1}^{\infty} a_j(X_{t-j})\right)$, a_j Lipschitz
- Mean fields type models $X_t = f\left(\xi_t, \sum_{s\ge1} a_s X_{t-s}\right)$, f Lipschitz

Integer valued models

Thining, Steutel & van Harn operator is defined as

$$a \circ X = sign(X) \sum_{i=1}^{|X|} Y_i \qquad \text{for } a > 0, \quad X \in \mathbb{Z},$$

$(Y_i)_i$ is iid, context-independent, $\mathbb{E} Y_0 = a$ (e.g. Poisson or Bernoulli).

- Galton-Watson process with immigration, INAR $X_t = a \circ X_{t-1} + \xi_t$
- Integral bilinear models $X_t = a \circ X_{t-1} + b \circ (\varepsilon_{t-1} X_{t-1}) + \varepsilon_t$
  Estimation from moments (Doukhan, Latour, Oraichi, 2006).
- INLARCH(∞) $X_t = \xi_t \left(a_0 + \sum_{j=1}^{\infty} a_j \circ X_{t-j} \right)$ QMLE (Latour, Truquet 2008).
- GLM integer models $X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t)$ with $\lambda_t = g(\lambda_{t-1}, X_{t-1}, \ldots)$ with Fokianos and Tjostheim, 2011 and with Fokianos and Rynkiewicz (2021).
- More recent papers on
  http://doukhan.perso.cyu.fr/publications.html

Existence of strictly stationary solutions, weak dependence properties
$\Longrightarrow$ limit theory in estimation procedures.

Allowing $X_t \leq 0$ also gives non-associated and perhaps non-mixing processes

Limit theorems are fundamental to prove consistencies

- Moment inequalities
  - for integer moments, Doukhan & Louhichi use combinatorial methods
  - for causal coefficients Louhichi, Prieur use Lindeberg method
  - for $(2+\delta)-$order Doukhan & Wintenberger extend Ibragimov (1975) argument
- Exponential inequalities
  - For iid rvs, Bernstein inequality writes $\mathbb{P}(S_n \geq t\sqrt{n}) \leq C \exp\left\{-\frac{t^2}{2\sigma^2+K\frac{t}{\sqrt{n}}}\right\}$
  - Doukhan, Louhichi use moment combinatorics to get $\leq Ce^{-c\sqrt{t}}$,
  - Doukhan, Neumann use cumulant techniques $\leq C \exp\left\{-\dfrac{t^2}{2\sigma^2+K(t/\sqrt{n})^\alpha}\right\}$,
  - Rio (2000) and Dedecker (1999) extend Nagaev-Fuk maximal inequalities
  - Dedecker & Prieur use coupling arguments under causality. See also Rio, Merlevède and Peligrad (2010).

Limits in distribution enable goodness of fit tests I

A) Donsker invariance principles,

X_n stationary, with $\mathbb{E}X_0 = 0$, with $\sigma^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(X_0, X_k) \geq 0$ (well defined), then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} X_k \xrightarrow{D[0,1]}_{n\to\infty} \sigma W_t$$

if one of those conditions holds

- $\mathbb{E}|X_0|^{2+\delta} < \infty$ and $\lambda(i) = O(i^{-a})$ for $a > 2 + 2/\delta$
- $\mathbb{E}|X_0|^{2+\delta} < \infty$ and $\kappa(i) = O(i^{-a})$ for $a > 2$
- $\mathbb{E}|X_0|^{2+\delta} < \infty$ and $\sum_{i>0} i^{1/\delta}\theta(i) < \infty$,
- $\mathbb{E}|X_0|^2 \log_+ |X_0| < \infty$ and $\theta(i) = O(a^i)$ for some $0 < a < 1$.

Dedecker, Doukhan, Louhichi, Prieur, Wintenberger

Limits in distribution enable goodness of fit tests II

B) Empirical Central Limit Theorem

X_n stationary, then $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\mathbf{1}(X_k \le x) - F(x)) \xrightarrow{D[\overline{\mathbb{R}}]}_{n\to\infty} Z(x)$ where $(Z(x))_{x\in\mathbb{R}}$ is the centered Gaussian process with covariance

$$\Gamma(x,y) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(\mathbf{1}(X_0 \le x), \mathbf{1}(X_k \le y))$$

if $F(x) \equiv x$, and a weak dependence condition is assumed

- $\theta(i) = O(i^{-a})$ for $a > 1$ (Dedecker and Prieur)
- $\lambda(i) = O(i^{-a})$ for $a > 15/2$ (under association: $a > 4$ is enough: Louhichi)
- $\eta(i) = O(i^{-a})$ for $a > 2 + 2\sqrt{2} \approx 4.8\cdots$ (Prieur)

Applications

- Estimation
  - Moment method for integer valued bilinear models (with Latour, Oraichi),
  - QMLE for ARCH(∞), INLARCH(∞)(Bardet, Latour, Truquet, Wintenberger)
  - Whittle estimator, empirical periodogram contrast (with Bardet, & León)
  - Kernel estimation $X_n = f(X_{n-1}, \ldots, X_{n-p}) + \xi_n g(X_{n-1}, \ldots, X_{n-q})$ (with Ango Nze, Dedecker, Louhichi, Prieur, Ragache, & Wintenberger), and prediction...
- Random fields, reliability of multicomponent systems (with Lang, Louhichi, Truquet, Ycart)
- Hard resampling is possible under nonparametric autoregression, since innovations dont need to have a density (with Neumann 2008, Neumann, Paparoditis, 2006)
- Stochastic algorithms, Sparsity, regression and density estimation (with Brandière, Alquier)
- Ripley statistics for point processes, uses spatial definitions for the dependence of such models (with Lang, 2016) ,we define weakly dependent point processes

A tool for CLT: Lindeberg Method

$Z_i \in \mathbb{R}^d$ 0-mean, $A_n = \sum_{i=1}^n \mathbb{E}\left(\|Z_i\|^{2+\delta}\right) < \infty, \quad 0 < \delta \leq 1$
for $Y_i \sim \mathcal{N}(0, \mathrm{Var}\ Z_i)$ independent and $f \in \mathcal{C}_b^3$ and $n \in \mathbb{N}^*$:

$$\Delta_n = \left|\mathbb{E}\big(f(Z_1 + \cdots + Z_n) - f(Y_1 + \cdots + Y_n)\big)\right| \quad (1)$$

Lemma 1 [standard Lindeberg Lemma under independence, 1922]

$$\Delta_n \leq 3\,\|f^{(2)}\|_\infty^{1-\delta}\,\|f^{(3)}\|_\infty^{\delta} \cdot A_n.$$

Lemma 2 [Dependent Lindeberg (Bardet, Doukhan, Lang & Ragache, 2007)]

Set $f(x) = e^{i<t,x>}$ for $t \in \mathbb{R}^d$, $T_t(n) = \sum_{j=1}^n \left|\mathrm{Cov}(e^{i<t,X_1+\cdots+X_{j-1}>}, e^{i<t,X_j>})\right|$ then

$$\Delta_n \leq T_t(n) + 3\|t\|^{2+\delta} A_n.$$

Kernel density estimation (a typical application)

$(X_i)_{i\in\mathbb{N}}$ stationary with marginal density f. $K:\mathbb{R}\to\mathbb{R}$ bounded Lipschitz, $\int_{-\infty}^{\infty} K(t)\,dt = 1$, $\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\left(\frac{x-X_i}{h_n}\right)$ for $x\in\mathbb{R}$, $h_n\to 0$, $nh_n\to\infty$.

Proposition 2

If $\|f\|_\infty<\infty$, $\sup_{i\neq j}\|f_{i,j}\|_\infty<\infty$ (joint marginal densities), then

$$\sqrt{nh_n}\left(\widehat{f}(x)-\mathbb{E}\widehat{f}(x)\right)\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}\left(0, f(x)\int_{\mathbb{R}}K^2(t)\,dt\right)$$

if for example $\theta(r)=O(r^{-\theta})$ with $\theta>3$, $h_n=o(1)$.

The random variables $Z_{i,n}=(U_{i,n}-\mathbb{E}U_{i,n})/\sqrt{nh}$ are pairwise asymptotically independent as $n\to\infty$ where $U_{i,n}=K\left(\frac{x-X_i}{h_n}\right)$. This will allow us to use directly our Lindeberg lemma.

Moment inequalities

In order to derive CLT for the partial sums a way is to make use of Bernstein blocks, this means splitting the indices in a sample $\{1, \ldots, n\}$ into k blocs distant at least q with $n \sim k(p+q)$. Then rvs are replaced by sums inside large blocks with size p while small blocks with size q are ignored.

Hence moments of partial sums are needed. Let $(X_t)_{t\geq 1}$ be a centered and stationary sequence:

$$\begin{aligned} M_p(n) &= |\mathbb{E}(X_1+\cdots+X_n)^p| \leq \sum_{1\leq i_1,\ldots,i_p\leq n} |\mathbb{E}(X_{i_1}\cdots X_{i_p})| \\ &\leq p! \sum_{1\leq i_1\leq\cdots\leq i_p\leq n} |\mathbb{E}(X_{i_1}\cdots X_{i_p})| \equiv p! A_p(n) \end{aligned}$$

The following inequality may essentially be found in Billingsley:

$$A_p(n) \leq C_p(n) + \sum_{k=2}^{p-2} A_k(n) A_{p-k}(n), \quad C_p(n) = (p-1)\sum_{r=1}^{n-1}(r+1)^{p-2} c_p(r)$$

$c_p(r) = \max \left| \mathrm{Cov}(X_{j_1}\cdots X_{j_k}, X_{j_{k+1}}\cdots X_{j_p})\right|$ where $j_1 \leq \cdots \leq j_k \leq j_k + r \leq j_{k+1} \leq \cdots \leq j_p \ldots$

Estimating a variance: D., Jakubowicz, León (2009) I

$$\text{If} \quad \frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k \to_{n\to\infty} \mathcal{N}_d(0,\Sigma), \quad \text{with} \quad \Sigma = \sum_{k=-\infty}^{\infty} \mathbb{E}X_0X_k'$$

Self-normalized results yield asymptotic confidence sets, Σ is estimated by:

- Spectrum: $\widehat{\Sigma} = \widehat{f}(0)$ if the matrix-spectral density is estimated
- Donsker: $\frac{1}{\sqrt{n}}\sum_{ns<i<nt} X_i \to Z(t) - Z(s)$ Brownian, $Z(1) \sim \mathcal{N}_d(0,\Sigma)$

$$\Delta_{j,n} = \frac{1}{\sqrt{n}}\sum_{i\in B_j} X_i \to Z(t_j) - Z(s_j) \qquad (B_j = [ns_j, nt_j]\cap \mathbb{N})$$

Then for suitable choices of F, and $0 = s_1 < t_1 \le s_2 < \cdots \le s_m < t_m = 1$

$$\widetilde{F}_n = \frac{1}{m}\sum_{j=1}^{m} F(\Delta_{j,n}) \to \mathbb{E}F\left(\mathcal{N}_d(0,\Sigma)\right)$$

Carlstein (1986) mixing, Peligrad-Shao (1995) ρ-mixing use both $t_j = s_{j+1}$

Estimating a variance: D., Jakubowicz, León (2009) II

In order to derive a self-normalized CLT, D., Jakubowicz, León (2009) set $t_i < s_{i+1}$ and, under weak dependence:

$$\frac{\sqrt{N_n}}{\sqrt{\left(\widehat{G}_n - \widehat{F}_n^2\right)^+}}\left(\widetilde{F}_n - \mathbb{E}F(\mathcal{N}_d(0,\Sigma))\right) \to \mathcal{N}(0,1), \qquad (G \equiv F^2)$$

Applications to
- Linear models with dependent inputs
- Sea waves modeling, $X_t = F(Y_t)$ for F approximately linear
- Crossing numbers of oscillatory systems

For such explicit examples for which such procedures is proved to be useful through simulation studies.

Ecodep

http://doukhan.perso.cyu.fr/ecodep.html

This is a project of ecology funded by CYU for 4 years, some details and some tasks are described on http://doukhan.perso.cyu.fr/abstract.html

People http://doukhan.perso.cyu.fr/members.html

Publications http://doukhan.perso.cyu.fr/publications.html

Related institutions http://doukhan.perso.cyu.fr/links.html

Regular seminar now at IHP on wednesday afternoon https://indico.math.cnrs.fr/category/621/

a special attention for the date of March 15 conference https://indico.math.cnrs.fr/event/9238/

Taylor's law

This law provides a qualitative property of probability distributions: consider distributions on $[0, \infty)$ such that the f discrepancy condition

$$\operatorname{Var} X = c(\mathbb{E}X)^{\alpha}$$

holds for fixed constants c, α in case X belongs to this family of distributions. So the problem turns to the asymptotic behaviours of the empirical counterpart $\widehat{T}$ of $c = \operatorname{Var} X/(\mathbb{E}X)^{\alpha}$. A test for the exponent α is obtained through a CLT for $\widehat{T}$ in de la Pena, Doukhan, Salhi (2022) JAP. For this one needs one sample of the distribution. If we have two samples then both c and α may be fitted in an ongoing project. In fact with de la Pena and Salhi we consider samples from a time series and in this case $\operatorname{Var} X$ is rather replaced by the standard limit variance in the CLT under weak dependence,

$$\sigma^2 = \sum_{j=-\infty}^{\infty} \operatorname{Cov}(X_0, X_j)$$

and we get it through a Bernstein block idea.
Even in the independent case the validity of the Taylor's law need a precise estimation of the centring; ongoing work Cohen, Doukhan, Truquet