An Outer Bracketing Approach for Empirical Process Central Limit Theorems of Weakly Dependent Data

Ruhr-University Bochum

Herold Dehling, Olivier Durieu, Marco Tusche

Background / Motivation

Empirical Processes Indexed by Classes of Sets

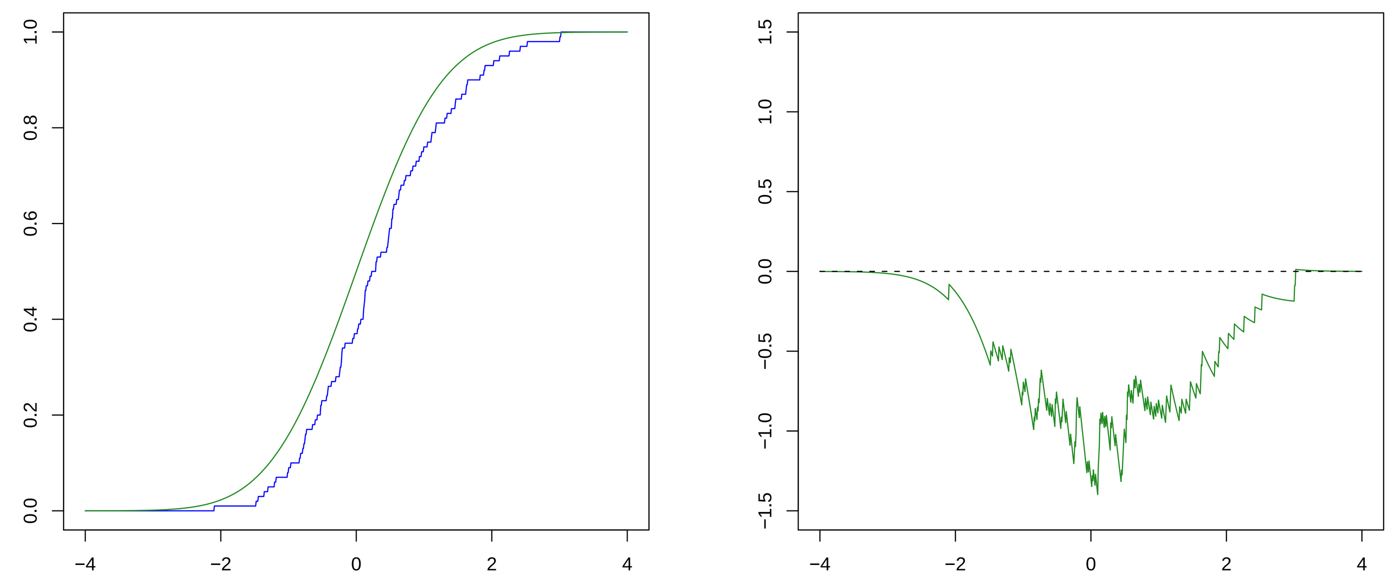
Let

- $(X_i)_{i\in\mathbb{N}}$ be stationary sequence of μ distributed random variables in $\mathbb{R}^d$ and
- $\mathcal{A}$ a collection of measurable sets in $\mathbb{R}^d$.

The empirical process $(U_n(A))_{A\in\mathcal{A}}$ given by

$$U_n(A) := n^{-1/2}\sum_{i=1}^{n}\big(\mathbf{1}\{X_i\in A\} - \mu(A)\big)$$

describes the probabilistic behavior as the empirical distribution $\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i\in\cdot\}$ convergences to μ in terms of *a process indexed by a class of sets* $A\in\mathcal{A}$.
We say that *an Empirical Process Central Limit Theorem holds (with respect to* $\mathcal{A}$*)*, if $(U_n(A))_{A\in\mathcal{A}}$ converges weakly to a Gaussian Process.



Classical Approach

- Verify the pointwise CLT, i.e. $U_n(A) \rightsquigarrow \mathcal{N}(0,\sigma_A)$ for all $A\in\mathcal{A}$
- Show tightness of $(U_n(A))_{A\in\mathcal{A}}$

To show this properties *one needs certain conditions on the processes* $\mathbf{1}\{X_i\in A\}$, $A\in\mathcal{A}$.

Our Setup

- No direct control over the processes $\mathbf{1}\{X_i\in A\}$, $A\in\mathcal{A}$
- Some control over the process under the class of Hölder functions $\mathcal{H}_\alpha$, equipped with the α-Hölder norm $\|f\|_\alpha := \|f\|_\infty + \sup_{x\neq y}|f(x)-f(y)|/|x-y|$

Let $\mu(f) := \mathbb{E}(f(X_0))$. As our central conditions, we assume that

(1) $n^{-1/2}\sum_{i=1}^{n}\big(f(X_i)-\mu(f)\big) \rightsquigarrow \mathcal{N}(0,\sigma_f)$ for all $f\in\mathcal{H}_\alpha$,

(2) for some fixed $M>0$, $a\in\mathbb{Z}$, we have for all $p\geq 1$ a moment bound

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}(f(X_i)-\mu(f))\right)^{2p}\right] \ll \sum_{i=1}^{p} n^i\,\mathbb{E}(|f(X_0)|)\log^{2p+ai}(\|f\|_\alpha+1)$$

for all $f\in\{g-h : g,h\in\mathcal{H}_\alpha, \|g\|_\infty, \|h\|_\infty \leq M\}$.

Main Result

Definition. For $l\leq u:\mathbb{R}^d\longrightarrow\mathbb{R}$ we define the *bracket*

$$[l,u] := \{f:\mathbb{R}^d\longrightarrow\mathbb{R} : l\leq f\leq u\}.$$

$[l,u]$ is a (ε,B)-bracket, if $\mathbb{E}|u(X_0)-l(X_0)|\leq\varepsilon$ and $\|l\|_\alpha, \|u\|_\alpha\leq B$.

The *bracketing number* $N(\varepsilon,B)$ is defined as the smallest number of brackets $[l,u]$ with $l,u\in\mathcal{H}_\alpha$, that are needed to cover $\{\mathbf{1}\{\cdot\in A\} : A\in\mathcal{A}\}$.

Theorem. Assume that **(1)** and **(2)** hold. If there is a $r>-1$, a $\gamma>\max\{2+a,1\}$, and a $C>0$ such that

(3) $$\int_0^1 \varepsilon^r \sup_{\varepsilon\leq\delta\leq 1} N^2\left(\delta, \exp\left(C\delta^{-\frac{1}{\gamma}}\right)\right)\,d\varepsilon < \infty,$$

then an Empirical Process Central Limit Theorem holds with respect to $\mathcal{A}$.

Applications

Our techniques can be applied in the situation of

- dynamical systems and Markov chains, where the Perron–Frobenius operator or the Markov operator satisfies a spectral gap condition,
- processes that feature the multiple mixing condition, i.e. $\exists\theta\in(0,1)\ \forall p\geq 1$ the exists a polynomial of bounded absolute degree such that

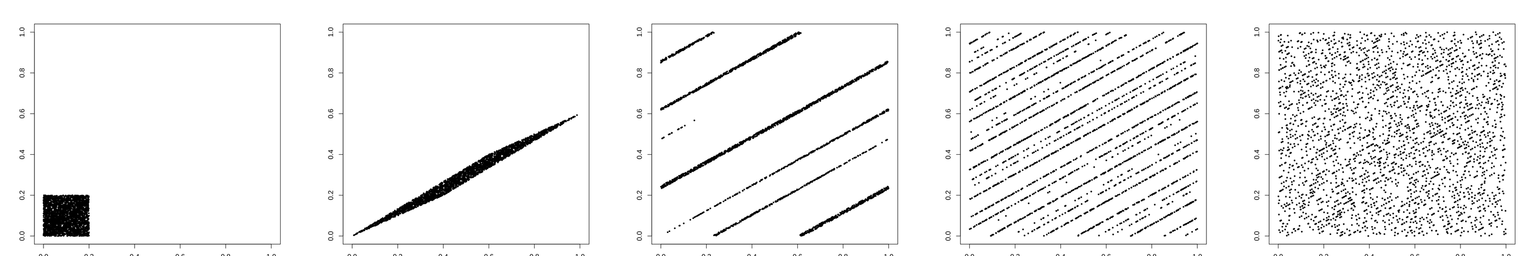
(4) $$\left|\mathbf{Cov}\left(f(X_{i_0})\cdots f(X_{i_{q-1}}), f(X_{i_q})\cdots f(X_{i_p})\right)\right| \leq P(i_1-i_0,\ldots,i_p-i_{p-1})\,\mathbb{E}|f(X_0)|\,\|f\|_\alpha^\ell\,\theta^{i_q-i_{q-1}}$$

for all $f\in\mathcal{B}$ with $\mathbb{E}(f(X_0))=0$, $\|f\|_\infty=1$, any integers $0\leq i_0<\ldots<i_p$, and any $q\in 1,\ldots,p$.

Example: Ergodic Automorphisms of the Multidimensional Torus

Let $T:\mathbb{T}^d\to\mathbb{T}^d$ be an ergodic automorphism on the d-dimensional torus $\mathbb{T}_d=\mathbb{R}_d/\mathbb{Z}_d$, which is identified with $[0,1]^d$, and $X_i := T^i(X_0)$, where X_0 is uniformly distributed on $\mathbb{T}^d$. Then an Empirical Process Central Limit Theorem holds with respect to

- the class of indicators of rectangles of $\mathbb{T}_d$,
- the class of indicators of Euclidean balls of $\mathbb{T}_d$,
- and the class of indicators of ellipsoids of bounded diameter of $\mathbb{T}_d$.



The graphics illustrate how the ergodic torus automorphism given by $T(x,y)=(3x+2y,2x+y) \mod 1$ acts on a sample of $3,000$ uniform distributed points in $[0,0.2]^2$ in four steps.

Generalizations

Our results also apply in more general cases. These are the following:

- *Weaker assumptions on dependence structure of* $(X_i)_{i\in\mathbb{N}}$
  - Replace $\log^{2p+ai}(\mathbf{id}+1)$ in condition **(2)** by faster increasing functions Φ_i, e.g. $\Phi_i=\mathbf{id}^i$. This allows one e.g. to treat processes that satisfy a weaker version of the multiple mixing condition **(4)**. However, in this situations one needs to use brackets with a stricter control of the α-Hölder norm (c.f. [2]).
  - Replace $\mathbb{E}|f(X_0)|$ in **(2)** by the $L^s(\mu)$-norm ($s\in\mathbb{N}^*$) of f. In this case one also has to work with brackets with a control of the $L^s(\mu)$-norm instead of the $L^1(\mu)$-norm (c.f. [3]).
- The $\mathbb{R}^d$ valued random variables X_i can be generalized to *random variables taking values in an arbitrary measurable space* $\mathcal{X}$ (c.f. [3]).
- Both $\mathcal{A}$ (or rather $\{\mathbf{1}\{\cdot\in\mathcal{A}\} : A\in\mathcal{A}\}$) and $\mathcal{H}_\alpha$ can be generalized to *more abstract spaces of functionals* (c.f. [3]).

Sequential Empirical Processes

Sequential empirical processes play an important role in determining the *asymptotic distribution of change-point tests*. A sequential empirical process is the process $(U_n(A,t))_{(A,t)\in\mathcal{A}\times[0,1]}$ given by

$$U_n(A,t) := n^{-1/2}\sum_{i=1}^{[tn]}\big(\mathbf{1}\{X_i\in A\}-\mu(A)\big).$$

Our techniques can also be applied to this situation. However, here one needs a multidimensional version of **(1)**: $\forall f_1,\ldots,f_k\in\mathcal{H}_\alpha, t_1,\ldots,t_k\in[0,1]$

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[nt_1]}\big(f_1(X_i)-\mu(f_1)\big), \ \ldots\ , \sum_{i=1}^{[nt_k]}\big(f_k(X_i)-\mu(f_k)\big)\right) \rightsquigarrow N(0,\Sigma),$$

where Σ may depend on $t_1,\ldots,t_k, f_1,\ldots,f_k$.
We can still work with the same moment bound **(2)** and we only need a slightly stronger entropy condition **(3)**, which still works with all our previous examples.

References

[1] **H. Dehling, O. Durieu and D. Volný (2009).** *"New techniques for empirical processes of dependent data"*. Stochastic Processes and their Applications, **119**, p. 3699-3718.

[2] **O. Durieu, M. Tusche (2013).** *"An empirical process central limit theorem for multidimensional dependent data"*. Journal of Theoretical Probability. DOI: 10.1007/s10959-012-0450-3.

[3] **H. Dehling, O. Durieu, M. Tusche.** *"Empirical Processes of Markov Chains and Dynamical Systems Indexed by Classes of Functions"*. Preprint. Online on http://arxiv.org/abs/1201.2256 [State: January 12., 2012]

DFG Deutsche Forschungsgemeinschaft

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

RUB