Reading Group: Time Series and ML (Week 1)

Presenter: Christis Katsouris
c.katsouris@soton.ac.uk

Department of Economics
School of Economic, Social and Political Sciences
Faculty of Social Sciences

University of Southampton

Reading Group: Time Series and Machine Learning
(School of Mathematical Sciences)

October 17, 2022

Outline

1. Introduction

Article: Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. The Annals of Probability, 94-116.

Author: Professor Bin Yu

- Professor of Statistics & Electrical Engineering and Computer Science
- Departments of Statistics & EECS at University of California, Berkeley

Journal: Annals of Probability

Key Words: Empirical Processes, Stationary Mixing Sequences, Rates of Convergence, Uniform Convergence, m-dependence, Invariance Principles.

1. Introduction

- Empirical Processes: VC (1971) showed the uniform convergence of the empirical processes indexed by V-C classes in the *i.i.d* case.
- Uniform convergence theorem specifically for bounded index classes.
- We use Polland's (1984) linear functional notation and use P instead of E to denote expectations such that

$$Pf = \int f dP = \mathbb{E}\left[f(X_1)\right] = \mathbb{E}\left[f(X_n)\right] \quad \text{for all } n \geq 1. \tag{1}$$

Mixing sequences and metric entropies. The size of the index class of the empirical process can be regulated through metric entropy conditions related to the empirical L^1 norm. Let $\underline{X} = (X_i)_{i \geq 1}$ be a strictly stationary and real-valued sequence with distribution P, which implies that $X_i.i \geq 1$, all have the same distribution P. We denote with

$$\sigma_\ell = \sigma(X_1, X_2, ..., X_\ell) \text{ and } \sigma'_{\ell+k} = \sigma(X_{\ell+k}, X_{\ell+k+1}, ...) \tag{2}$$

Definition

For any sequence $\underline{X}$, the $\beta-$mixing coefficient β_k is defined as follows:

$$\beta_k(\underline{X}) = \frac{1}{2}\sup\left\{\sum_{i=1}^{J}\sum_{j=1}^{J}|P(A_i \cap A_j) - P(A_i).P(B_j)|\right\} \tag{3}$$

where $\{A_i\}$ any finite partition in σ_ℓ and $\{B_j\}$ any finite partition in $\sigma'_{\ell+k}$, $\ell \geq 1$.

Definition

For any sequence $\underline{X}$, the $\phi-$mixing coefficient ϕ_k is defined as follows:

$$\phi_k(\underline{X}) = \sup\Big\{ P(B|A) - P(B)| : A \in \sigma_\ell, B \in \sigma'_{\ell+k}, \ell \geq 1 \Big\}. \tag{4}$$

Definition

The strong mixing coefficients $(a_n)_{n>0}$ of the sequence $(X_i)_{i\in\mathbb{Z}}$ is defined

$$a_n = \sup_{k\in\mathbb{Z}} \alpha\left(\mathcal{F}_k, \mathcal{F}_{k+n}\right) \tag{5}$$

where $\mathcal{F}_k = \sigma\left(X_i : i \leq k\right)$ and $\mathcal{F}_\ell = \sigma\left(X_i : i \geq \ell\right)$. Then, the stationary sequence $(X_i)_{i\in\mathbb{Z}}$ is called a strongly mixing sequence if $\lim_{n\to+\infty} \alpha_n = 0$.

1.1. Preliminaries

Definition (Covering number)

The covering number $N(\epsilon, d, \mathbb{F})$ related to a semimetric d on $\mathbb{F}$ is defined as below

$$N(\epsilon, d, \mathbb{F}) = \min_{m} \left\{ g_1, ..., g_m \text{ in } L^1(P), \text{ such that } \min_{1 \leq j \leq m} d(f, g_j) \leq \epsilon \right\}. \quad (6)$$

for any $f \in \mathbb{F}$. The quantity $\log N(\epsilon, d, \mathbb{F})$ is called the metric entropy at ϵ.

Consider two random semimetrics such that

$$\rho_{1,\mu_n}(f, g) = \frac{1}{n} \sum_{j=1}^{\mu_n} |Y_{j,f-g}| \quad (7)$$

$$\rho^{\star}_{1,\mu_n}(f, g) = \frac{1}{n} \sum_{j=1}^{\mu_n} |Z_{j,f-g}| \quad (8)$$

We shall compare these two semimetrics with the L^1 empirical metric

2. Main Results

The proof focuses on constructing independent block (IB) sequences. An IB sequence is constructed from the original mixing sequence such that the IB sequence is very close in distribution to the mixing sequence.

- We divide the n-sequence $X_n = (X_1, ..., X_n)$ into blocks of length a_n one after the other. Thus, we can construct an independent sequence of blocks where each block has the same distribution as one of the a_n-blocks of the original sequence.
- Thus, $X_n = (X_1, ..., X_n)$ is divided into $2\mu_n$ blocks of length a_n and the remainder block of length $n - 2\mu_n a_n$.

Generally, for $1 \leq j \leq \mu_n$ we have that

$$H_j = \{i : 2(j-1)a_n + 1 \leq i \leq (2j-1)a_n\}, \tag{10}$$

$$T_j = \{i : 2(j-1)a_n + 1 \leq i \leq 2ja_n\}, \tag{11}$$

Then, we denote the random variables corresponding to the H_j and T_j indices as:

$$X(H_j) = \{X_i, i \in H_j\}, \quad X(T_j) = \{X_i, i \in T_j\} \tag{12}$$

2. Main Results

Therefore, we consider the constructed independent a_n-sequence.

Theorem (Uniform Convergence for general families)

Assume that $\mathbb{F}$ *is a permissible index class of functions with an* $L^1(P)$ *envelope function* F. *If* $0 < r_\beta \leq 1$ *and for some* $0 < \alpha_0 < \frac{r_\beta}{r_\beta + 1}$, *and for some* $\epsilon > 0$ *it holds that*

$$\log\left[N(\epsilon, \rho_{1,n}, \mathbb{F})\right] = o_p(n^{\alpha_0}), \tag{13}$$

then for any given $\epsilon > 0$, *we have that*

$$\mathbb{P}\left(\sup_{f \in \mathbb{F}} |P_n f - Pf| > \epsilon\right) \to 0, \; \textit{i.e.}, \sup_{f \in \mathbb{F}} |P_n f - Pf| = 0_{\mathbb{P}}(1), \quad \textit{as } n \to \infty. \tag{14}$$

2. Main Results

Consider two random semimetrics such that

$$\rho_{1,\mu_n}(f,g) = \frac{1}{n}\sum_{j=1}^{\mu_n} |Y_{j,f-g}| \tag{15}$$

$$\rho^\star_{1,\mu_n}(f,g) = \frac{1}{n}\sum_{j=1}^{\mu_n} |Z_{j,f-g}| \tag{16}$$

We shall compare these two semimetrics with the L^1 empirical metric

$$\rho_{1,n}(f,g) = P_n|f-g| \tag{17}$$

Lemma

Suppose that $\mathbb{F}_M$ is a permissible bounded class, and $b_n = O(1)$, as $n \to \infty$. If $\mu_n b_n \to \infty$, then

$$\mathbb{P}\left(\sup_{f\in\mathbb{F}} |P_n f| \geq \epsilon b_n\right) \leq 2\mathbb{P}\left(\sup_{f\in\mathbb{F}} \left|P^\star_{1,\mu_n} f\right| \geq \frac{\epsilon}{4} b_n\right) + 2\mu_n\beta_{\alpha_n}. \tag{18}$$

2. Main Results (Proof)

Notice that the sum of f over the remainder block R_e is uniformly bounded by $M(2a_n)n^{-1} = O(\mu_n^{-1})$, which tends to zero faster than b_n since $\mu_n b_n \to \infty$, and X_{a_n} has the same distribution as

$$X_{1,a_n} = \{X_i : i \in T_j,\ 1 \le j \le \mu_n\} \tag{19}$$

Therefore, for n sufficiently large, we have that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{f\in\mathbb{F}} |P_n f| \ge \epsilon b_n\right) \\
&\quad \le \mathbb{P}\left(\sup_{f\in\mathbb{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{a_n}) + \frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{1,a_n})\right| > \frac{\epsilon}{2}\right) \\
&\quad \le \mathbb{P}\left(\sup_{f\in\mathbb{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{j,f}(X_{a_n})\right| > \frac{\epsilon}{4}\right) + \mathbb{P}\left(\sup_{f\in\mathbb{F}} \left|\frac{1}{n}\sum_{j=1}^{\mu_n} Y_{1,j,f}(X_{1,a_n})\right| > \frac{\epsilon}{4}\right)
\end{aligned}$$

Thus,

$$\mathbb{P}\left(\sup_{f\in\mathbb{F}}|P_n f| \geq \epsilon b_n\right) \equiv 2\mathbb{P}\left(\sup_{f\in\mathbb{F}}\left|\frac{1}{n}\sum_{j=1}^{\mu_n} Z_{j,f}(X_{a_n})\right| > \frac{\epsilon}{4}\right) + 2\mu_n\beta_{a_n}. \quad (20)$$

2. Main Results (Proof)

Lemma (4.3)

Assume that $\mathbb{F}$ *is a permissible index class and that* $\mathbb{F} = \mathbb{M}$. *Then,*

(i) *For* $b_n = O(1)$, *if* $\mu_n \beta_{a_n} = o(1)$ *and* $\log N(\epsilon, \rho_{1,n}, \mathbb{F}) = o_{\mathbb{P}}(b_n)$, *then*

$$\log N\left(\epsilon, \rho^{\star}_{1,\mu_n}, \mathbb{F}\right) = o_p(b_n) \tag{21}$$

(ii) *For any* $0 < s < r_\beta$ *and* $h_n \to \infty$ *as* $n \to \infty$, *let* $b_n = n^{\frac{-s}{1+s}} h_n$, $a_n = \left[n^{\frac{1}{1+s}}\right]$ *and* $\mu_n = \left[\frac{1}{2} n^{\frac{s}{1+s}}\right]$. *Then, for n large enough there exists a* $\delta(\epsilon) > 0$ *for which we have that*

$$\mathbb{P}\left(\sup_{f \in \mathbb{F}} \left|\mathbb{P}^{\star}_{\mu_n}\right| \geq \epsilon b_n\right) \leq 10 \exp\left\{-O(h_n)\right\} + \tag{22}$$

4. Appendix

- In terms of a FCLT, in which we mean that Donsker's normalized polygon line converges weakly in the Skorohod space $D([0,1])$ to some (possibly degenerate) Wiener measure.
- Recently, DMR (1994) improved on Ibragimov's CLT and Herrndorf's CLT: In particular, they obtained a sharp condition on the tail function $t \to \mathbb{P}(|X_0| > t)$ and on the mixing rate implying the CLT and the FCLT results.
- Notice that by sharp condition we mean that, given some rate of mixing and a tail function violating this condition, we mean that, one can construct a strictly stationary sequence $(X_i)_{i \in \mathbb{Z}}$ with corresponding tail function and mixing rate for which the CLT does not hold.

4. Appendix

A measurable function of the following form

$$f = \sum_{i=1}^{n} c_i I_{(a_i, b_i]} \tag{23}$$

is called a *simple function*. We can define the integral of $\int f d\mu$ of a simple function with respect to the measure μ by the following expression

$$\int f d\mu := \sum_{i=1}^{n} c_i \mu\left((a_i, b_i]\right) \tag{24}$$

when this is finite, we then say that f is $\mu-$measurable, and write $f \in L_1(\mu)$.