Reading Group: Time Series and ML (Week 3)

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)

November 18, 2022

Outline

Concentration Inequalities for Dependent Random Variables and Related Studies

- **Article I:** Sara Van De Geer (2002). On Hoeffding's inequality for dependent random variables. Empirical Process Techniques for Dependent Data (Springer Book).

- **Article II:** Kontorovich, L. A., Ramanan, K. (2008). Concentration inequalities for dependent random variables via the martingale method.. The Annals of Probability, 36(6), 2126-2158.

- **Article III:** Chang, J., Chen, X., Wu, M. (2022). Central limit theorems for high dimensional dependent data.. Forthcoming Bernoulli.

1. Introduction

We consider the concept of *pointwise inequalities*, i.e., inequalities that hold uniformly for any $\theta \in \Theta$.

Define the function

$$\psi_\alpha(x) := \exp\left(x^\alpha\right) - 1, \quad \text{for any} \quad x > 0. \tag{1}$$

For a real-valued random variable ξ, we define with

$$\|\xi\|_{\psi_\alpha} := \inf\left\{\lambda > 0 : \mathbb{E}\left[\psi_\alpha\left(\frac{|\xi|}{\lambda}\right)\right] \leq 1\right\} \tag{2}$$

Moreover, we write that $\xi \in \mathcal{L}^q$ for some $q > 0$ if it holds that

$$\|\xi\|_q := \left\{\mathbb{E}\left(|\xi|^q\right)\right\}^{1/q} \tag{3}$$

2. Main Definitions

Definition (Orlicz-norm)

For any convex function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\psi(0) = 0$ and $\psi(x) \to \infty$ as $x \to \infty$ and (real-valued) random variable X, we denote with $\|x\|_\psi$ the Orlicz-norm, which is defined by

$$\|X\|_\psi := \inf\left\{C > 0 : \mathbb{E}\left[\psi\left(\frac{|X|}{C}\right)\right] \leq 1\right\}. \tag{4}$$

- Denote the ℓ^p Orlicz-norm of X by $\|X\|_p$ for $p \in [0, +\infty)$ by setting $\psi(x) = x^p$ and $\|X\|_{e^\gamma}$ the exponential Oricz-norm for $\gamma > 0$ by setting $\psi(x) = \exp(x^\gamma) - 1$ for some $\gamma \geq 1$.
- The function $\psi(x)$ is the convex hull of $x \mapsto \exp(x^\gamma) - 1$ for some $\gamma \in (0, 1)$, which ensures convexity.
- Moreover, when $\boldsymbol{X}$ is a random vector, we define its Orlicz-norm by $\|X\|_\psi := \sup_{\|\boldsymbol{u}\| \leq 1} \|\boldsymbol{u}'\boldsymbol{X}\|_\psi$.

Central limit theorems for high dimensional dependent data

Recall that we define with $S_{n,x} = n^{-1/2} \sum_{t=1}^{n} X_t$. Let $\mathcal{G} \sim \mathcal{N}(0, \Xi)$ where $\Xi := \text{Cov}\left(n^{-1/2} \sum_{t=1}^{n} X_t\right)$. Without loss of generality we assume that $\mathcal{G}$ is independent of $\mathcal{X} = \{X_1, ..., X_n\}$. We write with $X_t = (X_{t,1}, ..., X_{t,p})'$.

Then, the long-run variance of the $j-$th coordinate marginal sequence $\{X_{t,j}\}_{t=1}^{n}$ is defined as below

$$V_{n,j} = \text{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t,j}\right). \tag{5}$$

Therefore, in order to determine the convergence rate of ρ_n for the $\alpha-$mixing sequence $\{X_t\}$, we impose additional regularity conditions.

2. Main Definitions

Assumption (Subexponential moment)

There exists a sequence of constants $B_n \geq 1$ and a universal constant $\gamma_1 \geq 1$ such that $\|X_{t,j}\|_{\psi_{\gamma_1}} \leq B_n$ for all $t \in [n]$ and $j \in [p]$.

Assumption (Decay of $\alpha-$mixing coefficients)

There exist some universal constants $K_1 > 1, K_2 > 0$ and $\gamma_2 > 0$ such that $\alpha_n(k) \leq K_1 e^{(-K_2 k^{\gamma_2})}$ for any $k \geq 1$.

Assumption (Non-degeneracy)

There exists a universal constant $K_3 > 0$ such that $\min_{j\in[p]} V_{n,j} \geq K_3$.

2. Main Definitions

The above condition assumes that the partial sum $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,j}$ is non-degenerated which is necessary to bound the probability of a Gaussian vector taking values in a small region. When $\{X_{t,j}\}_{t\geq 1}$ is stationary, then

$$V_{n,j} := \Gamma_j(0) + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\Gamma_j(k) \tag{6}$$

where $\Gamma_k(k) = \mathrm{Cov}\left(X_{1,j}, X_{k+1,j}\right)$ is the autocovariance of $\{X_{t,j}\}_{t\geq 1}$ at lag k.

Application: Orlicz norm Space

Example

Consider the inverse of the covariance matrix such that

$$\left|\left|\left|\widehat{\boldsymbol{\Sigma}}_i^{-1}\right|\right|\right|_\psi \leq C. \tag{7}$$

For the polynomial case, applying the union bound followed by Markov's inequality we conclude that

$$\max_i \left\|\widehat{\boldsymbol{\Sigma}}_i^{-1}\right\| \leq_{\mathbb{P}} n^{1/p} \quad \text{and} \quad \max_{i,t,j} \left|X_{i,t}^{(j)}\right| \leq_{\mathbb{P}} (nkT)^{1/p}. \tag{8}$$

Lemma

Let X and Y be random elements defined in the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in the metric space (S, d). Then for measurable A and $\delta \geq 0$

Application: Orlicz norm Space

Due to the fact that $x \mapsto \exp\left[\left(x/\|\|X\|\|_{e^\gamma}\right)^\gamma\right]$ is non-decreasing then

$$\mathbb{P}\left(|X| \geq x\right) = \mathbb{P}\Big(\exp\left[\left(|X|/\|\|X\|\|_{e^\gamma}\right)^\gamma\right] \geq \exp\left[\left(x/\|\|X\|\|_{e^\gamma}\right)^\gamma\right]\Big)$$
$$\leq \exp\left[-\left(x/\|\|X\|\|_{e^\gamma}\right)^\gamma\right] \mathbb{E}\exp\left[\left(|X|/\|\|X\|\|_{e^\gamma}\right)^\gamma\right].$$

It holds that,

$$\psi_{e^\gamma}(x) = K_\gamma x \mathbf{1}\left\{0 \leq x \leq a_\gamma\right\} + \left[\exp(x^\gamma) - 1\right]\mathbf{1}\left\{x \geq a_\gamma\right\} \tag{9}$$

where $K_\gamma := \frac{(\exp a_\gamma^\gamma - 1)}{a_\gamma}$ and a_γ is defined as below

$$a_\gamma := \inf\left\{x \in \mathbb{R}_+ : x \geq \left(\frac{1-\gamma}{\gamma}\right)^{1/\gamma}\right\} \tag{10}$$

Moreover, it holds that

$$\left(\frac{1-\gamma}{\gamma}\right)^{1/\gamma} \leq a_\gamma \leq \left(\frac{1}{\gamma}\right)^{1/\gamma}. \tag{11}$$

3. Hoeffding's inequality for dependent random variables

Consider the martingale sequence

$$S_n = \sum_{i=1}^{n} X_i, n \geq 1. \tag{12}$$

Consider the $\mathcal{F}_{i-1}$measurable random variables $K_i > 0$, for $i = 1, 2, \ldots$. Define with $B_0^2 = 0$ and for any $n \geq 1$ such that

$$B_n^2 = \sum_{i=1}^{n} K_i^2 \left\{ 1 + \mathbb{E}\left[\psi\left(\frac{|X_i|}{K_i} \right) \middle| \mathcal{F}_{i-1} \right] \right\} \tag{13}$$

3. Hoeffding's inequality for dependent random variables

Theorem

Let ψ *be an Orlicz function such that it holds that* $\sup_{x,y\to\infty}\psi(x)\psi(y)/\psi(cxy) < \infty$, *for some constant* c. *Suppose that* $\{Z_\theta : \theta \in \Theta\}$ *is a separable stochastic process indexed by* θ *in the pseudo-metric space* (Θ, τ). *Assume that*

$$\|Z_\theta - Z_\vartheta\|_\psi \leq C' \int_0^{\mathrm{diam}(\Theta)} \psi^{-1}\big(D(\delta)\big)d\delta \tag{14}$$

where $\mathrm{diam}(\Theta)$ *is the diameter of* Θ *and* $D(\delta)$ *is the* $\delta-$*packing number.*

3. Hoeffding's inequality for dependent random variables

Corollary

Let W_i be $\mathcal{F}_i-$measurable and $\mathbb{E}(W_i|\mathcal{F}_{i-1}) = 0$ for $i \geq 1$. Suppose that for some constant $c < \infty$ it holds that

$$\mathbb{E}\left(\psi\left(\frac{|W_i|}{c}\right)|\mathcal{F}_{i-1}\right) \leq 1, \quad \textit{almost surely } i = 1, 2, \ldots \tag{15}$$

Key Points:

- Partitioning entropy could be applied to nonstationary time series? This could be the case when considering a discretenized method, such as block of nonstationary time series (i.e., $m-$dependence).
- Notice that this paper doesn't have in depth explanation of the dependence structure. However, the Orlicz norm provides related moment condition for understanding the asymptotic behaviour.

3. Hoeffding's inequality for dependent random variables

- We define with $\phi(d)$ the following quantity

$$\phi(d) = \int_0^d H^{1/2}(\delta, d) d\delta \vee d := \min\left\{ \int_0^d H^{1/2}(\delta, d) d\delta, d \right\}, \quad (16)$$

- What type of dependence structure does the entropy integral $\phi(d)$ introduce? For example, what form this integral would have in the case of Garch processes or for the autoregressive model?
- Are there any related results to Hoeffding's inequality for $\beta-$mixing sequences? (e.g., Geometric ergodicity in autoregressive models)
- To derive the proofs of main results presented in the paper we use that $P(A) \leq \exp\left\{-\beta\alpha + 2\beta^2 b^2\right\}$
- Notice that $P(A')$ is considered to be a negligible probability.
- All probability bounds are derived with respect to S_n, which the sum of stationary martingale differences. Similarly, we can consider expressions for partial-sums or partial-sum self normalized processes.

3. Hoeffding's inequality for dependent random variables

- Related reference: Rademacher Complexity of Stationary Sequences.
- Define with $g(y_1, ..., y_n)$ a measurable function of the data, which for example could be extended to sample moments of estimators.
- Under the assumption of stationary sequences we assume that sub-Gaussianity condition holds in order to obtain probability bounds.
- Furthermore, an important related assumption is the Geometric ergodicity which along with $\beta-$mixing can facilitate the development of further the asymptotic theory in time series model. .
- The theoretical framework presented in the paper shows that the theory can be also extended to the case of M estimators (such as quantile autoregression) using suitable smoothing conditions and deriving the corresponding probability bounds.