Prediction of nonparametric regression models with long memory data.

HAMEL Elhadj ,
Laboratory of mathematics and Applications,
Hassiba Benbouali university of chlef Algeria .

abstract: In this work, the kernel estimator of regression function is investigated for stationary long memory (long range dependence) of the proposed predictor. Under general conditions , we establish the asymptotic normality and propose a confidence bands of these estimators .

Keyword : *regression function, Kernel estimation, Asymptotic normality ,prediction, Long memory,*

Related publications:

–Giraitis, L Koul, H L Surgailis, D. Asymptotic normality of regression estimators with long memory errors, Statist Probab Lett, 29: 317-335 (1996),.
–Palma. W. Long-Memory Time Series, Theory and Methods,, John Wiley, (2007).
– Wang. L, Wang. M, Asymptotics of estimators for nonparametric multivariate regression models with long memory, Appl. Math. J. Chinese Univ. 34(4): 403-422 (2019),

ha2.hamel@gmail.com

4th Conference on
Statistics and
Data Science
Salvador, Brazil (online)
December 1-3, 2022

Introduction & motivation

One of the most important problems in time series analysis is prediction of future observations. Namely, given the observed series $Z_1, Z_2, \ldots, Z_n$, the aim is to predict the unobserved value Z_{n+p}, for some integer $p \geq 1$.The classical approach to this problem is to find some nonparametric estimate of the regression function defined by m(x) = E(Y/X=x) where (X,Y) $\in R^p \times R$.

Kernel regression estimators have been widely studied in the literature when the data are (iid ,mexing and association condition …..) . Nadaraya and Watson (1964) , Schuster (1972) , Devroye and Györfi (1985) , Bosq (1998), Cai (2001), Roussas (2000) for associated data, Laib and .al(2010,2011) for ergodic data , Ferraty and al (2002,2007) for functional data and Wang and al (2019) when the error is long memory.

Long-memory (LM) or Long range dependence (LRD) is a phenomenon that is associated with strong correlations between the present and past values of a stochastic process. In recent years, LM has become an important tool for analyzing dependent data. They have played roles in various areas (physical sciences , finance , hydrology and climatology ….) see, for example , Beran (2013), Doukhan and al. (2003),and Giraitis and al . (2012) and references therein.

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Methods & Model

Definition :

for a stationary process (X_t), $t \in Z$, is said long-memory if the auto-covariance function **$\gamma(k) = cov(X_{t+k}, X_t)$** is not summable, i.e. $\sum_{k=1}^{\infty} \gamma(k) = +\infty.$

Model

Let $Z_1, Z_2, \ldots, Z_n$ be real-valued with long memory random variables forming a strictly stationary sequence, and let φ be a real-valued function defined on R. The prediction problem in a discrete parameter time series given by $X_i = (Z_i, \ldots, Z_{i+d-1}), Y_i = \varphi(Z_{i+d}), i \geq 1, d \geq 1$ $\quad \gamma_Z(i) = cov(Z_0, Z_i) = 0$

and the regression function $m(x) = \mathbb{E}[\varphi(Y_1) \mid X_1 = x] = \mathbb{E}[\varphi(Z_{d+1}) \mid Z_1, \ldots, Z_d], x \in \mathbb{R}^d$

The quantity **$\phi(Z_{j+d})$** is to be predicted on the basis of the past ***d*** r.v.'s ***$Z_1, Z_2, \ldots, Z_{j+d-1}$***. The proposed kernel estimator (Nadaraya-Watson) given by

$$m_n(x) = \frac{g_n(x)}{f_n(x)} \quad \text{where}$$

$$f_n(x) = \frac{1}{nh^d} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right), g_n(x) = \frac{1}{nh^d} \sum_{j=1}^{n} \varphi(Y_j) K\left(\frac{x - X_j}{h}\right)$$

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions

- The natural estimate $m_n^*(x)$ for $m(x)$ is defined by the conditional p.d.f (noted $w_n(.\,/x)$) of Y_1, given $X_1 = x$. where

$$m_n^*(x) = \int_{\mathbb{R}} \varphi(y) w_n(dy \mid x),$$

- It is shown, however that $m_n^*(x) - m_n(x) \underset{n\to\infty}{\to} 0, x \in \mathbb{R}$, pointwise,

- The result of this work is proving asymptotic normality of our kernel estimator.

$$\sqrt{nh^d}\,[m_n(x) - m(x)] \underset{n\to\infty}{\overset{d}{\to}} N\left(0, \sigma^2(x)\right), x \in \mathbb{R}$$

where

$$\sigma^2(x) = \sigma_\varphi^2(x) \int_{\mathbb{R}^d} \frac{K^2(t)dt}{f(x)}, (f(x) > 0),$$

and $\sigma_\varphi^2(x) = \mathcal{E}\left\{[\varphi(Y_1) - m(x)]^2 \mid X_1 = x\right\}$

جامعة حسيبة بن بوعلي الشلف
HASSIBA BENBOUALI UNIVERSITY OF CHLEF
1983

Conclusion : We study in this work the kernel estimator of regression function and solve the prediction problem in a discrete parameter time series under long memory dependence . And we can derive the confidence interval.

4th Conference on
Statistics and
Data Science
Salvador, Brazil (online)
December 1-3, 2022