Quaderni di Statistica
Vol. 1, 1999

Estimation of the asymptotic variance of Kernel smoothers for dependent data

Cira Perna, Francesco Giodano
Dipartimento di Scienze Economiche, Università degli Studi di Salerno
Centro di Specializzazione e Ricerche, Portici (NA)
perna@unisa.it; giordano@unisa.it;

Summary: The aim of this paper is to propose a procedure for the estimation of the asymptotic variance of kernel estimators for dependent data. The method is based on the assumption that the underlying process is stationary and φ-mixing. Under this assumption, the proposed approach, which is itself based on kernel smoothing, is proved to be consistent.

Keywords: Kernel estimator, φ-mixing processes, Autocovariance estimator.

1.Introduction

In the last years, kernel estimators have been found useful and successfully applied to estimate regression functions without reference to a specific parametric class. Most of the literature on this topic is based on the assumption that the observed data are independent (Hardle, 1990; Wand and Jones, 1995). However, there are many settings where it is reasonable to assume the existence of some kind of dependence in the data. This is particularly of interest in time series analysis (Gyorfi *et al.*, 1995) where the independence assumption is clear not acceptable.

In this paper we analyse the properties of kernel estimators in time series analysis. Under the assumption that the underlying process is stationary and φ-mixing, we propose a method to estimate the asymptotic variance of the kernel estimator.

The paper is organised as follows. In the next section kernel estimators are introduced in the context of time series and some conditions and theorems are stated. In section 3 the proposed approach is described. Finally, in section 4, some final comments are presented.

2. Kernel Estimators for dependent data

Let $\{Y_t\}_{t=1,n}$ be a process generated by the model:

$$Y_t = m(X_t) + \varepsilon_t \qquad (1)$$

where X_t is a non stochastic explanatory variable defined on a compact set $\aleph \equiv [a,b] \subset \Re$; $m(.)$ is a smooth function and $\{\varepsilon_t\}$ is a stationary and φ-mixing process (Billingsley, 1968). Then it is:

$$E(\varepsilon_t)=0;\ Var(\varepsilon_t)=\sigma^2<\infty\ Cov(\varepsilon_t, \varepsilon_{t+i})=\gamma(i) \qquad (2)$$

and

$$\varphi_k \to 0 \text{ as } k \to \infty$$

where the φ-mixing coefficients are defined by:

$$\varphi_k = \sup_n \sup_{P(A)>0} \left| P(B|A) - P(A) \right|$$

$\forall A \in \Im_{-\infty}^{n}$, $\forall B \in \Im_{n+k}^{\infty}$, with $\Im_n^m$ $(n, m \in Z \cup \{-\infty, +\infty\})$ being the σ-algebra generated by $\{\varepsilon_t, n \le t \le m\}$.

Throughout this paper we suppose also that:

$$\sum_{i=1}^{\infty} \varphi_i^{1/2} < \infty \tag{3}$$

This hypothesis implies some attractive properties for the moments of φ-mixing variables.

When $X_t = Y_{t-1}$, model (1) becomes an autoregressive non-linear model; it is easy to show that, in this case, all the hypotheses are verified.

In the most general form, the Priestley Chao kernel estimator of the function $m(.)$ is defined (Härdle, 1990), $\forall x \in \Re$, by

$$\hat{m}_h(x) = n^{-1} \sum_{t=1}^{n} K_h(x - X_t) Y_t \tag{4}$$

where the X_t are equally spaced in the compact set, h is the "bandwidth parameter", $K_h(x) = 1/hK(x/h)$ and $K(.)$ is a real bounded function.

The kernel function $K(.)$ and the bandwidth h will be supposed to satisfy the following conditions:

$$\int_{\chi} K(x)dx = 1; \qquad \int_{\chi} zK(z)dz = 0;$$

$$\int_{\chi} z^2 K(z)dz = d_K < \infty \qquad \int_{\chi} K^2(z)dz = c_K < \infty \tag{5}$$

$$h \to 0 \text{ as } n \to \infty; \; nh \to \infty \text{ as } n \to \infty \tag{6}$$

The properties of kernel estimators are stated in the following theorems. For the sake of brevity, the proofs are omitted. More details can be found in Härdle (1990) and Györfi *et al.* (1990).

Theorem 1 (Convergence in probability)

Suppose that model (1) holds where ε_t is a stationary and φ-mixing process. Under the hypotheses (3), (5) and (6), the estimator $\hat{m}(x)$ defined in (4) converges in probability to *m(x)*; that is:

$$\hat{m}(x) \xrightarrow{p} m(x)$$

Theorem 2 (Convergence in distribution)

Suppose that model (1) holds where ε_t is a stationary and φ-mixing process and *m*(.) is twice differentiable. Under the hypotheses (3), (5), (6) and if

$$h \approx n^{-1/5};$$

then

$$\sqrt{nh}(\hat{m}(x) - m(x))$$

converges in distribution to a Normal random variable; that is:

$$\sqrt{nh}(\hat{m}(x) - m(x)) \xrightarrow{D} N(\mu(x), V^2).$$

The mean is equal to:

$$\mu(x) = d_K\left(\frac{m''(x)}{2}\right)$$

and the variance is equal to:

$$V^2 = v^2 c_K$$

where

$$\nu^2 = \sigma^2 + 2\sum_{J=1}^{\infty}\gamma(j)$$

with $\gamma(j)$ being the autocovariance function at lag j of the error term

Following the lemma 1 in Ch. IV of Billingsley (1968), it is:

$$|\mathrm{E}(\varepsilon_t\, \varepsilon_{t+i})| < 2\varphi_t^{1/2}\ \mathrm{Var}(\varepsilon_t) \qquad (7)$$

so if the condition (3) holds, then $\sum_{J=1}^{\infty}\gamma(j)$ converges.

3.The estimation of the asymptotic variance

In order to estimate the asymptotic variance of $\hat{m}(x)$, it is necessary to determinate the quantities c_K and ν^2. The former depends only on the kernel function and then it can be analytically evaluated. Differently the latter depends on $\gamma(i)$ which has to be estimated. Furthermore an appropriate truncation lag M for the series $\gamma(i)$ must be determined.

The conditional autocovariance $\gamma(i;x)$ is defined as:

$$\gamma(i;x) = E\left[(Y_t - m(x))(Y_{t+i} - m(x)) \mid X = x\right]$$

where the dependence on x is due to fact that we are considering the estimate in a fixed point of the compact support.

A method to estimate this quantity is:

$$\hat{\gamma}(i;x) = \frac{1}{n}\sum_{t=1}^{n-i} W_{ti}(x;h)((Y_t - \hat{m}(x)(Y_{t+i} - \hat{m}(x))$$

We propose to estimate the weights using a kernel function i.e.:

$$W_{ti}(x;h) = \frac{\frac{1}{h} K\left(\frac{x - X_{ti}}{h}\right)}{\frac{1}{n-i} \sum_{j=1}^{n-i} \frac{1}{h} K\left(\frac{x - X_{ji}}{h}\right)}$$

where

$$| x - X_{ti} | = \max\{| x - X_t |, | x - X_{t+i} |\}.$$

In this case the kernel estimate is constructed by centring a scaled kernel at each observation The spread of the kernel is determined considering the maximum distance between the fixed point x and the observation at different lags.

The proposed estimator has some asymptotic properties stated in the following theorem.

Theorem 3

Suppose that model (1) holds where $\{\varepsilon_t\}$ is a stationary and φ-mixing process. Under the hypotheses (3), (5), (6), $\hat{\gamma}(i;x)$ is a consistent and asymptotically unbiased estimator; that is:

$$E[\hat{\gamma}(i;x) \mid X = x] \to \gamma(i;x)$$

$$Var[\hat{\gamma}(i;x) \mid X = x] \to 0$$

Sketch of proof

The theorem can be proved using the method proposed in Priestley (1981) and theorem 1. Under the given hypotheses and by means of lemma 1 in Ch. IV of Billingsley (1968), the result holds.

In order to approximate the series of $\gamma(i)$, it is necessary to determine an appropriate number M of significant autocovariance.

To do this we use an approach based on spectral theory.

Let $g(w)$ be the spectral density function:

$$g(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{+\infty}\gamma(s)\exp(-iws)$$

where $\gamma(s)$ is the autocovariance at lag s
Since it is:

$$2\pi g(0) = \sum_{s=-\infty}^{\infty}\gamma(s)$$

a consistent estimator of g(0) is:

$$\hat{g}(0) = \frac{1}{2\pi}\sum_{s=-M}^{M}w(s)\hat{\gamma}(s)$$

where $w(s)$ is defined as follows

$w(s)=w(-s)$; $w(0)=1$; $w(k)=0$ per $|k|>M$

If we consider the function:

$$w(s) = \begin{cases} 1 & \text{if } M \geq s \geq -M \\ 0 & \text{elsewhere} \end{cases} \qquad \text{with } M \in N.$$

it is:

$$2\pi\hat{g}(0) = \sum_{s=-M}^{M}\hat{\gamma}(s)$$

where the parameter M is such that:

$$\frac{M}{n} \rightarrow 0 \text{ as } n\rightarrow\infty.$$

In order to obtain a consistent estimator it is sufficient that $M=(n)^{1/2}$.

In this case the autocovariance estimates depend on the smoothing parameter h of the kernel so the rate of convergence of M must depend on that of h. It is easy to show that the rate of the convergence of $1/M$ must be not greater of that of h

We can suppose that:

$$M=[n^{1/5}]$$

4.Some concluding remarks

In the paper we have analysed the problems related to the use of kernel smoothing in the context of dependent data. To do this we have supposed a regression model with stationary and φ-mixing errors.

We have also proposed a particular kernel method for obtaining an estimator of the asymptotic variance. It is clear that it is possible to use the same estimator with a different spread in the kernel weights

Acknowledgments: The paper is supported by MURST98 "Modelli statistici per l'analisi delle serie temporali".
The work is joint responsability of the authors; F Giordano wrote sections 1, 2 and C. Perna wrote sections 3,4.

References

Billingsley P. (1968) *Convergence of Probability Measures* J. Wiley & Sons, New York

Györfi L. Härdle W. Sarda P. Vieu P. (1990) *Nonparametric Curve Estimation of Time Series* Springer-Verlag, Heidelberg

Härdle W. (1990) *Applied Nonparametric Regression* Cambridge University Press, Cambridge

Härdle W. Marron J.S. (1985) Optimal Bandwith Selection in Nonparametric Regression Function Estimation *Ann. Stat.* Vol. 13, N. 4, pp. 1465-81

Priestley M.B. (1981) *Spectral Analysis and Time Series* Academic Press, London, Vol. I and II

Wand M.P. Jones M.C. (1995) *Kernel smoothing*, Chapman & Hall, London