Statistics and Probability Letters 107 (2015) 369–377

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Moments, errors, asymptotic normality and large deviation principle in nonparametric functional regression

Gery Geenens *

School of Mathematics and Statistics, UNSW Australia, Sydney, Australia

ARTICLE INFO

Article history:
Received 6 July 2015
Received in revised form 9 September 2015
Accepted 9 September 2015
Available online 21 September 2015

Keywords:
Nonparametric regression
Functional regression
Scalar-on-function regression
Nadaraya–Watson estimator
Semi-metric

ABSTRACT

Recently, some nonparametric regression ideas have been extended to the functional context, allowing infinite-dimensional regressors. This paper gives a deep asymptotic study of the functional Nadaraya–Watson estimator, including moments of all orders, errors, asymptotic distribution and large deviation rate.

1. Introduction

Consider a functional variable $\mathcal{X}$, that is a random variable taking on values in some infinite-dimensional space. For instance, $\mathcal{X}$ can be a random function or a continuous time stochastic process, but many other types of random elements can be considered as 'functional'. The study of this kind of variable has become an important research field of statistics since the recent technological progress in measuring devices now allows us to observe spatio-temporal phenomena on arbitrarily fine grids, while their limits previously imposed to discretise any such phenomenon. See Ramsay and Silverman (2002, 2005) for a first review of functional data analysis. In this context, a problem of particular interest is the functional regression (also known as 'scalar-on-function regression'), which aims at linking a real response, say Y, to the functional regressor $\mathcal{X}$, as it is often required in numerous and various applications. This has been the case in medicine (Ratcliffe et al., 2002; Reiss and Ogden, 2010; Goldsmith et al., 2011; Huang et al., 2013), chemometrics (Goutis, 1998; Marx and Eilers, 1999; Ferraty et al., 2010a), climatology (Ferraty et al., 2005; Baíllo and Grané, 2009), and many others.

The first methods proposed to tackle this problem were essentially based on parametric assumptions (e.g. the functional linear model, as in Ramsay and Silverman (2005), Cardot et al. (1999), Fan and Zhang (2000) or Hall and Horowitz (2007) a.o., or the generalised functional linear model, as in James (2002) or Müller and Stadtmüller (2005)). Following some pioneer papers, the monographs of Ferraty and Vieu (2006) and Ferraty and Romain (2011) have, however, popularised the classical nonparametric regression model in the case of a functional regressor. Formally, this model is written

$$Y = \mu(\mathcal{X}) + \varepsilon, \tag{1.1}$$

where $Y \in \mathbb{R}$, $\mathcal{X}$ is assumed to belong to some semi-metric functional space $(\mathcal{S}, (\!(\cdot)\!))$, μ is some unknown operator from $\mathcal{S} \to \mathbb{R}$ satisfying mild regularity conditions and ε is some random disturbance such that $\mathbb{E}(\varepsilon|\mathcal{X}) = 0$ almost surely. This

* Tel.: +61 2 938 57032; fax: +61 2 9385 7123.
E-mail address: ggeenens@unsw.edu.au.

http://dx.doi.org/10.1016/j.spl.2015.09.015
0167-7152/

kind of nonparametric model totally avoids any hazardous rigid parametric assumption on μ. In fact, it is very difficult to figure out how this operator behaves on its infinite-dimensional support. Indeed, although some attempts at extending to the functional framework the usual regression visual tools, such as scatter-plots or residual plots, have been reported in the literature, those typically remain hard to handle and interpret. This leads to an increased risk of model misspecification compared to usual parametric modelling. By contrast, model (1.1) is as flexible as can be, hence its usefulness.

Note that working in a semi-metric space allows one to consider a broader range of topologies associated to the functional space. In particular, the 'classical' functional metric spaces (Hilbert or Banach) endowed with one of their 'usual' norm, for instance the space $L^2_{[0,1]}$ endowed with the L^2-norm $\|\chi\|_2 = \left(\int_0^1 \chi^2(t)dt\right)^{1/2}$, obviously enter the definition of a semi-metric space, but using a genuine semi-norm often leads to interesting results. For example, using some 'projection type' semi-metric may reduce the impact of the so-called 'curse of infinite dimensionality', see Ferraty and Vieu (2006, Chapter 13) and comments in Geenens (2011).

The estimation of μ from a sample of i.i.d. replications of $(\mathcal{X}, Y)$, say $\{(\mathcal{X}_k, Y_k),\ k = 1, \dots, n\}$, is of main interest. For any fixed $\chi \in \mathcal{S}$, this operator is such that $\mu(\chi) = \mathbb{E}(Y|\mathcal{X} = \chi)$, and therefore captures most of the effect of the regressor on the response. The basic nonparametric regression estimator, the so-called Nadaraya–Watson estimator from Nadaraya (1964) and Watson (1964), can readily be generalised to the functional regressor case, and is given by

$$\hat{\mu}(\chi) = \frac{\sum_{k=1}^{n} K((\!(\chi - \mathcal{X}_k)\!)/h)Y_k}{\sum_{k=1}^{n} K((\!(\chi - \mathcal{X}_k)\!)/h)}, \tag{1.2}$$

with K a kernel function from $[0, 1]$ to $\mathbb{R}$ and h a bandwidth. Since it has been introduced, this estimator has been subject to many studies, both theoretical and practical. In addition to the rates of convergence derived in Ferraty and Vieu (2006) and Masry (2005) established its asymptotic normality, while Ferraty et al. (2007) derived explicit expressions for its bias and its variance if $(\!(\cdot)\!)$ is a norm. See also Ferraty et al. (2006), Rachdi and Vieu (2007) and Burba et al. (2009), for related problems.

Actually, it turns out that a key element of all those results is the 'small ball probability' associated to the random process $\mathcal{X}$ and the semi-norm $(\!(\cdot)\!)$ at χ, i.e.

$$\phi_\chi(h) \doteq \mathbb{P}\left((\!(\mathcal{X} - \chi)\!) \le h\right). \tag{1.3}$$

The behaviour of this function when h decreases to zero plays the central role in the theoretical developments, as it quantifies how densely packed the data may be in the considered space with the given topology. Nevertheless, it is out of the scope of this paper to discuss the decay rate of $\phi_\chi(h)$ with respect to $\mathcal{X}$ and $(\!(\cdot)\!)$ (which is done e.g. in Ferraty et al. (2006, section 5)), and our results will be stated in terms of $\phi_\chi(h)$. In particular, the results stated in the above references imply that the optimal bandwidth, in the sense of minimum asymptotic mean squared error of the estimator (1.2), is such that

$$h \sim (n\phi_\chi(h))^{-1/2}. \tag{1.4}$$

Throughout the paper, we consider the case where $(\!(\cdot)\!)$ is a general semi-norm (therefore including norms). Similarly to Geenens (2014) in the vectorial context, we derive in Section 2 explicit expressions of any moment of $\hat{\mu}(\chi)$ of type

$$\mathbb{E}\left((\hat{\mu}(\chi) - \mu(\chi))^\gamma\right) \tag{1.5}$$

for any positive integer γ, and deduce expressions for the ensuing centred moments

$$\mathbb{E}\left((\hat{\mu}(\chi) - \mathbb{E}(\hat{\mu}(\chi)))^\gamma\right)$$

and absolute moments

$$\mathbb{E}\left(|\hat{\mu}(\chi) - \mu(\chi)|^\gamma\right),$$

that is L^γ-errors. Also, interesting observations and consequences are drawn from those results in terms of asymptotic normality and large deviation probability of the estimator. By doing so, this work complements the results of Masry (2005) and Ferraty et al. (2007) and provides new tools to future applied and theoretical studies involving the considered estimator.

2. Main results

2.1. Assumptions and notations

Consider the following case:

Assumption 1. The sample $\{(\mathcal{X}_k, Y_k),\ k = 1, \dots, n\}$ is made up of independent replications of $(\mathcal{X}, Y) \in \mathcal{S} \times \mathbb{R}$, such that $\mathbb{E}(Y|\mathcal{X} = \chi) = \mu(\chi)$.