Large Deviations Theorems in Nonparametric Regression on Functional Data

Mohamed Cherfi

L.S.T.A., Université Pierre et Marie Curie. 175, rue du Chevaleret, 8ème étage, bâtiment A,

75013 PARIS FRANCE.

Abstract.

In this paper we prove large deviations principles for the Nadaraya-Watson estimator of the regression of a real-valued variable with a functional covariate. Under suitable conditions, we show pointwise and uniform large deviations theorems with good rate functions.

AMS Subject Classifications: primary 60F10; 62G08

Keywords: Large deviation; nonparametric regression ; functional data.

1 Introduction

Let $\{(Y_i,\, X_i),\ i \geq 1\}$ be a sequence of independent and identically distributed random vectors. The random variables Y_i are real, with $\mathbb{E}|Y| < \infty$, and the X_i are random vectors with values in a semi-metric space $(\mathcal{X}, d(\cdot,\cdot))$.

Consider now the functional regression model,

$$Y_i := \mathbb{E}(Y|X_i) + \varepsilon_i = r(X_i) + \varepsilon_i \quad i = 1, \ldots, n, \tag{1}$$

where r is the regression operator mapping $\mathcal{X}$ onto $\mathbb{R}$, and the ε_i are real variables such that, for all i, $\mathbb{E}(\varepsilon_i|X_i) = 0$ and $\mathbb{E}(\varepsilon_i^2|X_i) = \sigma_\varepsilon^2(X_i) < \infty$. Note that in practice $\mathcal{X}$ is a normed space which can be of infinite dimension (e.g., Hilbert or Banach space) with norm $\|\cdot\|$ so that $d(x,x') = \|x - x'\|$, which is the case in this paper.

Ferraty and Vieu (2004) provided a consistent estimate for the nonlinear regression operator r, based on the usual finite-dimensional smoothing ideas, that is

$$\widehat{r}_n(x) := \frac{\sum_{i=1}^n Y_i K\left(\frac{\|x - X_i\|}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\|X_i - x\|}{h_n}\right)}, \tag{2}$$

where $K(\cdot)$ is a real-valued kernel and h_n the bandwidth, is a sequence of positive real numbers converging to 0 as $n \longrightarrow \infty$. Note that the bandwidth h_n depends on n, but we drop this index for simplicity. In what follows $K_h(u)$ stands for $K\left(\dfrac{u}{h}\right)$. The estimator defined in (2) is a generalization to the functional framework of the classical Nadaraya-Watson regression estimator. The asymptotic properties of this estimate have been studied extensively by several authors, we cite among others Ferraty *et al.* (2007), for a complete survey see the monograph by Ferraty and Vieu (2006).

The large deviations behavior of the Nadaraya-Watson estimate of the regression function, have been studied at first by Louani (1999), sharp results have been obtained by Joutard (2006) in the univariate framework. In the multidimensional case Mokkadem *et al.* (2008) obtained pointwise large and moderate deviations results for the Nadaraya-Watson and recursive kernel estimators of the regression.

E-mail address: mohamed.cherfi@upmc.fr

In this work, we are interested in the problem of establishing large deviations principles of the regression operator estimate $\widehat{r}_n(\cdot)$. The results stated in the paper deal with pointwise and uniform large deviations probabilities of $\widehat{r}_n(\cdot)$ from $r(\cdot)$. The organization of the paper is as follows, in Section 2 we will state pointwise and uniform large deviations results. The proofs are given in section 3.

2 Results

Let $F_x(h) = P[\|X_i - x\| \leq h]$, be the cumulative distribution of the real variable $W_i = \|X_i - x\|$. As in Ferraty *et al.* (2007), let φ be the real valued function defined by

$$\varphi(u) = \mathbb{E}\big\{r(X) - r(x)\big|\|X - x\| = u\big\}. \tag{3}$$

Before stating our results, we will consider the following conditions.

(C.1) The kernel K is positive, with compact support $[0, 1]$ of class $\mathcal{C}^1$ on $[0, 1)$, $K(1) > 0$ and its derivative K' exists on $[0, 1)$ and $K'(u) < 0$.

(C.2) K is Lipschitz.

(C.3) The operator r verifies the following Lipschitz property:

$$\forall (u, v) \in \mathcal{X}^2,\ \exists C,\ |r(u) - r(v)| \leq C\|u - v\|^{\beta} \tag{4}$$

(C.4) There exist three functions $\ell(\cdot)$, $\phi(\cdot)$ (supposed increasing and strictly positive and tending to zero as h goes to zero) and $\zeta_0(\cdot)$ such that

- (i) $F_x(h) = \ell(x)\phi(h) + o(\phi(h))$,
- (ii) for all $u \in [0, 1]$, $\lim_{h\to 0} \frac{\phi(uh)}{\phi(h)} =: \lim_{h\to 0} \zeta_h(u) = \zeta_0(u)$.

C.5 $\varphi'(0)$ exists.

There exist many examples fulfilling the decomposition mentioned in condition (C.4), see for instance Proposition 1 in Ferraty *et al.* (2007). The conditions stated above are classical in nonparametric estimation for functional data, see for instance Ferraty *et al.* (2007) and references cited therein.
Let now introduce the following functions,

$$I(t) = \exp\{-t\lambda K(1)\} - 1 + t\lambda \int_0^1 K'(u) \exp\{-t\lambda K(u)\}\zeta_0(u)\, \mathrm{d}u; \tag{5}$$

$\Gamma_x^+(\lambda) = \inf_{t>0}\{\ell(x)I(t)\}$; $\Gamma_x^-(\lambda) = \inf_{t>0}\{\ell(x)I(-t)\}$ and $\Gamma_x(\lambda) = \max\{\Gamma_x^+(\lambda); \Gamma_x^-(\lambda)\}$.
Let x be a an element of the functional space $\mathcal{X}$ and $\lambda > 0$. Our first theorem deals with pointwise large deviations probabilities.

Theorem 1 *Assume that the conditions* (C.1)–(C.5) *are satisfied. If* $n\phi(h) \longrightarrow \infty$, *then for any* $\lambda > 0$ *and any* $x \in \mathcal{X}$, *we have*

(a)
$$\lim_{n\to\infty} \frac{1}{n\phi(h)} \log P\big(\widehat{r}_n(x) - r(x) > \lambda\big) = \Gamma_x^+(\lambda) \tag{6}$$

(b)
$$\lim_{n\to\infty} \frac{1}{n\phi(h)} \log P\big(\widehat{r}_n(x) - r(x) < -\lambda\big) = \Gamma_x^-(\lambda) \tag{7}$$

(c)
$$\lim_{n\to\infty} \frac{1}{n\phi(h)} \log P\big(|\widehat{r}_n(x) - r(x)| > \lambda\big) = \Gamma_x(\lambda) \tag{8}$$

To establish uniform large deviations principles for the regression estimator we need the following assumptions. Let $\mathcal{C}$ be some compact subset of $\mathcal{X}$ and $B(z_k, \xi)$ a ball centered at $z_k \in \mathcal{X}$ with radius ξ, such that for any $\xi > 0$,

$$\mathcal{C} \subset \bigcup_{k=1}^{\tau} B(z_k, \xi), \tag{9a}$$

$$\exists \alpha > 0, \quad \exists C > 0, \quad \tau \xi^{\alpha} = C. \tag{9b}$$

The above conditions on the covering of the compact set $\mathcal{C}$ by a finite number of balls, the geometric link between the number of balls τ and the radius ξ are necessary to prove uniform convergence in the context of functional non-parametric regression and many functional non-parametric settings, see the discussion in Ferraty and Vieu (2008).

Before stating the Theorem about the uniform version of our result, we introduce the following function

$$g(\lambda) = \sup_{x \in \mathcal{C}} \Gamma_x(\lambda). \tag{10}$$

Theorem 2 *Assume that the conditions* (C.1)–(C.5) *are satisfied. If* $n\phi(h) \longrightarrow \infty$, *then for any compact set* $\mathcal{C} \subset \mathcal{X}$ *satisfying conditions (9) and for any* $\lambda > 0$,

$$\lim_{n \to \infty} \frac{1}{n\phi(h)} \log P\big(\sup_{x \in \mathcal{C}} |\widehat{r}_n(x) - r(x)| > \lambda\big) = g(\lambda) \tag{11}$$

3 Proofs

3.1 Proof of Theorem 1

We only prove the statement (6), (7) is derived in the same way.

(a) Write

$$Z_n = \sum_{i=1}^{n} \big\{ Y_i - r(x) - \lambda \big\} K_h(\|X_i - x\|)$$

Define $\Phi_x^{(n)}(t) := \mathbb{E} \exp(tZ_n)$ to be the moment generating function of Z_n. To prove the large deviations principles, we seek the limit of $\dfrac{1}{n\phi(h)} \log \Phi_x^{(n)}(t)$ as $n \longrightarrow \infty$.

Observe that

$$\Phi_x^{(n)}(t) = \left\{ 1 + \mathbb{E}\Big(\exp\{t[r(X_1) - r(x) - \lambda] K_h(\|X_1 - x\|)\} - 1 \Big) \right\}^n.$$

Using the definition of the function φ in (3), we can write

$$\begin{aligned}
\Phi_x^{(n)}(t) &= \left\{ 1 + \mathbb{E}\Big(\exp\{t[\varphi(\|X_1 - x\|) - \lambda] K_h(\|X_1 - x\|)\} - 1 \Big) \right\}^n, \\
&= \left\{ 1 + \int_0^h \Big(\exp\{t[\varphi(u) - \lambda] K_h(u)\} - 1 \Big) \, \mathrm{d}F_x(u) \right\}^n. \\
&= \left\{ 1 + \int_0^1 \Big(\exp\{t[\varphi(hu) - \lambda] K(u)\} - 1 \Big) \, \mathrm{d}F_x(hu) \right\}^n.
\end{aligned}$$

By (C.5), using a first order Taylor expansion of φ about zero, we obtain

$$\Phi_x^{(n)}(t) = \left\{ 1 + \int_0^1 \Big(\exp\{t[hu\varphi'(0) - \lambda + o(1)] K(u)\} - 1 \Big) \, \mathrm{d}F_x(hu) \right\}^n,$$

Integrating by parts and by (C.1), we have

$$\begin{aligned}\Phi_x^{(n)}(t) &= \Big\{1+\big(\exp\{t[h\varphi'(0)-\lambda]K(1)\}-1\big)F_x(h)\\ &- \int_0^1 t[h\varphi'(0)K(u)+K'(u)(uh\varphi'(0)-\lambda)]\\ &\quad \exp\{t[hu\varphi'(0)-\lambda]K(u)\}F_x(hu)\,\mathrm{d}u\Big\}^n.\end{aligned}$$

Therefore,

$$\begin{aligned}\log\Phi_x^{(n)}(t) &= n\log\Big\{1+\big(\exp\{t[h\varphi'(0)-\lambda]K(1)\}-1\big)F_x(h)\\ &- \int_0^1 t[h\varphi'(0)K(u)+K'(u)(uh\varphi'(0)-\lambda)]\\ &\quad \exp\{t[hu\varphi'(0)-\lambda]K(u)\}F_x(hu)\,\mathrm{d}u\Big\}.\end{aligned}$$

Using Taylor expansion of $\log(1+v)$ about $v=0$, we obtain

$$\begin{aligned}\log\Phi_x^{(n)}(t) &= n\Big\{\big(\exp\{t[h\varphi'(0)-\lambda]K(1)\}-1\big)F_x(h)\\ &- \int_0^1 t[h\varphi'(0)K(u)+K'(u)(uh\varphi'(0)-\lambda)]\\ &\quad \exp\{t[hu\varphi'(0)-\lambda]K(u)\}F_x(hu)\,\mathrm{d}u+O(h)\Big\}.\end{aligned}$$

Hence, from Assumption (C.4) (ii) it follows that

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n\phi(h)}\log\Phi_x^{(n)}(t) &= \ell(x)\Big\{\exp\{-t\lambda K(1)\}-1+t\lambda\int_0^1 K'(u)\exp\{-t\lambda K(u)\}\zeta_0(u)\,\mathrm{d}u\Big\}\\ &= \ell(x)I(t).\end{aligned}$$

Using Theorem of Plachky and Steinebach (1975), the proof of the theorem can be completed as in Louani (1998).

(c) Observe that for any $x\in\mathcal{X}$,

$$\max\{P(\widehat{r}_n(x)-r(x)>\lambda);\ P(\widehat{r}_n(x)-r(x)<-\lambda)\}\le P(|\widehat{r}_n(x)-r(x)|>\lambda)$$

and

$$P(|\widehat{r}_n(x)-r(x)|>\lambda)\le 2\max\{P(\widehat{r}_n(x)-r(x)>\lambda);\ P(\widehat{r}_n(x)-r(x)<-\lambda)\}.$$

Hence,

$$\Gamma_x(\lambda)\le\lim_{n\to\infty}\frac{1}{n\phi(h)}\log P(|\widehat{r}_n(x)-r(x)|>\lambda)\le\max\{\Gamma_x^+(\lambda);\ \Gamma_x^-(\lambda)\}=\Gamma_x(\lambda),$$

which complete the proof.

3.2 Proof of Theorem 2

First for any $x_0\in\mathcal{C}$, by Theorem 1 we have

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\phi(h)}\log P\big(\sup_{x\in\mathcal{C}}|\widehat{r}_n(x)-r(x)|>\lambda\big) &\ge \liminf_{n\to\infty}\frac{1}{n\phi(h)}\log P\big(|\widehat{r}_n(x_0)-r(x_0)|>\lambda\big)\\ &\ge \Gamma_{x_0}(\lambda).\end{aligned}$$

Hence

$$\liminf_{n\to\infty} \frac{1}{n\phi(h)} \log P\big(\sup_{x\in\mathcal{C}} |\widehat{r}_n(x) - r(x)| > \lambda\big) \geq g(\lambda). \tag{12}$$

To prove the reverse inequality, we note that by conditions (9) it follows

$$\sup_{x\in\mathcal{C}} |\widehat{r}_n(x) - r(x)| \leq \max_{1\leq k\leq\tau} \sup_{x\in B(z_k,\xi)} |\widehat{r}_n(x) - r(x)|. \tag{13}$$

Hence,

$$\sup_{x\in B(z_k,\xi)} |\widehat{r}_n(x) - r(x)| \leq \sup_{x\in B(z_k,\xi)} |\widehat{r}_n(x) - \widehat{r}_n(z_k)| + \sup_{x\in B(z_k,\xi)} |r(z_k) - r(x)| + |\widehat{r}_n(z_k) - r(z_k)|. \tag{14}$$

Using the fact that K is Lipschitz by condition (C.2), there exists $C > 0$ so that

$$\sup_{x\in B(z_k,\xi)} |\widehat{r}_n(x) - \widehat{r}_n(z_k)| \leq \frac{C\xi}{n\phi(h)h} \sum_{i=1}^{n} |Y_i|.$$

For n sufficiently large, we choose ξ according to the preassigned $\epsilon > 0$, so that

$$\sup_{x\in B(z_k,\xi)} |\widehat{r}_n(x) - \widehat{r}_n(z_k)| \leq \epsilon. \tag{15}$$

Moreover, r is Lipschitz, hence for a suitable choice of ξ

$$\sup_{x\in B(z_k,\xi)} |r(z_k) - r(x)| \leq \epsilon. \tag{16}$$

Finally, (13)-(16) yield

$$\sup_{x\in\mathcal{C}} |\widehat{r}_n(x) - r(x)| \leq \max_{1\leq k\leq\tau} \big\{|\widehat{r}_n(z_k) - r(z_k)| + 2\epsilon\big\}, \tag{17}$$

which implies

$$P\big(\sup_{x\in\mathcal{C}} |\widehat{r}_n(x) - r(x)| > \lambda\big) \leq \sum_{k=1}^{\tau} P\big(|\widehat{r}_n(z_k) - r(z_k)| > \lambda - 2\epsilon\big).$$

Thus,

$$P\big(\sup_{x\in\mathcal{C}} |\widehat{r}_n(x) - r(x)| > \lambda\big) \leq \tau \max_{1\leq k\leq\tau} P\big(|\widehat{r}_n(z_k) - r(z_k)| > \lambda - 2\epsilon\big).$$

It follows, by Theorem 1, that

$$\limsup_{n\to\infty} \frac{1}{n\phi(h)} \log P\big(\sup_{x\in\mathcal{C}} |\widehat{r}_n(x) - r(x)| > \lambda\big) \leq \inf_{t>0} \sup_{x\in\mathcal{C}} \ell(x) I_\epsilon(t),$$

where

$$I_\epsilon(t) = \exp\{-(t\lambda - 2\epsilon)K(1)\} - 1 + t(\lambda - 2\epsilon) \int_0^1 K'(u) \exp\{-t(\lambda - 2\epsilon)K(u)\}\zeta_0(u)\, \mathrm{d}u.$$

By continuity arguments, and the fact that

$$\inf_{t>0} \sup_{x\in\mathcal{C}} \ell(x) I(t) = \sup_{x\in\mathcal{C}} \inf_{t>0} \ell(x) I(t),$$

we obtain

$$\limsup_{n\to\infty} \frac{1}{n\phi(h)} \log P\big(\sup_{x\in\mathcal{C}} |\widehat{r}_n(x) - r(x)| > \lambda\big) \leq g(\lambda). \tag{18}$$

Combining (12) and (18), we see that the limit exists which is $g(\lambda)$.

References

Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *J. Nonparametric Stat.*, **16**, pp. 111–125.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics, NewYork.

Ferraty, F. and Vieu, P. (2008). Erratum of: Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametr. Stat.*, **20**, pp. 187–189.

Ferraty, F., Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: Inference and practical aspects. *Austral. New Zealand J. Statist.*, **49**, pp. 267–286.

Joutard, C. (2006). Sharp large deviations in nonparametric estimation. *J. Nonparam. Stat.*, **18**, pp. 293–306.

Louani, D. (1998). Large deviations limit theorems for the kernel density estimator. *Scandinavian Journal of Statistics*, **25**, pp. 243–253.

Louani, D. (1999). Some large deviations limit theorems in conditional nonparametric statistics. *Statistics*, **33**, pp. 171–196.

Mokkadem, A., Pelletier, M. and Thiam, B. (2008). Large and moderate deviations principles for kernel estimators of a multivariate regression. *Math. Methods Statist.*, **17**, pp. 146–172.

Plachky, D. and Steinebach, J. (1975). A theorem about probabilities of large deviations with an application to queuing theory. *Period. Math. Hungar.*, **6**, no. 4, 343–345.