Opuscula Math. 39, no. 5 (2019), 733–746
https://doi.org/10.7494/OpMath.2019.39.5.733 Opuscula Mathematica

LARGE AND MODERATE DEVIATION PRINCIPLES FOR NONPARAMETRIC RECURSIVE KERNEL DISTRIBUTION ESTIMATORS DEFINED BY STOCHASTIC APPROXIMATION METHOD

Yousri Slaoui

Communicated by Andrzej Kozek

Abstract. In this paper we prove large and moderate deviations principles for the recursive kernel estimators of a distribution function defined by the stochastic approximation algorithm. We show that the estimator constructed using the stepsize which minimize the Mean Integrated Squared Error (MISE) of the class of the recursive estimators defined by Mokkadem *et al.* gives the same pointwise large deviations principle (LDP) and moderate deviations principle (MDP) as the Nadaraya kernel distribution estimator.

Keywords: distribution estimation, stochastic approximation algorithm, large and moderate deviations principles.

Mathematics Subject Classification: 62G07, 62L20, 60F10.

1. INTRODUCTION

Let us first recall that a $\mathbb{R}^m$-valued sequence $(Z_n)_{n\geq 1}$ satisfies a LDP with speed (ν_n) and good rate function I if:

1. (ν_n) is a positive sequence such that $\lim_{n\to\infty} \nu_n = \infty$;
2. $I : \mathbb{R}^m \to [0,\infty]$ has compact level sets;
3. for every borel set $B \subset \mathbb{R}^m$,

$$\begin{aligned} -\inf_{x\in \overset{\circ}{B}} I(x) &\leq \liminf_{n\to\infty} \nu_n^{-1} \log \mathbb{P}\left[Z_n \in B\right] \\ &\leq \limsup_{n\to\infty} \nu_n^{-1} \log \mathbb{P}\left[Z_n \in B\right] \leq -\inf_{x\in\overline{B}} I(x), \end{aligned} \tag{1.1}$$

where $\overset{\circ}{B}$ and $\overline{B}$ denote the interior and the closure of B, respectively. Moreover, let (v_n) be a nonrandom sequence that goes to infinity; if $(v_n Z_n)$ satisfies a LDP, then (Z_n) is said to satisfy a MDP.

© Wydawnictwa AGH, Krakow 2019

Let $X_1, \ldots, X_n$ be independent, identically distributed of random variables, and let f and F denote respectively the probability density of X_1 and the distribution function of X_1. The LDP and MDP problems arise in the theory of statistical inference quite naturally. For estimation of the distribution function F, we apply a stochastic algorithm, which approximates the function F. In fact we apply a stochastic algorithm for search of zero of the function $h : y \to F(x) - y$ at a given point x. We thus proceed in the following way: (i) we set $F_0(x) \in [0, 1]$; (ii) for all $n \geq 1$, we set

$$F_n(x) = F_{n-1}(x) + \gamma_n W_n(x),$$

where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero and $W_n(x)$ can be interpreted as an "observation" of the function h at the point $F_{n-1}(x)$. We shall choose $W_n(x)$ such that $\mathbb{E}\left[W_n(x)|\mathcal{F}_{n-1}\right] = 0$, where $\mathcal{F}_{n-1}$ stands for the σ-algebra of the events occurring up the time $n-1$. To define $W_n(x)$, we follow the approach of [14, 15] and of [21] and introduce a kernel K satisfying $\int_{\mathbb{R}} K(x)dx = 1$, a function $\mathcal{K}$ defined by $\mathcal{K}(z) = \int_{-\infty}^{z} K(u)\, du$, and a bandwidth (h_n), which is a sequence of positive real numbers that goes to zero, and set

$$W_n(x) = \mathcal{K}\left(h_n^{-1}(x - X_n)\right) - F_{n-1}(x).$$

The stochastic approximation algorithm we introduce to recursively estimate the distribution function F at the point x can thus be written as

$$F_n(x) = (1 - \gamma_n)\, F_{n-1}(x) + \gamma_n \mathcal{K}\left(\frac{x - X_n}{h_n}\right). \tag{1.2}$$

Let us recall the estimators introduced in [9] to estimate recursively a probability density f at the point x which are given by

$$f_n(x) = (1 - \gamma_n)\, f_{n-1}(x) + \gamma_n h_n^{-1} K\left(\frac{x - X_n}{h_n}\right). \tag{1.3}$$

It is well known that for non-recursive kernel estimators the optimal speed of decrease of the window width is different for the density function and for the distribution estimation

Recently, LDP and MDP results have been proved for the following cases:

(a) the recursive density estimators defined by stochastic approximation method in [16].
(b) the nonrecursive Nadaraya's kernel distribution estimator ([10]) in [17].
(c) the recursive regression estimators defined by stochastic approximation method in [19, 20].
(d) the nonrecursive regression estimator Nadaraya–Watson ([11, 23]) in [7] and [5].

The purpose of this paper is to establish LDP and MDP for the recursive distribution estimators F_n defined by stochastic approximation algorithm (1.2).

We first establish pointwise LDP for the recursive kernel distribution estimators defined by the stochastic approximation algorithm (1.2). It turns out that the rate function depends on the choice of the stepsize (γ_n). In the first part of this paper we focus on the following two special cases: (1) $(\gamma_n) = \left(n^{-1}\right)$ and (2) $(\gamma_n) = \left(h_n \left(\sum_{k=1}^n h_k\right)^{-1}\right)$. The first one belongs to the subclass of recursive kernel density estimators which have a minimum MSE or MISE and the second choice belongs to the subclass of recursive kernel density estimators which have a minimum variance (see [9]).

We show that using the stepsize $(\gamma_n) = \left(n^{-1}\right)$ and bandwidths defined as $h_n = h(n)$ for all n, where h is a regularly varying function with exponent $(-a)$, $a \in]0,1[$, that the sequence $(F_n(x) - F(x))$ satisfies a LDP with speed (n) and the rate function defined as follows:

$$\begin{cases} I_x : t \to F(x) I\left(1 + \frac{t}{F(x)}\right) & \text{if } F(x) \neq 0, \\ I_x(0) = 0 \text{ and } I_x(t) = +\infty \text{ for } t \neq 0 & \text{if } F(x) = 0, \end{cases} \tag{1.4}$$

where $I(t) = t\ln t - t + 1$ is a conjugate function of $\psi(u) = \exp(u) - 1$. Moreover, we show that using the stepsize $(\gamma_n) = \left(h_n \left(\sum_{k=1}^n h_k\right)^{-1}\right)$, with bandwidths defined as $h_n = cn^{-a}$, with $a \in]0,1[$ and $c > 0$, that the sequence $(F_n(x) - F(x))$ satisfies a LDP with speed (n) and the rate function defined as follows:

$$\begin{cases} I_{x;a} : t \to F(x) I_a\left(\frac{1}{1-a} + \frac{t}{F(x)}\right) & \text{if } F(x) \neq 0, \\ I_{x;a}(0) = 0 \text{ and } I_{x;a}(t) = +\infty \text{ for } t \neq 0 & \text{if } F(x) = 0, \end{cases} \tag{1.5}$$

where

$$I_a(t) = \sup_{u \in \mathbb{R}} \{ut - \psi_a(u)\},$$

$$\psi_a(u) = \int_0^1 \left(\exp\left(us^{-a}\right) - 1\right) ds.$$

Our second aim is to provide pointwise MDP for the distribution estimator defined by the stochastic approximation algorithm (1.2). In this case, we consider more general stepsizes defined as $\gamma_n = \gamma(n)$ for all n, where γ is a regularly varying function with exponent $(-\alpha)$, $\alpha \in]1/2, 1]$. Throughout this paper we will use the following notation:

$$\xi = \lim_{n\to+\infty} (n\gamma_n)^{-1}. \tag{1.6}$$

For any positive sequence (v_n) satisfying

$$\lim_{n\to\infty} v_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \gamma_n v_n^2 = 0$$

and general bandwidths (h_n), we prove that the sequence

$$v_n (F_n(x) - F(x))$$

satisfies a LDP of speed $\left(1/\left(\gamma_n v_n^2\right)\right)$ and rate function $J_{\alpha;x}(.)$ defined by

$$\begin{cases} J_{\alpha;x} : t \to \frac{t^2(2-\alpha\xi)}{2F(x)} & \text{if } F(x) \neq 0, \\ J_{\alpha;x}(0) = 0 \text{ and } J_{\alpha;x}(t) = +\infty \text{ for } t \neq 0 & \text{if } F(x) = 0. \end{cases} \tag{1.7}$$

2. ASSUMPTIONS AND MAIN RESULTS

We define the following class of regularly varying sequences.

Definition 2.1. Let $\gamma \in \mathbb{R}$ and $(v_n)_{n\geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if

$$\lim_{n\to+\infty} n\left[1-\frac{v_{n-1}}{v_n}\right]=\gamma. \tag{2.1}$$

Condition (2.1) was introduced by [3] to define regularly varying sequences (see also [1]), and by [8] in the context of stochastic approximation algorithms. Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^\gamma(\log n)^b$, $n^\gamma(\log\log n)^b$, and so on.

2.1. POINTWISE LDP FOR THE DISTRIBUTION ESTIMATOR DEFINED BY THE STOCHASTIC APPROXIMATION ALGORITHM (1.2)

2.1.1. Choices of (γ_n) minimizing the MISE of f_n

It was shown in [9] that to minimize the MISE of the recursive kernel density estimators noted f_n, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and must satisfy $\lim_{n\to\infty} n\gamma_n = 1$. The most simple example of stepsize belonging to $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = 1$ is $(\gamma_n) = \left(n^{-1}\right)$. For this choice of stepsize, the estimator F_n defined by (1.2) can be rewritten as

$$F_n(x)=\frac{1}{n}\sum_{k=1}^{n}\mathcal{K}\left(\frac{x-X_k}{h_k}\right).$$

This estimator was considered by [4].

To establish pointwise LDP for F_n in this case, we assume the following assumptions:

(L1) $K:\mathbb{R}\to\mathbb{R}$ is a bounded and integrable function satisfying $\int_{\mathbb{R}} K(z)\,dz = 1$, and $\int_{\mathbb{R}} zK(z)\,dz = 0$.
(L2) (i) $(h_n) \in \mathcal{GS}(-a)$ with $a \in]0,1[$.
(ii) $(\gamma_n) = \left(n^{-1}\right)$.

The following theorem gives the pointwise LDP for F_n in this case.

Theorem 2.2 (Pointwise LDP for Isogai and Hirose estimator)**.** *Let Assumptions* (L1) *and* (L2) *hold and assume that F is continuous at x. Then, the sequence* $(F_n(x) - F(x))$ *satisfies a LDP with speed* (n) *and rate function defined by* (1.4).

2.1.2. Choices of (γ_n) minimizing the variance of f_n

It was shown in [9] that to minimize the asymptotic variance of the recursive kernel density estimators f_n, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and must satisfy $\lim_{n\to\infty} n\gamma_n = 1-a$. The most simple example of stepsize belonging to $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty} n\gamma_n = 1 - a$ is $(\gamma_n) = \left((1-a)\,n^{-1}\right)$, an other stepsize satisfying this conditions is $(\gamma_n) = \left(h_n\left(\sum_{k=1}^n h_k\right)^{-1}\right)$, in this case the estimator F_n defined by (1.2) can be rewritten as

$$F_n(x) = \frac{1}{\sum_{k=1}^n h_k} \sum_{k=1}^n h_k \mathcal{K}\left(\frac{x - X_k}{h_k}\right).$$

Moreover, in order to establish pointwise LDP for F_n in this case, we assume that:

(L3) (i) $(h_n) = (cn^{-a})$ with $a \in]0, 1[$ and $c > 0$.
(ii) $(\gamma_n) = \left(h_n\left(\sum_{k=1}^n h_k\right)^{-1}\right)$.

The following theorem gives the pointwise LDP for F_n in this case.

Theorem 2.3 (Pointwise LDP)**.** *Let Assumptions* (L1) *and* (L3) *hold and assume that F is continuous at x. Then, the sequence $(F_n(x) - F(x))$ satisfies a LDP with speed (n) and rate function defined by* (1.5).

2.2. POINTWISE MDP FOR THE DISTRIBUTION ESTIMATOR DEFINED BY THE STOCHASTIC APPROXIMATION ALGORITHM (1.2)

Let (v_n) be a positive sequence. We assume that

(M1) $K : \mathbb{R} \to \mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)\,dz = 1$, and $\int_{\mathbb{R}} zK(z)\,dz = 0$ and $\int_{\mathbb{R}} z^2 |K(z)|\,dz < \infty$.
(M2) (i) $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in]1/2, 1]$.
(ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in]0, \alpha[$.
(iii) $\lim_{n\to\infty} (n\gamma_n) \in]\min\{2a, (\alpha + a)/2\}, \infty]$.
(M3) F is bounded, twice differentiable, and $F^{(2)}(x)$ is bounded.
(M4) $\lim_{n\to\infty} v_n = \infty$ and $\lim_{n\to\infty} \gamma_n v_n^2 = 0$.

The following theorem gives the pointwise MDP for F_n.

Theorem 2.4 (Pointwise MDP for the recursive estimators defined by (1.2))**.** *Let Assumptions* (M1)–(M4) *hold and assume that F is continuous at x. Then, the sequence $(F_n(x) - F(x))$ satisfies a MDP with speed $\left(1/\left(\gamma_n v_n^2\right)\right)$ and rate function $J_{\alpha;x}$ defined in* (1.7).

3. CONCLUSION

The purpose in this paper is to prove LDP and MDP for the recursive kernel estimators of a distribution function defined by the stochastic approximation algorithm introduced by Slaoui ([18]).

Moreover, we showed that the estimator constructed using the stepsize which minimize the (MISE) of the recursive estimators defined by the stochastic approximation algorithm ([18]) gives the same pointwise LDP and MDP as the nonrecursive Nadaraya's distribution kernel estimator.

In conclusion, the proposed method allowed us to obtain quite similar results as the nonrecrsive Nadaraya's distribution kernel estimator. Moreover, we plan to extend the i.i.d relationship to the Markovian context (see [6, 13] and [22]).

4. PROOFS

Throughout this section we use the following notation:

$$\Pi_n = \prod_{j=1}^{n} (1-\gamma_j), \tag{4.1}$$

$$Y_n = \mathcal{K}\left(\frac{x-X_n}{h_n}\right). \tag{4.2}$$

Let us first state the following technical lemma, which is repeatedly applied throughout the proofs.

Lemma 4.1 ([9, Lemma 2]). *Let $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$, and $m > 0$ such that $m - v^*\xi > 0$, where ξ is defined in* (1.6). *We have*

$$\lim_{n\to+\infty} v_n \Pi_n^m \sum_{k=1}^{n} \Pi_k^{-m} \gamma_k v_k^{-1} = (m - v^*\xi)^{-1}.$$

Moreover, for all positive sequences (α_n) such that $\lim_{n\to+\infty} \alpha_n = 0$, and for all $\delta \in \mathbb{R}$,

$$\lim_{n\to+\infty} v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \gamma_k v_k^{-1} \alpha_k + \delta\right] = 0.$$

Note that, in view of (1.2), we have

$$\begin{aligned}
F_n(x) - F(x) &= (1-\gamma_n)\left(F_{n-1}(x) - F(x)\right) + \gamma_n\left(Y_n - F(x)\right) \\
&= \sum_{k=1}^{n-1}\left[\prod_{j=k+1}^{n}(1-\gamma_j)\right]\gamma_k\left(Y_k - F(x)\right) + \gamma_n\left(Y_n - F(x)\right) \\
&\quad + \left[\prod_{j=1}^{n}(1-\gamma_j)\right]\left(F_0(x) - F(x)\right) \\
&= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\left(Y_k - F(x)\right) + \Pi_n\left(F_0(x) - F(x)\right).
\end{aligned}$$

It follows that

$$\mathbb{E}\left[F_n(x)\right] - F(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\left(\mathbb{E}\left[Y_k\right] - F(x)\right) + \Pi_n\left(F_0(x) - F(x)\right).$$

Then, we can write that

$$F_n(x) - \mathbb{E}\left[F_n(x)\right] = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left(Y_k - \mathbb{E}\left[Y_k\right]\right).$$

Let (Ψ_n) and (B_n) be the sequences defined as

$$\begin{aligned} \Psi_n(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k \left(Y_k - \mathbb{E}\left[Y_k\right]\right), \\ B_n(x) &= \mathbb{E}\left[F_n(x)\right] - F(x). \end{aligned}$$

We have

$$F_n(x) - F(x) = \Psi_n(x) + B_n(x). \tag{4.3}$$

Theorems 2.2, 2.3 and 2.4 are consequences of (4.3) and the following propositions.

Proposition 4.2 (Pointwise LDP and MDP for (Ψ_n))**.**

1. *Under the assumptions* (L1) *and* (L2)*, the sequence* $(F_n(x) - \mathbb{E}(F_n(x)))$ *satisfies a LDP with speed* (n) *and rate function* I_x*.*
2. *Under the assumptions* (L1) *and* (L3)*, the sequence* $(F_n(x) - \mathbb{E}(F_n(x)))$ *satisfies a LDP with speed* (n) *and rate function* $I_{x;a}$*.*
3. *Under the assumptions* (M1)–(M4)*, the sequence* $(v_n \Psi_n(x))$ *satisfies a LDP with speed* $\left(1/\left(\gamma_n v_n^2\right)\right)$ *and rate function* $J_{\alpha;x}$*.*

The proof of the following proposition is given in [18, p. 319].

Proposition 4.3 (Convergence rate of (B_n))**.** *Let* (M1)–(M3) *hold. If* f' *is continuous at* x*, then the following assertions are satisfied.*

1. *If* $a \leq \alpha/3$*, then*

$$B_n(x) = O\left(h_n^2\right).$$

2. *If* $a > \alpha/3$*, then*

$$B_n(x) = o\left(\sqrt{\gamma_n h_n}\right).$$

Set $x \in \mathbb{R}$. Since the assumptions of Theorems 2.2 and 2.3 guarantee that $\lim_{n\to\infty} B_n(x) = 0$, Theorem 2.2 (respectively Theorem 2.3) is a straightforward consequence of the application of Part 1 (respectively of Part 2) of Proposition 4.2. Moreover, under the assumptions of Theorem 2.4, we have by application of Proposition 4.3, $\lim_{n\to\infty} v_n B_n(x) = 0$. Theorem 2.4 follows immediately from Part 3 of Proposition 4.2.

We now state a preliminary lemma, which will be used in the proof of Proposition 4.2.

For any $u \in \mathbb{R}$, set

$$\Lambda_{n,x}(u) = \gamma_n v_n^2 \log \mathbb{E}\left[\exp\left(\frac{u}{\gamma_n v_n}\Psi_n(x)\right)\right],$$
$$\Lambda_x^{L,1}(u) = F(x)\left(\psi(u) - u\right),$$
$$\Lambda_x^{L,2}(u) = F(x)\left(\psi_a(u) - u\right),$$
$$\Lambda_x^{M}(u) = \frac{u^2}{2\left(2-\alpha\xi\right)}F(x).$$

Lemma 4.4 (Convergence of $\Lambda_{n,x}$ when $v_n \equiv 1$)**.**

1. *Let Assumptions* (L1) *and* (L2) *hold, assume that F is continuous at x, then for all $u \in \mathbb{R}$*
$$\lim_{n\to\infty}\Lambda_{n,x}(u) = \Lambda_x^{L,1}(u).$$
2. *Let Assumptions* (L1) *and* (L3) *hold, assume that F is continuous at x, then for all $u \in \mathbb{R}$*
$$\lim_{n\to\infty}\Lambda_{n,x}(u) = \Lambda_x^{L,2}(u).$$

Lemma 4.5 (Convergence of $\Lambda_{n,x}$ when $v_n \to \infty$)**.** *Let Assumptions* (M1)–(M4) *hold, assume that F is continuous at x, then for all $u \in \mathbb{R}$*
$$\lim_{n\to\infty}\Lambda_{n,x}(u) = \Lambda_x^{M}(u).$$

Our proofs are now organized as follows: Lemmas 4.4 and 4.5 are proved in Section 4.1 and Proposition 4.2 in Section 4.2.

4.1. PROOF OF LEMMAS 4.4 AND 4.5

Set $u \in \mathbb{R}$, $u_n = u/v_n$ and $a_n = \gamma_n^{-1}$. We have

$$\begin{aligned}\Lambda_{n,x}(u) &= \frac{v_n^2}{a_n}\log\mathbb{E}\left[\exp\left(u_n a_n \Psi_n(x)\right)\right]\\ &= \frac{v_n^2}{a_n}\log\mathbb{E}\left[\exp\left(u_n a_n \Pi_n \sum_{k=1}^n \Pi_k^{-1}a_k^{-1}\left(Y_k - \mathbb{E}\left[Y_k\right]\right)\right)\right]\\ &= \frac{v_n^2}{a_n}\sum_{k=1}^n \log\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)\right] - uv_n\Pi_n\sum_{k=1}^n \Pi_k^{-1}a_k^{-1}\mathbb{E}\left[Y_k\right].\end{aligned}$$

By the Taylor expansion, there exists $c_{k,n}$ between 1 and $\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)\right]$ such that

$$\begin{aligned}\log\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)\right] &= \mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right) - 1\right]\\ &\quad - \frac{1}{2c_{k,n}^2}\left(\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right) - 1\right]\right)^2,\end{aligned}$$

and $\Lambda_{n,x}$ can be rewritten as

$$\begin{aligned}\Lambda_{n,x}(u) &= \frac{v_n^2}{a_n}\sum_{k=1}^{n}\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\\ &- \frac{v_n^2}{2a_n}\sum_{k=1}^{n}\frac{1}{c_{k,n}^2}\left(\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\right)^2\\ &- uv_n\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}a_k^{-1}\mathbb{E}\left[Y_k\right].\end{aligned}$$

Let us first prove Lemma 4.5. We consider $v_n \to \infty$ as n goes to infinity. The Taylor expansion implies the existence of $c'_{k,n}$ between 0 and $u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k$ such that

$$\begin{aligned}\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right] &= u_n\frac{a_n\Pi_n}{a_k\Pi_k}\mathbb{E}\left[Y_k\right]\\ &+ \frac{1}{2}\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}\right)^2\mathbb{E}\left[Y_k^2\right]\\ &+ \frac{1}{6}\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}\right)^3\mathbb{E}\left[Y_k^3e^{c'_{k,n}}\right].\end{aligned}$$

Therefore,

$$\begin{aligned}\Lambda_{n,x}(u) &= \frac{1}{2}u^2a_n\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}a_k^{-2}\mathbb{E}\left[Y_k^2\right]+\frac{1}{6}u^2u_na_n^2\Pi_n^3\sum_{k=1}^{n}\Pi_k^{-3}a_k^{-3}\mathbb{E}\left[Y_k^3e^{c'_{k,n}}\right]\\ &- \frac{v_n^2}{2a_n}\sum_{k=1}^{n}\frac{1}{c_{k,n}^2}\left(\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\right)^2 \qquad (4.4)\\ &= \frac{1}{2}u^2F(x)a_n\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}a_k^{-1}\gamma_k+R_{n,x}^{(1)}(u)+R_{n,x}^{(2)}(u)\end{aligned}$$

with

$$\begin{aligned}R_{n,x}^{(1)}(u) &= u^2a_n\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}a_k^{-1}\gamma_k\int_{\mathbb{R}}K(z)\mathcal{K}(z)\left[F(x-zh_k)-F(x)\right]dz,\\ R_{n,x}^{(2)}(u) &= \frac{1}{6}\frac{u^3}{v_n}a_n^2\Pi_n^3\sum_{k=1}^{n}\Pi_k^{-3}a_k^{-3}\mathbb{E}\left[Y_k^3e^{c'_{k,n}}\right]\\ &- \frac{v_n^2}{2a_n}\sum_{k=1}^{n}\frac{1}{c_{k,n}^2}\left(\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\right)^2.\end{aligned}$$

Since F is continuous, we have $\lim_{k\to\infty}|F(x-zh_k)-F(x)|=0$, and thus, by the dominated convergence theorem, (M1) implies that

$$\lim_{k\to\infty}\int_{\mathbb{R}}K(z)\mathcal{K}(z)\left|F(x-zh_k)-F(x)\right|dz=0.$$

Since $(a_n) \in \mathcal{GS}(\alpha)$, and $\lim_{n\to\infty}(n\gamma_n) > \alpha/2$, Lemma 4.1 then ensures that

$$a_n\Pi_n^2\sum_{k=1}^n \Pi_k^{-2}a_k^{-1}\gamma_k = (2-\alpha\xi)^{-1} + o(1), \tag{4.5}$$

it follows that $\lim_{n\to\infty}\left|R_{n,x}^{(1)}(u)\right| = 0$.

Moreover, in view of (4.2), we have $|Y_k| \leq \|\mathcal{K}\|_\infty$, then

$$c'_{k,n} \leq \left|u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right| \leq |u_n|\,\|\mathcal{K}\|_\infty. \tag{4.6}$$

Note that $\mathbb{E}|Y_k|^3 \leq 3\|F\|_\infty\int_{\mathbb{R}}|K(z)|\left|\mathcal{K}^2(z)\right|dz$. Hence, using Lemma 4.1 and (4.6), there exists a positive constant c_1 such that, for n large enough,

$$\left|\frac{u^3}{v_n}a_n^2\Pi_n^3\sum_{k=1}^n\Pi_k^{-3}a_k^{-3}\mathbb{E}\left[Y_k^3e^{c'_{k,n}}\right]\right| \leq c_1e^{|u_n|\|\mathcal{K}\|_\infty}\frac{u^3}{v_n}\|F\|_\infty\int_{\mathbb{R}}|K(z)|\left|\mathcal{K}^2(z)\right|dz \tag{4.7}$$

which goes to 0 as $n\to\infty$ since $v_n\to\infty$. Moreover, Lemma 4.1 ensures that

$$\begin{aligned}
&\left|\frac{v_n^2}{2a_n}\sum_{k=1}^n\frac{1}{c_{k,n}^2}\left(\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\right)^2\right| \\
&\leq \frac{v_n^2}{2a_n}\sum_{k=1}^n\left(\mathbb{E}\left[\exp\left(u_n\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\right)^2 \\
&\leq \frac{u^2}{2}\|f\|_\infty^2 a_n\Pi_n^2\sum_{k=1}^n\Pi_k^{-2}a_k^{-1}\gamma_kh_k + o\left(a_n\Pi_n^2\sum_{k=1}^n\Pi_k^{-2}a_k^{-1}\gamma_kh_k\right) \\
&= o(1).
\end{aligned} \tag{4.8}$$

The combination of (4.7) and (4.8) ensures that $\lim_{n\to\infty}\left|R_{n,x}^{(2)}(u)\right| = 0$. Then, we obtain from (4.4) and (4.5), $\lim_{n\to\infty}\Lambda_{n,x}(u) = \Lambda_x^M(u)$. Which proves Lemma 4.5.

Let us now prove Lemma 4.4. We have $v_n \equiv 1$. Then, it follows from (4.4) that

$$\begin{aligned}
\Lambda_{n,x}(u) = {}& \frac{1}{a_n}\sum_{k=1}^n\mathbb{E}\left[\exp\left(u\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right] \\
&- \frac{1}{2a_n}\sum_{k=1}^n\frac{1}{c_{k,n}^2}\left(\mathbb{E}\left[\exp\left(u\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\right)^2 \\
&- u\Pi_n\sum_{k=1}^n\Pi_k^{-1}a_k^{-1}\mathbb{E}[Y_k].
\end{aligned}$$

Moreover, using integration by parts, we get

$$\begin{aligned}\Lambda_{n,x}(u) = uF(x)\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k \int_{\mathbb{R}} K(z)\left(\exp\left(uV_{n,k}\mathcal{K}(z)\right)-1\right)dz \\ - R_{n,x}^{(3)}(u) + R_{n,x}^{(4)}(u),\end{aligned} \tag{4.9}$$

with

$$\begin{aligned} V_{n,k} &= \frac{a_n\Pi_n}{a_k\Pi_k}, \\ R_{n,x}^{(3)}(u) &= \frac{1}{2a_n}\sum_{k=1}^{n}\frac{1}{c_{k,n}^2}\left(\mathbb{E}\left[\exp\left(u\frac{a_n\Pi_n}{a_k\Pi_k}Y_k\right)-1\right]\right)^2, \\ R_{n,x}^{(4)}(u) &= u\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\int_{\mathbb{R}}K(z)\left(\exp\left(uV_{n,k}\mathcal{K}(z)\right)-1\right)\left[F(x-zh_k)-F(x)\right]dz. \end{aligned} \tag{4.10}$$

It follows from (4.8) that $\lim_{n\to\infty}\left|R_{n,x}^{(3)}(u)\right| = 0$.

Farther, since $|e^t-1| \leq |t|\,e^{|t|}$, we have

$$\left|R_{n,x}^{(4)}(u)\right| \leq u^2 e^{|u|\|\mathcal{K}\|_\infty}\gamma_n^{-1}\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2\int_{\mathbb{R}}|K(z)|\,|\mathcal{K}(z)|\,|F(x-zh_k)-F(x)|\,dz.$$

Moreover, in view of Lemma 4.1 the sequence $\left(\gamma_n^{-1}\Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2\right)$ is bounded. Then, the dominated convergence theorem ensures that $\lim_{n\to\infty} R_{n,x}^{(4)}(u) = 0$.

Then, it follows from (4.9) that

$$\lim_{n\to\infty}\Lambda_{n,x}(u) = \lim_{n\to\infty} uF(x)\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\int_{\mathbb{R}}K(z)\left(\exp\left(uV_{n,k}\mathcal{K}(z)\right)-1\right)dz. \tag{4.11}$$

Let us now prove the first part of Lemma 4.4. In view of (L2), it follows from (4.1) that

$$\frac{\Pi_n}{\Pi_k} = \frac{k}{n},$$

and from (4.10),

$$V_{n,k} = 1.$$

Consequently, it follows from (4.11) that

$$\begin{aligned}\lim_{n\to\infty}\Lambda_{n,x}(u) &= uF(x)\int_{\mathbb{R}}K(z)\left(\exp\left(u\mathcal{K}(z)\right)-1\right)dz \\ &= F(x)\left(\exp(u)-1-u\right) \\ &= \Lambda_x^{L,1}(u).\end{aligned}$$

This concludes the proof of the first part of Lemma 4.4.

Let us now prove the second part of Lemma 4.4. In view of (L3), it follows from (4.1) that we have

$$\frac{\Pi_n}{\Pi_k} = \frac{\gamma_n}{\gamma_k}\frac{h_k}{h_n},$$

and from (4.10),

$$V_{n,k} = \frac{h_k}{h_n}.$$

Consequently, it follows from (4.11) and from some analysis considerations that

$$\begin{aligned}
\lim_{n\to\infty} \Lambda_{n,x}(u) &= uF(x)\int_0^1\int_{\mathbb{R}} s^{-a}K(z)\left(\exp\left(us^{-a}\mathcal{K}(z)\right)-1\right)dzds \\
&= F(x)\left(\int_0^1 \left(\exp\left(us^{-a}\right)-1\right)ds - u\right) \\
&= \Lambda_x^{L,2}(u),
\end{aligned}$$

and thus Lemma 4.4 is proved.

4.2. PROOF OF PROPOSITION 4.2

To prove Proposition 4.2, we apply Lemmas 4.4 and 4.5 and the following result, which can be deduced from Lemma 3.5 in [12].

Lemma 4.6. *Let (Z_n) be a sequence of real random variables, (ν_n) a positive sequence satisfying $\lim_{n\to\infty}\nu_n = +\infty$, and suppose that there exists some convex non-negative function Γ defined on $\mathbb{R}$ such that*

$$\forall u \in \mathbb{R}: \lim_{n\to\infty}\frac{1}{\nu_n}\log\mathbb{E}\left[\exp\left(u\nu_n Z_n\right)\right] = \Gamma(u).$$

If the Legendre function Γ^* of Γ is a strictly convex function, then the sequence (Z_n) satisfies a LDP of speed (ν_n) and good rate function Γ^*.

In our framework, when $v_n \equiv 1$ and $\gamma_n = n^{-1}$, we take $Z_n = F_n(x) - \mathbb{E}(F_n(x))$, $\nu_n = n$ and $\Gamma = \Lambda_x^{L,1}$. In this case, the Legendre transform of $\Gamma = \Lambda_x^{L,1}$ is the rate function $I_x : t \to F(x)I\left(1+\frac{t}{F(x)}\right)$, since ψ is strictly convex, then its Cramer transform I is a good rate function on $\mathbb{R}$ (see [2]). Farther, when $v_n \equiv 1$ and $\gamma_n = h_n\left(\sum_{k=1}^n h_k\right)^{-1}$, with $h_n = cn^{-a}$, $a \in \,]0,1[$ and $c > 0$, and we take $Z_n = F_n(x) - \mathbb{E}(F_n(x))$, $\nu_n = n$ and $\Gamma = \Lambda_x^{L,2}$. In this case, the Legendre transform of $\Gamma = \Lambda_x^{L,2}$ is the rate function $I_{x;a} : t \to F(x)I_a\left(\frac{1}{1-a}+\frac{t}{F(x)}\right)$, since ψ_a is strictly convex, then its Cramer transform I_a is a good rate function on $\mathbb{R}$ (see [2]). Otherwise, when $v_n \to \infty$ we take $Z_n = v_n\left(F_n(x) - \mathbb{E}(F_n(x))\right)$, $\nu_n = 1/\left(\gamma_n v_n^2\right)$ and $\Gamma = \Lambda_x^M$; Γ^* is then the quadratic rate function $J_{\alpha;x}$ defined in (1.7) and thus Proposition 4.2 follows.

Acknowledgements

The author would like to thank the Editor and the referees for their very helpful comments, which led to considerable improvement of the original version of the paper and a more sharply focused presentation.

REFERENCES

[1] R. Bojanic, E. Seneta, *A unified theory of regularly varying sequences*, Math. Z. **134** (1973) 2, 91–106.

[2] A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, Applications of Mathematics, New York, 1998.

[3] J. Galambos, E. Seneta, *Regularly varying sequences*, Proc. Amer. Math. Soc **41** (1973) 1, 110–116.

[4] E. Isogai, K. Hirose, *Nonparametric recursive kernel estimators of a distribution functions*, Bull. Inform. Cybernet. **26** (1994) 1–2, 87–99.

[5] C. Joutard, *Sharp large deviations in nonparametric estimation*, J. Nonparametr. Stat. **18** (2006) 3, 293–306.

[6] F. Klebaner, R. Liptser, *Large deviations for past-dependent recursions*, Probl. Inf. Transm. **32** (1996) 4, 23–34.

[7] D. Louani, *Some large deviations limit theorems in conditionnal nonparametric statistics*, Statistics **33** (1999) 2, 171–196.

[8] A. Mokkadem, M. Pelletier, *A companion for the Kiefer–Wolfowitz–Blum stochastic approximation algorithm*, Ann. Statist. **35** (2007) 4, 1749–1772.

[9] A. Mokkadem, M. Pelletier, Y. Slaoui, *The stochastic approximation method for the estimation of a multivariate probability density*, J. Statist. Plann. Inference **139** (2009) 7, 2459–2478.

[10] E.A. Nadaraya, *Some new estimates for distribution functions*, Theory Probab. Appl. **9** (1964) 3, 497–500.

[11] E.A. Nadaraya, *On estimating regression*, Theory Probab. Appl. **9** (1964) 1, 141–142.

[12] A.A. Puhalskii, *The method of stochastic exponentials for large deviations*, Stochastic Process. Appl. **54** (1994) 1, 45–70.

[13] A.A. Puhalskii, *Large deviations for stochastic processes*, LMS/EPSRC Short Course: Stochastic Stability, Large Deviations and Coupling Methods, Heriot-Watt University, Edinburgh, September 4–6, 2006.

[14] P. Révész, *Robbins–Monro procedure in a Hilbert space and its application in the theory of learning processes I*, Studia Sci. Math. Hung. **8** (1973) 1, 391–398.

[15] P. Révész, *How to apply the method of stochastic approximation in the non-parametric estimation of a regression function*, Math. Operationsforsch. Statist., Ser. Statistics **8** (1977) 1, 119–126.

[16] Y. Slaoui, *Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method*, Serdica Math. J. **39** (2013) 1, 53–82.

[17] Y. Slaoui, *Large and moderate deviation principles for kernel distribution estimator*, Int. Math. Forum **9** (2014) 18, 871–890.

[18] Y. Slaoui, *The stochastic approximation method for the estimation of a distribution function*, Math. Methods Statist. **23** (2014) 4, 306–325.

[19] Y. Slaoui, *Large and moderate deviation principles for averaged stochastic approximation method for the estimation of a regression function*, Serdica Math. J. **41** (2015) 2–3, 307–328.

[20] Y. Slaoui, *Moderate deviation principles for recursive regression estimators defined by stochastic approximation method*, Int. J. Math. Stat. **16** (2015) 2, 51–60.

[21] A.B. Tsybakov, *Recurrent estimation of the mode of a multidimensional distribution*, Problems Inform. Transmission **8** (1990) 3, 119–126.

[22] S.R.S. Varadhan, *Large Deviations*, Ann. Probab. **36** (2008) 2, 397–419.

[23] G.S. Watson, *Smooth regression analysis*, Sankhya A **26** (1990) 4, 359–372.

Yousri Slaoui
Yousri.Slaoui@math.univ-poitiers.fr

University of Poitiers
Laboratoire de Mathématiques et Applications
UMR 7348 du CNRS
Téléport 2 – BP 30179
11 Boulevard Marie et Pierre Curie
86962 Futuroscope Chasseneuil, France

Received: April 24, 2017.
Revised: January 29, 2019.
Accepted: March 30, 2019.