Theory of Stochastic Processes
Vol. 23 (39), no. 1, 2018, pp. 18–52

ARTURO KOHATSU-HIGA, NICOLAS VAYATIS, AND KAZUHIRO YASUDA

TUNING OF A BAYESIAN ESTIMATOR UNDER DISCRETE TIME OBSERVATIONS AND UNKNOWN TRANSITION DENSITY

We study the asymptotic behaviour of a Bayesian parameter estimation method on a compact one-dimensional parameter space. The estimation procedure is considered under discrete observations and unknown transition density. Here, we observe the data with constant time steps and the transition density of the data is approximated by using a kernel density estimation method applied to the Monte Carlo simulations of approximations of the theoretical random variables generating the observations. We estimate the error between the theoretical estimator, which assumes the knowledge of the transition density and its approximation which uses the simulation. We prove the strong consistency of the approximated estimator and find the order of the error. Most importantly, we give a parameter tuning result which relates the number of data, the weak error in the approximation process, the number of the Monte-Carlo simulations and the bandwidth size of the kernel density estimation. A guiding example for this situation is the use of Monte Carlo simulations of the Euler scheme for Bayesian estimation in a diffusion setting.

1. Introduction

We consider a parameter estimation problem using Bayesian inference under discrete observations taken at constant time intervals. That is, our purpose is to estimate the posterior expectation of some function f given the data;

$$E_N[f] := E_\theta[f|Y_0, ..., Y_N] := \frac{\int f(\theta)\phi_\theta(Y_0^N)\pi(\theta)d\theta}{\int \phi_\theta(Y_0^N)\pi(\theta)d\theta}, \tag{1}$$

where $Y_0, Y_1, ..., Y_N$ are observed data,

$$\phi_\theta(Y_0^N) = \phi_\theta(Y_0, ..., Y_N) = \mu_\theta(Y_0)\prod_{j=1}^{N} p_\theta(Y_{j-1}, Y_j)$$

is the joint density of $(Y_0, Y_1, ..., Y_N)$. We assume that Y forms a stationary α-mixing Markov chain and μ_θ is the stationary distribution. The prior distribution density of the parameter θ is denoted by π and p_{θ_0} denotes the transition density of Y, where θ_0 is the true value of the parameter associated to the data $\{Y_i\}$.

Method of estimations based on the formula 1 are fairly common in Bayesian statistics and is also a very simple case of a filtering method. This problem has been studied in both frameworks by Cano, Kessler, Salmeron [6] and Del Moral, Jacod, Protter [10].

In most applications p_θ is not known. Therefore many different ad-hoc methods have been developed to deal with the estimation problem. In this article, we study this problem from a theoretical point of view in the case where p_θ is approximated using Monte Carlo simulations.

In such a situation, there are two approximations taking place. The first is the approximation of the process that generates Y. For example, in the case that Y is generated using a stochastic differential equation (SDE), then one classical approximation is the Euler-Maruyama scheme for the SDE. This approximation scheme has as approximation

2010 *Mathematics Subject Classification.* 65C30, 62F15 .
Key words and phrases. Diffusion process, Bayesian estimation, Monte Carlo simulation .

parameter m which is the number of time steps used in the approximation (or the number of random numbers required for the simulation of one path).

Monte Carlo simulations of this approximation are performed n times and are used in order to generate an approximation of the transition density. A classical method to carry out this approximation is the Gaussian kernel density estimator. This approximation requires an approximation parameter, h, called window size. It is well known that in order for the approximation to converge to p_θ a correct choice of m, n and h as functions of N is needed. This procedure is usually called "parameter tuning". All the articles on filter approximation known to the authors deal with convergence results that do not lead to a tuning result.

In mathematical terms the approximating estimator can be expressed as

$$\hat{E}^n_{N,m}[f] := \frac{\int f(\theta)\hat{\phi}^N_\theta(Y^N_0)\pi(\theta)d\theta}{\int \hat{\phi}^N_\theta(Y^N_0)\pi(\theta)d\theta}, \tag{2}$$

where $\hat{\phi}^N_\theta(Y^N_0) := \mu_\theta(Y_0)\prod_{j=1}^N \hat{p}^N_\theta(Y_{j-1},Y_j)$ and $\hat{p}^N_\theta(y,z) = \frac{1}{nh}\sum_{k=1}^n K(\frac{X^{y,(k)}_{(m)}(\theta)-z}{h})$, where K is some suitable probability kernel (say, the density of a standard Gaussian random variable) and $X^{y,(k)}_{(m)}(\theta)$ is a process which is used in order to approximate the transition density function p_θ. Note that for simplicity we put only N for index in the right-hand side of (2), but $\hat{p}^N_\theta$ obviously depends on all parameters m, n, h. Through this expression, we can simulate an approximation of the posterior expectation.

One may also consider other possibilities for the approximation of the transition density p_θ. For example, the exact simulation which was introduced by the paper of Beskos et al. [2] for one dimensional SDE's. In this paper, we discuss the case of one-dimensional stochastic process and a scalar parameter in order to avoid complicated notations and for the restriction of the space, but we can extend our results to multi-dimensional settings.

Another remark is that in (2), we have integrals with respect to θ which need to be approximated, but the approximation is not related to the number of observed data, therefore we have not included it in this paper.

The main goal of the present paper is to prove that there exists a choice of m, $n = O(N^{\alpha_1})$ and $h = O(N^{-\alpha_2})$, so that we obtain that rate of convergence of $\hat{E}^n_{N,m}[f]$ to $E_N[f]$ is $N^{-\frac{1}{2}}$ a.s. The value of m is determined so as the weak rate of convergence of $X^{y,(1)}_{(m)}(\theta)$ to the law with density p_θ is close to $N^{-\frac{1}{2}}$. At present our results are somewhat theoretical, but this tuning procedure is important in practice as it shows that there are cases where the above convergence is not satisfied.

Our proof points toward the theoretical issues behind the complex tuning that have to be tackled in this situation. We give an explicit tuning results for the case of smooth diffusions in Theorem 3.1. (See Theorem 6.12.)

Note that in the present problem as N increases the number of approximated transition densities increase and therefore there is the potential of error dispersion and therefore tuning becomes an important problem. To solve the problem, we essentially use the Laplace method.

As pointed above the standard example for this setting is the diffusion case. Due to space constraints, we do not give the verification that the various assumed hypotheses are satisfied in this case. Instead, we refer the reader to Kohatsu-Higa et. al. [17] where the Ornstein-Uhlenbeck process case is treated and Kohatsu-Higa and Yasuda [18] where the case of the Euler scheme is considered in detail.

The goal of this article is to point at reasons why an ad-hoc tuning may or may not work. This can be clearly seen in the calculations related to the Theorem 6.12 where the tuning takes place. An example was treated under a different light by Cano, Kessler, Salmeron [6]. We also remark that there are many other methods that have been proven to be more efficient in practice without theoretical proofs. Our study is a first step to study these algorithms from a theoretical point of view.

Idea of the Proof. In order to guide the reader through the arguments we give a brief explanation of the argument that we will use. The first step is to study the exploding behavior of the numerators of (1) and (2) by writing the integral in exponential form.

This calculation generates a main term appearing from the ergodic theorem applied to the sequence $\{Y_i\}_{i\in\mathbb{N}}$. A second error term appears due to the central limit theorem. In Proposition 4.1 one proves that the second error term behaves asymptotically close to $N^{\frac{1}{2}}$ while the main term behaves asymptotically as N. This property is important in order to prove Theorem 2.3 which gives the rate of convergence. The condition that ensures that the main term behaves asymptotically like N is the identifiability condition (Assumption 2.2 (4)). To prove that the remainder is of order $N^{-\frac{1}{2}}$ requires the use of the central limit theorem for α mixing sequences (see Section 5.3 for the case of (1)).

To do the same for the integrals in (2) is much more complicated as it involves also the Monte Carlo simulations. Therefore we need to assure that the Monte Carlo simulation is probabilistically speaking close to the density of the approximation process. Therefore the tuning procedure naturally appears. This is reflected in the hypothesis made in 2.2 (6) which is the only assumption involving the Monte Carlo approximation. We remark that this condition requires the differentiability of the approximating process with respect to θ. The other conditions in hypothesis 2.2 (6) also ensure the closeness of the approximating density to the transition density of the sequence $\{Y_i\}_{i\in\mathbb{N}}$.

The main tuning appears when we have to verify Assumption 2.2 (6). This is studied in detail in Section 6. In order to prove assumption 2.2 (6), we first use a Borel-Cantelli argument in order to limit the values of the sequence $\{Y_i\}_{i\in\mathbb{N}}$ to a compact set. In order to bound the denominators, we require lower bound conditions on the approximating densities. Finally, in order to find the rate of convergence of the various differences between the Monte Carlo simulations and the approximating density, we need to first use Borel-Cantelli and continuity arguments to get rid of the supremums in (y, z). Finally using exponential inequalities we obtain a rate of convergence. Some of these ideas have been taken from kernel density convergence arguments that can be found in Bosq [5].

These Borel-Cantelli arguments require convergence of certain probabilities which lead naturally to the final tuning requirements that appear in equations (35) and (36).

This paper is structured as follows; In Section 2, we will give our framework and state our first goal, i.e. finding the rate of convergence of (2) toward (1). This long list of conditions refer to what may be considered as a regular case. There are many other variations that maybe entertained using the same method but this will require much more space. For the same reason, we have detached the problem from the study of the approximation procedure of the process generating the data. In Section 3, we explain the meaning of the hypotheses in the case that Y is generated by a diffusion and its transition density is approximated using Monte Carlo simulations of the Euler-Maruyama scheme. In Section 4, we will use the Laplace method in order to prove the rate of convergence stated in Section 2. This proof uses some estimations of various error processes which are stated in Proposition 4.1. This proposition plays an important role and it is the core of the paper. These estimates on error processes are stated in two levels, one is on the level of the error between the density of the process underlying the observations and the density of its approximation and another is on the level of the approximative density and the simulations. The second proves to be more challenging than the first and it is in this second error the tuning process appears. This is condensed in Assumption 2.2 (6). In Section 5, we will show the four estimations required in Proposition 4.1. Finally in Section 6, we give some smoothness conditions that ensure that Assumption 2.2 (6) is satisfied. We obtain our main result Theorem 6.12 which gives two tuning requirements, (35) and (36), for the parameters m, n, h.

We close the article with some appendices where we collect some technical results used in the article.

2. Framework and a First Convergence Result

Notation: We denote by $C(A;B)$ the space of continuous functions from A to B, where A and B are sets. Denote by $C^{k_1,k_2,k_3}(K_1 \times K_2 \times K_3; B)$ the space of functions from $K_1 \times K_2 \times K_3$ to B that are k_i-times continuously differentiable in the interior of K_i and continuous on K_i, where $k_1, k_2, k_3 \in \mathbb{N}$ and K_1, K_2, K_3 are sets. When a subscript b is added it means that the functions are also continuous on $K_1 \times K_2$ and bounded $C_b^{0,0}(K_1 \times K_2; B)$. Finally, $\mathcal{B}(A)$ denotes the Borel σ-field on A.

2.1. Framework.

We consider the following problem: Let $\theta_0 \in \Theta := [\theta^l, \theta^u]$, $(\theta^l < \theta^u)$ be a parameter that we want to estimate $\theta_0 \in \dot{\Theta}$, where $\dot{\Theta}$ denotes the interior of the set Θ and $\Theta_0 = \Theta - \{\theta_0\}$.

In order to frame the problem in a proper mathematical setting, we will use three separate probability spaces.

In the first space, $(\Omega, \mathcal{F}, P_{\theta_0})$, we will define the observation data process Y (see (1)).

In the second space, $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, we will define the family of processes $X(\theta)$ which are realizations of the Markov chain with transition density p_θ. This space is needed in order to prove asymptotic properties of the estimators.

Finally, in the third probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$, we will define the simulations (see (2) and $X^{y,(k)}(\theta)$ right after this equation).

- **(i). (Observation process)** For $\Delta > 0$, fixed let $\{Y_{i\Delta}\}_{i=0,1,\dots,N}$ be a sequence of $N+1$-observations of a stationary Markov chain having transition density $p_{\theta_0}(y,z)$, $y, z \in \mathbb{R}$ and invariant measure μ_{θ_0}. This sequence is defined on the probability space $(\Omega, \mathcal{F}, P_{\theta_0})$. We write $Y_i := Y_{i\Delta}$ for $i = 0, 1, \dots, N$.
- **(ii). (Model process)** Denote by $X^y(\theta)$, $y \in \mathbb{R}, \theta \in \Theta$ be a family of random variables defined on the probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ such that its law is given by $p_\theta(y,z)$.
- **(iii).** Denote by $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ the probability space where one generates the simulation of the approximation to the process X^y.
- **(iv). (Approximating process)** Denote by $X^y_{(m)}(\theta)$ the approximation to $X^y(\theta)$, which is defined on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$. The parameter that determines the quality of the approximation is given by large values of $m \equiv m(N)$. Denote by $\tilde{p}_\theta^N(y,z) = \tilde{p}_\theta^N(y,z;m(N))$ the transition density for the process $X^y_{(m)}(\theta)$.

  In the case, that Y is a diffusion process with coefficients that depend on the parameter θ then $X^y_{(m)}(\theta)$ may denote the associated Euler-Maruyama scheme with step $\frac{\Delta}{m}$ starting at y and same coefficients which therefore also depend on θ. For more details, see section 3.
- **(v). (Approximated transition density)** Set $\mathbb{R}_+ = [0, \infty)$. Let $K \in C^2(\mathbb{R}; \mathbb{R}_+)$ (usually called kernel), which satisfies $\int K(x)dx = 1$. Denote by $\hat{p}_\theta^N(y,z)$, the kernel density estimate of $\tilde{p}_\theta^N(y,z)$ based on $n \equiv n(N)$ simulated i.i.d. copies of $X^y_{(m)}(\theta)$ which are defined on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ and denoted by $X^{y,(k)}_{(m)}(\theta, \cdot)$, $k = 1, \dots, n$; for $h \equiv h(N) > 0$, define the approximated transition density as

$$\hat{p}_\theta^N(y,z) := \hat{p}_\theta^N(y,z;\hat{\omega};m(N),h(N),n(N)) := \frac{1}{n(N)h(N)} \sum_{k=1}^{n(N)} K\left(\frac{X^{y,(k)}_{(m(N))}(\theta,\hat{\omega}) - z}{h(N)}\right).$$

(vi). For given m, we introduce the mean of the approximated transition density over all trajectories with respect to the kernel K;

$$\bar{p}_\theta^N(y,z) := \bar{p}_\theta^N(y,z;m(N),h(N)) := \hat{E}\left[\hat{p}_\theta^N(y,z)\right] = \hat{E}\left[\frac{1}{h(N)}K\left(\frac{X^{y,(1)}_{(m(N))}(\theta,\cdot)-z}{h(N)}\right)\right],$$

where $\hat{E}$ means the expectation with respect to $\hat{P}$.

As it can be deduced from the above set-up, we have preferred to state our problem in abstract terms without explicitly defining the dynamics that generate $X^y(\theta)$ or how the approximation $X^y_{(m)}(\theta)$ is defined. All the properties that will be required for p_θ and $\tilde{p}_\theta^N$ will be satisfied for an appropriate subclass of diffusion processes.

Remark 2.1. *Without loss of generality, we can consider the product of the above three probability spaces so that all random variables are defined on the same probability space. We do this without any further mentioning.*

Our purpose is to estimate the posterior expectation for some function $f \in C^1(\Theta)$ given the data;

$$E_N[f] := E_\theta[f|Y_0,...,Y_N] = \frac{I_N(f)}{I_N(1)} := \frac{\int f(\theta)\phi_\theta(Y_0^N)\pi(\theta)d\theta}{\int \phi_\theta(Y_0^N)\pi(\theta)d\theta},$$

where

$$\phi_\theta(Y_0^N) = \phi_\theta(Y_0,...,Y_N) = \mu_\theta(Y_0)\prod_{j=1}^N p_\theta(Y_{j-1},Y_j)$$

is the joint density of $(Y_0, Y_1, ..., Y_N)$.

We propose to estimate this quantity on the basis of simulated instances of the process;

$$\hat{E}^n_{N,m}[f] := \frac{\hat{I}^n_{N,m}(f)}{\hat{I}^n_{N,m}(1)} := \frac{\int f(\theta)\hat{\phi}_\theta^N(Y_0^N)\pi(\theta)d\theta}{\int \hat{\phi}_\theta^N(Y_0^N)\pi(\theta)d\theta},$$

where $\hat{\phi}_\theta^N(Y_0^N) := \mu_\theta(Y_0)\prod_{j=1}^N \hat{p}_\theta^N(Y_{j-1},Y_j)$.

2.2. **A First Convergence Result.**

Assumption 2.2. *We assume the following*

(1). (Observation process) *$\{Y_i\}_{i=0,1,...,N}$ is an α-mixing process with $\alpha_n = O(n^{-p})$ for some $p > 6$.*

(2). (The prior distribution density) *The prior distribution density $\pi \in C(\Theta)$, and for all $\theta \in \Theta$, $\pi(\theta) > 0$.*

(3). (Density regularity) *The transition densities $p, \bar{p}^N \in C^{3,0,0}(\Theta \times \mathbb{R}^2; \mathbb{R}_+)$, and for all $\theta \in \Theta$, $y, z \in \mathbb{R}$, we have that $\min\left\{p_\theta(y,z), \bar{p}_\theta^N(y,z)\right\} > 0$. And p_θ admits an invariant measure $\mu \in C_b^{0,0}(\Theta \times \mathbb{R}; \mathbb{R}_+)$, and for all $\theta \in \Theta$, $\mu_\theta(y) > 0$ for every $y \in \mathbb{R}$.*

(4). (Identifiability) *Assume that there exist $c_1 : \mathbb{R} \to (0,\infty)$ such that for all $\theta \in \Theta$,*

$$\inf_N \int |p_\theta(y,z) - p_{\theta_0}(y,z)|dz \geq c_1(y)|\theta - \theta_0|,$$

and $C_1(\theta_0) := \int c_1(y)^2\mu_{\theta_0}(y)dy \in (0,+\infty)$.

(5). (Regularity of the log-density) *We assume that there exists $\delta > 0$ such that for $q_\theta = p_\theta, \bar{p}_\theta^N$*

$$\sup_N \sup_{\theta\in\Theta} \iint \left(\frac{\partial^i}{\partial\theta^i}\ln q_\theta(y,z)\right)^{4+\delta} p_{\theta_0}(y,z)\mu_{\theta_0}(y)dydz < +\infty, \ \ for\ i = 0,1,2,$$

where $\frac{\partial^0}{\partial\theta^0} q_\theta = q_\theta$.

(6). (Parameter tuning) *We assume the following boundedness for some* $\epsilon > 0$

$$\sup_N \sup_{\theta\in\Theta} \left| \frac{N^{-\epsilon}}{\sqrt{N}} \sum_{i=0}^{N-1} \left(\frac{\partial}{\partial\theta} \ln \hat{p}^N_\theta(Y_i, Y_{i+1}) - \frac{\partial}{\partial\theta} \ln \bar{p}^N_\theta(Y_i, Y_{i+1}) \right) \right| < +\infty \ a.s. \tag{3}$$

$$\sup_N \sup_{\theta\in\Theta} \left| \frac{N^{-\epsilon}}{\sqrt{N}} \sum_{i=0}^{N-1} \left(\frac{\partial}{\partial\theta} \ln \bar{p}^N_\theta(Y_i, Y_{i+1}) - \frac{\partial}{\partial\theta} \ln p_\theta(Y_i, Y_{i+1}) \right) \right| < +\infty \ a.s. \tag{4}$$

The first goal is to prove the following result.

Theorem 2.3. *Under Assumption 2.2 and for* $f \in C^1(\Theta)$, *there exists some positive finite random variables* Ξ_1 *and* Ξ_2 *which do not depend on* N *such that for any* $\varepsilon > 0$

$$|E_N[f] - f(\theta_0)| \le \frac{\Xi_1}{N^{\frac{1}{2}-\varepsilon}} \ a.s., \ and \ \left|\hat{E}^n_{N,m}[f] - f(\theta_0)\right| \le \frac{\Xi_2}{N^{\frac{1}{2}-\varepsilon}} \ a.s.,$$

and therefore

$$\left|E_N[f] - \hat{E}^n_{N,m}[f]\right| \le \frac{\Xi_1 + \Xi_2}{N^{\frac{1}{2}-\varepsilon}} \ a.s.$$

Our final goal is to prove that there is a choice for α_1 and α_2 with $m = \sqrt{N}$, $n = C_1 N^{\alpha_1}$ and $h = C_2 N^{-\alpha_2}$ under which the above assumptions are satisfied and therefore the above result can be applied. This result is obtained in Theorem 6.12. See also Theorem 3.1 for the case of SDEs.

3. Understanding the hypotheses in the diffusion case

In this section, we give a brief description of how to interpret the different hypotheses and how they are verified in the particular case of diffusions. We only give brief comments on these matters and we refer the reader to the detailed articles [17] and [18].

In this section, the data is obtained from a diffusion of the type

$$Y_t = Y_0 + \int_0^t b(\theta_0, Y_s)ds + \int_0^t \sigma(\theta_0, Y_s)dW_s, \ t \ge 0.$$

In order to simplify the situation, we consider the one dimensional situation on a compact parameter space. So that $b, \ \sigma : [\theta^l, \theta^u] \times \mathbb{R} \to \mathbb{R}$ are smooth functions with bounded derivatives. Suppose that the diffusion satisfies sufficient conditions for existence and regularity of its invariant measure (see e.g. [7] and [8]) and that it is α-mixing. Furthermore, we assume that the process Y is stationary.

Additionally, we require ellipticity conditions so that upper and lower bounds for transition densities can be obtained. Then $X^y(\theta) \equiv X^y_\Delta(\theta)$ denotes a copy of the underlying random variable. That is,

$$X^y_t(\theta) = y + \int_0^t b(\theta, X^y_s(\theta))ds + \int_0^t \sigma(\theta, X^y_s(\theta))dW_s, \ t \ge 0.$$

Then $X^y_{(m)}(\theta) \equiv \bar{X}^y_\Delta(\theta)$ where the Euler-Maruyama scheme is defined as

$$\bar{X}^y_t(\theta) = y + \int_0^t b(\theta, \bar{X}^y_{\eta(s)}(\theta))ds + \int_0^t \sigma(\theta, \bar{X}^y_{\eta(s)}(\theta))dW_s, \ t \ge 0.$$

Here $\eta(s) = \sup\{\frac{i\Delta}{m}; \ \frac{i\Delta}{m} < s\}$. One may consider higher order weak schemes in order to improve the performance of the method which would obviously lead to much more complicated estimates. For convergence properties concerning $\bar{X}$, we refer the reader to Bally and Talay [1] and Guyon [14]. The kernel K is usually chosen to be a Gaussian density with mean 0 and variance 1.

The identifiability condition (4), is needed in order to be able to obtain that the density can be used to discern the value of θ from the observations. This type of assumption is natural in statistics.

Assumption 2.2 (5) will be satisfied under enough regularity of the transition density function p_θ and its approximation $\bar{p}_\theta^N$. This is usually obtained using Malliavin Calculus techniques.

Assumption 2.2 (6) will be crucial in what follows and it is this property that will determine the rate of convergence and the tuning properties. This is the only hypothesis that involves $\hat{p}_\theta^N$, which is random. In particular, obtaining a lower bound for $\hat{p}_\theta^N$ will be the important problem to solve.

This will be further discussed in Section 6. In fact, we have the following theorem.

Theorem 3.1. *Assume the following*

(1) *The parameter N is large enough with $m = \sqrt{N}$, $n = C_1 N^{\alpha_1}$ and $h = C_2 N^{-\alpha_2}$.*
(2) *There exists some constant $c_2 > 0$, $E[\exp(\frac{3}{c_2} Y_1^2)] < \infty$ holds.*
(3) *There exists some constants $\varphi_1, \varphi_2 > 0$ such that*

$$\inf_{(x,y,\theta);\|(x,y)\| \le \sqrt{c_2 \ln N}} \bar{p}_\theta^N(x,y) \wedge p_\theta(x,y) \ge \varphi_1 \exp\left(-\frac{\varphi_2 c_2 \ln N}{\Delta}\right),$$

where c_2 is the same as the above.
(4) *Finally, assume that $\alpha_1 > 8\alpha_2 + 1 + \frac{2\varphi_2 c_2}{\Delta}$.*

Then Assumption 2.2 (6) holds.

Idea of Proof. In fact, as b and σ are smooth with bounded derivatives all the conditions in Theorem 6.12 are satisfied by choosing $\gamma_3 > 1$, $\dot{\gamma}_6 > 1$ and noting that conditions **(iii)**, **(viii)** and **(ix)** are satisfied with r_3, $\dot{r}_6$ and $\dot{q}_6$ big enough (therefore (35) is satisfied). All the other conditions in Theorem 6.12 can be verified by using the smoothness of b and σ. □

Therefore roughly speaking, $a_1(N) = \frac{1}{m(N)} + h(N)^2$.

Finally, we remark that one also needs to approximate the invariant measure but this problem can be solved with an extra term. The quality of approximation is studied in Talay [21] and the references therein.

4. Proof of Theorem 2.3

We start introducing some notation; let p and q be positive functions of two variables. Define

$$H(p,q) := \iint \Big(\ln p(y,z)\Big) q(y,z) \mu_{\theta_0}(y) dy dz.$$

We also let

$$\begin{aligned}
Z_N(\theta) &:= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \Big\{ \ln p_\theta(Y_i, Y_{i+1}) - H(p_\theta, p_{\theta_0}) \Big\}, \\
\varepsilon(\theta) &:= H(p_\theta, p_{\theta_0}) - H(p_{\theta_0}, p_{\theta_0}), \\
\beta_N(\theta) &:= Z_N(\theta) - Z_N(\theta_0), \\
\zeta(\theta) &:= \mu_\theta(Y_0)\pi(\theta). \\
R_N^1(\theta) &:= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \left(\ln \hat{p}_\theta^N (Y_i, Y_{i+1}) - \ln \bar{p}_\theta^N (Y_i, Y_{i+1})\right), \\
R_N^2(\theta) &:= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \left(\ln \bar{p}_\theta^N (Y_i, Y_{i+1}) - \ln p_\theta (Y_i, Y_{i+1})\right),
\end{aligned}$$

Some of notations above are related to the information theory and its approximations, for example $\varepsilon(\theta)$ is the Kullback-Leibler divergence and so on.

The following proposition states the key properties that are needed to achieve the proof of Theorem 2.3.

Proposition 4.1. *Under Assumption 2.2,*

(i). *There exist strictly negative constants c_1, c_2 such that the following inequality is satisfied*

$$c_1 \leq \inf_N \inf_{\theta \in \Theta_0} \frac{\varepsilon(\theta)}{(\theta-\theta_0)^2} \leq \sup_N \sup_{\theta \in \Theta_0} \frac{\varepsilon(\theta)}{(\theta-\theta_0)^2} \leq c_2 < 0.$$

(ii). *There exist $\epsilon \in (0, \frac{1}{2})$ and a random variable M on $(\Omega \times \hat{\Omega}, \mathcal{F} \otimes \hat{\mathcal{F}}, P_{\theta_0} \otimes \hat{P})$ such that*

$$\sup_N \sup_{\theta \in \Theta_0} \left| \frac{N^{-\epsilon}\beta_N(\theta)}{\theta - \theta_0} \right| \leq M \ a.s.$$

(iii). *For any $i = 1, 2$ and $\epsilon \in (0, \frac{1}{2})$ there exists a random variable M on $(\Omega \times \hat{\Omega}, \mathcal{F} \otimes \hat{\mathcal{F}}, P_{\theta_0} \otimes \hat{P})$ such that*

$$\sup_N \sup_{\theta \in \Theta_0} \left| \frac{N^{-\epsilon}(R_N^i(\theta) - R_N^i(\theta_0))}{\theta - \theta_0} \right| \leq M \ a.s.$$

We will give the proof of this proposition in section 5.

Proof of Theorem 2.3. We decompose the approximation error as follows;

$$E_N[f] - \hat{E}^n_{N,m}[f] = \left(\frac{I_N(f) - f(\theta_0) I_N(1)}{I_N(1)} \right) - \left(\frac{\hat{I}^n_{N,m}(f) - f(\theta_0)\hat{I}^n_{N,m}(1)}{\hat{I}^n_{N,m}(1)} \right) =: \Delta_1 - \Delta_2.$$

First we consider Δ_1: We will prove that there exists some positive random variable C_1 such that

$$|\Delta_1| \leq \frac{C_1}{N^{\frac{1}{2}-\epsilon}} \ a.s. \tag{5}$$

Indeed, using the definitions provided at the beginning of this section, we can write $I_N(f)$ as follows;

$$I_N(f) = e^{NH(p_{\theta_0}, p_{\theta_0}) + \sqrt{N} Z_N(\theta_0)} \int_{\Theta_0} f(\theta) e^{N\varepsilon(\theta) + \sqrt{N}\beta_N(\theta)} \zeta(\theta) d\theta. \tag{6}$$

Here, we perform the following change of variables; $\theta = \theta_0 + \frac{r}{\sqrt{N}}$. Then,

$$\begin{aligned}
&|I_N(f) - f(\theta_0) I_N(1)| \\
&= \left| e^{NH(p_{\theta_0}, p_{\theta_0}) + \sqrt{N} Z_N(\theta_0)} \int_{\Theta_0} (f(\theta) - f(\theta_0)) e^{N\varepsilon(\theta) + \sqrt{N}\beta_N(\theta)} \zeta(\theta) d\theta \right| \\
&= e^{NH(p_{\theta_0}, p_{\theta_0}) + \sqrt{N} Z_N(\theta_0)} \\
&\times \left| \int_{\sqrt{N}(\theta^l - \theta_0)}^{\sqrt{N}(\theta^u - \theta_0)} \left(f\left(\theta_0 + \frac{r}{\sqrt{N}}\right) - f(\theta_0) \right) e^{N\varepsilon(\theta_0 + \frac{r}{\sqrt{N}}) + \sqrt{N}\beta_N(\theta_0 + \frac{r}{\sqrt{N}})} \zeta\left(\theta_0 + \frac{r}{\sqrt{N}}\right) \frac{dr}{\sqrt{N}} \right| \\
&= \frac{e^{NH(p_{\theta_0}, p_{\theta_0}) + \sqrt{N} Z_N(\theta_0)}}{N} \\
&\times \left| \int_{\sqrt{N}(\theta^l - \theta_0)}^{\sqrt{N}(\theta^u - \theta_0)} r f'(\xi_{r,N}) e^{N\varepsilon(\theta_0 + \frac{r}{\sqrt{N}}) + \sqrt{N}\beta_N(\theta_0 + \frac{r}{\sqrt{N}})} \zeta\left(\theta_0 + \frac{r}{\sqrt{N}}\right) dr \right|.
\end{aligned}$$

In the above, we have used the mean-value theorem: there exists a constant $\xi_{r,N}$ which is between $\theta_0 + \frac{r}{\sqrt{N}}$ and θ_0. Set $\Theta_N = [\sqrt{N}(\theta^l - \theta_0), \sqrt{N}(\theta^u - \theta_0)]$. Now we separate the above integral into two parts:

$$
\begin{aligned}
&= \frac{e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{N} \left| \int_{\{|r|\le cN^a\}\cap\Theta_N} \right. \\
&\left. + \int_{\{|r|>cN^a\}\cap\Theta_N} rf'(\xi_{r,N})\, e^{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})+\sqrt{N}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)dr \right|,
\end{aligned} \tag{7}
$$

where a and c are two positive constants, which will be defined later.

We start considering the first term in (7), by dividing it by $I_N(1)$, we have, using that $\{|r| \le cN^a\} \cap \Theta_N \subset \Theta_N$:

$$
\begin{aligned}
&\frac{e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{NI_N(1)} \\
&\times \left| \int_{\{|r|\le cN^a\}\cap\Theta_N} rf'(\xi_{r,N})\, e^{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})+\sqrt{N}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)dr \right| \\
&\le \frac{\|f'\|_\infty}{\sqrt{N}} \int_{\{|r|\le cN^a\}\cap\Theta_N} |r| \frac{e^{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})+\sqrt{N}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)}{\int_{\sqrt{N}(\theta^l-\theta_0)}^{\sqrt{N}(\theta^u-\theta_0)} e^{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})+\sqrt{N}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)dr} dr \\
&\le \frac{cN^a\|f'\|_\infty}{\sqrt{N}} \int_{\{|r|\le cN^a\}\cap\Theta_N} \frac{e^{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})+\sqrt{N}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)}{\int_{\sqrt{N}(\theta^l-\theta_0)}^{\sqrt{N}(\theta^u-\theta_0)} e^{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})+\sqrt{N}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)dr} dr \\
&\le \frac{c\|f'\|_\infty}{N^{\frac{1}{2}-a}}.
\end{aligned} \tag{8}
$$

For the moment we let $a > 0$ be such that $a > \epsilon$.

For the second term in (7), we again separate the integral into two parts:

$$
\begin{aligned}
&\frac{e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{N} \left| \int_{\{r>cN^a\}\cap\Theta_N} \right. \\
&\left. + \int_{\{r<-cN^a\}\cap\Theta_N} rf'(\xi_{r,N})\, e^{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})+\sqrt{N}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)dr \right| \\
&\le \frac{\|f'\|_\infty\|\zeta\|_\infty e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{N} \left\{ \int_{r>cN^a} + \int_{r<-cN^a} \right\} |r|\, e^{c_2(r)r^2+d_2(r)N^\epsilon r}dr
\end{aligned}
$$

where we have set $c_2(r) = \frac{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})}{r^2}$ and $d_2(r) = \frac{N^{\frac{1}{2}-\epsilon}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}{r}$. From Proposition 4.1 (i) and (ii), we have

$$
\begin{aligned}
&\le \frac{\|f'\|_\infty\|\zeta\|_\infty e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{N} \\
&\times \left\{ \int_{r>cN^a} re^{-|c_2|r^2+MN^\epsilon r}dr - \int_{r<-cN^a} re^{-|c_2|r^2-MN^\epsilon r}dr \right\}.
\end{aligned} \tag{9}
$$

For the first integral term in (9), we have

$$\begin{aligned}&\frac{\|f'\|_\infty\|\zeta\|_\infty e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{N}\int_{r>cN^a} re^{-|c_2|r^2+MN^\epsilon r}dr\\&=\frac{\|f'\|_\infty\|\zeta\|_\infty e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{N}\\&\times\left\{\frac{1}{2|c_2|}e^{-|c_2|(cN^a)^2+MN^\epsilon(cN^a)}+\frac{MN^\epsilon}{2|c_2|}\int_{cN^a}^\infty e^{-|c_2|r^2+MN^\epsilon r}dr\right\}.\end{aligned}$$

For $cN^a \geq \frac{MN^\epsilon}{2|c_2|}$, we have, from Lemma 7.3,

$$\begin{aligned}&\leq\frac{\|f'\|_\infty\|\zeta\|_\infty e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{N}\\&\times\left\{\frac{1}{2|c_2|}e^{-|c_2|(cN^a)^2+MN^\epsilon(cN^a)}+\frac{MN^\epsilon}{2|c_2|}\frac{e^{-|c_2|((cN^a)^2-\frac{MN^\epsilon}{|c_2|}cN^a)}}{\sqrt{|c_2|}}\right\}\\&=\frac{\|f'\|_\infty\|\zeta\|_\infty e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{2|c_2|N}\\&\times\left\{1+\frac{MN^\epsilon}{\sqrt{|c_2|}}\right\}e^{-|c_2|c^2N^{2a}+cMN^{\epsilon+a}}.\end{aligned}\tag{10}$$

By using a symmetric argument, we have the same results for the second term in (9), namely we have, from Lemma 7.3, for $cN^a \geq \frac{MN^\epsilon}{2|c_2|}$,

$$\begin{aligned}&-\int_{\{r<-cN^a\}\cap\Theta_N} re^{-|c_2|r^2-MN^\epsilon r}dr\\&\leq\frac{\|f'\|_\infty\|\zeta\|_\infty e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{2|c_2|N}\left\{1+\frac{MN^\epsilon}{\sqrt{|c_2|}}\right\}e^{-|c_2|c^2N^{2a}-cMN^{\epsilon+a}}.\end{aligned}\tag{11}$$

From now on, we estimate the denominator $I_N(1)$ from below using similar arguments. Set $\theta=\theta_0+\frac{r}{\sqrt{N}}$ and $\Theta'_N=[\sqrt{N}(\theta^l-\theta_0),\sqrt{N}(\theta^u-\theta_0)]$.

$$I_N(1)=\frac{e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{\sqrt{N}}\int_{\sqrt{N}(\theta^l-\theta_0)}^{\sqrt{N}(\theta^u-\theta_0)} e^{c'_2(r)r^2+d'_2(r)N^\epsilon r}\zeta\left(\theta_0+\frac{r}{\sqrt{N}}\right)dr,$$

where set $c'_2(r)=\frac{N\varepsilon(\theta_0+\frac{r}{\sqrt{N}})}{r^2}$ and $d'_2(r)=\frac{N^{\frac{1}{2}-\epsilon}\beta_N(\theta_0+\frac{r}{\sqrt{N}})}{r}$, $\epsilon>0$. From Assumption 2.2 (2) and (3), there exists some random variable $\kappa>0$ on $(\Omega,\mathcal{F},P_{\theta_0})$ such that $\zeta\geq\kappa$, then we have, for N large enough, using Proposition 4.1 (i) and (ii),

$$I_N(1)\geq\frac{\kappa e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)}}{\sqrt{N}}\int_{\Theta'_N} e^{c_1r^2-MrN^\epsilon}dr.$$

Now we just compute the integral explicitly, using the change of variables $y=r-\frac{MN^\epsilon}{2c_1}$, in order to find a lower bound as follows:

$$\int_{\Theta'_N} e^{c_1r^2-MrN^\epsilon}dr=e^{\frac{M^2N^{2\epsilon}}{2c_1}}\int_{\Theta''_N} e^{c_1y^2}dy\geq\frac{\sqrt{\pi}}{\sqrt{2c_1}}e^{\frac{M^2N^{2\epsilon}}{2|c_1|}}.$$

Here $\Theta''_N=[\sqrt{N}(\theta^l-\theta_0)-\frac{MN^\epsilon}{2c_1},\sqrt{N}(\theta^u-\theta_0)-\frac{MN^\epsilon}{2c_1}]$.

Taking the quotient between (10), (11) and the above lower bound, we get

$$\leq\frac{\|f'\|_\infty\|\zeta\|_\infty|c_1|}{\sqrt{\pi}\kappa|c_2|\sqrt{N}}\left\{1+\frac{MN^\epsilon}{\sqrt{|c_2|}}\right\}e^{-|c_2|c^2N^{2a}+cMN^{\epsilon+a}-\frac{M^2N^{2\epsilon}}{2c_1}}.$$

Putting together the above bound with (8) and as $a > \epsilon$, one obtains the announced rate. This finishes the proof of (5). Similarly we can prove that there exists some positive random variable C_2, independent of N, such that

$$|\Delta_2| \leq \frac{C_2}{N^{\frac{1}{2}-a}} \ a.s.$$

The only important point in the treatment of this term is to first note that instead of the decomposition (6), we will have

$$\hat{I}_N(f) = e^{NH(p_{\theta_0},p_{\theta_0})+\sqrt{N}Z_N(\theta_0)} \int_{\Theta_0} f(\theta) e^{N\varepsilon(\theta)+\sqrt{N}\beta_N(\theta)+\sqrt{N}(R^1_N(\theta)+R^2_N(\theta))}\zeta(\theta)d\theta$$

Therefore we obtain our conclusion if one follows the same calculations as above and further uses the result in Proposition 4.1 (iii). □

5. Proof of Proposition 4.1

5.1. Proof of the upper estimate for Proposition 4.1 (i).

Proposition 5.1. *Under Assumption 2.2 (3), (4) and (5), there exists some strictly negative constant c_2 such that*

$$\sup_{\theta\in\Theta_0} \frac{\varepsilon(\theta)}{(\theta-\theta_0)^2} \leq c_2 < 0.$$

Proof. From Exercise 1.22 (c) in pp.353 of Eggermont, LaRiccia [11], we have the following generalization of Pinsker's inequality; for all $\theta \in \Theta_0$,

$$0 \leq \frac{1}{2}\left(\int |p_\theta(y,z) - p_{\theta_0}(y,z)|\, dz\right)^2 \leq \int \ln \frac{p_{\theta_0}(y,z)}{p_\theta(y,z)} p_{\theta_0}(y,z)dz.$$

The finiteness and good definition of the above upper bound follows from Assumption 2.2 (3) and (5). Therefore from the definition of ε, we have that for all $\theta \in \Theta_0$,

$$\begin{aligned}\varepsilon(\theta) &= \int \left\{ -\int \left(\ln \frac{p_{\theta_0}(y,z)}{p_\theta(y,z)} \right) p_{\theta_0}(y,z)dz \right\} \mu_{\theta_0}(y)dy \\ &\leq -\frac{1}{2}\int \left(\int |p_\theta(y,z) - p_{\theta_0}(y,z)|dz \right)^2 \mu_{\theta_0}(y)dy \leq 0.\end{aligned}$$

From the identifiability condition (Assumption 2.2 (4)), we obtain the following; for all $\theta \in \Theta_0$,

$$\varepsilon(\theta) \leq -\frac{1}{2}\int c(y)^2(\theta-\theta_0)^2 \mu_{\theta_0}(y)dy = -\frac{1}{2}C(\theta_0)(\theta-\theta_0)^2.$$

Hence we have

$$\sup_{\theta\in\Theta_0} \frac{\varepsilon(\theta)}{(\theta-\theta_0)^2} \leq -\frac{1}{2}C(\theta_0) < \infty.$$

□

5.2. Proof of the lower estimate for Proposition 4.1 (i).

First we give a useful lemma for the first derivative of $H(p_\theta, p_{\theta_0})$ in θ. Its proof is straightforward.

Lemma 5.2. *Let q be a transition density, which depends on a parameter θ. We assume that for all $\theta \in \Theta$,*

$$\frac{\partial}{\partial\theta}\iint (\ln q(y,z;\theta))\, q(y,z;\theta_0)\mu_{\theta_0}(y)dydz = \iint \left(\frac{\partial}{\partial\theta}\ln q(y,z;\theta)\right) q(y,z;\theta_0)\mu_{\theta_0}(y)dydz,$$

and the following exchange of derivative and integral

$$\iint \frac{\partial}{\partial\theta} q(y,z;\theta)\Big|_{\theta=\theta_0} \mu_{\theta_0}(y)dydz = \int \frac{\partial}{\partial\theta}\int q(y,z;\theta)dz\Big|_{\theta=\theta_0} \mu_{\theta_0}(y)dy.$$

Then

$$\frac{\partial}{\partial\theta}\iint \left(\ln q(y,z;\theta)\right) q(y,z;\theta_0)\mu_{\theta_0}(y)dydz\Big|_{\theta=\theta_0} = 0.$$

Proposition 5.3. *Under Assumption 2.2 (3), (4) and (5), there exists some strictly negative constant c_1 such that*

$$-\infty < c_1 \le \inf_{\theta\in\Theta} \frac{\varepsilon(\theta)}{(\theta-\theta_0)^2} < 0.$$

Proof. For $\theta \in \Theta_0$, we apply the Taylor expansion to $\theta \mapsto H(p_\theta, p_{\theta_0})$ around θ_0 and by Lemma 5.2,

$$H\left(p_\theta, p_{\theta_0}\right) = H\left(p_{\theta_0}, p_{\theta_0}\right) + \frac{1}{2}(\theta-\theta_0)^2 \frac{\partial^2}{\partial\theta^2} H\left(p_\theta, p_{\theta_0}\right)\Big|_{\theta=\theta(\gamma)}, \tag{12}$$

where $\theta(\gamma) := \gamma\theta + (1-\gamma)\theta_0$, for some $\gamma \in (0,1)$.

From (12), we have

$$\varepsilon(\theta) = H\left(p_\theta, p_{\theta_0}\right) - H\left(p_{\theta_0}, p_{\theta_0}\right) = \frac{1}{2}(\theta-\theta_0)^2 \frac{\partial^2}{\partial\theta^2} H\left(p_\theta, p_{\theta_0}\right)\Big|_{\theta=\theta(\gamma)}.$$

From Assumption 2.2 (5), $\frac{\partial^2}{\partial\theta^2} H(p_\theta, p_{\theta_0})$ satisfies;

$$\sup_{\theta\in\Theta}\left|\frac{\partial^2}{\partial\theta^2} H\left(p_\theta, p_{\theta_0}\right)\right| < \infty.$$

Finally from Proposition 5.1, we have

$$0 > \inf_{\theta\in\Theta} \frac{\varepsilon(\theta)}{(\theta-\theta_0)^2} \ge -\frac{1}{2}\sup_{\theta\in\Theta}\left|\frac{\partial^2}{\partial\theta^2} H\left(p_\theta, p_{\theta_0}\right)\right| > -\infty.$$

□

5.3. **Proof of Proposition 4.1 (ii).** Our next goal is to prove the uniform estimates for β_N. The difference with the previous section lies on the fact that now these quantities are random. Therefore one naturally is lead to the consideration of limit theorems in the space of continuous functions in θ with the supremum norm.

In what follows we use the following notation for a sequence of strongly mixing sequence of random variables valued on the Banach space of continuous functions on $[\theta_l, \theta_u]$ with the maximum norm, denoted by $\|\cdot\|$.

$$\eta_i(\theta) = \frac{\ln\left(p_\theta/p_{\theta_0}(Y_i, Y_{i+1})\right) - \left(H(p_\theta, p_{\theta_0}) - H(p_{\theta_0}, p_{\theta_0})\right)}{\theta-\theta_0}, \quad \theta \in [\theta_l, \theta_u].$$

As p_θ is C^1 in θ, we have

(13)

$$\begin{aligned}
&\sup_N \sup_{\theta\in\Theta_0} \left| \frac{N^{-\epsilon}\beta_N(\theta)}{\theta-\theta_0} \right| \\
&= \sup_N \sup_{\theta\in\Theta_0} \frac{1}{N^{1/2+\epsilon}} \left| \sum_{i=0}^{N-1} \frac{\ln p_\theta(Y_i,Y_{i+1}) - \ln p_{\theta_0}(Y_i,Y_{i+1}) - (H(p_\theta,p_{\theta_0}) - H(p_{\theta_0},p_{\theta_0}))}{\theta-\theta_0} \right| \\
&= \sup_N \sup_{\theta\in\Theta_0} \frac{1}{N^{1/2+\epsilon}} \left| \sum_{i=0}^{N-1} \frac{\int_0^1 \partial_\theta \ln p_{(1-\alpha)\theta_0+\alpha\theta}(Y_i,Y_{i+1}) d\alpha(\theta-\theta_0)}{\theta-\theta_0} \right. \\
&\qquad \left. - \frac{\int_0^1 \partial_\theta H(p_{(1-\alpha)\theta_0+\alpha\theta}, p_{\theta_0}) d\alpha(\theta-\theta_0)}{\theta-\theta_0} \right| \\
&= \sup_N \sup_{\theta\in\Theta_0} \frac{1}{N^{1/2+\epsilon}} \left| \sum_{i=0}^{N-1} \int_0^1 \left(\partial_\theta \ln p_{(1-\alpha)\theta_0+\alpha\theta}(Y_i,Y_{i+1}) - \partial_\theta H(p_{(1-\alpha)\theta_0+\alpha\theta}, p_{\theta_0}) \right) d\alpha \right|.
\end{aligned}$$

Therefore an equivalent way of setting the random variable, η_i, is

$$\eta_i(\theta) := \int_0^1 \left(\partial_\theta \ln p_{\tilde\theta(\alpha)}(Y_i,Y_{i+1}) - \partial_\theta H(p_{\tilde\theta(\alpha)}, p_{\theta_0}) \right) d\alpha,$$

where $\tilde\theta(\alpha) := (1-\alpha)\theta_0 + \alpha\theta$. Note that $E[\eta_i(\theta)] = 0$ holds.

In order to carry out the proof we will follow similar steps as in the proof of Theorem 3 in Dehling [9] which also uses an argument which appears in Proposition 2.2 in Kuelbs and Philips [19]. To apply their arguments, one needs to have moment estimates which can be easily obtained but they need to be explicit in order to work with the Banach space of continuous functions on $[\theta_l, \theta_u]$ with the supremum norm. Here, we will also use the moment norm notation $\|\eta\|_p := E\left[|\eta|^p\right]^{1/p}$ for a real valued random variable η.

Lemma 5.4. *Under Assumption 2.2 (5), one has that for $\delta \in [0,2]$*

$$\left\| \sup_\theta \sum_{j=1}^N \eta_j(\theta) \right\|_{2+\delta} \le CN^{1/2}.$$

Here C is an explicit constant that depends only on δ, θ_l, θ_u, $A_r(\alpha) := \sum_{i=0}^\infty (i+1)^{r/2-1}[\alpha(i)]^{\frac{2}{r+2}}$, $r = 2, 4$ and finally the constants in the Assumption 2.2 (5).

Proof. In fact, for any $\delta \ge 0$, using the fundamental theorem of calculus and Hölder's inequality

$$\begin{aligned}
\sup_\theta \left| \sum_{j=1}^N \eta_j(\theta) \right|^{2+\delta} &\le 2^{1+\delta} \sup_\theta \left| \sum_{j=1}^N \int_{\theta_l}^\theta \partial_\theta \eta_j(\theta) d\theta \right|^{2+\delta} + 2^{1+\delta} \left| \sum_{j=1}^N \eta_j(\theta_l) \right|^{2+\delta} \\
&\le 2^{1+\delta}(\theta_u-\theta_l)^{1+\delta} \int_{\theta_l}^{\theta_u} \left| \sum_{j=1}^N \partial_\theta \eta_j(\theta) \right|^{2+\delta} d\theta + 2^{1+\delta} \left| \sum_{j=1}^N \eta_j(\theta_l) \right|^{2+\delta}. \qquad (14)
\end{aligned}$$

The second term on the right hand side of the above expression can be dealt with usual estimates for the moments of mean zero random variables which are α-mixing of the required order as explained in Theorem 1 in Yokoyama [22].

Rather than following the general path in that article, we will only use the case for the integer power of 4 proved in Theorem 1, case (i), as the constant is explicit and easy to understand in that case. Therefore we let $\delta \le 2$ and then

$$E\left[\left|\sum_{j=1}^{N}\partial_\theta\eta_j(\theta)\right|^{2+\delta}\right] \le \left\|\sum_{j=1}^{N}\partial_\theta\eta_j(\theta)\right\|_4^{2+\delta}$$
$$\le K_{\alpha,3}^{\frac{2+\delta}{4}} E\left[|\partial_\theta\eta_j(\theta)|^{4+\delta}\right]^{\frac{2+\delta}{4+\delta}} n^{\frac{2+\delta}{2}},$$
$$K_{\alpha,3} := 36A_4(\alpha) + 288(A_2(\alpha))^2.$$

Therefore putting these estimates together in (14), we obtain that

$$E\left[\sup_\theta\left|\sum_{j=1}^{N}\eta_j(\theta)\right|^{2+\delta}\right]$$
$$\le C(\delta,\theta_u,\theta_l)\left(K_{\alpha,3}^{\frac{2+\delta}{4}} N^{\frac{2+\delta}{2}}\int_{\theta_l}^{\theta_u} E\left[|\partial_\theta\eta_j(\theta)|^{4+\delta}\right]^{\frac{2+\delta}{4+\delta}} d\theta + E\left[|\partial_\theta\eta_j(\theta_l)|^{2+\delta}\right]\right),$$
$$C(\delta,\theta_u,\theta_l) := 2^{1+\delta}\left(1+(\theta_u-\theta_l)^{1+\delta}\right).$$

Using the finiteness for $E\left[|\partial_\theta\eta_j(\theta)|^r\right]$ for $r = 4+\delta, 2+\delta$ in Assumption 2.2 (5) and the above argument one concludes that for any $N \in \mathbb{N}$

$$E\left[\sup_\theta\left|\sum_{j=1}^{N}\eta_j(\theta)\right|^{2+\delta}\right] \le CN^{1+\delta/2}. \tag{15}$$

Note that the finiteness of $A_2(\alpha)$ and $A_4(\alpha)$ follows from Assumption 2.2 (1).

□

We will be using in what follows the following result.

Theorem 5.5. *(Dehling [9], Theorem 2) Let $\{\eta_j,\ j \ge 1\}$ be a weakly stationary strong mixing sequence of random variables with values in the separable Banach space X with norm $\|\cdot\|$ such that $E[\eta_j] = 0$ for each $j \ge 1$ and $\sup_{j\ge1} E[\|\eta_j\|^{2+\delta}] \le \rho_{2+\delta} < \infty$ for some $0 < \delta < \frac{2}{3}$ and suppose that the mixing rate is*

$$\alpha(n) = O(n^{-(1+\varepsilon)(1+2/\delta)}) \quad as\ n\to\infty\ for\ some\ \varepsilon \in (0,1].$$

Let P_m be a sequence of bounded operators on X with m-dimensional range satisfying

$$\sup_{\|x\|=1}\|P_m x\| = O(m^r) \quad as\ m\to\infty\ for\ some\ r>0, \tag{16}$$

and uniformly on a and N

$$E\left[\left\|N^{-1/2}\sum_{j=a+1}^{a+N}(\eta_j - P_m\eta_j)\right\|^2\right] = O(m^{-s}) \quad as\ m\to\infty\ for\ some\ s>0. \tag{17}$$

Then there exists a covariance operator T which converges absolutely such that the Banach valued Gaussian random variable $B(T)$ with covariance structure T satisfies the following law distance estimate for $\kappa = \frac{s\delta\varepsilon}{200(2+\varepsilon)(9+3r+s)}$

$$\mathcal{M}\left(n^{-1/2}\sum_{j=1}^{n}\eta_j, B(T)\right) = O((1+\rho_{2+\delta}^{1/3})n^{-\kappa}) \quad as\ n\to\infty.$$

Here $\mathcal{M}$ stands for the Lévy-Prohorov distance between probability measures and the covariance operator T is given for $f, g \in B^*$ by the following absolutely convergence sum

$$T(f,g) = E\left[fg(\eta_1)\right] + \sum_{k\geq 2} E\left[f(\eta_1)g(\eta_k)\right] + \sum_{k\geq 2} E\left[g(\eta_1)f(\eta_k)\right].$$

Lemma 5.6. *Assume the conditions stated in Lemma 5.4. Then the conditions stated in Theorem 5.5 are satisfied with $s = 3$, any $r > 0$ and $\delta \in (1/3, 2/3)$.*

Proof. The boundedness of $\sup_{j\geq 1} E[\|\eta_j\|^{2+\delta}]$ follows from Lemma 5.4 for $N = 1$ and the stationarity hypothesis for Y. The α-mixing condition follows from Assumption 2.2 (take e.g $\varepsilon = 1$ and any $\delta \in (0, 2/3)$).

For $m \in \mathbb{N}$ and $i = 0, 1, ..., m$, set $\theta_i := \theta_l + \frac{i}{m}(\theta_u - \theta_l)$. Let P_m be a projection which is defined as

$$P_m x(\theta) = \frac{m}{\theta_u - \theta_l}(x(\theta_{i+1}) - x(\theta_i))(\theta - \theta_i) + x(\theta_i) \quad \textit{when } \theta \in [\theta_i, \theta_{i+1}),$$

for $x \in C(\Theta_0)$. Then note that $P_m x$ is continuous in θ and that for any $m \in \mathbb{N}$, $\frac{\|P_m x\|}{\|x\|} \leq 1$ holds, therefore $\{P_m\}$ is a sequence of bounded operators. Therefore the condition (16) with any $r > 0$ (in fact it is bounded).

Now we start to consider the assumption (17) for the projection operator P_m. Without loss of generality, we consider the case $a = 0$.

(18)

$$\begin{aligned}
&E\left[\sup_\theta \left|\sum_{j=1}^N (\eta_j(\theta) - P_m\eta_j(\theta))\right|^2\right] \\
&= E\left[\max_{i=0,1,\dots,m-1} \sup_{\theta\in[\theta_i,\theta_{i+1})} \left|\sum_{j=1}^N \left\{(\eta_j(\theta) - \eta_j(\theta_i)) \right.\right.\right. \\
&\left.\left.\left. - \frac{m}{\theta_u - \theta_l}\left(\eta_j(\theta_{i+1}) - \eta_j(\theta_i)\right)(\theta - \theta_i)\right\}\right|^2\right] \\
&= E\left[\max_{i=0,1,\dots,m-1} \sup_{\theta\in[\theta_i,\theta_{i+1})} \left|\sum_{j=1}^N \left\{\int_{\theta_i}^{\theta} \partial_\theta\eta_j(\beta)\,d\beta - \frac{m(\theta-\theta_i)}{\theta_u-\theta_l}\int_{\theta_i}^{\theta_{i+1}} \partial_\theta\eta_j(\gamma)\,d\gamma\right\}\right|^2\right] \\
&\leq E\left[\max_{i=0,1,\dots,m-1} \sup_{\theta\in[\theta_i,\theta_{i+1})} \left|\sum_{j=1}^N \left\{\int_{\theta_i}^{\theta} (\partial_\theta\eta_j(\beta) - \partial_\theta\eta_j(\theta))\,d\beta \frac{\theta_u - \theta_l - m(\theta-\theta_i)}{\theta_u-\theta_l} \right.\right.\right. \\
&\left.\left.\left. - \frac{m(\theta-\theta_i)}{\theta_u-\theta_l}\int_{\theta}^{\theta_{i+1}} (\partial_\theta\eta_j(\gamma) - \partial_\theta\eta_j(\theta))\,d\gamma\right\}\right|^2\right] \\
&\leq 2E\left[\max_{i=0,1,\dots,m-1} \sup_{\theta\in[\theta_i,\theta_{i+1})} \left\{\left|\sum_{j=1}^N \int_{\theta_i}^{\theta}\int_{\beta}^{\theta} \partial_\theta^2\eta_j(\gamma)\,d\gamma d\beta\right|^2 \right.\right. \\
&\left.\left. + \left|\sum_{j=1}^N \int_{\theta}^{\theta_{i+1}}\int_{\theta}^{\gamma} \partial_\theta^2\eta_j(\beta)\,d\beta d\gamma\right|^2\right\}\right]
\end{aligned}$$

$$\leq 2E\left[\max_{i=0,1,\dots,m-1}\sup_{\theta\in[\theta_i,\theta_{i+1})}\left\{\left|\int_{\theta_i}^{\theta}(\gamma-\theta_i)\sum_{j=1}^{N}\partial_\theta^2\eta_j\left(\gamma\right)d\gamma\right|^2\right.\right.$$
$$\left.\left.+\left|\int_{\theta}^{\theta_{i+1}}(\theta_{i+1}-\beta)\sum_{j=1}^{N}\partial_\theta^2\eta_j\left(\beta\right)d\beta\right|^2\right\}\right].$$

Let us compute the first term above. The second is dealt with similarly. In fact,

$$E\left[\max_{i=0,1,\dots,m-1}\sup_{\theta\in[\theta_i,\theta_{i+1})}\left|\int_{\theta_i}^{\theta}(\gamma-\theta_i)\sum_{j=1}^{N}\partial_\theta^2\eta_j\left(\gamma\right)d\gamma\right|^2\right]$$
$$\leq E\left[\max_{i=0,1,\dots,m-1}\sup_{\theta\in[\theta_i,\theta_{i+1})}(\theta-\theta_i)^2\left\{\int_{\theta_i}^{\theta}\left|\sum_{j=1}^{N}\partial_\theta^2\eta_j\left(\gamma\right)\right|d\gamma\right\}^2\right]$$
$$\leq E\left[\max_{i=0,1,\dots,m-1}\sup_{\theta\in[\theta_i,\theta_{i+1})}(\theta_{i+1}-\theta_i)^3\int_{\theta_l}^{\theta_u}\left|\sum_{j=1}^{N}\partial_\theta^2\eta_j\left(\gamma\right)\right|^2d\gamma\right]$$
$$\leq\frac{(\theta_u-\theta_l)^3}{m^3}\int_{\theta_l}^{\theta_u}E\left[\left|\sum_{j=1}^{N}\partial_\theta^2\eta_j\left(\gamma\right)\right|^2\right]d\gamma.$$

We would like to prove this upper boundedness using Yokoyama [22]. To use it, we need $E[\partial_\theta^2\eta_j(\gamma)]=0$ for all $\gamma\in[\theta_l,\theta_u]$. In fact, using Assumption 2.2, (5), we have that the expectation and the partial derivative in θ are exchangable and therefore

$$E[\partial_\theta^2\eta_j(\gamma)]=E\left[\partial_\theta^2\int_0^1\left(\partial_\theta\ln p_{\tilde{\theta}(\alpha)}(Y_i,Y_{i+1})-\partial_\theta H(p_{\tilde{\theta}(\alpha)},p_{\theta_0})\right)d\alpha\right]$$
$$=\partial_\theta^2\int_0^1E\left[\partial_\theta\ln p_{\tilde{\theta}(\alpha)}(Y_i,Y_{i+1})-\partial_\theta H(p_{\tilde{\theta}(\alpha)},p_{\theta_0})\right]d\alpha$$
$$=\partial_\theta^2\int_0^1\left\{\partial_\theta E\left[\ln p_{\tilde{\theta}(\alpha)}(Y_i,Y_{i+1})\right]-\partial_\theta H(p_{\tilde{\theta}(\alpha)},p_{\theta_0})\right\}d\alpha=0.$$

Then, from the satarionarity and eq. (4.2) in Yokoyama [22], we have

$$E\left[\sum_{j=1}^{N}\partial_\theta^2\eta_j\left(\gamma\right)^2+2\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}\partial_\theta^2\eta_j\left(\gamma\right)\partial_\theta^2\eta_k\left(\gamma\right)\right]$$
$$=NE\left[\partial_\theta^2\eta_1\left(\gamma\right)^2\right]+2\sum_{j=1}^{N-1}(N-j)E\left[\partial_\theta^2\eta_1\left(\gamma\right)\partial_\theta^2\eta_{j+1}\left(\gamma\right)\right]$$
$$\leq NE\left[\partial_\theta^2\eta_1\left(\gamma\right)^2\right]+2N\sum_{j=1}^{N-1}\left|E\left[\partial_\theta^2\eta_1\left(\gamma\right)\partial_\theta^2\eta_{j+1}\left(\gamma\right)\right]\right|$$
$$\leq NE\left[\partial_\theta^2\eta_1\left(\gamma\right)^2\right]+24NE\left[\left|\partial_\theta^2\eta_1\left(\gamma\right)\right|^{2+\delta}\right]^{\frac{2}{2+\delta}}\sum_{i=0}^{\infty}\alpha(i)^{\frac{\delta}{2+\delta}},$$

for any $\delta \in (0,2]$ where we need that $\sum_{i=0}^{\infty} \alpha(i)^{\frac{\delta}{2+\delta}} < \infty$ and for all $\gamma \in [\theta_l, \theta_u]$, $E\left[\left|\partial_\theta^2 \eta_1(\gamma)\right|^{2+\delta}\right]^{\frac{2}{2+\delta}} < \infty$ holds. Therefore we have for $\delta = 2$

$$\frac{(\theta_u-\theta_l)^3}{m^3}\int_{\theta_l}^{\theta_u} E\left[\left|\sum_{j=1}^{N}\partial_\theta^2\eta_j(\gamma)\right|^2\right]d\gamma$$
$$\leq N\frac{(\theta_u-\theta_l)^3}{m^3}\left\{\int_{\theta_l}^{\theta_u} E\left[\partial_\theta^2\eta_1(\gamma)^2\right]d\gamma + 24\int_{\theta_l}^{\theta_u} E\left[\left|\partial_\theta^2\eta_1(\gamma)\right|^4\right]^{\frac{1}{2}}d\gamma\sum_{i=0}^{\infty}\alpha(i)^{\frac{1}{2}}\right\}.$$

Using the proof in Lemma 5.4, we have that

$$\int_{\theta_l}^{\theta_u} E\left[\partial_\theta^2\eta_1(\gamma)^2\right]d\gamma < \infty \text{ and } \int_{\theta_l}^{\theta_u} E\left[\left|\partial_\theta^2\eta_1(\gamma)\right|^4\right]^{\frac{1}{2}}d\gamma < \infty$$

holds. Therefore, we have proved the uniform boundedness in θ for the first term of eq. (18). The second term of eq. (18) can be proved similarly.

From the above estimate one can easily obtain that

$$E\left[\left\|N^{-1/2}\sum_{j=1}^{N}(\eta_j - P_m\eta_j)\right\|^2\right] = O(m^{-3}) \quad as\ m\to\infty.$$

□

Theorem 5.7. *For any $a > 0$, we have that*

$$P\left(N^{-1/2-a}\sup_\theta\sum_{j=1}^{N}\eta_j(\theta)\to 0, \quad as\ N\to\infty\right) = 1.$$

Proof. Before starting the proof, we remark in order to prove the statement in Theorem 5.7, we can assume without loss of generality that T is non-degenerate.

In fact, consider a sequence of i.i.d. Gaussian r.v. β_j. Then repeat the same proof to follow for both $n^{-1/2-a}\sum_{j=1}^{n}(\eta_j+\beta_j)$ and $n^{-1/2-a}\sum_{j=1}^{n}\beta_j$ which will converge to zero almost surely if we prove that their corresponding covariance operators are non-degenerate. Note that for the first sum, we have

$$T_1(f,g) = E\left[fg(\eta_1+\beta_1)\right] + \sum_{k\geq 2}E\left[f(\eta_1+\beta_1)g(\eta_k+\beta_k)\right] + \sum_{k\geq 2}E\left[g(\eta_1+\beta_1)f(\eta_k+\beta_k)\right].$$

Using the linearity of f and g, the fact that Banach valued Gaussian random variables have mean zero and the independence between the sequences η_j, β_j (as well as within all β_j) we obtain

$$T_1(f,g) = T(f,g) + E\left[fg(\beta_1)\right].$$

Therefore the non-degeneracy follows.

The proof as previously announced, although long, it uses basic ideas which are explained in Dehling between other references. The idea is to separate the sum $\sum_{j=1}^{N}\eta_j(\theta)$ into blocks so that half of the blocks will be negligible and the other blocks will converge to the Gaussian law on the Banach space of continuous functions. The removal of negligible blocks will allow us to obtain the convergence almost surely. This idea of using "Bernstein" blocks is very old in probability theory and it can be found for example in Section 18.1 of Ibragimov and Linnik [15].

Let us denote the blocks, following Dehling [9], page 423, as H_k and I_k which are subsets of consecutive indices so that $\#(H_k) = n_k = [k^{1+\beta'}]$, $\#(I_k) = [k^{\frac{2+\beta'}{4}}]$. Here $\beta' := \beta^{-1} > 0$, where $s = 3$ and $r > 0$ can be chosen freely (recall Lemma 5.6) and

$$\beta := \frac{s\lambda}{9+3r+s}$$
$$\lambda := \frac{\delta\epsilon}{200(2+\epsilon)}.$$

The sum $\sum_{j=1}^{N} \eta_j(\theta)$ is then divided using these blocks. The main blocks are defined as $Y_k(\theta) := \sum_{\nu\in H_k} \eta_\nu(\theta)$.

Furthermore, the above defined constants appear when one wants to apply Theorem 2 of Dehling which it further requires that $\alpha(k) \leq Ck^{-(1+\epsilon)(1+\frac{2}{\delta})}$ for some $\epsilon \in (0,1]$ and $\delta \in (0, \frac{2}{3})$. Finally as $\|\eta_\nu\|_{2+\delta} \leq \|\eta_\nu\|_4 \leq \rho$ for some constant ρ then for all $k \in \mathbb{N}$,

$$\mathcal{M}\left(\mathcal{L}\left(n_k^{-1/2} Y_k\right), N(0,T)\right) \leq C(1+\rho^{\frac{2+\delta}{3}}) n_k^{-\beta}.$$

Here without loss of generality, we assume that T is a well defined non-trivial covariance operator of the Gaussian law $N(0,T)$ in the corresponding Banach space. Here $\mathcal{M}$ denotes the Lévy-Prohorov distance between probability laws.

Now the argument follows by proving that one can apply Borel-Cantelli lemma for the study of the almost sure convergence of $n_k^{-1/2-a} Y_k$ for any $a > 0$. In fact, using the definition of the Prohorov distance and Chebyshev's inequality together with Fernique's theorem we have that for any $r \geq C(1+\rho^{\frac{2+\delta}{3}}) n_k^{-\beta}$

$$\begin{aligned} P(\|n_k^{-1/2} Y_k\| \geq n_k^a) \leq & P(\|N(0,T)\| \geq n_k^a - r) + r \\ \leq & c\exp\left(-c(n_k^a - r)^2\right) + r. \end{aligned}$$

Therefore, as $n_k \leq Ck^{1+\beta'}$,

$$\sum_{k\in\mathbb{N}} P(\|n_k^{-1/2} Y_k\| \geq n_k^a) \leq 2C(1+\rho^{\frac{2+\delta}{3}}) \sum_{k\in\mathbb{N}} n_k^{-\beta} < \infty.$$

From Borel-Cantelli's lemma then we conclude that $\|n_k^{-1/2} Y_k\| \leq n_k^a$, for all k sufficiently large almost surely.

Now, in order to deal with the smaller blocks I_j, we have to use a similar technique as in Kuelbs and Philipp [19], Proposition 2.2. That is, define for $t_k := \sum_{j=1}^{k} \#(H_j \cup I_j)$,

$$F_k(r,s) := \left\| \sum_{\nu=t_k+1+r}^{t_k+r+s} \eta_\nu \right\|.$$

Note that there exists two positive constants c_0 and c_1 such that $t_k \in [c_0 k^{2+\beta'}, c_1 k^{2+\beta'}]$. Suppose that $t_k \leq N < t_{k+1}$ and let $\bar{n}_k \equiv \bar{n}_k(N) := \max\{n; 2^n \leq N - t_k\}$. If we write $N - t_k = \sum_{l=0}^{\bar{n}_k} \varepsilon_l 2^l$, $\varepsilon_l \in \{0,1\}$ in its dyadic expansion we obtain from a combinatorics lemma in S. and L. Gaal [13], Lemma 3.10, that there exists $0 \leq m_l < 2^{n-l}$ (which depends on the sequence ε) so that

$$F_k(0, N - t_k) \leq \sum_{l=0}^{\bar{n}_k} F_k(m_l 2^{l+1}, 2^l).$$

With the introduction of this notation we now claim that our objective is to prove that

$$\max_{t_k < N \leq t_k + [k^{\frac{2+\beta'}{4}}]} \left\| \sum_{\nu=t_k+1}^{N} \eta_\nu \right\| \leq t_k^{1/2-\gamma}, \tag{19}$$

for any $\gamma \in (0, 1/2)$ to be explicitly determined in the proof.

To prove the above statement, define the sets

$$G_k(m,l) := \left\{ F_k(m2^{l+1}, 2^l) \geq t_k^{\frac{1-\gamma}{2}} \right\}$$
$$G_k := \cup_{l \leq \bar{n}_k} \cup_{m < 2^{\bar{n}_k - l}} G_k(m,l).$$

Then using Chebyshev's inequality and the moment bounds in Lemma 5.4, we obtain

$$\begin{aligned} P(G_k(m,l)) =& P\left(\left\| \sum_{\nu = t_k + m2^{l+1}+1}^{t_k + m2^{l+1} + 2^l} \eta_\nu \right\| \geq t_k^{\frac{1-\gamma}{2}} \right) \\ \leq & C2^{2l} t_k^{-2(1-\gamma)}. \end{aligned}$$

Therefore

$$\begin{aligned} P(G_k) \leq & C t_k^{-2(1-\gamma)} \sum_{l=1}^{\bar{n}_k} 2^{\bar{n}_k + l} \leq C k^{-2(2+\beta')(1-\gamma)} 2^{2\bar{n}_k} \\ \leq & C k^{-2(2+\beta')(1-\gamma)} (N - t_k)^2 \leq C k^{-2(2+\beta')(1-\gamma)} (t_{k+1} - t_k)^2 \\ \leq & C k^{-2+2(2+\beta')\gamma}. \end{aligned}$$

If we choose any $\gamma < \frac{1}{2(2+\beta')}$, we have that $\sum_k P(G_k) < \infty$ therefore by Borel-Cantelli's lemma we obtain that for k sufficiently large, and for all $l \leq \bar{n}_k$ and all $m \leq 2^{\bar{n}_k - l}$, then $F_k(m2^{l+1}, 2^l) < t_k^{\frac{1-\gamma}{2}}$ almost surely. Therefore for all k sufficiently large, we have that for any $\gamma' < \gamma$

$$F_k(0, N - t_k) \leq \sum_{l=0}^{\bar{n}_k} F_k(m_l 2^{l+1}, 2^l) \leq \bar{n}_k t_k^{\frac{1-\gamma}{2}} \leq C t_k^{\frac{1-\gamma'}{2}}.$$

Note that here we have used the fact that $\bar{n}_k \leq C \log_2(t_k)$. In fact,

$$\begin{aligned} \bar{n}_k \leq \log_2(N - t_k) \leq \log_2(t_{k+1} - t_k) = \log_2(\#(H_k \cup I_k)) \leq \log_2(2k^{1+\beta'}) \\ \leq 1 + \frac{1+\beta'}{2+\beta'} \log_2(c_0^{-1} t_k). \end{aligned}$$

From the above, we can conclude the claimed statement in (19).

Now, we can conclude the proof, using the above estimates for each sum in H_j and I_j as follows for some a' to be chosen later

$$\begin{aligned} t_k^{-\frac{1}{2} - a'} \left\| \sum_{j=1}^k \sum_{\nu \in H_j \cup I_j} \eta_\nu \right\| \leq & t_k^{-\frac{1}{2} - a'} \sum_{j=1}^k \left(n_j^{1/2 + a} + t_j^{1/2 - \gamma} \right) \\ \leq & C t_k^{-\frac{1}{2} - a'} \left(k^{(1+\beta')(1/2+a)+1} + k^{(2+\beta')(1/2-\gamma)+1} \right) \\ \leq & C t_k^{-\frac{1}{2} - a'} \left(t_k^{\frac{(1+\beta')(1/2+a)+1}{2+\beta'}} + t_k^{\frac{(2+\beta')(1/2-\gamma)+1}{2+\beta'}} \right). \end{aligned}$$

Therefore in order to obtain that the above converges to zero almost surely, we need to have that a' satisfies

$$a' > \max \left\{ \frac{2(1+\beta')a + 1}{2(2+\beta')}, \frac{1}{2+\beta'} - \gamma \right\}.$$

Finally in order to prove that a' can be chosen as small as possible, we recall that $\beta' = \beta^{-1} = \frac{9+3r+s}{s\lambda}$, therefore taking r large enough, β' can be large as desired. Therefore the result will follow by taking a as small as needed. □

5.4. **Proof of Proposition 4.1 (iii).** The fact that R_N^1 also includes randomness coming from the simulation process makes the proof of the estimate in Proposition 4.1 (iii) particularly difficult. Instead of dealing with it in full generality, we made a strong assumption. That is, the hypothesis (6) in Assumption 2.2. Later in Section 6 we give conditions in order to verify this hypothesis. This hypothesis expresses the main parameter tuning between the parameters n, h and N. Note that

$$J_N^1(\theta,\omega) := \frac{R_N^1(\theta,\omega) - R_N^1(\theta_0,\omega)}{\theta-\theta_0}$$
$$= \begin{cases} \dfrac{1}{\sqrt{N}} \displaystyle\sum_{i=0}^{N-1} \frac{1}{\theta-\theta_0} \left\{ \ln \frac{\hat{p}_\theta^N}{\bar{p}_\theta^N}(Y_i,Y_{i+1}) - \ln \frac{\hat{p}_{\theta_0}^N}{\bar{p}_{\theta_0}^N}(Y_i,Y_{i+1}) \right\}, & \theta \neq \theta_0 \\ \dfrac{1}{\sqrt{N}} \displaystyle\sum_{i=0}^{N-1} \frac{\partial}{\partial\theta} \ln \frac{\hat{p}_\theta^N}{\bar{p}_\theta^N}(Y_i,Y_{i+1}) \Big|_{\theta=\theta_0}, & \theta=\theta_0. \end{cases}$$

$$J_N^2(\theta,\omega) := \frac{R_N^2(\theta,\omega) - R_N^2(\theta_0,\omega)}{\theta-\theta_0}.$$

Proposition 5.8. *Under Assumption 2.2 (6), we have for $i = 1, 2$*

$$\sup_N \sup_{\theta\in\Theta} N^{-\epsilon} \left| J_N^i(\theta,\omega) \right| < +\infty, \ a.s.$$

Proof. Using the mean value theorem for $q_\theta = \hat{p}_\theta^N, \ \bar{p}_\theta^N$, and the assumptions we have

$$J_N^1(\theta,\omega) = \frac{N^{-\epsilon}}{\sqrt{N}} \sum_{i=0}^{N-1} \int_0^1 \left(\frac{\partial}{\partial\theta} \ln \hat{p}^N_{t\theta+(1-t)\theta_0}(Y_i,Y_{i+1}) - \frac{\partial}{\partial\theta} \ln \bar{p}^N_{t\theta+(1-t)\theta_0}(Y_i,Y_{i+1}) \right) dt$$
$$\leq \sup_N \sup_{\theta\in\Theta} \left| \frac{N^{-\epsilon}}{\sqrt{N}} \sum_{i=0}^{N-1} \left(\frac{\partial}{\partial\theta} \ln \hat{p}_\theta^N(Y_i,Y_{i+1}) - \frac{\partial}{\partial\theta} \ln \bar{p}_\theta^N(Y_i,Y_{i+1}) \right) \right| < +\infty \ a.s. \tag{20}$$

The proof for J_N^2 is similar and omitted. □

6. Main Theorem: Parameter Tuning and Assumption 2.2 (6)

This section is devoted to show that Assumption 2.2 (6) is satisfied under sufficient smoothness hypothesis on the random variables and processes that appear in the problem as well as a certain parameter tuning condition.

The conditions in (6) are for the comparison between $\hat{p}$ and $\bar{p}$ and then between $\bar{p}$ and p. The second is easier to deal with than the first. Therefore we only perform the first and leave the second for the reader. For the study of the second we will only give some remarks in Section 6.5

In order to understand the role of all the approximation parameters, we rewrite $\hat{p}_\theta^N$ and $\bar{p}_\theta^N$ as follows

$$\hat{p}_\theta^N(y,z) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{X_{(m)}^{y,(k)}(\theta,\omega) - z}{h} \right), \ \bar{p}_\theta^N(y,z) = E\left[\frac{1}{h} K\left(\frac{X_{(m)}^{y,(1)}(\theta,\cdot) - z}{h} \right) \right].$$

The idea in order to obtain the property (3) is to first restrict to a compact set of values for the random variables $Y_i, i = 0, ..., N-1$. This is obtained using an exponential type Chebyshev's inequality and the Borel-Cantelli Lemma.

Lemma 6.1. *Assume the following hypothesis*

(H0). *The expectation,* $m_{c_1} := E[e^{c_1|Y_1|^2}]$ *is finite for some constant* $c_1 > 0$. *Furthermore let* $a_N \geq \theta^u - \theta^l$ *be a sequence of strictly positive numbers such that*

$$\sum_{N=1}^{\infty} N \exp\left(-c_1 a_N^2\right) < \infty.$$

Then we have by Borel-Cantelli's lemma that for a.s. $\omega \in \Omega$, *there exists* $N_0(\omega)$ *big enough such that for any* $N \geq N_0$, *we have* $\max_{i=1,...,N} |Y_i| < a_N$. *That is, for* $A_N := \{\omega \in \Omega;\ \exists i = 1, ..., N \text{ s.t. } |Y_i| > a_N\}$, *we have* $P\left(\limsup_{N\to\infty} A_N\right) = 0$.

The decomposition that we will use in order to prove (3) is as follows

$$B^N := \left\{(\mathbf{x}, \theta) = (y, z, \theta) \in \mathbb{R}^2 \times \Theta;\ \|\mathbf{x}\| < a_N\right\},$$

where $\|\cdot\|$ is the max-norm. Then, note that

$$\begin{aligned}
&\sup_{(\mathbf{x},\theta)\in B^N} \left| \frac{\partial_\theta \hat{p}_\theta^N}{\hat{p}_\theta^N}(y,z) - \frac{\partial_\theta \bar{p}_\theta^N}{\bar{p}_\theta^N}(y,z) \right| \\
&\leq \sup_{(\mathbf{x},\theta)\in B^N} \left| \frac{\partial_\theta \hat{p}_\theta^N(y,z)}{\bar{p}_\theta^N(y,z)} - \frac{\partial_\theta \bar{p}_\theta^m(y,z)}{\bar{p}_\theta^N(y,z)} \right| + \sup_{(\mathbf{x},\theta)\in B^N} \left| \frac{\partial_\theta \hat{p}_\theta^N(y,z)}{\hat{p}_\theta^N(y,z)} - \frac{\partial_\theta \hat{p}_\theta^N(y,z)}{\bar{p}_\theta^N(y,z)} \right| \\
&\leq \frac{\sup_{(\mathbf{x},\theta)\in B^N} \left|\partial_\theta \hat{p}_\theta^N(y,z) - \partial_\theta \bar{p}_\theta^N(y,z)\right|}{\inf_{(\mathbf{x},\theta)\in B^N} \bar{p}_\theta^N(y,z)} \\
&+ \sup_{(\mathbf{x},\theta)\in B^N} \left| \frac{\partial_\theta \hat{p}_\theta^N(y,z)}{\hat{p}_\theta^N(y,z)} \right| \frac{\sup_{(\mathbf{x},\theta)\in B^N} \left|\hat{p}_\theta^N(y,z) - \bar{p}_\theta^N(y,z)\right|}{\inf_{(\mathbf{x},\theta)\in B^N} \bar{p}_\theta^N(y,z)} \\
&=: \frac{A}{B} + C\frac{D}{B}, \qquad (21)
\end{aligned}$$

where we remark that

$$\begin{aligned}
\partial_\theta \ln \hat{p}_\theta^N(Y_i, Y_{i+1}) &= \frac{\partial_\theta \hat{p}_\theta^N(Y_i, Y_{i+1})}{\hat{p}_\theta^N(Y_i, Y_{i+1})}, \\
\partial_\theta \ln \bar{p}_\theta^N(Y_i, Y_{i+1}) &= \frac{\partial_\theta \bar{p}_\theta^N(Y_i, Y_{i+1})}{\bar{p}_\theta^N(Y_i, Y_{i+1})}, \\
\partial_\theta \hat{p}_\theta^N(y,z) &= \frac{1}{nh^2} \sum_{k=1}^{n} K'\left(\frac{X_{(m)}^{y,(k)}(\theta,\omega) - z}{h}\right) \partial_\theta X_{(m)}^{y,(k)}(\theta,\omega), \\
\partial_\theta \bar{p}_\theta^N(y,z;\omega) &= E\left[\frac{1}{h^2} K'\left(\frac{X_{(m)}^{y,(k)}(\theta,\cdot) - z}{h}\right) \partial_\theta X_{(m)}^{y,(k)}(\theta,\cdot)\right].
\end{aligned}$$

Therefore in order to prove the finiteness of (3), we need to bound $\sqrt{N}\left(\frac{A}{B} + C\frac{D}{B}\right)$. This will be done in a series of Lemmas using Borel-Cantelli arguments together with the modulus of continuity for the quantities $\bar{p}_\theta^N$ and $\hat{p}_\theta^N$. First, we start analyzing the difficult term: $C\frac{D}{B}$.

6.1. **Upper bound for $C\frac{D}{B}$ in (21).** We work in this section under the following hypotheses:

(H1). Assume that there exist some positive constants φ_1, φ_2, where φ_1 is independent of N and φ_2 is independent of N and Δ, such that the following holds;

$$\inf_{(\mathbf{x},\theta)\in B^N} \bar{p}_\theta^N(y,z) \geq \varphi_1 \exp\left(-\frac{\varphi_2 a_N^2}{\Delta}\right).$$

(H2). Assume that the kernel K is the Gaussian kernel; $K(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$.

(H3). Assume that for some constant $r_3 > 0$ and a sequence $\{b_{3,N}; N \in \mathbb{N}\} \subset [1, \infty)$, we have that $\sum_{N=1}^{\infty} \frac{n a_N^{2r_3} E[|Z_{3,N}(\cdot)|^{r_3}]}{(h^2 b_{3,N})^{r_3}} < \infty$, where

$$Z_{3,N}^{(k)}(\omega) := a_N^{-2} \left(\sup_{(\mathbf{x},\theta)\in B^N} \left| X_{(m)}^{y,(k)}(\theta,\omega) \right| + 1 \right) \sup_{(\mathbf{x},\theta)\in B^N} \left| \partial_\theta X_{(m)}^{y,(k)}(\theta,\omega) \right|. \tag{22}$$

(H4). Assume that for some constant $r_4 > 0$ and a sequence $\{b_{4,N}; N \in \mathbb{N}\} \subset [1, \infty)$, we have that $\sum_{N=1}^{\infty} \frac{n E[|Z_{4,N}(\cdot)|^{r_4}]}{(b_{4,N})^{r_4}} < \infty$, where

$$Z_{4,N}^{(k)}(\omega) := a_N^{-1} \left(\sup_{(\mathbf{x},\theta)\in B^N} \left| \partial_y X_{(m)}^{y,(k)}(\theta;\omega) \right| + \sup_{(\mathbf{x},\theta)\in B^N} \left| \partial_\theta X_{(m)}^{y,(k)}(\theta;\omega) \right| + 1 \right). \tag{23}$$

(H5). Assume that there exists some positive constant $C_5 > 0$ such that for all $y, z \in \mathbb{R}$, $m \in \mathbb{N}$ and $\theta \in \Theta$,

$$\left|\partial_y \bar{p}_\theta^N(y,z)\right|, \ \left|\partial_z \bar{p}_\theta^N(y,z)\right|, \ \left|\partial_\theta \bar{p}_\theta^N(x,y)\right| \le C_5 < +\infty.$$

(H6). Assume that η_N and ν_N are sequences of positive numbers so that

$$\sum_{N=1}^{\infty} \nu_N^3 \exp\left(-\frac{(\eta_N)^2 n h^2}{64\|K\|_\infty}\right) < \infty,$$

where $\|\cdot\|_\infty$ denotes the sup-norm.

(H7). Assume that $X_{(m)}^{y,(1)}(\theta)$ is once differentiable with respect to y and θ a.s.

Note that from assumption **(H1)**, we have a lower bound for B in (21).

6.2. **Upper bound of C in (21).**

Lemma 6.2. *Assume hypotheses **(H2)**, **(H3)** and **(H7)**, then we have that*

$$P\left(\limsup_{N\to\infty} \left\{ \sup_{(\mathbf{x},\theta)\in B^N} \left| \frac{\partial_\theta \hat{p}_\theta^N(y,z)}{\hat{p}_\theta^N(y,z)} \right| > b_{3,N} \right\} \right) = 0.$$

Proof. Consider

$$\sup_{(\mathbf{x},\theta)\in B^N} \left| \frac{\partial_\theta \hat{p}_\theta^N(y,z)}{\hat{p}_\theta^N(y,z)} \right| = \frac{1}{h} \sup_{(\mathbf{x},\theta)\in B^N} \left| \frac{\sum_{k=1}^n K \cdot \frac{K'}{K}\left(\frac{X_{(m)}^{y,(k)}(\theta,\omega)-z}{h}\right) \partial_\theta X_{(m)}^{y,(k)}(\theta,\omega)}{\sum_{k=1}^n K\left(\frac{X_{(m)}^{y,(k)}(\theta,\omega)-z}{h}\right)} \right|$$

$$\le \frac{1}{h} \sup_{(\mathbf{x},\theta)\in B^N} \max_{k=1,\dots,n} \left| \frac{K'}{K}\left(\frac{X_{(m)}^{y,(k)}(\theta,\omega)-z}{h}\right) \partial_\theta X_{(m)}^{y,(k)}(\theta,\omega) \right|. \tag{24}$$

Note that under hypothesis **(H2)**, $\frac{K'}{K}(x) = -x$. So we have

$$\begin{aligned}
(24) &\le \frac{1}{h^2} \sup_{(\mathbf{x},\theta)\in B^N} \max_{k=1,\dots,n} \left\{ \left(\left| X_{(m)}^{y,(k)}(\theta,\omega) \right| + |z| \right) \left| \partial_\theta X_{(m)}^{y,(k)}(\theta,\omega) \right| \right\} \\
&\le \frac{1}{h^2} \max_{k=1,\dots,n} \sup_{(\mathbf{x},\theta)\in B^N} \left\{ \left(\left| X_{(m)}^{y,(k)}(\theta,\omega) \right| + |z| \right) \left| \partial_\theta X_{(m)}^{y,(k)}(\theta,\omega) \right| \right\} \\
&\le \frac{a_N^2}{h^2} \max_{k=1,\dots,n} \left\{ Z_{3,N}^{(k)}(\omega) \right\}.
\end{aligned}$$

where we have used the definition (22). Define the set

$$B_{a_N,n}^m := \left\{ \frac{a_N^2}{h^2} \max_{k=1,\dots,n} Z_{3,N}^{(k)}(\omega) > b_{3,N} \right\}.$$

Note that $\{Z_{3,N}^{(k)}(\omega)\}_{k\in\mathbb{N}}$ is a sequence of i.i.d. r.v.'s, then from the Chebyshev's inequality, we have

$$\sum_{N=1}^{\infty} P\left(B_{a_N,n}^m\right) \le \sum_{N=1}^{\infty}\sum_{k=1}^{n} P\left(Z_{3,N}^{(k)}(\omega) > \frac{h^2}{a_N^2} b_{3,N}\right) \le \sum_{N=1}^{\infty} \frac{n a_N^{2r_3} E\left[|Z_{3,N}(\cdot)|^{r_3}\right]}{(h^2 b_{3,N})^{r_3}} < \infty,$$

where the above follows due to hypothesis **(H3)**. Then by the Borel-Cantelli lemma we have the conclusion. □

The above Borel-Cantelli argument is used repeatedly in what follows. We will use it from now on, without giving further details.

6.3. **Upper bound of D in (21).** In this section, we use the modulus of continuity for $\hat{p}^N$ and $\bar{p}^N$ in order to find an upper bound for B. For $\nu_N \in \mathbb{N}$, set

$$B_{l_1 l_2}^N := \left\{(\mathbf{x},\theta) \in \mathbb{R}^2 \times \Theta;\ \|\mathbf{x} - \mathbf{x}_{l_1}^N\| \le \frac{a_N}{\nu_N},\ |\theta - \theta_{l_2}^N| \le \frac{\theta^u - \theta^l}{\nu_N}\right\},$$
$$l_1 = 1,\cdots,\nu_N^2,\ l_2 = 1,\cdots,\nu_N,$$

such that $\mathring{B}_{l_1 l_2}^N \cap \mathring{B}_{l_1' l_2'}^N = \emptyset$ $((l_1,l_2) \neq (l_1',l_2'))$ and appropriate set of points $\mathbf{x}_{l_1}^N$, $\theta_{l_2}^N$, $l_1 = 1,...,\nu_N^2$ and $l_2 = 1,...,\nu_N$ such that $\cup_{l_1=1}^{\nu_N^2} \cup_{l_2=1}^{\nu_N} B_{l_1 l_2}^N = \overline{B^N}$. Then

$$\sup_{(\mathbf{x},\theta)\in B^N} \left|\hat{p}_\theta^N(\mathbf{x}) - \bar{p}_\theta^N(\mathbf{x})\right| = \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\hat{p}_\theta^N(\mathbf{x}) - \bar{p}_\theta^N(\mathbf{x})\right|$$
$$\le \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\hat{p}_\theta^N(\mathbf{x}) - \hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right| + \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \left|\hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right) - \bar{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right| \tag{25}$$
$$+ \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\bar{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right) - \bar{p}_\theta^N(\mathbf{x})\right|.$$

Now, consider the first term of (25).

Lemma 6.3. *Under **(H2)**, **(H4)** and **(H7)**, we have that*

$$P\left(\limsup_{N\to\infty}\left\{\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\hat{p}_\theta^N(\mathbf{x}) - \hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right| > \frac{\|K'\|_\infty}{h^2}\frac{a_N^2}{\nu_N}(b_{4,N}+1)\right\}\right) = 0.$$

Proof. If $(\mathbf{x},\theta) = (y,z,\theta) \in B_{l_1 l_2}^N$, then we have

$$\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\hat{p}_\theta^N(\mathbf{x}) - \hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right|$$
$$= \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\frac{1}{nh}\sum_{k=1}^{n}\left\{K\left(\frac{X_{(m)}^{y,(k)}(\theta;\omega) - z}{h}\right) - K\left(\frac{X_{(m)}^{y_{l_1}^N,(k)}(\theta_{l_2}^N;\omega) - z_{l_1}^N}{h}\right)\right\}\right|$$
$$\le \frac{\|K'\|_\infty}{h^2} \max_{k=1,\dots,n} \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|X_{(m)}^{y,(k)}(\theta;\omega) - X_{(m)}^{y_{l_1}^N,(k)}(\theta_{l_2}^N;\omega) - \left(z - z_{l_1}^N\right)\right|$$
$$= \frac{\|K'\|_\infty}{h^2} \max_{k=1,\dots,n} \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le\nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\left(y - y_{l_1}^N\right)\int_0^1 \partial_y X_{(m)}^{\varepsilon y + (1-\varepsilon)x_{l_1}^N,(k)}(\theta;\omega)\,d\varepsilon\right.$$
$$\left. + \left(\theta - \theta_{l_2}^N\right)\int_0^1 \partial_\theta X_{(m)}^{y_{l_1}^N,(k)}\left(\varepsilon\theta + (1-\varepsilon)\theta_{l_2}^N;\omega\right)d\varepsilon - \left(z - z_{l_1}^N\right)\right|$$

$$\le \frac{\|K'\|_\infty}{h^2}\frac{a_N}{\nu_N}\max_{k=1,\dots,n}\left\{\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}}\sup_{|y-y_{lN}|\le \frac{a_N}{\nu_N}}\left|\partial_y X^{y,(k)}_{(m)}(\theta;\omega)\right|\right.$$

$$\left.+\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}}\sup_{|\theta-\theta^N_{l_2}|\le \frac{\theta^u-\theta^l}{\nu_N}}\left|\partial_\theta X^{y^N_{l_1},(k)}_{(m)}(\theta;\omega)\right|+1\right\}$$

$$\le \frac{\|K'\|_\infty}{h^2}\frac{a_N^2}{\nu_N}\left\{\max_{k=1,\dots,n} Z^{(k)}_{4,N}(\omega)+1\right\}.$$

Here we have used definition (23) and that $a_N \ge \theta^u-\theta^l$. The proof finishes by using Markov's inequality with Borel-Cantelli Lemma as in the proof of Lemma 6.2. ☐

Now we consider the third term in (25).

Lemma 6.4. *Assume **(H5)**, then,*

$$\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}}\left|\bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)-\bar{p}^N_\theta(\mathbf{x})\right|\le 3C_5\frac{a_N}{\nu_N}.$$

Proof. From the mean value theorem and **(H5)**, we have

(26)

$$\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}}\left|\bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)-\bar{p}^N_\theta(\mathbf{x})\right|$$

$$=\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}}\left|\left(y-y^N_{l_1}\right)\int_0^1\partial_y\bar{p}^N_\theta\left(\varepsilon y+(1-\varepsilon)y^N_{l_1},z\right)d\varepsilon\right.$$

$$\left.+\left(z-z^N_{l_1}\right)\int_0^1\partial_z\bar{p}^N_\theta\left(y^N_{l_1},\varepsilon z+(1-\varepsilon)z^N_{l_1}\right)d\varepsilon+\left(\theta-\theta^N_{l_2}\right)\int_0^1\partial_\theta\bar{p}^N_{\varepsilon\theta+(1-\varepsilon)\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)d\varepsilon\right|$$

$$\le\frac{a_N}{\nu_N}\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}}\left\{\sup_{0\le\varepsilon\le1}\left|\partial_y\bar{p}^N_\theta\left(\varepsilon y+(1-\varepsilon)y^N_{l_1},z\right)\right|\right.$$

$$\left.+\sup_{0\le\varepsilon\le1}\left|\partial_z\bar{p}^N_\theta\left(y^N_{l_1},\varepsilon z+(1-\varepsilon)z^N_{l_1}\right)\right|+\sup_{0\le\varepsilon\le1}\left|\partial_\theta\bar{p}^N_{\varepsilon\theta+(1-\varepsilon)\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right|\right\}.$$

From here the result follows. ☐

Finally, we consider the second term of (25).

Lemma 6.5. *Assume **(H2)** and that η_N satisfies **(H6)**, then we have that*

$$P\left(\limsup_{N\to\infty}\left\{\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\left|\hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)-\bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right|>\eta_N\right\}\right)=0.$$

Proof. Set

$$W^{j,\mathbf{x}}_{m,h}(\theta;\omega):=\frac{1}{h}K\left(\frac{X^{y,(j)}_{(m)}(\theta;\omega)-z}{h}\right)-\frac{1}{h}E\left[K\left(\frac{X^{y,(1)}_{(m)}(\theta;\cdot)-z}{h}\right)\right].$$

Note that $\{W^{j,\mathbf{x}}_{m,h}(\theta;\omega)\}_{j\in\mathbb{N}}$ is a sequence of i.i.d. r.v. with $E\left[W^{j,\mathbf{x}}_{m,h}(\theta;\omega)\right]=0$. For all $\mathbf{x}\in\mathbb{R}^2,\ \theta\in\Theta,\ m\in\mathbb{N}$ and $h>0$,

(27) $$\sup_{j=1,\dots,n}\left|W^{j,\mathbf{x}}_{m,h}(\theta;\omega)\right|\le\frac{2}{h}\|K\|_\infty=:b_h.$$

Set

$$S^{n,\mathbf{x}}_{m,h}(\theta;\omega) := \sum_{j=1}^{n} W^{j,\mathbf{x}}_{m,h}(\theta;\omega).$$

If we use inequality (1.25) of pp.27 in Bosq [5] with $q = \frac{n}{2}$, then we have, for all $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{N=1}^{\infty} P\left(\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \left|\hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right) - \bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right| > \eta_N\right)\\
&\le \sum_{N=1}^{\infty}\sum_{l_2=1}^{\nu_N}\sum_{l_1=1}^{\nu_N^2} P\left(\left|\hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right) - \bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right| > \eta_N\right)\\
&= \sum_{N=1}^{\infty}\sum_{l_2=1}^{\nu_N}\sum_{l_1=1}^{\nu_N^2} P\left(\left|S^{n,\mathbf{x}}_{m,h}(\theta^N_{l_2};\omega)\right| > n\eta_N\right)\\
&\le \sum_{N=1}^{\infty}\sum_{l_2=1}^{\nu_N}\sum_{l_1=1}^{\nu_N^2} 4\exp\left(-\frac{(\eta_N)^2 nh^2}{64\|K\|_\infty}\right)\\
&= \sum_{N=1}^{\infty} \nu_N^3 4\exp\left(-\frac{(\eta_N)^2 nh^2}{64\|K\|_\infty}\right).
\end{aligned}$$

Finally, from the Borel-Cantelli lemma the result follows. □

Now we can conclude this section with the following upper bound for $C\frac{D}{B}$.

Theorem 6.6. *Assume conditions* ***(H2)****,* ***(H4)****,* ***(H5)****,* ***(H6)*** *and* ***(H7)****, then for any ω, there exists $N_0 \equiv N_0(\omega)$ such that for all $N \ge N_0$ we have that*

$$\sup_{(\mathbf{x},\theta)\in B^N}\left|\hat{p}^N_\theta(\mathbf{x}) - \bar{p}^N_\theta(\mathbf{x})\right| \le \frac{2\|K'\|_\infty}{h^2}\frac{a_N^2}{\nu_N}b_{4,N} + \eta_N + 3C_5\frac{a_N}{\nu_N}.$$

Therefore if we also assume ***(H1)*** *and* ***(H3)****, then we have*

$$C\frac{D}{B} \le b_{3,N} \times \frac{1}{\varphi_1}\exp\left(\frac{\varphi_2 a_N^2}{\Delta}\right) \times \left(\frac{2\|K'\|_\infty}{h^2}\frac{a_N^2}{\nu_N}b_{4,N} + \eta_N + 3C_5\frac{a_N}{\nu_N}\right).$$

6.4. **Upper bound for $\frac{A}{B}$ in (21).** The proof in this case is simpler on the one hand because many of the previous estimates can be used. On the other hand, when considering the analogous result of Lemma 6.5 for the derivatives of $\hat{p}_\theta$, the proof has to be reworked as the condition analogous to (27) can only be obtained with a random upper bound. Therefore, we only briefly sketch the results when the proofs are similar to previous ones. From the hypothesis **(H1)**, we have

$$\frac{A}{B} \le \frac{1}{\varphi_1}e^{\frac{\varphi_2 a_N^2}{\Delta}} \times \sup_{(\mathbf{x},\theta)\in B^N}\left|\partial_\theta\hat{p}^N_\theta(y,z) - \partial_\theta\bar{p}^N_\theta(y,z)\right|.$$

Here we consider the above sup-term as before. We use the same notations as the previous section.

$$
\begin{aligned}
&\sup_{(\mathbf{x},\theta)\in B^N} \left|\partial_\theta \hat{p}_\theta^N(\mathbf{x}) - \partial_\theta \bar{p}_\theta^N(\mathbf{x})\right| \\
&\le \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}} \left|\partial_\theta \hat{p}_\theta^N(\mathbf{x}) - \partial_\theta \hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right| + \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \left|\partial_\theta \hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right) - \partial_\theta \bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right| \\
&\quad + \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}} \left|\partial_\theta \bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right) - \partial_\theta \bar{p}_\theta^N(\mathbf{x})\right|.
\end{aligned} \tag{28}
$$

As in previous sections, if $(\mathbf{x},\theta)=(y,z,\theta)\in B^N_{l_1 l_2}$, then from **(H2)**, we have

$$
\begin{aligned}
&\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}} \left|\partial_\theta \hat{p}_\theta^N(\mathbf{x};\omega) - \partial_\theta \hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right| \\
&\le \frac{\|K'\|_\infty \vee \|K''\|_\infty}{h^3}\frac{a_N^2}{\nu_N} \max_{k=1,\dots,n}\left\{\dot{Z}^{(k)}_{4,N}(\omega)\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\dot{Z}^{(k)}_{4,N}(\omega) := a_N^{-1}\Bigg(& h \sup_{(\mathbf{x},\theta)\in B^N}\left|\partial_y\partial_\theta X^{y,(k)}_{(m)}(\theta;\omega)\right| + h \sup_{(\mathbf{x},\theta)\in B^N}\left|\partial_\theta\partial_\theta X^{y,(k)}_{(m)}(\theta;\omega)\right| \\
& + \left(Z^{(k)}_{4,N}+1\right)\sup_{(\mathbf{x},\theta)\in B^N}\left|\partial_\theta X^{y,(k)}_{(m)}(\theta;\omega)\right|\Bigg).
\end{aligned}
$$

Note that $\{\dot{Z}^{(k)}_{4,N}(\cdot)\}_{k\in\mathbb{N}}$ is a sequence of i.i.d. random variables. Then we set the following hypothesis and obtain the following lemma which is the parallel of Lemma 6.3.

(H4') Assume that for some constant $\dot{r}_4>0$ and a sequence $\{\dot{b}_{4,N}; N\in\mathbb{N}\}\subset[1,\infty)$, we have that $\sum_{N=1}^\infty \frac{nE\left[\left|\dot{Z}_{4,N}(\cdot)\right|^{\dot{r}_4}\right]}{(\dot{b}_{4,N})^{\dot{r}_4}}<\infty$.

(H7') Assume that $X^{y,(1)}_{(m)}(\theta)$ is twice differentiable with respect to θ a.s. and $\partial_\theta X^{y,(1)}_{(m)}(\theta)$ is differentiable with respect to y a.s.

Lemma 6.7. *Under **(H2)**, **(H4')**, **(H7)** and **(H7')**, we have that*

$$
\begin{aligned}
P\Bigg(\limsup_{N\to\infty}\Bigg\{ & \max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}} \left|\partial_\theta \hat{p}_\theta^N(\mathbf{x}) - \partial_\theta \hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right| \\
& \ge \frac{\|K'\|_\infty \vee \|K''\|_\infty}{h^3}\frac{a_N^2}{\nu_N}\dot{b}_{4,N}\Bigg\}\Bigg) = 0.
\end{aligned}
$$

Next we set the following hypothesis;

(H5'). Assume that there exists some positive constant $\dot{C}_5>0$ such that for all $y,z\in\mathbb{R}$, $m\in\mathbb{N}$ and $\theta\in\Theta$,

$$
\left|\partial_y\partial_\theta \bar{p}_\theta^N(y,z)\right|,\ \left|\partial_z\partial_\theta \bar{p}_\theta^N(y,z)\right|,\ \left|\partial^2_\theta \bar{p}_\theta^N(y,z)\right| \le \dot{C}_5 < +\infty.
$$

Lemma 6.8. *Assume **(H5')**. Then, we have*

$$
\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}} \left|\partial_\theta \bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right) - \partial_\theta \bar{p}_\theta^N(\mathbf{x})\right| \le 3\dot{C}_5\frac{a_N}{\nu_N}.
$$

Proof. Using a similar argument as in the proof of Lemma 6.4, we consider the third term of (28).

$$\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}} \left|\partial_\theta \bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)-\partial_\theta \bar{p}^N_\theta\left(\mathbf{x}\right)\right|$$
$$\le \frac{a_N}{\nu_N}\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B^N_{l_1 l_2}} \left\{\sup_{0\le\varepsilon\le 1}\left|\partial_y\partial_\theta \bar{p}^N_\theta\left(\varepsilon y+(1-\varepsilon)y^N_{l_1},z\right)\right|\right.$$
$$\left.+\sup_{0\le\varepsilon\le 1}\left|\partial_z\partial_\theta \bar{p}^N_\theta\left(y^N_{l_1},\varepsilon z+(1-\varepsilon)z^N_{l_1}\right)\right|+\sup_{0\le\varepsilon\le 1}\left|\partial^2_\theta \bar{p}^N_{\varepsilon\theta+(1-\varepsilon)\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right|\right\}.$$

□

Finally, we consider the second term of (28). Set

$$\dot{W}^{j,\mathbf{x}}_{m,h}\left(\theta;\omega\right):=\frac{1}{h^2}K'\left(\frac{X^{y,(j)}_{(m)}(\theta;\omega)-z}{h}\right)\partial_\theta X^{y,(j)}_{(m)}(\theta;\omega)$$
$$-\frac{1}{h^2}E\left[K'\left(\frac{X^{y,(1)}_{(m)}(\theta;\cdot)-z}{h}\right)\partial_\theta X^{y,(1)}_{(m)}(\theta;\cdot)\right].$$

Note that $\{\dot{W}^{j,\mathbf{x}}_{m,h}\left(\theta;\omega\right)\}_{j\in\mathbb{N}}$ is a sequence of i.i.d. r.v. with $E\left[\dot{W}^{j,\mathbf{x}}_{m,h}\left(\theta;\cdot\right)\right]=0$. To study this term, we assume:

(H6'). There exists $\dot{C}_6>0$ and $\dot{\alpha}_6>0$ and a sequence of positive numbers $\dot{b}_{6,N}$ such that $\sum_{N=1}^\infty \nu_N^3\exp\left(-\frac{n(\eta_N)^2h^4}{2\|K'\|^2_\infty(\dot{b}_{6,N})^2a_N^2}\right)\left\{1+\frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}}\right\}^n<\infty$.

(H6a'). Assume that there exists $\dot{r}_6>0$ and that the sequence of positive numbers $\dot{b}_{6,N}$ in **(H6')** satisfy $\sum_{N=1}^\infty n\frac{E\left[\left|\dot{Z}_{6,N}(\cdot)\right|^{\dot{r}_6}\right]}{(\dot{b}_{6,N})^{\dot{r}_6}}<\infty$ for

$$\dot{Z}^{(j)}_{6,N}(\omega):=a_N^{-1}\sup_{(\mathbf{x},\theta)\in B^N}\left\{\left|\partial_\theta X^{y,(j)}_{(m)}(\theta;\omega)\right|+E\left[\left|\partial_\theta X^{y,(1)}_{(m)}(\theta;\cdot)\right|\right]\right\}.$$

(H6b'). Assume that for some $\dot{q}_6>1$, $\sup_{N\in\mathbb{N}}E\left[\left|\dot{Z}_{6,N}(\cdot)\right|^{\dot{q}_6}\right]<+\infty$ and for $\dot{\alpha}_6>0$, $\dot{C}_6>0$ and $\dot{b}_{6,N}$ given in **(H6')** the following is satisfied

$$\left(\frac{\eta_N h^2}{(\|K'\|_\infty\dot{b}_{6,N})^2a_N}\exp\left(-\frac{(\eta_N)^2}{2(\frac{\|K'\|_\infty}{h^2}\dot{b}_{6,N}a_N)^2}\right)\right)^{\dot{q}_6}\le\frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}},$$
$$\sup_N\frac{\eta_N h^2}{\dot{b}_{6,N}a_N}<\infty.$$

The following result is analogous to Lemma 6.5. The proof requires a further use of Borel-Cantelli's lemma.

Lemma 6.9. *Assume* ***(H6')****,* ***(H6a')*** *and* ***(H6b')****. Then for a.s. ω, there exists $N\equiv N(\omega)$ big enough such that*

$$\max_{\substack{1\le l_1\le \nu_N^2\\ 1\le l_2\le \nu_N}}\left|\partial_\theta\hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)-\partial_\theta\bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right|\le\eta_N.$$

Proof. First, note that

$$\sup_{(\mathbf{x},\theta)\in B^N}\left|\dot{W}^{j,\mathbf{x}}_{m,h}(\theta;\omega)\right| \leq \frac{\|K'\|_\infty \dot{Z}^{(j)}_{6,N}(\omega)a_N}{h^2}.$$

Define

$$D^m_{a_N} := \left\{\max_{j=1,\dots,n}\dot{Z}^{(j)}_{6,N}(\omega) > \dot{b}_{6,N}\right\}.$$

Note that $\{\dot{Z}^{(j)}_{6,N}(\omega)\}_{j\in\mathbb{N}}$ is a sequence of i.i.d. random variables.

From Chebyshev's inequality, we have by **(H6a')**

$$\sum_{N=1}^\infty P\left(D^m_{a_N}\right) \leq \sum_{N=1}^\infty n\frac{E\left[\left|\dot{Z}^{(j)}_{6,N}(\cdot)\right|^{\dot{r}_6}\right]}{(\dot{b}_{6,N})^{\dot{r}_6}} < \infty.$$

Therefore, by Borel-Cantelli's lemma we have that

$$P\left(\limsup_{N\to\infty} D^m_{a_N}\right) = 0.$$

Now, set

$$\tilde{S}^{n,\mathbf{x}}_{m,h}(\theta;\omega) := n\left(\partial_\theta\hat{p}^N_\theta(\mathbf{x}) - \partial_\theta\bar{p}^N_\theta(\mathbf{x})\right) = \sum_{j=1}^n \dot{W}^{j,\mathbf{x}}_{m,h}(\theta;\omega).$$

Therefore using Lemma 7.2 in the Appendix with $X_1(\omega) := \dot{W}^{\mathbf{x}}_{m,h}(\theta;\omega)$, $\varepsilon := \eta_N$, $f_n := \|K'\|_\infty h^2\dot{b}_{6,N}a_N$ and $\dot{C}_{\dot{q}_6} = \frac{\|K'\|^{\dot{q}_6}_\infty \sup_{N\in\mathbb{N}} E[|\dot{Z}_{6,N}|^{q_1}]a_N^{\dot{q}_6}}{h^{2\dot{q}_6}}$ and the hypothesis **(H6b')**, we obtain

$$\begin{aligned}
&\sum_{N=1}^\infty P\Bigg(\max_{\substack{1\leq l_1\leq\nu_N^2\\1\leq l_2\leq\nu_N}}\left|\partial_\theta\hat{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right) - \partial_\theta\bar{p}^N_{\theta^N_{l_2}}\left(\mathbf{x}^N_{l_1}\right)\right| > \eta_N;\\
&\qquad\qquad \left|\dot{W}^{j,\mathbf{x}}_{m,h}(\theta;\omega)\right| \leq \frac{\|K'\|_\infty}{h^2}\dot{b}_{6,N}a_N,\ j=1,\cdots,n\Bigg)\\
&\leq \sum_{N=1}^\infty\sum_{l_2=1}^{\nu_N}\sum_{l_1=1}^{\nu_N^2} P\left(\left|\tilde{S}^{n,\mathbf{x}}_{m,h}(\theta)\right| > n\eta_N;\ \left|\dot{W}^{j,\mathbf{x}}_{m,h}(\theta)\right| \leq \frac{\|K'\|_\infty}{h^2}\dot{b}_{6,N}a_N,\ j=1,\cdots,n\right)\\
&\leq 2\sum_{N=1}^\infty \nu_N^3\exp\left(-\frac{n(\eta_N)^2h^4}{2\|K'\|^2_\infty(\dot{b}_{6,N})^2a_N^2}\right)\left\{1+\frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}}\right\}^n.
\end{aligned}$$

Finally, from hypothesis **(H6)** and Borel-Cantelli's lemma, the conclusion follows. □

Theorem 6.10. *Assume conditions **(H1)**, **(H2)**, **(H4')**, **(H5')**, **(H6')**, **(H6a')**, **(H6b')**, **(H7)** and **(H7')**. Then we have that for a.s. ω, there exists $N_0 \equiv N_0(\omega)$ such that for all $N \geq N_0$ we have*

$$\frac{A}{B} \leq \frac{1}{\varphi_1}e^{\frac{\varphi_2 a_N^2}{\Delta}} \times \left(\frac{\|K'\|_\infty \vee \|K''\|_\infty}{h^3}\frac{a_N^2}{\nu_N}\dot{b}_{4,N} + \eta_N + 3\dot{C}_5\frac{a_N}{\nu_N}\right).$$

Finally collecting all our results together, we have (see Theorem 6.6).

Theorem 6.11. *Assume conditions **(H0)**, **(H1)**, **(H2)**, **(H3)**, **(H4)**, **(H5)**, **(H6)**, **(H4')**, **(H5')**, **(H6')**, **(H6a')**, **(H6b')**, **(H7)** and **(H7')**. Then for a.s. ω, there exists*

$N_0 \equiv N_0(\hat{\omega})$ *such that for all* $N \geq N_0$ *we have*

$$N^{-1/2-\epsilon}\left(\frac{A}{B}+C\frac{D}{B}\right) \leq 6N^{-1/2-\epsilon}\frac{1}{\varphi_1}e^{\frac{\varphi_2 a_N^2}{\Delta}} \times\left(\frac{\|K'\|_\infty \vee \|K''\|_\infty}{h^2}\frac{a_N^2}{\nu_N}\left(\frac{\dot{b}_{4,N}}{h}+b_{4,N}b_{3,N}\right)+\left(\eta_N+\frac{C_5\vee\dot{C}_5}{\nu_N}a_N\right)b_{3,N}\right).$$

6.5. **The treatment of $J_N^2(\theta)$.** As dealing with $J_N^2(\theta)$ follows the same steps taken throughout this section with simplifications we will just remark here these and the hypotheses to be added.

Overall the argument will follow in a similar way but replacing instead of $\hat{p}_\theta^N(y,z)$, $p_\theta(y,z)$.

First note that although we may still use the decomposition (21), there is no explicit expression for $\partial_\theta p_\theta(y,z)$.

Now one adds to hypothesis **(H1).**

$$\inf_{(\mathbf{x},\theta)\in B^N} p_\theta(y,z) \geq \varphi_1 \exp\left(-\frac{\varphi_2 a_N^2}{\Delta}\right).$$

Similarly, to **(H5).** one adds that for all $y,z\in\mathbb{R},\ m\in\mathbb{N}$ and $\theta\in\Theta$,

$$|\partial_y p_\theta(y,z)|,\ |\partial_z p_\theta(y,z)|,\ |\partial_\theta p_\theta(x,y)| \leq C_5 < +\infty.$$

Assume that for each $y,z\in\mathbb{R}$, there exist a positive constant C and such that

$$\left|p_\theta(y,z)-\bar{p}_\theta^N(y,z)\right|+\left|\partial_\theta p_\theta(y,z)-\partial_\theta\bar{p}_\theta^N(y,z)\right| \leq Ca_1(N), \tag{29}$$

where $a_1(N)\to 0$ as $N\to\infty$.

The above assumption is usually obtained using Malliavin Calculus techniques (in a non-straightforward manner) as in Bally and Talay [1] or Guyon [14]. Usually the choice $m(N)=\sqrt{N}$ will satisfy the above assumption

Now instead of the Lemma 6.2, one has the trivial bound

$$\sup_{(\mathbf{x},\theta)\in B^N}\left|\frac{\partial_\theta p_\theta(y,z)}{\hat{p}_\theta(y,z)}\right| \leq C_5\varphi_1^{-1}\exp\left(\frac{\varphi_2 a_N^2}{\Delta}\right).$$

Using Assumption (29), one can obtain the analogous result of Theorem 6.6, which gives

$$C\frac{D}{B} \leq \frac{1}{\varphi_1^2}\exp\left(2\frac{\varphi_2 a_N^2}{\Delta}\right)a_1(N).$$

Similarly,

$$\frac{A}{B} \leq \frac{1}{\varphi_1}e^{\frac{\varphi_2 a_N^2}{\Delta}}a_1(N).$$

As we will see later we will choose $a_N := \sqrt{c_2\ln(N)}$. Therefore in order to assure (4), we need that

$$\sup_N N^{1/2-\epsilon}\exp\left(2\frac{\varphi_2 a_N^2}{\Delta}\right)a_1(N) < \infty. \tag{30}$$

This can be achieved by choosing the parameters $h(N)$ and $m(N)$ appropriately as proven in [1].

6.6. **Main Theorem: Tuning for n and h.** In this section, we rewrite the previous hypothesis **(H0)-(H6b')** in a simpler form, so that we can verify them easily in examples, such as the case of smooth diffusions.

We need to find now a sequence of values for n and h such that all the hypothesis in the previous Theorem are satisfied and that the upper bound is uniformly bounded in N. Now, we rewrite the needed conditions that are related to the parameters n and h. We assume stronger hypothesis that may help us understand better the existence of the right choice of parameters n and h.

As we are only interested in the relationship between n and h with N, we will denote by C_1, C_2 etc., various constants that may change from one equation to the next. These constants depend on K, Δ and Θ. They are independent of n, h and N but they depend continuously on other parameters. We will assume the existence of some sequences of strictly positive numbers which are bigger than 1.

(i). There exists some positive constant $C_{K,\Delta,\Theta} \geq 0$, which depends on K, Δ, Θ, and is independent of N such that

$$N^{1/2-\epsilon} e^{\frac{\varphi_2 a_N^2}{\Delta}} \times \left(\frac{a_N^2}{\nu_N h^2} \left(\frac{\dot{b}_{4,N}}{h} + b_{4,N} b_{3,N} \right) + \left(\eta_N + \frac{a_N}{\nu_N} \right) b_{3,N} \right) \leq C_{K,\Delta,\Theta}. \tag{31}$$

(ii). (Borel-Cantelli for Y_i, **(H0)**) Assume that $m_{c_1} := E[e^{c_1 |Y_1|^2}] < +\infty$ for some constant $c_1 > 0$ and $\{a_N\}_{N\in\mathbb{N}} \subset [\theta^u - \theta^l, \infty)$ is a sequence such that for the same c_1, $\sum_{N=1}^{\infty} \frac{N}{\exp(c_1 a_N^2)} < +\infty$.

(iii). (Borel-Cantelli for $Z_{3,N}^{(k)}(\omega)$, **(H3)**) For some $r_3 > 0$ and $b_{3,N} \geq 1$,

$$\sum_{N=1}^{\infty} \frac{n a_N^{2r_3}}{(h^2 b_{3,N})^{r_3}} < +\infty \text{ and } \sup_{N\in\mathbb{N}} E\left[|Z_{3,N}(\cdot)|^{r_3}\right] < +\infty.$$

(iv). (Borel-Cantelli for $Z_{4,N}^{(k)}(\omega)$, **(H4)**) For some $r_4 > 0$ and $b_{4,N} \geq 1$,

$$\sum_{N=1}^{\infty} \frac{n}{(b_{4,N})^{r_4}} < +\infty \text{ and } \sup_{N\in\mathbb{N}} E\left[|Z_{4,N}(\cdot)|^{r_4}\right] < +\infty.$$

(v). (Borel-Cantelli for $|\hat{p}_\theta^N(\mathbf{x}) - \bar{p}_\theta^N(\mathbf{x})|$, **(H6)**)

$$\sum_{N=1}^{\infty} \nu_N^3 \exp\left(-\frac{(\eta_N)^2 n h^2}{64\|K\|_\infty} \right) < +\infty.$$

(vi). (Borel-Cantelli for $\dot{Z}_{4,N}^{(k)}(\omega)$, **(H4')**) For some $\dot{r}_4 > 0$ and $\dot{b}_{4,N} \geq 1$,

$$\sum_{N=1}^{\infty} \frac{n}{(\dot{b}_{4,N})^{\dot{r}_4}} < +\infty \text{ and } \sup_{N\in\mathbb{N}} E\left[\left|\dot{Z}_{4,N}(\cdot)\right|^{\dot{r}_4}\right] < +\infty.$$

(vii). (Borel-Cantelli for $|\partial_\theta \hat{p}_\theta^N(\mathbf{x}) - \partial_\theta \bar{p}_\theta^N(\mathbf{x})|$, **(H6')**) For some $\dot{\alpha}_6 > 0$, and a constant $\dot{C}_6$,

$$\sum_{N=1}^{\infty} \nu_N^3 \exp\left(-\frac{n(\eta_N)^2 h^4}{2\|K'\|_\infty^2 (\dot{b}_{6,N})^2 a_N^2} \right) \left\{ 1 + \frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}} \right\}^n < +\infty.$$

(viii). (Borel-Cantelli for $\dot{Z}_{6,N}^{(k)}(\omega)$, **(H6a')**) For some $\dot{r}_6 > 0$ and $\dot{b}_{6,N} \geq 1$,

$$\sup_N \frac{\eta_N h^2}{\dot{b}_{6,N} a_N} < \infty \text{ and } \sum_{N=1}^{\infty} \frac{n}{(\dot{b}_{6,N})^{\dot{r}_6}} < +\infty \text{ and } \sup_N E\left[\left|\dot{Z}_{6,N}(\cdot)\right|^{\dot{r}_6}\right] < +\infty.$$

(ix). (**(H6b')**) For some $\dot{q}_6 > 1$,

$$\left(\frac{\eta_N h^2}{(\|K'\|_\infty (\dot{b}_{6,N})^2 a_N} \exp\left(-\frac{(\eta_N)^2}{2(\frac{\|K'\|_\infty}{h^2} \dot{b}_{6,N} a_N)^2} \right) \right)^{\dot{q}_6} \leq \frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}}$$

$$\text{and } \sup_{N \in \mathbb{N}} E\left[\left| \dot{Z}_{6,N}(\cdot) \right|^{\dot{q}_6} \right] < +\infty$$

where $\dot{C}_6$ and $\dot{\alpha}_6$ are the same as **(vii)** above.

6.6.1. *Parameter Tuning.* Choose $a_N := \sqrt{c_2 \ln N}$ for some positive constant c_2, $n = C_1 N^{\alpha_1}$ for $\alpha_1, C_1 > 0$, and $h = C_2 N^{-\alpha_2}$ for $\alpha_2, C_2 > 0$.

For **(ii)** to be satisfied, we need to have

$$\sum_{N=1}^{\infty} \frac{N}{\exp(c_1 c_2 \ln N)} = \sum_{N=1}^{\infty} \frac{1}{N^{c_1 c_2 - 1}}.$$

Then we need $c_1 c_2 - 1 > 1 \Leftrightarrow c_1 > \frac{2}{c_2}$. Note that if we choose c_2 as large enough, we can choose c_1 as small enough.

Next we substitute $a_N = \sqrt{c_2 \ln N}$ into (31). We have

$$N^{\frac{\varphi_2 c_2}{\Delta} + \frac{1}{2} - \epsilon} \left(\frac{c_2 \ln N}{h^2 \nu_N} \left(\frac{\dot{b}_{4,N}}{h} + b_{4,N} b_{3,N} \right) + b_{3,N} \left(\eta_N + \frac{\sqrt{c_2 \ln N}}{\nu_N} \right) \right) \leq C_{K,\Delta,\Theta}. \tag{32}$$

For **(iii)**, we assume that there exists some $\gamma_3 > 1$, $r_3 > 0$ and some constant $C_3 \neq 0$ such that

$$\frac{n(c_2 \ln N)^{r_3}}{(h^2 b_{3,N})^{r_3}} = \frac{C_3}{N^{\gamma_3}} \text{ and therefore } b_{3,N} = \frac{C_3 (N^{\gamma_3} n)^{\frac{1}{r_3}} c_2 \ln N}{h^2}.$$

For **(iv)**, we assume that there exists some $\gamma_4 > 1$, $r_4 > 0$ and some constant $C_4 \neq 0$ such that

$$\frac{n}{(b_{4,N})^{r_4}} = \frac{C_4}{N^{\gamma_4}} \text{ and therefore } b_{4,N} = (C_4 n N^{\gamma_4})^{\frac{1}{r_4}}.$$

For **(vi)**, we assume that there exists some $\dot{\gamma}_4 > 1$, $\dot{r}_4 > 0$ and some constant $\dot{C}_4 \neq 0$ such that

$$\frac{n}{(\dot{b}_{4,N})^{\dot{r}_4}} = \frac{\dot{C}_4}{N^{\dot{\gamma}_4}} \text{ and therefore } \dot{b}_{4,N} = \left(\dot{C}_4 n N^{\dot{\gamma}_4} \right)^{\frac{1}{\dot{r}_4}}.$$

For **(viii)**, we assume that there exists some $\dot{\gamma}_6 > 1$, $\dot{r}_6 > 0$ and some constant $\dot{C}_6 \neq 0$ such that

$$\frac{n}{(\dot{b}_{6,N})^{\dot{r}_6}} = \frac{\dot{C}_6}{N^{\dot{\gamma}_6}} \text{ and therefore } \dot{b}_{6,N} = \left(\dot{C}_6 n N^{\dot{\gamma}_6} \right)^{\frac{1}{\dot{r}_6}}.$$

Furthermore note that as $\dot{b}_{6,N} \wedge a_N \to \infty$ as $N \to \infty$, then condition **(vii)** implies condition **(v)**. Note that all $\min(b_{3,N}, b_{4,N}, \dot{b}_{4,N}, \dot{b}_{6,N}) \uparrow \infty$ as $N \uparrow \infty$.

Taking (32) into consideration, we set ν_N and η_N as follows; (here we assume that N is large enough)

- $\nu_N := C_{K,\Delta,\Theta} N^{\frac{\varphi_2 c_2}{\Delta} + \frac{1}{2} - \epsilon} \frac{c_2 \ln N}{h^3} \max\left\{ \left(n N^{\dot{\gamma}_4} \right)^{\frac{1}{\dot{r}_4}}, \frac{(N^{\gamma_3} n)^{\frac{1}{r_3}} c_2 \ln N}{h} (n N^{\gamma_4})^{\frac{1}{r_4}} \right\} \to \infty,$
- $\eta_N := \dfrac{C_{K,\Delta,\Theta} h^2}{N^{\frac{\varphi_2 c_2}{\Delta} + \frac{1}{2} - \epsilon} (N^{\gamma_3} n)^{\frac{1}{r_3}} c_2 \ln N} \to 0.$

Note in particular that this choice ensures that (32) is satisfied. And therefore **(i)** also follows. To finish, we only need to check conditions **(vii)** and **(ix).** Instead of separating the study of the above parameters in cases, we prefer to use the following notation $\nu_N = C_\nu N^{\delta_1} (\ln N)^{\delta_2} h^{-\delta_3}$, $\eta_N = C_{b^6} h^{\delta_4} N^{-\delta_5} n^{-\delta_6} (\ln N)^{-\delta_7}$ for some positive constants C_ν, C_{b^6}, δ_1,...,δ_7. From **(vii)**, we need, for some $\lambda_6 > 1$ and some constant $\ddot{C}_6 \neq 0$,

$$\nu_N^3 \exp\left(-\frac{n(\eta_N)^2 h^4}{2\|K'\|_\infty^2 (\dot{b}_{6,N})^2 a_N^2}\right)\left\{1 + \frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}}\right\}^n \leq \frac{\ddot{C}_6}{N^{\lambda_6}}. \tag{33}$$

As we are only proving the existence of appropriate parameters n and h so that the conditions **(i)**-**(ix)** are satisfied, we are going to ignore certain constants putting them together under the notation C. Replacing all the above values found so far we have that the inequality (33) is equivalent to

$$\begin{aligned} &N^{\alpha_1(1-2\delta_6)} N^{-4\alpha_2} N^{-2\alpha_2\delta_4} N^{-2\delta_5} \left(N^{\alpha_1} N^{\dot{\gamma}_6}\right)^{-\frac{2}{\dot{r}_6}} (\ln N)^{-2\delta_7 - 1} \\ &\geq C \ln\left(N^{\lambda_6} C_\nu N^{3\delta_1} (\ln N)^{3\delta_2} h^{-3\delta_3} \left\{1 + \frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}}\right\}^n\right). \end{aligned}$$

Therefore the condition needed here is that

$$\alpha_1(1 - 2\delta_6) - 4\alpha_2 - 2\alpha_2\delta_4 - 2\delta_5 - 2\frac{\alpha_1 + \dot{\gamma}_6}{\dot{r}_6} > 0. \tag{34}$$

Finally, we consider **(ix)**;

$$\left(\frac{C_2^2 N^{-2\alpha_2 - 2\frac{\alpha_1+\dot{\gamma}_6}{\dot{r}_6}} \eta_N}{(\|K'\|_\infty \left(\dot{C}_6 C_1\right)^{\frac{1}{\dot{r}_6}})^2 \sqrt{c_2 \ln N}} \exp\left(-\frac{C_2^4 N^{-4\alpha_2 - 2\frac{\alpha_1+\dot{\gamma}_6}{\dot{r}_6}} (\eta_N)^2}{2(\|K'\|_\infty \left(\dot{C}_6 C_1\right)^{\frac{1}{\dot{r}_6}})^2 c_2 \ln N}\right)\right)^{\dot{q}_6} \leq \frac{\dot{C}_6}{N^{\alpha_1(1+\dot{\alpha}_6)}}.$$

As $\eta_N \to 0$ as $N \to \infty$, it is enough to have

$$\left(4\alpha_2 + 2\frac{\alpha_1 + \dot{\gamma}_6}{\dot{r}_6} + \frac{\varphi_2 c_2}{\Delta} + \frac{1}{2} - \epsilon + \frac{\gamma_3}{r_3} + \frac{\alpha_1}{r_3}\right)\dot{q}_6 > \alpha_1, \tag{35}$$

which has to be satisfied together with (34) which we recall is

$$\alpha_1\left(1 - \frac{2}{r_3} - \frac{2}{\dot{r}_6}\right) > 8\alpha_2 + 1 - 2\epsilon + \frac{2\varphi_2 c_2}{\Delta} + \frac{2\gamma_3}{r_3} + 2\frac{\dot{\gamma}_6}{\dot{r}_6}. \tag{36}$$

Notice that the above two inequalities will always be satisfied if $\dot{q}_6$ is chosen big enough. Furthermore the only condition needed of η_N for the all the above conditions to hold is that $\eta_N \to 0$ as $N \to \infty$. Putting all the above calculations together, we obtain the following result.

Theorem 6.12. *Assume that the constants are chosen so as to satisfy* $c_1 > \frac{2}{c_2}$, (30), *(35) and (36). Assume that* **(H7)** *and* **(H7')**. *And also assume that the moment conditions stated in **(ii)**, **(iii)**, **(iv)**, **(vi)**, **(viii)** and **(ix)** above are satisfied. Then **(H0)**, **(H3)**, **(H4)**, **(H4')**, **(H6)**, **(H6')**, **(H6a')** and **(H6b')** are satisfied. Furthermore, if we assume **(H1)**, **(H2)**, **(H5)**, **(H5')**, then Assumption 2.2 (6) is satisfied.*

Furthermore if all other conditions on Assumption 2.2 are satisfied, then there exist some positive finite random variables Ξ_1 *and* Ξ_2 *such that*

$$|E_N[f] - f(\theta_0)| \leq \frac{\Xi_1}{N^{1/2-\epsilon}} \text{ a.s., and } \left|E_{N,m}^n[f] - f(\theta_0)\right| \leq \frac{\Xi_2}{N^{1/2-\epsilon}} \text{ a.s.,}$$

and

$$\left|E_N[f] - E^n_{N,m}[f]\right| \le \frac{\Xi_1 + \Xi_2}{N^{1/2-\epsilon}} \ a.s.$$

In fact, we remark that we are able to simplify the inequalities (35) and (36) to the above $\alpha_1 > 8\alpha_2 + 1 + \frac{2\varphi_2 c_2}{\Delta}$ if one can freely choose the constants $r_3,\ r_4,\ \dot{r}_4,\ \dot{r}_6$ and $\dot{q}_6$ due to the existence of all moments associated with the processes in the hypotheses **(iii)**, **(iv)**, **(vi)** **(viii)** and **(ix)**. Remember that φ_2 is the constant which was introduced in the lower bound of $\bar{p}^N_\theta$ in assumption **(H1)** and c_1 is the constant related to the integrability condition in Hypothesis **(H0)**. Hence from the assumptions $c_1 c_2 > 2$ and $\alpha_1 > 8\alpha_2 + 1 + \frac{2\varphi_2 c_2}{\Delta}$, we can find that c_1 and φ_2 are connected through the parameter c_2.

7. Appendix

7.1. Refinements of Markov's inequalities.

In this section we state a refinement of Markov's inequality that is applied in this article. For $\lambda > 0$, let $S_n := \sum_{i=1}^n X_i$ where X_i is a sequence of i.i.d. r.v.'s with $E[X_i] = 0$.

Lemma 7.1. *Let X be a random variable with $E[X] = 0$. Then, for $\lambda \in \mathbb{R}$, $c > 0$ and $p := P(|X| < c)$, we have*

$$E\left[e^{\lambda X}\mathbf{1}\left(|X| < c\right)\right] \le -\frac{e^{\lambda c} - e^{-\lambda c}}{2c} E[X\mathbf{1}(|X| \ge c)] + p e^{\frac{\lambda^2 c^2}{2}}.$$

Proof. From the convexity of the exponential function, we have, for $a \le x \le b$,

$$e^{\lambda x} \le \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}.$$

Now we let $a = -c$ and $b = c$ for some positive constant c. Then

$$\begin{aligned} E\left[e^{\lambda X}\mathbf{1}\left(|X| < c\right)\right] &\le \frac{cp - E[X\mathbf{1}(|X| < c)]}{2c} e^{-\lambda c} + \frac{E[X\mathbf{1}(|X| < c)] + cp}{2c} e^{\lambda c} \\ &= \frac{e^{\lambda c} - e^{-\lambda c}}{2c} E[X\mathbf{1}(|X| < c)] + p\frac{e^{\lambda c} + e^{-\lambda c}}{2}. \end{aligned}$$

The conclusion follows using that $E[X] = 0$ and analyzing the function $\ln(e^x + e^{-x}) - \ln 2$, which gives

$$\frac{e^{\lambda c} + e^{-\lambda c}}{2} \le e^{\frac{\lambda^2 c^2}{2}}.$$

□

Lemma 7.2. *Let $q_1^{-1} + q_2^{-1} = 1$ and assume that $E[|X_i|^{q_1}] < \bar{C}_{q_1}$, then for all $0 < \varepsilon < 1$ and $\{f_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ satisfying that $\varepsilon f_n^{-1} \le K$ we have*

$$\begin{aligned} &P\left(|S_n| > n\varepsilon; |X_i| < f_n,\ i = 1, \cdots, n\right) \\ &\le 2e^{-\frac{n\varepsilon^2}{2f_n^2}} \left\{1 + (q_2 - 1)\left(q_2^{-1} K_1 \frac{\varepsilon}{f_n^2} \bar{C}_{q_1}^{q_1^{-1}} e^{-\frac{\varepsilon^2}{2f_n^2}}\right)^{q_1}\right\}^n. \end{aligned} \tag{37}$$

Here $K_1 = \max\left\{1, \frac{e^K - e^{-K}}{2K}\right\}$.

Proof. By Markov's inequality, we have that for $\lambda > 0$,

$$A := P\left(S_n > t; |X_i| < f_n,\ i = 1, \cdots, n\right) \le e^{-\lambda t} \prod_{i=1}^n E\left[e^{\lambda X_i}\mathbf{1}\left(|X_i| < f_n\right)\right]. \tag{38}$$

From Lemma 7.1, we have

$$A \le e^{-\lambda t}\left\{-\frac{e^{\lambda f_n} - e^{-\lambda f_n}}{2f_n} E[X_i\mathbf{1}(|X_i| \ge f_n)] + p_n e^{\frac{\lambda^2 f_n^2}{2}}\right\}^n,$$

where $p_n := P(\omega \in \Omega;\ |X_1(\omega)| < f_n)$. Here we take $\lambda = \frac{\varepsilon}{f_n^2}$ and $t = n\varepsilon$; to obtain

$$A \le e^{-\frac{n\varepsilon^2}{f_n^2}} \left\{ -\frac{e^{\frac{\varepsilon}{f_n}} - e^{-\frac{\varepsilon}{f_n}}}{2f_n} E[X_i \mathbf{1}(|X_i| \ge f_n)] + p_n e^{\frac{\varepsilon^2}{2f_n^2}} \right\}^n$$

Furthermore if we assume that $\varepsilon f_n^{-1} \le K$ then

$$\frac{e^{\frac{\varepsilon}{f_n}} - e^{-\frac{\varepsilon}{f_n}}}{2\varepsilon f_n^{-1}} \le \frac{e^K - e^{-K}}{2K} =: K_1.$$

Therefore by Cauchy-Schwarz's inequality with $q_1^{-1} + q_2^{-1} = 1$, we have

$$\left| -\frac{e^{\frac{\varepsilon}{f_n}} - e^{-\frac{\varepsilon}{f_n}}}{2f_n} E[X_i \mathbf{1}(|X_i| \ge f_n)] + p_n e^{\frac{\varepsilon^2}{2f_n^2}} \right| \le K_1 \frac{\varepsilon}{f_n^2} \bar{C}_{q_1}^{q_1^{-1}} (1-p_n)^{q_2^{-1}} + p_n e^{\frac{\varepsilon^2}{2f_n^2}}.$$

Next, we consider the function $g(x) = \beta(1-x)^{q_2^{-1}} + \alpha x$ for $x \in [0,1]$. This function has its absolute maximum at $x^* = 1 - \left(\frac{\beta}{q_2\alpha}\right)^{q_1}$ and its maximum value is given by $\max_{x\in[0,1]} g(x) = \alpha \left(1 + (q_2 - 1)\left(\frac{\beta}{q_2\alpha}\right)^{q_1}\right)$. Therefore if applied to the above inequality, we have (37).

For $P(S_n < -t;\ |X_i| < f_n,\ i = 1, \cdots, n)$, we can apply the same argument, and from

$$\begin{aligned} &P\left(|S_n| < t;\ |X_i| < f_n,\ i = 1, \cdots, n\right) \\ &= P\left(S_n > t;\ |X_i| < f_n,\ i = 1, \cdots, n\right) + P\left(S_n < -t;\ |X_i| < f_n,\ i = 1, \cdots, n\right), \end{aligned}$$

we can obtain our conclusion. □

7.2. **An application of Komatsu inequality.**

Lemma 7.3. *Let c be a positive constant and M be a positive random variable.*

(i). *For fixed ω and all $a \ge \frac{M(\omega)}{2c}$, we have*

$$\frac{e^{-c(a^2 - \frac{M}{c}a)}}{c(a - \frac{M}{2c}) + \sqrt{c^2(a - \frac{M}{2c})^2 + 2c}} \le \int_a^\infty e^{-cx^2 + Mx} dx \le \frac{e^{-c(a^2 - \frac{M}{c}a)}}{c(a - \frac{M}{2c}) + \sqrt{c^2(a - \frac{M}{2c})^2 + c}}.$$

Proof. Set $y = \sqrt{2c}(x - \frac{M}{2c})$. We can rewrite the middle term as follow:

$$\int_a^\infty e^{-cx^2 + Mx} dx = e^{\frac{M^2}{4c}} \int_a^\infty e^{-c(x - \frac{M}{2c})^2} dx = \frac{e^{\frac{M^2}{4c}}}{\sqrt{2c}} \int_{\sqrt{2c}(a - \frac{M}{2c})}^\infty e^{-\frac{y^2}{2}} dy.$$

From Komatsu's inequality in p.17 of Itô and McKean [16], we have

$$\begin{aligned} &\frac{e^{\frac{M^2}{4c}}}{\sqrt{2c}} \frac{2e^{-c(a - \frac{M}{2c})^2}}{\sqrt{2c}(a - \frac{M}{2c}) + \sqrt{2c(a - \frac{M}{2c})^2 + 4}} \\ &\le \int_a^\infty e^{-cx^2 + Mx} dx \le \frac{e^{\frac{M^2}{4c}}}{\sqrt{2c}} \frac{2e^{-c(a - \frac{M}{2c})^2}}{\sqrt{2c}(a - \frac{M}{2c}) + \sqrt{2c(a - \frac{M}{2c})^2 + 2}}. \end{aligned}$$

Now we have obtained the inequality. □

8. Acknowledgements

The first author would like to thank Alex Kulik to point out a mistake in an earlier version of this article. As a consequence conditions that were quoted in the application of these results which appear in [17] and [18] have to be changed accordingly.

References

1. V. Bally, D. Talay, *The law of the Euler scheme for stochastic differential equations II. Convergence rate of the density*, Monte Carlo Methods Appl. **2** (1996), no. 2, 93–128.
2. A. Beskos, O. Papaspiliopoulos, G.O. Roberts, P. Fearnhead, *Exact and Computationally Efficient Likelihood-based Estimation for Discretely Observed Diffusion Processes*, J. R. Statist. Soc. B **68** (2006), Part 3, 333–382.
3. P. Billingsley, *Probability and measure*, John Wiley & Sons, 1979.
4. P. Billingsley, *Convergence of Probability Measures (Second Edition)*, John Wiley & Sons, 1999.
5. D. Bosq, *Non-parametric Statistics for Stochastic Processes. Estimation and Prediction (second edition)*, Springer-Verlag, 1998.
6. J.A. Cano, M. Kessler, D. Salmeron, *Approximation of the posterior density for diffusion processes*, Statist. Probab. Lett. **76** (2006), no. 1, 39–44.
7. G. Da Prato, *An Introduction to Infinite-Dimensional Analysis*, Springer, 2006.
8. G. Da Prato, J.Zabczyk, *Ergodicity for Infinite Dimensional Systems*, London Mathematical Society. Lecture Notes Series, 1996.
9. H. Dehling, *Limit theorems for sums of weakly dependent Banach space valued random variables*, Z. Wahrsch. Verw. Gebiete **63** (1983), 393–432.
10. P. Del Moral, J. Jacod, P. Protter, *The Monte-Carlo method for filtering with discrete-time observations*, Probab. Theory and Related Fields **120** (2001), 346–368.
11. P.P.B. Eggermont, V.N. LaRiccia, *Maximum Penalized Likelihood Estimation; Volume I: Density Estimation*, Springer-Verlag, New York, 2001.
12. S.N. Ethier, T.G. Kurtz, *Markov processes; Characterization and convergence*, Wiley Series in Probability and Mathematical Statistics, Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986.
13. S. Gál and L. Gál, *The discrepancy of the sequence* $\{(2^n x)\}$, Koninklijke Nederlandse Akademie van Wetenschappen Proceedings, Seres A, Indagationes Mathematicae **67** (1964), 129–143.
14. J. Guyon, *Euler scheme and tempered distributions*, Stochastic Process. Appl. **116** (2006), no. 6, 877–904.
15. I.A. Ibragimov, Yu.V. Linnik. *Independent and Stationary Sequences of Random Variables* Wolters-Noordhoff Publishing Groningen, 1971.
16. S. Itô, H.P. McKean, *Diffusion processes and their Sample Paths*, Springer-Verlag, 1974.
17. A. Kohatsu-Higa, N. Vayatis, K. Yasuda, *Strong consistency of Bayesian estimator for the Ornstein-Uhlenbeck process.* Stochastic Analysis with Financial Applications, Progress in Probability **65** (2011), 145–167.
18. A. Kohatsu-Higa, K. Yasuda, *An Ornstein-Uhlenbeck type process which satisfies sufficient conditions for a simulation based filtering procedure*, Malliavin Calculus and Stochastic Analysis, 173–193, Springer Proceedings in Mathematics & Statistics book series (PROMS, volume 34), 2013.
19. Kuelbs, J. and Philipp, Walter. Almost sure invariance principles for partial sums of mixing B-valued random variables. Ann. Probab. **8** (1980), 1003–1036.
20. D.W. Stroock, S.R.S. Varadhan, *Multidimensional diffusion processes*, Springer-Verlag, 1979.
21. D. Talay, *Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure and discretization by the implicit Euler scheme*, Markov Processes and Related Fields **8** (2002), 163–198.
22. R. Yokoyama, *Moment Bounds for Starionary Mixing Sequences*, Z. Wahrsch. Verw. Gebiete **52** (1980), 45–57.

Ritsumeikan University

ENS de Cachan

Hosei University
E-mail address: k_yasuda@hosei.ac.jp