# Strong consistency of Bayesian estimator under discrete observations and unknown transition density *

A. Kohatsu-Higa †, N. Vayatis ‡, Kazuhiro Yasuda §

September 18, 2010

**Abstract**

We consider the asymptotic behavior of a Bayesian parameter estimation method under discrete stationary observations. We suppose that the transition density of the data is unknown, and therefore we approximate it using a kernel density estimation method applied to the Monte Carlo simulations of approximations of the theoretical random variables generating the observations. In this article, 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 number of time-steps used in the approximation process, the number of the Monte-Carlo simulations and the bandwidth size of the kernel density estimation.

## 1 Introduction

We consider a parameter estimation method of Bayesian type under discrete observations. That is, our goal is to estimate the posterior expectation of some function $f$ given the observed data $Y_0^N = (Y_0, Y_1, ..., Y_N)$;

$$E_N[f] := E_\theta[f|Y_0, ..., Y_N] := \frac{\int f(\theta)\phi_\theta(Y_0^N)\pi(\theta)d\theta}{\int \phi_\theta(Y_0^N)\pi(\theta)d\theta}, \tag{1.1}$$

where $\phi_\theta(Y_0^N)$ is the joint-density of the model process for $Y_0^N = (Y_0, Y_1, ..., Y_N)$ and $\pi(\theta)$ is a prior distribution. This method can be applied when the joint-density $\phi_\theta(Y_0^N)$ is known a priori. In this article, we suppose that this is not the case.

Suppose that $Y_i$ is a stationary Markov chain therefore $\phi_\theta(y_0^N) = \mu_\theta(y_0)\prod_{i=0}^{N-1} p_\theta(y_i, y_{i+1})$, where $\mu_\theta$ is a probability density function which is the invariant measure of $Y_i$ and $p_\theta(y,z)$ is the transition density from $y$ to $z$. But this expression (1.1) is still theoretical, as we do not know the transition density $p_\theta$. So we propose to estimate this quantity based on the simulation of the underlying process. Usually the underlying process is also approximated using, for example, the Euler-Maruyama scheme in the case that $Y$ is generated by a diffusion.

*The full paper [8] is still in preparation. If interested, you can request a copy by sending e-mail to k_yasuda@hosei.ac.jp.

†Osaka University & CREST.

‡ENS de Cachan.

§Hosei University.

Then we approximate the transition density of the Euler-Maruyama approximation through the kernel density estimation method. Under these settings, we consider an approximation of the posterior expectation (1.1):

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

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 $\hat{p}^N_\theta(y,z)$ is an approximation of the transition density obtained using the kernel density estimation method. The problem we want to address is that there exists an appropriate choice of all the parameters that appear in the approximation of the posterior expectation so that convergence is obtained. In particular, the issue that is new in this research in comparison with previous results is that we give an estimate of how good the approximation of the transition density has to be as the amount of observations increase to infinity (therefore the number of arguments in $q_\theta$ tend to infinity). In reality 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 [10] and the references therein.

Approximating posterior distribution can also be interpreted as a type of filtering problem for a diffusion process. This approach is considered in a general framework by Del Moral, Jacod, Protter [6]. Cano et al. [5] considered this problem based on Bayesian inference for a stochastic differential equation with parameter $\theta$ under similar settings to ours. They use direct calculations through an error estimation between the transition density of the true stochastic differential equation and the transition density of the Euler-Maruyama approximation given by Bally and Talay [1]. They prove that for fixed $N$ the approximate posterior distribution which uses the exact density of the Euler approximation converges to the exact posterior distribution as the number of steps increase.

Yasuda [12] improved this result and gave a general framework where the rate of convergence is $N^{-1/2}$ under a condition of approximation for the simulated density and the density of the approximation. This condition (**(6)-(b)** in this article) hides the tuning relationship between $m$ and $N$. The result is stated in general terms so that it can be adapted to various diffusion cases.

In this article, we consider the problem of approximation the transition densities using the kernel density estimation method and therefore we clarify the tuning process that is required for the convergence of approximation of posterior expectations. Therefore, we give an explicit expression that shows how to choose parameters (number of Monte-Carlo simulation $n$, bandwidth size of the kernel density estimation $h$ and number of approximation parameter of the process $m$) based on the number of observations $N$.

In this paper, we assume that observation data $Y_0, Y_1, ..., Y_N$ comes from a stationary and $\alpha$-mixing process and the time interval between $Y_i$ and $Y_{i+1}$ is fixed for all $i$. We give the relationship between the number of data and the parameters of the transition density approximation and we estimate the error as follows;

$$\left|E_N[f] - \hat{E}^n_{N,m}[f]\right| \le \frac{\Xi}{\sqrt{N}} \quad \text{a.s. } \omega, \hat{\omega},$$

where $\Xi$ is some positive random variable with $\Xi < +\infty$ a.s. $\omega, \hat{\omega}$ (here $\omega$ denotes the randomness associated with the data and $\hat{\omega}$ denotes the randomness associated with the simulations). And also we prove that $E_N[f]$ converges to $f(\theta_0)$ with order $\frac{1}{\sqrt{N}}$, where $\theta_0$ is the true value.

This paper is structured as follows. In Section 2, we will give the setting of our problem and precisely state our main theorem. In Section 3, we will prove our main theorem stated in Section 2 by using Laplace method dividing the proof in four estimations expressed in Proposition 3.1. This decomposition plays a central role in the proof. In Section 4, we will show how to deal with each term in the decomposition. In Section 5, we will give the tuning result which states the relationship between the number of Monte-Carlo simulation $n$ and the bandwidth size $h$ with related to $N$ so that a certain rate of convergence is achieved.

The proofs of various statements are involved and therefore we have tried to give the main line of thought in the proofs in this article. One can find proofs of Theorems, Propositions and Lemmas of this paper in Kohatsu-Higa et al. [8] and Yasuda [12]. And also one can find an example, the Ornstein-Uhlenbeck process case, of this paper in Kohatsu-Higa et al. [9].

# 2 Settings and Main Theorem

## 2.1 Settings

First we recall the definition of $\alpha$-mixing process;

**Definition 2.1** *(Billingsley [2], p.315) For a sequence $X_1, X_2, ...$ of random variables, let $\alpha_n$ be a number such that*

$$|P(A \cap B) - P(A)P(B)| \leq \alpha_n$$

*for $A \in \sigma(X_1, ..., X_k)$, $B \in \sigma(X_{k+n}, X_{k+n+1}, ...)$ and $k, n \in \mathbb{N}$. Suppose that $\alpha_n \to 0$ as $n \to \infty$, then we call the sequence $X_1, X_2, ...$, $\alpha$-mixing process.*

For convenience, we set $\alpha_0 = 1$.

Here we consider the following setting: Let $\theta_0 \in \Theta := [\theta^l, \theta^u]$, $(\theta^l < \theta^u)$ be a parameter that we want to estimate, where $\Theta$ is a compact subset in $\mathbb{R}$ and $\theta_0 \in \dot{\Theta}$, where $\dot{\Theta}$ denotes the interior of the set $\Theta$ and $\Theta_0 = \Theta - \{\theta_0\}$. Let $(\Omega, \mathcal{F}, P_{\theta_0})$, $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ and $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ be three probability spaces, where the probability measure $P_{\theta_0}$ is parametrized by $\theta_0$. $\Delta > 0$ is a fixed parameter that represents the time between observations. The probability space $(\Omega, \mathcal{F}, P_{\theta_0})$ is used for the observations, $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ is used for the process that defines the process with law $P_\theta$ and the space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ is used for the simulations that are used in estimating the transition density.

**(i).** **(Observation process)** Let $\{Y_{i\Delta}\}_{i=0,1,...,N}$ be a sequence of $N+1$-observations of a Markov chain having transition density $p_{\theta_0}(y, z)$, $y, z \in \mathbb{R}$ and invariant measure $\mu_{\theta_0}$. This sequence is defined on the probability space $(\Omega, \mathcal{F}, P_{\theta_0})$. We write $Y_i := Y_{i\Delta}$ for $i = 0, 1, ..., N$.

**(ii).** **(Model process)** Denote by $X^y(\theta)$ a random variable defined on the probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ such that its law is given by $p_\theta(y, \cdot)$.

**(iii).** Denote by $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ the probability space which generates the simulation of the random variable $X^y(\theta)$.

**(iv).** **(Approximating process)** Denote by $X^y_{(m)}(\theta)$ the simulation of an approximation of the model process $X^y(\theta)$, which is defined on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$. $\hat{\mathcal{H}}_t := \hat{\mathcal{F}}_t \otimes \bar{\mathcal{F}}$ satisfies the usual condition. $m$ is the parameter that determines the quality of the approximation. $\tilde{p}^N_\theta(y, \cdot) = \tilde{p}^N_\theta(y, \cdot; m(N))$ is the density for the random variable $X^y_{(m)}(\theta)$.

**(v).** **(Approximated transition density)** Let $K : \mathbb{R} \to \mathbb{R}_+$ be a kernel which satisfies $\int K(x)dx = 1$ and $K(x) > 0$ for all $x$. Denote by $\hat{p}^N_\theta(y, z)$, the kernel density estimation of $\tilde{p}^N_\theta(y, z)$ based on $n$ simulation (independent) of $X^y_{(m)}(\theta)$ which are defined on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$. The outcomes of $n$ simulations are denoted by $X^{y,(k)}_{(m)}(\theta)$, $k = 1, ..., n$. For $h(N) > 0$,

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

**(vi).** For given $m$, we introduce the "average" transition density over all trajectories with respect to the kernel $K$;

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

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 $X_{(m)}^y(\theta)$ is defined. This is done in order not to obscure the arguments that follow and to avoid making the paper excessively long. All the properties that will be required for $p_\theta$ are $\tilde{p}_\theta^N$ that will be satisfied for a subclass of diffusion processes.

**Remark 2.2** **(i).** *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.*

**(ii).** *Note that from the definition and the properties of the kernel $K$, $\overline{p}_\theta^N(y,\cdot)$ satisfies for all $\theta \in \Theta$ and $y \in \mathbb{R}$*

$$\int \overline{p}_\theta^N(y,z)dz = 1 \quad and \quad \overline{p}_\theta^N(y,z) > 0 \quad for\ all\ z \in \mathbb{R}.$$

Our purpose is to estimate the posterior expectation of 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 $\phi_\theta(Y_0^N) = \phi_\theta(Y_0,...,Y_N) = \mu_{\theta_0}(Y_0)\prod_{j=1}^N p_{\theta_0}(Y_{j-1},Y_j)$ is the joint density of $(Y_0,Y_1,...,Y_N)$.

We propose to estimate this quantity based on the simulation of the process:

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

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 Main Theorem

**Assumption 2.3** *We assume the following*

**(1).** **(Observation process)** $\{Y_i\}_{i=0,1,...,N}$ *is an $\alpha$-mixing process with $\alpha_n = O(n^{-5})$.*

**(2).** **(The prior distribution)** *The prior distribution $\pi$ is continuous in $\theta$. And the support is $\Theta$, that is, for all $\theta \in \Theta$, $\pi(\theta) > 0$.*

**(3).** **(Density regularity)** *The transition densities $p, \overline{p}^N \in C^{2,0,0}(\Theta\times\mathbb{R}^2;\mathbb{R}_+)$, and for all $\theta \in \Theta$, $y,z \in \mathbb{R}$, we have that $\min\left\{p_\theta(y,z), \overline{p}_\theta^N(y,z)\right\} > 0$. And $p_\theta$ 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* $c_1$ *and* $c_2 : \mathbb{R} \to (0, \infty)$ *such that for all* $\theta \in \Theta$*,*

$$\inf_N \int |q^i_\theta(y,z) - q^i_{\theta_0}(y,z)|dz \geq c_i(y)|\theta - \theta_0|,$$

*and* $C_i(\theta_0) := \int c_i(y)^2 \mu_{\theta_0}(y)dy \in (0, +\infty)$ *for* $i = 1, 2$ *and* $q^1_\theta = p_\theta$ *and* $q^2_\theta = \overline{p}^N_\theta$*.*

**(5). (Regularity of the log-density)** *We assume that for* $q_\theta = p_\theta, \overline{p}^N_\theta$

$$\sup_N \sup_{\theta\in\Theta} \iint \left( \frac{\partial^i}{\partial \theta^i} \ln q_\theta(y,z) \right)^{12} p_{\theta_0}(y,z)\mu_{\theta_0}(y)dydz < +\infty, \quad for\ i = 0, 1, 2,$$

$$\sup_N \sup_{\theta\in\Theta} \left| \frac{\partial^2}{\partial \theta^2} \iint (\ln q_\theta(y,z))\, \overline{p}^N_{\theta_0}(y,z)\mu_{\theta_0}(y)dydz \right| < +\infty,$$

$$\sup_N \sup_{\theta\in\Theta} \iint \left| \frac{\partial^i}{\partial \theta^i} \ln q_\theta(y,z) \right| \overline{p}^N_{\theta_0}(y,z)\mu_{\theta_0}(y)dydz < +\infty, \quad for\ i = 0, 1,$$

*where* $\frac{\partial^0}{\partial \theta^0} q_\theta = q_\theta$*.* [1]

**(6). (Parameter tuning)**

**(a).** *We assume the following boundedness condition;*

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

**(b).** *Assume that for each* $y, z \in \mathbb{R}$*, there exists a factor* $C^N_1(y,z)$ *and* $c_1(y,z)$ *such that*

$$\left| p_{\theta_0}(y,z) - \overline{p}^N_{\theta_0}(y,z) \right| \leq C^N_1(y,z)a_1(N),$$

*where* $\sup_N C^N_1(y,z) < +\infty$ *and* $a_1(N) \to 0$ *as* $N \to \infty$*, and*

$$C^N_1(y,z)a_1(N)\sqrt{N} < c_1(y,z),$$

*where* $c_1$ *satisfies the following;*

$$\sup_N \sup_{\theta\in\Theta} \iint \left| \frac{\partial}{\partial \theta} \ln \overline{p}^N_\theta(y,z) \right| c_1(y,z)\mu_{\theta_0}(y)dydz < +\infty.$$

**(c).** *There exist some function* $g^N : \mathbb{R}^2 \to \mathbb{R}$ *and constant* $a_2(N)$*, which depends on* $N$*, such that for all* $y, z \in \mathbb{R}$

$$\sup_{\theta\in\Theta} \left| \frac{\partial}{\partial \theta} \ln \overline{p}^N_\theta(y,z) - \frac{\partial}{\partial \theta} \ln p_\theta(y,z) \right| \leq |g^N(y,z)|a_2(N),$$

*where* $\sup_N E_{\theta_0}[|g^N(Y_0, Y_1)|^4] < +\infty$ *and* $a_2(N) \to 0$ *as* $N \to \infty$*.*

**Remark 2.4** **(i).** *Assumption 2.3* **(4)** *is needed in order to be able to obtain that the density can be used in order to discern the value of* $\theta$ *from the observations. This type of assumption is natural in statistics and can be assured in the case of one dimensional stochastic differential equations under differentiability of the coefficients. Assumption 2.3* **(5)** *will be satisfied under enough regularity of the transition density function* $p_\theta$ *and its approximation* $\overline{p}^N_\theta$*. This can be achieved with Malliavin Calculus techniques in the case of diffusion equations.*

[1] The power 12 is needed to prove a central limit theorem (Proposition 4.4.).

**(ii).** *Assumption 2.3* **(6)-(a)** *will be crucial in what follows and it is the property that will determine the rate of convergence and the tuning properties. Note that all other hypothesis deal with the transition density $p_\theta$ or the average of its approximation $\overline{p}_\theta^N$. Therefore the needed properties essentially follow from similar properties of $p_\theta$ and some limit arguments. Assumption 2.3* **(6)-(a)** *is the only condition that deals with the approximation itself $\hat{p}_\theta^N$, 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 5.*

**(iii).** *In this problem, we need to study two approximation problems, one is a difference between the transition densities of the observation process and the approximated process, and the other is a difference of the transition density of the approximated process and the expectation for the approximation based on kernel density estimation. Assumption 2.3* **(6)-(b)** *and* **(c)** *state the rate of convergence of the density of the approximation and its derivatives. This problem was studied in Bally, Talay [1] and Guyon [7], and the second approximation problem can be dealt using the kernel density estimation theory. For example, in the case that the data $\{Y_i\}$ comes from a stochastic differential equation, the Euler-Maruyama approximation is most commonly used. And in that case, we have the following results. Assume that the drift and diffusion coefficients belong to $C_b^\infty(\mathbb{R})$, and the diffusion coefficient satisfies the uniform ellipticity condition. Then for $\alpha, \beta \in \mathbb{N}$, there exist $c_1 \geq 0$ and $c_2 > 0$ such that for all $N \geq 1$, $\Delta \in (0, 1]$ and $y, z \in \mathbb{R}$,*

$$\partial_y^\alpha \partial_z^\beta \tilde{p}_{\theta_0}^N(y,z) - \partial_y^\alpha \partial_z^\beta p_{\theta_0}(y,z) \leq \frac{1}{m(N)} \partial_y^\alpha \partial_z^\beta \pi_{\theta_0}(\Delta, y, z) + r_{\theta_0}^{m(N)}(\Delta, y, z),$$

*where let $D$ be a differential operator and $p_{\theta_0}^t(y,z)$ be the transition density of $Y_t$, set*

$$\pi_{\theta_0}(\Delta, y, z) := \int_0^\Delta \int_{-\infty}^\infty p_{\theta_0}^s(y,w) D(p_{\theta_0}^{\Delta-s}(\cdot, z))(w) dw ds,$$

*and*

$$\left| r_{\theta_0}^{m(N)}(\Delta, y, z) \right| \leq c_1 \frac{1}{m(N)^2} \frac{1}{\Delta^{\frac{\alpha+\beta+5}{2}}} \exp\left( -\frac{c_2 |y-z|^2}{\Delta} \right).$$

*For more details, see Proposition 1 in Guyon [7].*

*For the second problem, in the case that the kernel $K$ satisfies $\int xK(x)dx = 0$ and $\int x^2 K(x)dx < +\infty$, if $\tilde{p}_{\theta_0}^N(y,z)$ is uniformly bounded in $y, z \in \mathbb{R}$ and $N \in \mathbb{N}$ and twice continuously differentiable with respect to $z$ for all $y \in \mathbb{R}$ and $N \in \mathbb{N}$, then we have, for all $y \in \mathbb{R}$ and $N \in \mathbb{N}$,*

$$\overline{p}_{\theta_0}^N(y,z) - \tilde{p}_{\theta_0}^N(y,z) = \frac{1}{2} h(N)^2 \frac{\partial^2 \tilde{p}_{\theta_0}^N}{\partial z^2}(y,z) \int x^2 K(x) dx + o(h(N)^2).$$

*For more details, see Wand, Jones [11].*

*Therefore roughly speaking, $a_1(N) = \frac{1}{m(N)} + h(N)^2$. And the choice for $m(N) = \sqrt{N}$ will satisfy Assumption 2.3* **(6)-(b)**.

Now we state the main result of the paper.

**Theorem 2.5** *Under Assumption 2.3, there exists some positive random variable $\Xi$ such that $\Xi < +\infty$ a.s. and*

$$\left| E_N[f] - \hat{E}_{N,m}^n[f] \right| \leq \frac{\Xi}{\sqrt{N}} \text{ a.s.}$$

*And also, we have*

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

*where* $\Xi_1$ *is some positive random variable with* $\Xi_1<+\infty$ *a.s. and* $\Xi_2$ *is some positive random variable with* $\Xi_2<+\infty$ *a.s.*

# 3 Idea of the Proof of Theorem 2.5

First we introduce some notation. Let $p,q:\mathbb{R}^2\to\mathbb{R}_+$ be strictly positive functions of two variables. Then let

$$\begin{aligned}
H(p,q) &:= \iint\Big(\ln p(y,z)\Big)q(y,z)\mu_{\theta_0}(y)dydz,\\
Z_N(\theta) &:= \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}\Big\{\ln p_\theta(Y_i,Y_{i+1})-H(p_\theta,p_{\theta_0})\Big\},\\
\varepsilon(\theta) &:= H(p_\theta,p_{\theta_0})-H(p_{\theta_0},p_{\theta_0}),\\
\beta_N(\theta) &:= Z_N(\theta)-Z_N(\theta_0).
\end{aligned}$$

And also we set;

$$\begin{aligned}
\bar{Z}_N(\theta) &:= \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}\Big(\ln\hat{p}^N_\theta(Y_i,Y_{i+1})-H\left(\bar{p}^N_\theta,\bar{p}^N_{\theta_0}\right)\Big),\\
\bar{\varepsilon}^N(\theta) &:= H\left(\bar{p}^N_\theta,\bar{p}^N_{\theta_0}\right)-H\left(\bar{p}^N_{\theta_0},\bar{p}^N_{\theta_0}\right),\\
\bar{\beta}_N(\theta) &:= \bar{Z}_N(\theta)-\bar{Z}_N(\theta_0).
\end{aligned}$$

Set $\Theta_0:=\Theta\setminus\{\theta_0\}$. The following proposition states the properties that are needed to achieve the proof of Theorem 2.5

**Proposition 3.1** *Under Assumption 2.3, we have the following results.*

**(i).** *There exist some strictly negative constants* $c_1,c_2$ *such that*

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

**(ii).** *There exist some random variables* $d_1,d_2$ *such that*

$$d_1\le\inf_N\inf_{\theta\in\Theta_0}\frac{\beta_N(\theta)}{\theta-\theta_0}\le\sup_N\sup_{\theta\in\Theta_0}\frac{\beta_N(\theta)}{\theta-\theta_0}\le d_2\quad a.s.$$

**(iii).** *There exist some strictly negative constants* $c_3,c_4$ *such that*

$$c_3\le\inf_N\inf_{\theta\in\Theta_0}\frac{\bar{\varepsilon}^N(\theta)}{(\theta-\theta_0)^2}\le\sup_N\sup_{\theta\in\Theta_0}\frac{\bar{\varepsilon}^N(\theta)}{(\theta-\theta_0)^2}\le c_4<0.$$

**(iv).** *There exist some random variables* $d_3,d_4$ *such that*

$$d_3\le\inf_N\inf_{\theta\in\Theta_0}\frac{\bar{\beta}_N(\theta)}{\theta-\theta_0}\le\sup_N\sup_{\theta\in\Theta_0}\frac{\bar{\beta}_N(\theta)}{\theta-\theta_0}\le d_4\quad a.s.$$

We will give an idea of the proof of this proposition after this section. We give the proof of Theorem 2.5 using the results of Proposition 3.1.

*Idea of the Proof of Theorem 2.5.* We decompose the approximation error as follows;

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

The goal is then to prove that there exists some random variable $C_1$ and $C_2$ such that

$$\left|\frac{I_N(f) - f(\theta_0)I_N(1)}{I_N(1)}\right| \leq \frac{C_1}{\sqrt{N}} \quad a.s., \quad \text{and} \quad \left|\frac{\hat{I}^n_{N,m}(f) - f(\theta_0)\hat{I}^n_{N,m}(1)}{\hat{I}^n_{N,m}(1)}\right| \leq \frac{C_2}{\sqrt{N}} \quad a.s.$$

Indeed, we can write $I_N(f)$ and $\hat{I}^n_{N,m}(f)$ as follows;

$$\begin{aligned} I_N(f) &= e^{NH(p_{\theta_0}, p_{\theta_0}) + \sqrt{N}Z_N(\theta_0)} \int_{\Theta_0} f(\theta) e^{N\varepsilon(\theta) + \sqrt{N}\beta_N(\theta)} \mu_\theta(Y_0)\pi(\theta)d\theta, \\ I^n_{N,m}(f) &= e^{NH(\bar{p}^N_{\theta_0}, \bar{p}^N_{\theta_0}) + \sqrt{N}\bar{Z}_N(\theta_0)} \int_{\Theta_0} f(\theta) e^{N\bar{\varepsilon}(\theta) + \sqrt{N}\bar{\beta}_N(\theta)} \mu_\theta(Y_0)\pi(\theta)d\theta. \end{aligned}$$

Then by using the Laplace method and Proposition 3.1, we have our conclusion. In fact, as $N$ goes to infinity the leading term in the quotients $\frac{I_N(f)}{I_N(1)}$ and $\frac{\hat{I}^n_{N,m}(f)}{\hat{I}^n_{N,m}(1)}$ are determined by $\varepsilon(\theta)$ and $\bar{\varepsilon}(\theta)$ due to Proposition 3.1. Their behaviors are similar to Gaussian integrals where the variance tends to zero and therefore the integrals will tend to the value in their "mean" which in this case is $\theta_0$ as it follows from Proposition 3.1. Details can be found in Kohatsu-Higa et al. [8]. ■

# 4 Proof of Proposition 3.1

## 4.1 Ideas for the proof of Proposition 3.1 (i) and (iii)

In order to prove the upper bounds for the statements in Proposition 3.1 (i) and (iii), one uses the Pinsker's inequality;

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

and the identifiability condition. Therefore under Assumptions 2.3 **(4)**, we obtain the upper bound in (i) and under Assumption 2.3 **(4)** and **(5)**, we obtain the upper bound in (iii).

For the lower bounds, we give a useful lemma for the first derivative of $H(p_\theta, p_{\theta_0})$ in $\theta$;

**Lemma 4.1** *Let q be a transition density, which depends on a parameter $\theta$. We assume that for all $\theta \in \Theta$,*

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

*Then*

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

Using the above Lemma together with Taylor's expansion we obtain the lower bound for (i) under Assumption 2.3 **(3)**, **(4)** and **(5)**, and under Assumption 2.3 **(4)** and **(5)** for (iii).

## 4.2 Ideas for the proof of Proposition 3.1 (ii)

In this section, we consider Proposition 3.1 (ii). We prove this boundedness by using a central limit theorem in $C(\Theta;\mathbb{R}^\infty)$. Note that $C(\Theta;\mathbb{R}^\infty)$ is a complete and separable metric space with metric $\eta$ defined as follows; For $x=(x_1,x_2,...), y=(y_1,y_2,...)\in C(\Theta;\mathbb{R}^\infty)$,

$$\eta(x,y) := \sup_{\theta\in\Theta}\sum_{i=1}^{\infty}\frac{1}{2^i}\left\{|x_i(\theta)-y_i(\theta)|\wedge 1\right\}.$$

For $N\in\mathbb{N}$, set

$$\tilde{\beta}_N(\theta) := \begin{cases} \dfrac{\beta_N(\theta)}{\theta-\theta_0} = \dfrac{Z_N(\theta)-Z_N(\theta_0)}{\theta-\theta_0} & \text{if } \theta\neq\theta_0, \\ \dfrac{\partial}{\partial\theta}Z_N(\theta_0) & \text{if } \theta=\theta_0. \end{cases}$$

And set $\gamma_N := (\tilde{\beta}_N, \tilde{\beta}_{N-1}, ..., \tilde{\beta}_1, 0, ...)\in C(\Theta;\mathbb{R}^\infty)$. The reason why we need to use $\mathbb{R}^\infty$ in the above setting can be seen from the following Lemmas. In fact, in order to prove the boundedness of $\tilde{\beta}_N(\theta)$ in $N$, we need to consider another random vector which has the same "joint"-distribution as $\gamma_N$ when we apply the Skorohod representation theorem.

The idea of the proof consists of proving that the sequence $\gamma_N$ converges weakly in $C(\Theta;\mathbb{R}^\infty)$. Therefore the limit $\gamma=(\gamma^1,\gamma^2,\cdots)$ should satisfy that there exist some random variables $d_1,d_2$ such that

$$d_1 \le \inf_{\theta\in\Theta}\gamma^1(\theta) \le \sup_{\theta\in\Theta}\gamma^1(\theta)\le d_2 \quad a.s.$$

Now, without loss of generality we can use the Skorohod representation theorem, so that the first component $\gamma_N^1=\tilde{\beta}_N$ of $\gamma_N$ will satisfy a similar property. And finally, from the convergence, we obtain the boundedness of $\tilde{\beta}_N$ in $N$. In particular, we use the following lemmas.

**Lemma 4.2** *Let* $X=(X_1,X_2,...), Y=(Y_1,Y_2,...)\in C(\Theta;\mathbb{R}^\infty)$ *be random variables such that* $X\overset{d}{=}Y$. *Then we have*

$$\sup_n\sup_{\theta\in\Theta}|X_n(\theta)| \overset{d}{=} \sup_n\sup_{\theta\in\Theta}|Y_n(\theta)|.$$

**Lemma 4.3** *Let* $(S,\|\cdot\|)$ *be a complete, separable metric normed space. Let* $X$ *be an* $S$*-valued random variable on a probability space* $(\Omega,\mathcal{F},P)$ *and let* $Y$ *be an* $S$*-valued random variable on a probability space* $(\Omega',\mathcal{F}',P')$. *Supposed that* $X\overset{d}{=}Y$ *and there exists an* $\mathbb{R}^+$*-valued random variable* $d$ *on* $(\Omega',\mathcal{F}',P')$ *such that*

$$\|Y(\omega')\|\le d(\omega') \qquad \text{for all } \omega'\in\Omega'.$$

*Then there exists a positive random variable* $M$ *on* $(\Omega,\mathcal{F},P)$ *such that*

$$\|X(\omega)\|\le M(\omega) \quad \text{a.s. } \omega\in\Omega.$$

In order to prove the weak convergence of $\gamma_N$ we extend Theorem 7.1 and Theorem 7.3 in Billingsley [3] which imply that it is enough to prove convergence of marginals and tightness of the sequence $\gamma_N$.

In order to prove convergence of marginals, we use Assumption 2.3 **(1)** and **(5)**, so that for every $r\in\mathbb{N}$ and $\theta_1,...,\theta_r\in\Theta$, $(\gamma_N(\theta_1),...,\gamma_N(\theta_r))$ converges weakly. The proof uses the Crámer-Wold device and the following extension of the Central limit theorem for $\alpha$-mixing processes (the proof is an extension of Theorem 27.5 in Billingsley [2], p.316.)

**Proposition 4.4** *Suppose that $X_1, X_2, ...$ is stationary and $\alpha$-mixing with $\alpha_n = O(n^{-5})$, where we set $\alpha_0 = 1$ and $f$ is a $\mathcal{B}(\mathbb{R}^2)$-measurable function which satisfies $E[f(X_0, X_1)] = 0$ and $E[f(X_0, X_1)^{12}] < +\infty$. If we set $f_i := f(X_i, X_{i+1})$ and $S_n := f_1 + \cdots + f_n$, then*

$$\frac{1}{n}Var(S_n) \longrightarrow \sigma^2 := E\left[f_1^2\right] + 2\sum_{k=1}^{\infty} E\left[f_1 f_{1+k}\right],$$

*where the series converges absolutely. And $\frac{S_n}{\sqrt{n}} \Rightarrow M$, where $M$ is a normal distributed random variable with mean 0 and variance $\sigma^2$. If $\sigma = 0$, we define $M = 0$.*

To prove the tightness of $\gamma_N$, we use Assumption 2.3 **(1)**, **(3)** and **(5)**. One needs to prove that for all $\varepsilon > 0$,

$$\lim_{\delta\to 0}\limsup_{N\to\infty} P_{\theta_0}\left(\sup_{|\theta-\theta'|\le\delta}\sum_{i=1}^{N}\frac{1}{2^i}\left\{\left|\tilde{\beta}_{N-i+1}(\theta) - \tilde{\beta}_{N-i+1}(\theta')\right| \wedge 1\right\} \ge \varepsilon\right) = 0.$$

In the proof, the basic ingredient is the Garsia, Rodemich, Rumsey lemma.

## 4.3 Decomposition for the Estimation for Proposition 3.1 (iv)

Set

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

$$J_N^2(\theta) := \begin{cases} \sqrt{N}\displaystyle\iint\frac{\ln\overline{p}_\theta^N - \ln\overline{p}_{\theta_0}^N}{\theta-\theta_0}(y,z)\left\{\overline{p}_{\theta_0}^N - p_{\theta_0}\right\}(y,z)\mu_{\theta_0}(y)dydz, & \theta \neq \theta_0 \\ \sqrt{N}\displaystyle\iint\frac{\partial}{\partial\theta}\ln\overline{p}_\theta^N(y,z)\Big|_{\theta=\theta_0}\left\{\overline{p}_{\theta_0}^N - p_{\theta_0}\right\}(y,z)\mu_{\theta_0}(y)dydz, & \theta = \theta_0, \end{cases}$$

$$J_N^3(\theta) := \begin{cases} \dfrac{1}{\sqrt{N}}\displaystyle\sum_{i=0}^{N-1}\left\{\frac{\ln\overline{p}_\theta^N - \ln\overline{p}_{\theta_0}^N}{\theta-\theta_0}(Y_i, Y_{i+1}) \right. \\ \qquad\left. - \displaystyle\iint\frac{\ln\overline{p}_\theta^N - \ln\overline{p}_{\theta_0}^N}{\theta-\theta_0}(y,z)p_{\theta_0}(y,z)\mu_{\theta_0}(y)dydz\right\}, & \theta \neq \theta_0 \\ \dfrac{1}{\sqrt{N}}\displaystyle\sum_{i=0}^{N-1}\left\{\frac{\partial}{\partial\theta}\ln\overline{p}_\theta^N(Y_i, Y_{i+1})\Big|_{\theta=\theta_0} - \iint\frac{\partial}{\partial\theta}\ln\overline{p}_\theta^N(y,z)\Big|_{\theta=\theta_0}p_{\theta_0}(y,z)\mu_{\theta_0}(y)dydz\right\}, & \theta = \theta_0, \end{cases}$$

Then we define:

$$\frac{\overline{\beta}_N(\theta)}{\theta-\theta_0} := J_N^1(\theta) - J_N^2(\theta) + J_N^3(\theta).$$

We can easily prove boundedness of the first term $J_N^1(\theta)$ from Assumption 2.3 **(6)-(a)**, and also we can easily prove boundedness of $J_N^2(\theta)$ from Assumption 2.3 **(6)-(b)**. Finally we consider the third term $J_N^3(\theta)$. To prove boundedness of $J_N^3(\theta)$, we prove a weak convergence by using the similar argument as Section 4.2. Then we can prove Proposition 3.1 (iv).

# 5 Parameter Tuning and the Assumption 2.3 (6)-(a)

This section is devoted to proving that Assumption 2.3 **(6)-(a)** is satisfied under sufficient smoothness hypothesis on the random variables and processes that appear in the problem as well as the correct parameter tuning. That is, we need to prove that the following condition (Assumption 2.3 **(6)-(a)** in Section 2.2) is satisfied

$$\sup_N \sup_{\theta\in\Theta} \left| \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \left( \frac{\partial}{\partial\theta} \ln \hat{p}_\theta^N (Y_i, Y_{i+1}) - \frac{\partial}{\partial\theta} \ln \overline{p}_\theta^N (Y_i, Y_{i+1}) \right) \right| < +\infty \quad a.s. \tag{5.1}$$

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

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

Here $m \equiv m(N)$, $n \equiv n(N)$ and $h \equiv h(N)$ are parameters that depend on $N$. $n$ is the number of Monte Carlo simulations used in order to estimate the density and $m$ is the parameter of approximation (in the Euler scheme this is the number of time steps used in the simulation of $X_{(m)}^{y,(1)}(\theta)$) and $h$ is the bandwidth size associated to the kernel density estimation method. In this sense we will always think of hypotheses in terms of $N$ although we will drop them from the notation and just use $m, n$ and $h$. The goal of this section is to prove that under certain hypotheses, there is a choice of $m, n$ and $h$ that ensures that condition (5.1) is satisfied.

As the main problem is to obtained upper and lower bounds for $\hat{p}_\theta^N$ for random arguments, we will first restrict the values for the random variables $Y_i$, $i = 0, ...., N-1$ to a compact set. This is obtained using an exponential type Chebyshev's inequality and the Borel-Cantelli Lemma (Theorem 4.3 in pp.53 of Billingsley [2]) as follows

**Lemma 5.1** *Assume the following hypothesis*

**(H0).** $m_{c_1} := \sup_i E[e^{c_1|Y_i|^2}] < \infty$ *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 that for a.s.* $\omega \in \Omega$, *there exists N big enough such that* $\max_{i=1,\dots,N} |Y_i| < a_N$. *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 (5.1) is the same decomposition as in the proof of Theorem 3.2 in pp.73 of Bosq [4]. That is,

$$\begin{aligned} &\sup_{\theta\in\Theta} \sup_{|x|,|y|\leq a_N} \left| \frac{\partial_\theta \hat{p}_\theta^N}{\hat{p}_\theta^N}(x,y) - \frac{\partial_\theta \overline{p}_\theta^N}{\overline{p}_\theta^N}(x,y) \right| \\ &\leq \frac{\sup_{\theta\in\Theta} \sup_{|x|,|y|\leq a_N} \left| \partial_\theta \hat{p}_\theta^N(x,y) - \partial_\theta \overline{p}_\theta^N(x,y) \right|}{\inf_{\theta\in\Theta} \inf_{|x|,|y|\leq a_N} \overline{p}_\theta^N(x,y)} \\ &\quad + \sup_{\theta\in\Theta} \sup_{|x|,|y|\leq a_N} \left| \frac{\partial_\theta \hat{p}_\theta^N}{\hat{p}_\theta^N}(x,y) \right| \frac{\sup_{\theta\in\Theta} \sup_{|x|,|y|\leq a_N} \left| \hat{p}_\theta^N(x,y) - \overline{p}_\theta^N(x,y) \right|}{\inf_{\theta\in\Theta} \inf_{|x|,|y|\leq a_N} \overline{p}_\theta^N(x,y)} =: \frac{A}{B} + C\frac{D}{B}, \end{aligned} \tag{5.2}$$

where we remark that

$$\partial_\theta \hat{p}^N_\theta(x,y) = \frac{1}{nh^2}\sum_{k=1}^{n} K'\left(\frac{X^{y,(k)}_{(m)}(\theta)-z}{h}\right)\partial_\theta X^{y,(k)}_{(m)}(\theta),$$

$$\partial_\theta \overline{p}^N_\theta(x,y) = E\left[\frac{1}{h^2}K'\left(\frac{X^{y,(k)}_{(m)}(\theta)-z}{h}\right)\partial_\theta X^{y,(k)}_{(m)}(\theta)\right].$$

Therefore in order to prove the finiteness of (5.1), 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 of the quantities $\overline{p}^N_\theta$ and $\hat{p}^N_\theta$. First, we start analyzing the difficult part $C\frac{D}{B}$.

## 5.1 Upper bound for $C\frac{D}{B}$ in (5.2)

We work in this section under the following hypotheses:
**(H1).** Assume that there exist some positive constants $\varphi_1,\ \varphi_2$, where $\varphi_1$ is independent of $N$ and $\varphi_2$ is independent of $N$ and $\Delta$, such that the following holds;

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

where we set

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

where $\|\cdot\|$ is the max-norm.
**(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 $\sum_{N=1}^{\infty}\frac{n a_N^{2r_3} E\left[\left|Z_{3,N}(\cdot)\right|^{r_3}\right]}{(h^2 b_{3,N})^{r_3}} < \infty$, where

$$Z^{(k)}_{3,N}(\omega) := a_N^{-2}\left(\sup_{(\mathbf{x},\theta)\in B^N}\left|X^{x,(k)}_{(m)}(\theta,\omega)\right|+1\right)\sup_{(\mathbf{x},\theta)\in B^N}\left|\partial_\theta X^{x,(k)}_{(m)}(\theta,\omega)\right|.$$

**(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 $\sum_{N=1}^{\infty}\frac{nE\left[\left|Z^{(k)}_{4,N}(\cdot)\right|^{r_4}\right]}{(b_{4,N})^{r_4}} < \infty$, where

$$Z^{(k)}_{4,N}(\omega) := a_N^{-1}\left(\sup_{(\mathbf{x},\theta)\in B^N}\left|\partial_x X^{x,(k)}_{(m)}(\theta;\omega)\right| + \sup_{(\mathbf{x},\theta)\in B^N}\left|\partial_\theta X^{x,(k)}_{(m)}(\theta;\omega)\right|\right).$$

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

$$\left|\partial_x\overline{p}^N_\theta(x,y)\right|,\ \left|\partial_y\overline{p}^N_\theta(x,y)\right|,\ \left|\partial_\theta\overline{p}^N_\theta(x,y)\right| \leq C_5 < +\infty.$$

**(H6).** Assume that $\eta_N$ and $\nu_N$ are sequences of positive numbers so that

$$\sum_{N=1}^{\infty}\nu_N^3\exp\left(-\frac{(\eta_N)^2 nh^2}{16\|K\|_\infty}\right) < \infty,$$

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

Note that from the assumption **(H1)**, we have a lower bound of $B$ in (5.2).

**Lemma 5.2** *Assume hypothesis* **(H2)** *and* **(H3)***, then we have that*

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

The proof of the Lemma follows by rewriting the fraction $\frac{\partial_\theta \hat{p}_\theta^N}{\hat{p}_\theta^N}$ and show that it can be bounded using the upper bound of $\frac{K'}{K}$ and the derivative of the approximating process with respect to $\theta$.

## 5.2 Upper bound of $D$ in (5.2)

In this section, we use the modulus of continuity of $\hat{p}_\theta^N$ and $\overline{p}_\theta^N$ in order to find an upper bound for $B$. Similar ideas appear in Theorem 2.2 in pp.49 of Bosq [4].

**Lemma 5.3** *Set*

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

*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*

$$\begin{aligned}
\sup_{(\mathbf{x},\theta)\in B^N}\left|\hat{p}_\theta^N(\mathbf{x})-\overline{p}_\theta^N(\mathbf{x})\right| &= \max_{\substack{1\le l_1\le\nu_N^2\\1\le l_2\le\nu_N}}\sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N}\left|\hat{p}_\theta^N(\mathbf{x})-\overline{p}_\theta^N(\mathbf{x})\right|\\
&\le \max_{\substack{1\le l_1\le\nu_N^2\\1\le l_2\le\nu_N}}\sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N}\left|\hat{p}_\theta^N(\mathbf{x})-\hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right|+\max_{\substack{1\le l_1\le\nu_N^2\\1\le l_2\le\nu_N}}\left|\hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)-\overline{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right|\\
&\quad+\max_{\substack{1\le l_1\le\nu_N^2\\1\le l_2\le\nu_N}}\sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N}\left|\overline{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)-\overline{p}_\theta^N(\mathbf{x})\right|. \qquad (5.3)
\end{aligned}$$

The proof of the above Lemma is straightforward. We consider the first term of (5.3).

**Lemma 5.4** *Under* **(H2)** *and* **(H4)***, then we have that*

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

Now we consider the third term in (5.3).

**Lemma 5.5** *Assume* **(H5)***, then,*

$$\max_{\substack{1\le l_1\le\nu_N^2\\1\le l_2\le\nu_N}}\sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N}\left|\overline{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)-\overline{p}_\theta^N(\mathbf{x})\right|\le 3C_5\frac{a_N}{\nu_N}$$

Finally, we consider the second term of (5.3).

**Lemma 5.6** *Assume* **(H2)** *and that* $\eta_N$ *satisfies* **(H6)***, then we have that*

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

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

**Theorem 5.7** *Assume conditions* **(H2)**, **(H4)**, **(H5)** *and* **(H6)***, then there exists* $N$ *big enough such that*

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

## 5.3 Upper bound for $\frac{A}{B}$ in (5.2)

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 5.6 for the derivatives, $\partial_\theta \hat{p}_\theta^N$ and $\partial_\theta \overline{p}_\theta^N$, has to be reworked as the condition $\|\hat{p}_\theta^N\|_\infty < +\infty$ and $\|\overline{p}_\theta^N\|_\infty < +\infty$ are not valid. From the condition **(H1)**, we have

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

Here we consider the sup-term as before.

**Lemma 5.8** *We use the same notations as the previous section. Then we have*

$$\begin{aligned}
&\sup_{\theta\in\Theta} \sup_{|x|,|y|\le a_N} \left|\partial_\theta \hat{p}_\theta^N(\mathbf{x}) - \partial_\theta \overline{p}_\theta^N(\mathbf{x})\right| \\
&\le \max_{\substack{1\le l_1\le \nu_N^2 \\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\partial_\theta \hat{p}_\theta^N(\mathbf{x}) - \partial_\theta \hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right| + \max_{\substack{1\le l_1\le \nu_N^2 \\ 1\le l_2\le \nu_N}} \left|\partial_\theta \hat{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right) - \partial_\theta \overline{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right| \\
&\quad + \max_{\substack{1\le l_1\le \nu_N^2 \\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\partial_\theta \overline{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right) - \partial_\theta \overline{p}_\theta^N(\mathbf{x})\right|. \qquad (5.4)
\end{aligned}$$

As in previous sections, we define that

$$\begin{aligned}
\dot{Z}_{4,N}^{(k)} := a_N^{-1}\Bigg( & h \sup_{\theta\in\Theta} \sup_{|x|\le a_N} \left|\partial_x \partial_\theta X_{(m)}^{x,(k)}(\theta)\right| + h \sup_{\theta\in\Theta} \sup_{|x|\le a_N} \left|\partial_\theta \partial_\theta X_{(m)}^{x,(k)}(\theta)\right| \\
& + \left(Z_{4,N}^{(k)} + 1\right) \sup_{\theta\in\Theta} \sup_{|x|\le a_N} \left|\partial_\theta X_{(m)}^{x,(k)}(\theta)\right|\Bigg).
\end{aligned}$$

Note that $\{\dot{Z}_{4,N}^{(k)}\}_{k=1,2,\ldots}$ is a sequence of independent and identically distributed random variables. Then we set the following hypothesis.
**(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 $\sum_{N=1}^{\infty} \frac{nE\left[\left|\dot{Z}_{4,N}^{(k)}\right|^{\dot{r}_4}\right]}{(\dot{b}_{4,N})^{\dot{r}_4}} < \infty$.

**Lemma 5.9** *Under the above hypothesis and* **(H4')**, *we have that*

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

We consider the third term of (5.4).

$$\begin{aligned}
&\max_{\substack{1\le l_1\le \nu_N^2 \\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \left|\partial_\theta \overline{p}_{\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right) - \partial_\theta \overline{p}_\theta^N(\mathbf{x})\right| \\
&\le \frac{a_N}{\nu_N} \max_{\substack{1\le l_1\le \nu_N^2 \\ 1\le l_2\le \nu_N}} \sup_{(\mathbf{x},\theta)\in B_{l_1 l_2}^N} \Bigg\{ \sup_{0\le\varepsilon\le 1} \left|\partial_x \partial_\theta \overline{p}_\theta^N\left(\varepsilon x + (1-\varepsilon) x_{l_1}^N, y\right)\right| + \sup_{0\le\varepsilon\le 1} \left|\partial_y \partial_\theta \overline{p}_\theta^N\left(x_{l_1}^N, \varepsilon y + (1-\varepsilon) y_{l_1}^N\right)\right| \\
&\qquad + \sup_{0\le\varepsilon\le 1} \left|\partial_\theta \partial_\theta \overline{p}_{\varepsilon\theta+(1-\varepsilon)\theta_{l_2}^N}^N\left(\mathbf{x}_{l_1}^N\right)\right| \Bigg\}.
\end{aligned}$$

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

$$\left|\partial_x \partial_\theta \overline{p}_\theta^N(x,y)\right|,\ \left|\partial_y \partial_\theta \overline{p}_\theta^N(x,y)\right|,\ \left|\partial_\theta^2 \overline{p}_\theta^N(x,y)\right| \le \dot{C}_5 < +\infty.$$

**Lemma 5.10** *Assume* **(H5')**. *Then, we have*

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

Finally, we consider the second term of (5.4). Set

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

Note that $\{\dot{W}^{(j),\mathbf{x}}_{m,h}(\theta)\}_{j=1,2,\dots}$ is a sequence of independent and identically distributed random variables with $E\left[\dot{W}^{(j),\mathbf{x}}_{m,h}(\theta)\right] = 0$.

**(H6').** There exists $\dot{C}_6 > 0$ and $\dot{\alpha}_6 > 0$ and a sequence of positive numbers $\dot{b}_{6,N}$ such that $\sum_{N=1}^\infty \nu_N^3 \exp\left(-\frac{n(\eta_N)^2 h^4}{2\|K'\|^2_\infty (\dot{b}_{6,N})^2 a_N^2}\right)\left\{1 + \frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}}\right\}^n < \infty$.
**(H6a').** Assume that there exists $\dot{r}_6 > 0$ and a sequence of positive numbers $\dot{b}_{6,N}$ in **(H6')** satisfy, such that $\sum_{N=1}^\infty n\frac{E\left[\left|\dot{Z}_{6,N}\right|^{\dot{r}_6}\right]}{(\dot{b}_{6,N})^{\dot{r}_6}} < \infty$ for

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

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

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

Applying Lemma 6.2 in Appendix, we obtain the following important Lemma.

**Lemma 5.11** *Assume* **(H6)**, **(H6a')** *and* **(H6b')**. *Then there exists $N$ big enough such that*

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

**Theorem 5.12** *Assume conditions* **(H1)**, **(H2)**, **(H4')**, **(H5')**, **(H6')**, **(H6a')** *and* **(H6b')**. *Then we have that for a.s. $\omega$, there exists $N_0 \equiv N_0(\omega)$ such that for all $N \ge N_0$ we have*

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

Finally putting all our results together, we have (see Theorem 5.7).

**Theorem 5.13** *Assume conditions* **(H0)**, **(H1)**, **(H2)**, **(H3)**, **(H4)**, **(H5)**, **(H6)**, **(H4')**, **(H5')**, **(H6')**, **(H6a')** *and* **(H6b')**. *Then for a.s. $\omega$, there exists $N_0 \equiv N_0(\hat{\omega})$ such that for all $N \ge N_0$ we have*

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

## 5.4 Conclusion: Tuning for $n$ and $h$

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 $N$. We rewrite some 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 next. These constants depend on $K$, $\Delta$ and $\Theta$. They are independent of $n$, $h$ and $N$ but they depend continuously on other parameters.

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

$$\sqrt{N} e^{\frac{\varphi_2 a_N^2}{\Delta}} \times \left( \frac{a_N^2}{\nu_N h^2} \left( b_{4,N} b_{3,N} + \frac{\dot{b}_{4,N}}{h} \right) + \left( \eta_N + \frac{a_N}{\nu_N} \right) b_{3,N} \right) \leq C_{K,\Delta,\Theta}. \tag{5.5}$$

**(ii).** (Borel-Cantelli for $Y_i$, **(H0)**)
$m_{c_1} := E[e^{c_1|Y_1|^2}] < +\infty$ 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 $r_3 > 0$,

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

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

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

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

$$\sum_{N=1}^{\infty} \nu_N^3 \exp\left( -\frac{(\eta_N)^2 n h^2}{16\|K\|_\infty} \right) < +\infty.$$

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

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

**(vii).** (Borel-Cantelli for $|\partial_\theta \hat{p}_\theta^N(\mathbf{x}) - \partial_\theta \overline{p}_\theta^N(\mathbf{x})|$, **(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 $\dot{Z}^{(k)}_{6,N}(\omega)$, **(H6a')**) For some $\dot{r}_6 > 0$,

$$\sum_{N=1}^{\infty} \frac{n}{(\dot{b}_{6,N})^{\dot{r}_6}} < +\infty \quad \text{and} \quad \sup_{N\in\mathbb{N}} E\left[\left|\dot{Z}_{6,N}(\cdot)\right|^{\dot{r}_6}\right] < +\infty \quad \text{for each fixed } m \in \mathbb{N}.$$

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

$$\left(\frac{\eta_N h^2}{(\|K'\|_\infty \dot{b}_{6,N})^2 a_N} \exp\left(-\frac{(\eta_N)^2}{2(\frac{\|K'\|_\infty}{h^2}\dot{b}_{6,N}a_N)^2}\right)\right)^{\dot{q}_6} \le \frac{\dot{C}_6}{n^{1+\dot{\alpha}_6}} \quad \text{and} \quad \sup_{N\in\mathbb{N}} E\left[\left|\dot{Z}_{6,N}(\cdot)\right|^{\dot{q}_6}\right] < +\infty.$$

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

### 5.4.1 Parameter Tuning

Set $a_N := \sqrt{c_2 \ln N}$ for some positive constant $c_2$. Set $n = C_1 N^{\alpha_1}$ 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 $c_1 > \frac{2}{c_2}$. Note that if we choose $c_2$ large enough, then $c_1$ can be chosen as small as needed.

With the above specifications, we can check that all the conditions in Section 5.4 are satisfied if the following parameter condition is satisfied for $N$ sufficiently large

$$\left(4\alpha_2 + 2\frac{\alpha_1 + \dot{\gamma}_6}{\dot{r}_6} + \frac{\varphi_2 c_2}{\Delta} + \frac{1}{2} + \frac{\gamma_3}{r_3} + \frac{\alpha_1}{r_3}\right)\dot{q}_6 > \alpha_1, \tag{5.6}$$

which has to be satisfied together with

$$\alpha_1\left(1 - \frac{2}{r_3} - \frac{2}{\dot{r}_6}\right) > 8\alpha_2 + 1 + \frac{4\varphi_2 c_2}{\Delta} + \frac{2\dot{\gamma}_3}{r_3} + 2\frac{\dot{\gamma}_6}{\dot{r}_6}. \tag{5.7}$$

Notice that the above two inequalities will be satisfied if we take $c_2$ or $c$ big enough. Here for **(iii)**, we assume that there exist some $r_3 > 0$, $\gamma_3 > 1$ and some constant $C_3 \neq 0$ such that

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

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

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

For **(iv)** and **(vi)**, we assume that for $g = b_{4,N},\ \dot{b}_{4,N}$, there exists some $r_g > 0,\ \gamma_g > 1$ and some constant $C_g \neq 0$ such that

$$\frac{n}{(g)^{r_g}} = \frac{C_g}{N^{\gamma_g}}.$$

These constants do not appear in the main restriction but the fact that the convergence is of the above form is used in the proof of the following theorem.

**Theorem 5.14** *Assume that the constants are chosen so as to satisfy* $c_1 > \frac{2}{c_2}$, *(5.6) and (5.7). 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)** *and* **(H5')**, *then Assumption 2.3* **(6)-(a)** *is satisfied.*

*And also if all other conditions on Assumption 2.3 are satisfied then there exist some positive finite random variables* $\Xi$, $\Xi_1$ *and* $\Xi_2$ *such that*

$$|E_N[f] - f(\theta_0)| \leq \frac{\Xi_1}{\sqrt{N}} \ a.s. \quad and \quad \left|E^n_{N,m}[f] - f(\theta_0)\right| \leq \frac{\Xi_2}{\sqrt{N}} \ a.s.,$$

*and therefore*

$$\left|E_N[f] - E^n_{N,m}[f]\right| \leq \frac{\Xi}{\sqrt{N}} \ a.s.$$

**Remark 5.15** **(i).** *In (5.7),* $r_3$ *and* $\dot{r}_6$ *represent moment conditions on the derivatives of* $X^x_{(m)}(\theta)$, $\varphi_2^{-1}$ *represents the variance of* $X^x_{(m)}(\theta)$, $\Delta$ *represents the length of the time interval between observations. Finally* $c_2 > 2c_1^{-1}$ *expresses a moment condition on* $Y_i$. *In (5.6), recall that* $\dot{q}_6$ *determines a moment condition on* $X^x_{(m)}(\theta)$.

**(ii).** *Roughly speaking, if* $r_3, \dot{r}_6$ *and* $\dot{q}_6$ *are big enough (which implies a condition on n), and we choose* $\alpha_1 > 8\alpha_2 + 1 + \frac{4\varphi_2 c_2}{\Delta}$, $m = \sqrt{N}$, $h = C_2 N^{-\alpha_2}$ *and* $n = C_1 N^{\alpha_1}$, *then Assumption 2.3* **(6)-(a)** *and* **(6)-(b)** *are satisfied. Then conditions contain the main tuning requirements.*

# 6 Appendix

## 6.1 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 independent and identically distributed random variables with $E[X_i] = 0$.

**Lemma 6.1** *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*

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

**Lemma 6.2** *Let* $q_1^{-1} + q_2^{-1} = 1$ *and assume that* $E[|X_i|^{q_1}] < \overline{C}_{q_1}$, *then for all* $0 < \varepsilon < 1$ *and* $\{f_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+$ *satisfying that* $\varepsilon f_n^{-1} \leq K$ *we have*

$$P(|S_n| > n\varepsilon; |X_i| < f_n, \ i = 1, \cdots, n) \leq 2e^{-\frac{n\varepsilon^2}{2f_n^2}} \left\{1 + (q_2 - 1)\left(q_2^{-1} K_1 \frac{\varepsilon}{f_n^2} \overline{C}_{q_1}^{q_1^{-1}} e^{-\frac{\varepsilon^2}{2f_n^2}}\right)^{q_1}\right\}^n. \qquad (6.1)$$

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

# 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, no. 2, 93-128, 1996.

[2] P. Billingsley, *Probability and measure*, John Wiley & Sons, 1979.

[3] P. Billingsley, *Convergence of Probability Measures (Second Edition)*, John Wiley & Sons, 1999.

[4] D. Bosq, *Non-parametric Statistics for Stochastic Processes. Estimation and Prediction (second edition)*, Springer-Verlag, 1998.

[5] J.A. Cano, M. Kessler, D. Salmeron, *Approximation of the posterior density for diffusion processes*, Statist. Probab. Lett. 76, no. 1, 39-44, 2006.

[6] P. Del Moral, J. Jacod, P. Protter *The Monte-Carlo method for filtering with discrete-time observations*, Probab. Theory and Related Fields, 120, 346-368, 2001.

[7] J. Guyon, *Euler scheme and tempered distributions*, Stochastic Process. Appl., 116, no. 6, 877-904, 2006.

[8] A. Kohatsu-Higa, N. Vayatis, K. Yasuda, *Tuning of a Bayesian Estimator under Discrete Time Observations and Unknown Transition Density*, in preparation.

[9] A. Kohatsu-Higa, N. Vayatis, K. Yasuda, *Strong consistency of Bayesian estimator for the Ornstein-Uhlenbeck process*, in preparation.

[10] 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, 163-198, 2002.

[11] M.P. Wand, M.C. Jones, *Kernel Smoothing*, Chapman & Hall, 1995.

[12] K. Yasuda, *Kernel Density Estimation; The Malliavin-Thalmaier formula and Bayesian parameter estimation*, Ph.D. thesis, 2008.