Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process

Charlotte Dion \(^1\) · Sarah Lemler \(^2\)

Received: 24 October 2019 / Accepted: 9 April 2020 / Published online: 7 May 2020
© Springer Nature B.V. 2020

Abstract

We consider a 1-dimensional diffusion process \(X\) with jumps. The particularity of this model relies in the jumps which are driven by a multidimensional Hawkes process denoted \(N\). This article is dedicated to the study of a nonparametric estimator of the drift coefficient of this original process. We construct estimators based on discrete observations of the process \(X\) in a high frequency framework with a large horizon time and on the observations of the process \(N\). The proposed nonparametric estimator is built from a least squares contrast procedure on subspace spanned by trigonometric basis vectors. We obtain adaptive results that are comparable with the one obtained in the nonparametric regression context. We finally conduct a simulation study in which we first focus on the implementation of the process and then on showing the good behavior of the estimator.

Keywords Nonparametric estimator · Model selection · Diffusion · Hawkes process

1 Introduction

Boosted by applications in neuroscience and finance this paper is dedicated to a jump-diffusion model. We focus on a new kind of diffusion with jumps driven by a multidimensional Hawkes process. The process \(X\) of interest is solution of the following equation

\[
dX_t = b(X_t)dt + \sigma(X_t)dW_t + a(X_t-) \sum_{j=1}^{M} dN_{t}^{(j)},
\]

(1.1)

where \(X_{t^-}\) denotes the process of left limits, \(N = (N^{(1)}, \ldots, N^{(M)})\) is a \(M\)-dimensional Hawkes process with (conditional) intensity function \(\lambda\) (that will be detailed in the next section) and \(W\) is the standard Brownian motion independent of \(N\). This process is a diffusion process between the jumps of \(N\).
The first probabilistic properties of this process have been established in Dion et al. (2019). The present paper is part of the continuation of this first work and focuses on the estimation part. Since this model is new, no statistical inference results have yet been carried out on it. In this paper, we focus on the nonparametric estimation of the drift function which leads the dynamic of the trajectories and which is already a real challenge. An estimator obtained from a minimization of a modified least squares contrast is proposed. The difficulty is to control the variance of the estimators taking into account the jumps in order to establish an oracle-type inequality for the empirical risk.

1.1 Motivation

A main motivation is to find a good model to describe neurons membrane potentials between spikes. A spike is defined as an electrical signal that exceeds a certain threshold. Considering one neuron, jump diffusion models are often used to fit the dynamic of the membrane potential of a neuron (see e.g. Jahn et al. 2011). Indeed, according to the literature (see e.g. Höpfner 2007) diffusion processes seem suitable to model the membrane potential of a neuron between successive spikes. Then, Poissonian firing can be added (with a state dependent firing rate) to the diffusion process in order to describe the whole dynamic of the neuron including its own spikes. In the present work, we keep the diffusion model to describe the segments of the membrane potential between spikes of a fixed neuron, but, enriched by a Hawkes jump term, representing the influence of a network of neurons surrounding the one neuron of interest in the study. In fact, we take advantage of both kind of data available due to the biologists. On one hand, we can recorded the membrane potential of one neuron at high frequency over a long time interval (intracellular recording), this signal is continuous. On the other hand, we are also able to record the spike trains occurrence of \( M \) neurons (extracellular recording), which are the sequence of times when the \( M \) neurons spike; we obtain a discrete signal. The network activity is captured in our model (1.1) by the multidimensional Hawkes process. This hybrid model, in which the dynamic of the membrane potential jumps when one of the \( M \) neurons around spikes, allows to combine the wealth of both kind of data by using them all together in a single model.

But we also believe in the usefulness of this new model to describe many other phenomena. In financial mathematics for example, this model could enrich the collection of models for the stochastic volatility of an asset price. The asset price along time \( S_t \) can be modeled by a stochastic differential equation and the volatility coefficient \( \sigma_t \) (see e.g. Renault and Touzi 1996) can be written as \( f(X_t) \) where \( X_t \) satisfies equation (1.1). In this case, the arrival of economic news are well described by the Hawkes process and affects the dynamic of asset prices (see e.g. Rambaldi et al. 2015). In Aït-Sahalia et al. (2015) the authors consider the model they called “Mutually exciting Jump-Diffusion” for the asset log-returns where the jumps are driven by an exponential Hawkes process.

The idea here is different from interaction diffusion models (see Carmona et al. 2013; Gobet and Matulewicz 2016) because it takes into account a continuous signal and a discrete signal all together. For what concerns for example the neuronal application, the data are the jump processes observed from several neurons and the membrane potential of one neuron.

The main contribution of this model is that the Hawkes process has a special time dependence structure: the conditional intensity \( t \mapsto \lambda_t \) is a random predictable function that depends on the past before time \( t \). Nevertheless, we do not assume that the jump intensity of \( N \) at time \( t \) depends on the state of the process \( X_t \).
1.2 State of the art

Diffusions find wide use in the modeling of dynamic phenomena in continuous time. Diffusion processes (without jumps) are Markovian processes, used in applied problems as medical sciences (see e.g. Donnet and Samson 2013), physics (see e.g. Parisi and Sourlas 1992) and financial mathematics (see e.g. Vasicek 1977; El Karoui et al. 1997). Then, jump-diffusion processes are born for example in the risk management context (see Tankov and Viotchkova 2009) driven by a Poisson process or a Lévy process (Masuda 2007; Mancini 2009; Schmisser 2014b). In neuroscience the trajectories of the membrane potential of one fixed neuron have been modeled through different diffusion processes (Höpfner 2007; Jahn et al. 2011; Dion 2014). Many other models exist, taking into account other quantities additionally to the membrane potential (for example the Hodgkin-Huxley model in the recent work Höpfner et al. 2016).

The Hawkes process (Hawkes 1971) comes later analogically to auto-regressive model. It has first been introduced to model earthquakes and their aftershocks in seismology (Vere-Jones and Ozaki 1982). This point process generalizes the Poisson process. It has been used in genomics (Bonnet et al. 2018) and more recently it has been considered to model the intensity of spike trains of neuronal networks (Reynaud-Bouret et al. 2013; Ditlevsen and Löcherbach 2017). In this last context the interaction function are often piecewise constant, see e.g. (Brémaud and Massoulié 1996).

In this last case, it is a multidimensional Hawkes also called mutually exciting point processes that is considered to model the interactions between several neurons (see Brémaud and Massoulié 1996; Delattre et al. 2016, for theoretical studies). This model can deal with observations which represent events associated to agents or nodes on a given network, and that arrives randomly through time but that are not stochastically independent. There is a large literature in the financial side on Hawkes processes, among others recently Jaisson and Rosenbaum (2015), Bacry et al. (2015). Moreover, social networks are nowadays often modeled using a Hawkes process (see e.g. Lukasik et al. 2016). The strength of the multidimensional Hawkes process is mainly the time dependency of each process, which is called sometimes “self-excitation” and modeled by the kernel function. The inference of this process has been investigated parametrically for example in Bacry et al. (2015), nonparametrically in Bacry et al. (2012), Lemonnier and Vayatis (2014), Kirchner (2017) and through neighbor method in Hansen et al. (2015) and Duarte et al. (2016).

In this paper we focus on the case of linear exponential Hawkes process. It has several advantageous properties. First theoretically this choice implies the Markov property for the intensity process of \( N \). Secondly, the expected value of arbitrary functions of the point process is explicit (see Bacry et al. 2015) and it can be exactly simulated (Dassios and Zhao 2013). The allows a clustering representation of the Hawkes process (see Hawkes and Oakes 1974; Reynaud-Bouret and Roy 2007) and non-negativity of the kernel functions eliminates the inhibition behavior. Besides, it is in this context that ergodicity and \( \beta \)-mixing properties have been proved for \( X \) in Dion et al. (2019). The following results are based on these characteristics.

1.3 Main contribution

Statistical inference for diffusions with jumps contains many challenges. Model (1.1) is presented in detail the next section. It is decomposed in three terms: a drift term, a diffusion term and a jump term. The jump term is contained in the multidimensional Hawkes process. The
model have continuous-time dynamic but it has to be inferred from discrete-time observations (whether one or more jumps are likely to have occurred between any two consecutive observation times). We assume to observe the process $X$ at equidistant times with time step $\Delta$ which goes to zero when the number of observations $X_{k,\Delta}$ named $n$ goes to infinity. Besides, the horizon time $T = n\Delta$ goes to infinity with $n$. We also assume to observe the jumps times of $N$ on $[0, T]$.

The main purpose of this paper is to propose a nonparametric estimator for the drift function $b$ in this new framework. There is a few frequentist nonparametric work where the drift function of jump-diffusion processes is estimated (Shimizu and Yoshida 2006; Hoffmann 1999; Mancini 2009; Schmisser 2014a; Amorino and Gloter 2018), but to the best of our knowledge it is the first time that jumps are driven by a Hawkes process. Nevertheless, it is a well known problem for diffusion processes: one can cite for example Bibby and Sørensen (1995) and Kessler and Sørensen (1999) for martingale estimation functions, Gobet et al. (2004) in the low frequency context, Hoffmann (1999), Comte et al. (2007) for least squares contrast estimator.

To make our input in this statistical field, we focus on the nonparametric estimation of the drift function named $b$, when the coupled process $(X, \lambda)$ is ergodic, stationary and exponentially $\beta$-mixing. The proposed method is based on a simple penalized mean square approach. The particularity relies in the contrast function which takes into account the difficulty of the additional Hawkes process. The estimators are chosen to belong to finite-dimensional spaces of variable dimension. An upper bound for the empirical risk of the collection of estimator is obtained. Then, we deal with an adapted selection strategy for the dimension. The final estimator achieves for example the classical rates of convergence for Besov regularity (without logarithmic loss) under some assumptions.

### 1.4 Plan of the paper

First, the model is introduced in Section 2. Then, the estimator is constructed and studied along Sect. 3. A numerical study is presented in Sect. 4. We lead a discussion in Sect. 5. Lastly, the proofs are detailed in Sect. 6. In this last section some more technical lemmas are exposed.

### 2 Framework and assumptions

#### 2.1 The Hawkes process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let us define the Hawkes process on $\mathbb{R}^+$-time via conditional stochastic intensity representation. We denote the $M$-dimensional point process $N_t = (N_t^{(1)}, \ldots, N_t^{(M)})$. Its intensity is a vector of non-negative stochastic intensity functions given by a collection of baseline intensities, which are positive constants $\xi_j, 1 \leq j \leq M$, and $M \times M$ interaction functions $h_{i, j} : \mathbb{R}^+ \rightarrow \mathbb{R}^+, 1 \leq i, j \leq M$, which are measurable functions. Let moreover $n^{(i)}, 1 \leq i \leq M$, be discrete point measures on $\mathbb{R}^-$ satisfying that

$$\int_{\mathbb{R}^-} h_{i, j}(t-s)n^{(i)}(ds) < \infty \text{ for all } t \geq 0.$$ 

We interpret them as initial condition of our process. The linear Hawkes process with parameters $(\xi_i, h_{ij})_{1 \leq i, j \leq M}$ and with initial condition $n^{(i)}, 1 \leq i \leq M$, is a multivariate counting
process $N_t$, $t \geq 0$, such that $\mathbb{P}$—almost surely, for all $i \neq j$, $N^{(i)}$ and $N^{(j)}$ never jump simultaneously, and for all $1 \leq i \leq M$, the compensator of $N_t^{(i)}$ is given by $\Lambda_t^{(j)} := \int_0^t \lambda_s^{(i)} ds$ where

$$\lambda_t^{(j)} = \xi_j + \sum_{i=1}^M \int_{0}^{t-} h_{i,j}(t-u) dN_u^{(i)} + \sum_{i=1}^M \int_{-\infty}^{0} h_{i,j}(t-u) dN_u^{(i)}.$$ (2.1)

Here $N_t^{(j)}$ is the cumulative number of events in the $j$-th component at time $t$ and $dN_t^{(j)}$ represents the number of points in $[t, t+dt]$. The associated measure can also be written $N^{(j)}(dt) = \sum_{k \in \mathbb{N}} \delta_{\tilde{\iota}_k^{(j)}}(dt)$. This intensity process $\lambda$ of the counting process $N$ is the $\tilde{F}_t$-predictable process such that $(\tilde{N}_t = N_t - \Lambda_t)$ is a $\tilde{F}_t$-local martingale with $\tilde{F}_t = \sigma \{ N_s, 0 \leq s \leq t \} = \sqrt{M} \tilde{\iota}_t^{(j)}$ the history of the process $N$ (see Daley and Vere-Jones 2007).

If the functions $h_{i,j}$ are locally integrable, the existence of a process $(N_t^{(j)})_{t \geq 0}$ with prescribed intensity on finite time intervals follows from standard arguments (see e.g. Delattre et al. 2016). Then, $\tilde{\iota}_j$ is the exogenous intensity and we denote $(T_k^{(j)})_{k \geq 1}$ the non-decreasing jump times of the process $N^{(j)}$.

The interaction functions $h_{i,j}$ represent the influence of the past activity of subject $i$ on the subject $j$, the parameter $\xi_j$ is the spontaneous rate and is used to take into account all the unobserved signals. In the following we focus on the exponential kernel functions defined by

$$h_{i,j} : \mathbb{R}^+ \to \mathbb{R}^+, \quad h_{i,j}(t) = c_{i,j} e^{-\alpha t}, \quad \alpha > 0, \quad c_{i,j} > 0, \quad 1 \leq i, j \leq M.$$ (2.2)

The conditional intensity process $(\lambda_t)$ is then Markovian. Before the first occurrence and when $n^{(i)} \equiv 0$ in (2.1), which means that the Hawkes process has empty history., all the point processes $N^{(j)}$ behave like homogeneous Poisson processes with constant intensity $\tilde{\iota}_j$. But as soon as an occurrence appears for a process $N^{(i)}$, then this affects all the process by increasing the conditional intensity via the interaction functions.

### 2.2 General model assumptions

We are now able to write the process as $M + 1$ stochastic equations

$$\begin{align*}
d\lambda_t^{(j)} &= -\alpha (\lambda_t^{(j)} - \xi_j) dt + \sum_{i=1}^M c_{i,j} dN_t^{(i)}, \quad j = 1, \ldots, M \\
\lambda_t^{(j)} &= b(X_t) dt + \sigma(X_t) dW_t + a(X_t) \sum_{j=1}^M dN_t^{(j)}
\end{align*}$$

with $\lambda_0^{(j)}$, $X_0$ random variables independent from the other variables, then we see that $(\lambda_t^{(1)}, \ldots, \lambda_t^{(M)}, X_t)$ is a Markovian process for the general filtration

$${\mathcal{F}}_t = \sigma(\xi_j, W_s, N_s^{(j)}, j = 1, \ldots, M, 0 \leq s \leq t).$$

The process is observed at high frequency on the time interval $[0, T]$, the observations are denoted $X_0, X_\Delta, \ldots, X_{n\Delta}$, with $\Delta \to 0$, $n \to \infty$ and $T := n\Delta \to \infty$, and the jumps times $T_k^{(j)}, j = 1, \ldots, M$ on $[0, T]$.

The size parameter $M$ is fixed and finite all along and asymptotic properties are obtained when $T \to \infty$.

**Assumption 2.1** Assumptions on the coefficients of $X$.

1. $a, b, \sigma$ are globally Lipschitz, and $b$ and $\sigma$ are of class $C^2$.\[\text{\textcopyright} Springer\]
2. There exist positive constants \( c, q \) such that for all \( x \in \mathbb{R} \), \(|b'(x)| + |\sigma'(x)| \leq c\) and \(|b''(x)| + |\sigma''(x)| \leq c(1 + |x|^q)\).

3. There exist positive constants \( a_1 \) and \( \sigma_0, \sigma_1 \), such that \(|a(x)| < a_1\) and \(0 < \sigma_0 < \sigma(x) < \sigma_1\) for all \( x \in \mathbb{R} \).

4. There exist \( d \geq 0, r > 0 \) such that for all \( x \) satisfying that \(|x| > r\), we have \( xb(x) \leq -dx^2\).

Under the three first assumptions, Eq. (1.1) admits a unique strong solution (the proof can be adapted from (Le Gall 2010) under the Lipschitz assumption on function \( a \)). The fourth additional assumption is classical in the study on the longtime behavior of \( X \) (Has’minskii 1980; Veretennikov 1997) and ensure its ergodicity (see Dion et al. 2019).

**Assumption 2.2** Assumptions on the kernels.

1. Let \( H \) be the matrix with entries \( H_{i,j} = \int_0^\infty h_{i,j}(t)\,dt = c_{i,j}/\alpha \), \( 1 \leq i, j \leq M \). The matrix \( H \) has a spectral radius smaller than 1.

2. The offspring matrix \( H \) is invertible.

3. For all \( i, j \in \{1, \ldots, M\}, \alpha \geq c_{i,j} \).

Assumption 2.2.1 implies that \((N_t)\) admits a version with stationary increments (see Brémaud and Massoulié 1996, for e.g.). In the following we always consider that this assumption holds. Also we consider that \((N_t)\) correspond to the asymptotic limit and \((\lambda_t)\) is a stationary process. The second Assumption 2.2.2. is required to ensure the positive Harris recurrence of the coupled process \((X_t, \lambda_t)\) discussed in the next section. The last Assumption is needed to prove Proposition 2.5. For example, under the condition \(\sum_{i,j} c_{i,j}^2 \leq 1/\alpha\) Assumptions 2.2.1. and 3. hold.

Besides, as shown in Bacry et al. (2013), under Assumption 2.2 it comes

\[ \mathbb{E}[\lambda_t] = (I_M - H)^{-1}\xi, \]

where \( I_M \) the identity matrix of size \( M \), \( \xi = (\xi_1, \ldots, \xi_M) \), and for \( \delta > 0 \),

\[ \mathbb{E}[N_{t+\delta} - N_t] = \delta(I_M - H)^{-1}\xi. \] (2.3)

Let us note here that in Dion et al. (2019) a slightly more general model is studied. In particular, the Hawkes process in no longer linear and the exponential kernel functions are allowed to be positive or negative (depending on the sign of \( c_{i,j} \)).

### 2.3 First results on the model

We refer to the paper Dion et al. (2019) for the proofs of the following probabilistic results.

**Theorem 2.3** (Dion et al. 2019) Under Assumptions 2.1, 2.2 the process \((X_t, \lambda_t)_{t \geq 0}\) is positive Harris recurrent with unique invariant measure \( \pi \). In particular, for any starting point \((x, y)\) and any positive measurable function \( g : \mathbb{R} \times \mathbb{R}^M \to \mathbb{R}_+ \), as \( T \to \infty \), \( \mathbb{P}_{(x,y)} \)-almost surely,

\[ \frac{1}{T} \int_0^T g(X_t, \lambda_t)\,dt \to \pi(g), \]

where \( \pi(g) = \int g(u, v)\pi(du, dv) \).

\( \mathbb{C} \) Springer
Moreover, in Dion et al. (2019), the proven Foster–Lyapunov type condition implies that for all \( t \geq 0 \), \( \mathbb{E}[X_t^4] < \infty \).

**Assumption 2.4** \((X_0, \lambda_0)\) has distribution \( \pi \).

The process \((X_t, \lambda_t)\) is then in its stationary regime in the following. Note that \((X_t)\) is non-Markovian but invariant anyway.

Recall that the \( \beta \)-mixing coefficient of the stationary Markovian process \((Z_t) = (X_t, \lambda_t)\) is given by

\[
\beta_{Z}(t) = \int \|P_t(z, \cdot) - \pi\|_{TV} \pi(dz)
\]

where \( \|\mu\|_{TV} := \sup_{g \colon |g| \leq 1} \mu(g) \). Moreover, \( \beta_{X}(t) = \int_{\mathbb{R} \times \mathbb{R}^M} \|P^1_t(z, \cdot) - \pi^X\|_{TV} \pi(dz) \)

where \( P^1_t(z, \cdot) \) is the projection on \( X \) of \( P_t(z, \cdot) = P_t(z, dx \times \mathbb{R}^M) \), similarly \( \pi^X(dx) = \pi(dx \times \mathbb{R}^M) \) is the projection on the coordinate \( X \). Then, under the previous Assumptions 2.1, 2.2 and 2.4, according to Dion et al. (2019) Theorem 4.9 the process \((Z_t) = (X_t, \lambda_t)\) is exponentially \( \beta \)-mixing and there exist some constants \( K, \theta > 0 \) such that

\[
\beta_{X}(t) \leq \beta_{Z}(t) \leq Ke^{-\theta t}. \tag{2.4}
\]

The following result is analogous to Proposition A obtained in Gloter (2000). It is very useful for the following estimation part.

**Proposition 2.5** Grant Assumptions 2.1, 2.2 and 2.4. For all \( t \geq 0, 0 < \delta < 1 \), there exists a constant \( C(M, a_1) > 0 \) such that for \( p = 2 \) or \( p = 4 \) we have

\[
\mathbb{E} \left[ \sup_{s \in [t, t+\delta]} |X_s - X_t|^p \right] \leq C(M, a_1)\delta.
\]

The proofs are relegated in Sect. 6. This result is not surprising. Indeed, the jumps part of the process is responsible of the degradation of the result that one can obtain for a diffusion process which would have been a control of order \( \delta^{p/2} \).

The following section is dedicated to the estimation of the function \( b \). We assume that the coefficient \( a \) is known and the function \( b \) is unknown.

### 3 Estimation procedure of the drift function

The functions \( b \) is estimated only on a compact set \( A \) of \( \mathbb{R} \). It follows from Proposition 4.7 of Dion et al. (2019) that \( \pi^X \) the projection of the invariant measure \( \pi \) onto the \( X \)–coordinate is bounded from below on any compact subset of \( \mathbb{R} \), and we denote:

\[
0 < \pi_0 \leq \pi^X(x). \tag{3.1}
\]

Note that \( \pi^X \) is also bounded from above if \( a \equiv 0 \) or if \( h_{i,j} \equiv 0 \) for all \( i, j \). Here if \( b \) is bounded then the result also holds (see details in Dion et al. 2019, it comes from the bounds of the transition densities given in Gobet (2002)). Moreover, we denote: \( \|\cdot\|_{\pi^X} \) the \( L^2 \)-norm of \( L^2(A, \pi^X(x) dx) \).
3.1 Non-adaptive estimator and space of approximation

From the observations, the following increments are available:

\[ Y_{k\Delta} := \frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta} = b(X_{k\Delta}) + I_{k\Delta} + Z_{k\Delta} + T_{M,k\Delta} \]  

(3.2)

with

\[ I_{k\Delta} := \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta})) ds, \quad Z_{k\Delta} := \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s, \]  

(3.3)

\[ T_{M,k\Delta} := \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} a(X_s-) \sum_{j=1}^{M} dN_s^{(j)}. \]  

(3.4)

Considering the Doob-Meyer decomposition, the \( T_{M,k\Delta} \) term can be decomposed in

\[ T_{M,k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} a(X_s-) \sum_{j=1}^{M} \lambda_s^{(j)} ds + \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} a(X_s-) \sum_{j=1}^{M} d\tilde{N}_s^{(j)}. \]

Following Hansen et al. (2015) the second term of this decomposition can be seen as a “noise” term. In decomposition (3.2) the first term is the term of interest. The second term is negligible, the third term is a centered noise term. But the fourth term \( T_{M,k\Delta} \) is then not centered and not negligible. The real regression-type equation is actually:

\[ Y_{k\Delta} - T_{M,k\Delta} = b(X_{k\Delta}) + I_{k\Delta} + Z_{k\Delta}. \]

Based on these variables, we propose a nonparametric estimation procedure for the drift function \( b \) on a closed interval \( A \). We consider \( S_m \) a linear subspace of \( L^2(A) \) such that \( S_m = \text{span}(\varphi_1, \ldots, \varphi_{D_m}) \) of dimension \( D_m \) where \( (\varphi_k) \) is an orthonormal basis of \( L^2(A) \). We denote \( S_n = \bigcup_{m \in M_n} S_m \) where \( M_n \) is a set of indexes for the model collection. The contrast function is defined for \( t \in S_n \), by

\[ \gamma_{n,M}(t) := \frac{1}{n} \sum_{k=1}^{n} (U_{k\Delta} - t(X_{k\Delta}))^2, \quad U_{k\Delta} := Y_{k\Delta} - T_{M,k\Delta}. \]  

(3.5)

The associated mean squares contrast estimator is

\[ \hat{b}_m := \arg\min_{t \in S_m} \gamma_{n,M}(t). \]  

(3.6)

Let us comment this nonparametric mean squares contrast \( \gamma_{n,M} \). In the literature of nonparametric frequentist drift estimation from discrete data (Hoffmann 1999; Comte et al. 2007) the regression is done on the random variables \( Y_{k\Delta} \)’s. We consider rather a regression on \( U_{k\Delta} \) which depends on \( T_{M,k\Delta} \) to take into account the jumps of the process. This new contrast leads to a better approximation of the coefficient \( b \) according to formula (3.2) when the coefficient \( a \) is known. For example in the neuronal context presented in Introduction, \( a \) can be assumed to be linear at first to represent that the impact of the neuronal network surrounding the neuron of interest is linear in the position of it. Then, as the jump process is observed and also \( X \) at discrete times, the coefficient \( T_{M,k\Delta} \) can be approximated.
Remark 3.1 The estimator \( \hat{b}_m \) is defined by \( \sum_{\ell=1}^{D_m} \hat{a}_\ell \psi_\ell \) and \( \alpha \) is the solution of the equation \( \Phi_m^t U = (\Phi_m \Phi_m) \alpha \) where \( \Phi_m \) is the Gram matrix \( \Phi_m = (\phi_j(X_i))_{i,j=1,\ldots,n; \ j=1,\ldots,D_m} \) of size \( n \times D_m \) and \( U = (U_\Delta, U_{2\Delta}, \ldots, U_{n\Delta}) \). Thus, the solution may not be unique if the kernel of \( \Phi_m \) is not reduced to \( \{0\} \). Nevertheless, the sequence \( (\hat{b}_m(X_\Delta), \ldots, \hat{b}_m(X_{n\Delta})) \) is unique and it is this sequence that we take into account in the following empirical risk.

The spaces of approximation \( S_m \) must satisfy the following key properties.

Assumption 3.2 Assumptions on the subspaces.
1. \( \exists \phi_1 > 0, \forall t \in S_m, \|t\|_{\infty}^2 \leq \phi_1 D_m \|t\|^2 \)
2. The spaces \( S_m \) are nested spaces of dimension \( D_m \leq D_n \) such that there exists \( S_n \) of dimension \( D_n \) such that \( S_m \subset S_{m+1} \subset S_n \) for all \( m \in \mathcal{M}_n \).
3. There exists a finite positive constant \( \phi_{\pi X} \) such that

\[
\phi_{\pi X} := \sup_{\pi \in \mathcal{M}_n} \int_A \|\phi_j(x)\|_{\pi X} \, dx = \sup_{j=1,\ldots,D_n} \|\phi_j\|_{\pi X}^2.
\]

Note that Assumption 3.2.3 is automatically true for example for the trigonometric basis (also called Fourier basis).

3.2 Risk bound

The next paragraph is dedicated to the main result obtained on the non-adaptive estimator of the drift function on \( A \): \( \hat{b}_m \) for all model \( m \). We consider the empirical risk \( \mathbb{E}[\|t - b\|_{\pi n}^2] \) define through the following empirical squared norm:

\[
\|t\|_{\pi n}^2 := \frac{1}{n} \sum_{k=1}^{n} t^2(X_{k\Delta}).
\]

Proposition 3.3 Under Assumptions 2.1, 2.2, 2.4 and 3.2, if \( \Delta \to 0 \) and \( (n\Delta)/(\ln^2(n)) \to \infty \) when \( n \to \infty \), if \( D_n \leq O(\sqrt{n}/\ln(n)) \), then, the estimator \( \hat{b}_m \) of \( b \) on \( A \) given by Eq. (3.6) satisfies

\[
\mathbb{E}[\|\hat{b}_m - b\|_{\pi n}^2] \leq 13 \inf_{t \in S_m} \|b - t\|_{\pi X}^2 + C_1 \frac{D_m}{n\Delta} + C_2 \Delta + \frac{C_3}{n\Delta}
\]

with \( C_1 \) depending on \( \phi_1, \sigma_1, \) and \( C_2 \) depending on \( \sigma_1, C_3 \) depending on \( a_1, \sigma_1, M \).

Let us precise here that the notation \( b \) is for \( b1_A \) as \( b_m \) and \( \hat{b}_m \) are \( A \)-supported. The upper bound of Proposition 3.3 is decomposed into different types of error. The first term is the bias term which naturally decreases with the dimension of the space of approximation \( D_m \). The second term is the variance term or the estimation error which increases with \( D_m \). The third and fourth terms come from the discretization error.

Moreover, if \( \pi X \), is bounded from above by \( \pi_1 \) on \( A \), then we can provide the classical nonparametric rate of convergence depending only on \( n, \Delta \). Indeed, let us assume that \( b \) belongs to a Besov ball denote \( B_{\alpha,2,\infty}(A) \) (see the reference book DeVore and Lorentz 1993), where \( \alpha \) measures the regularity of the function. Then, taking \( t = b_m \) the projection of \( b \) on \( S_m \), \( \|b - b_m\|^2 \leq C(\alpha, L)D_m^{-2\alpha} \), and if \( n\Delta^2 \to 0 \) and choosing \( D_m = (n\Delta)^{1/(2\alpha+1)} \) leads to

\[
\mathbb{E}[\|\hat{b}_m - b\|_{\pi n}^2] = O((n\Delta)^{-2\alpha/(2\alpha+1)}).
\]
This is the optimal nonparametric rate of convergence in the case of estimation of the drift without the additional jumps coming from the multivariate Hawkes process (see Hoffmann 1999).

**Remark 3.4** Besides, let us emphasized that the result is expressed for the expectation of the empirical norm because of the difficulty of the problem. Indeed, as only one observation of the process is available on $[0, T]$, it is tricky to obtain theoretical guarantees for the expectation of the integrated norm. Indeed, we are not exactly in the regression context because the observations $(X_k \Delta, U_k \Delta)k$ are dependent. Thus, the consistency results are obtained for the empirical norm as in Baraud et al. (2001a, b); Comte et al. (2007). While, in the (independent) regression context, a truncated version of the estimator is convergent of the integrated norm (see Baraud 2002 or Györfi et al. 2006 Chapter 12).

Nevertheless, if we allow ourselves to make repeated observations of the process (which is beyond the scope of the paper), then, shorting the collection of models, might provide the consistency of the estimator for an integrated norm (see the recent work Comte and Genon-Catalot 2019).

The objective of the following section is to obtain an oracle-type inequality for the empirical risk for a final estimator where the dimension is chosen through a data-driven procedure.

### 3.3 Adaptation procedure

We must define a criterion in order to select automatically the best dimension $D_m$ (the best model) in the sense of the empirical risk. This procedure should be adaptive, meaning independent of $b$ and dependent only on the observations. The final chosen model minimizes the following criterion

$$
\hat{m} := \arg\min_{m \in \mathcal{M}_n} \left\{ \gamma_{n, m}(\hat{b}_m) + \text{pen}(m) \right\}
$$

(3.7)

with $\text{pen}(\cdot)$ the increasing function on $D_m$ given by

$$
\text{pen}(m) := \rho \sigma_2^2 \frac{D_m}{n \Delta},
$$

(3.8)

where $\rho$ is a universal constant that has to be calibrated for the problem (see Sect. 4).

**Theorem 3.5** Under Assumptions 2.1, 2.2, 2.4 and 3.2, if $\Delta \to 0$ and $(n\Delta)/(\ln^2(n)) \to \infty$ when $n \to \infty$, if $D_n \leq O(\sqrt{n\Delta}/\ln(n))$, then, for the final estimator $\hat{b}_m$ of $b$ on $A$ given by (3.6) and (3.7), there exists $\rho_0$ such that for $\rho \geq 7 \rho_0$,

$$
\mathbb{E}[\|\hat{b}_m - b\|^2_n] \leq C \inf_{m \in \mathcal{M}_n} \left\{ \inf_{t \in S_m} \|b - t\|^2_{X} + \text{pen}(m) \right\} + C' \left( \frac{1}{\Delta n} + \Delta \right),
$$

where $\text{pen}(\cdot)$ is given by (3.8), and $C$ is a numerical constant independent from $n$ and $D_m$ and $C'$ depends on the parameters of the model.

The result teaches us that the estimator $\hat{b}_m$ realizes automatically the best compromise between the bias term and the penalty term which has the same order than the variance term. Again a similar result is obtained for diffusion processes in Comte et al. (2007) which confirms that when the jump process is observed the estimation of the drift function is not modified.
4 Numerical experiments

In this section we describe the protocol of simulations which has been used to evaluate the presented methodology.

4.1 Implementation of the process and chosen examples

We simulate the Hawkes process \(N\) with \(M = 2\) and here we denote \((T_k)_k\) the sequence of jump times of the aggregate process (all the jump times of \(N\) sorted and gathered). For this simulation we use a Thinning method. One could also use the \texttt{R}-package \texttt{hawkes} for classical kernels or the package \texttt{tick} for Python language. Also, an exact simulation procedure is possible for the exponential Hawkes (see e.g. Dassios and Zhao 2013). The intensity process is written as

\[
\lambda_t^{(j)} = \xi_j + (\lambda_0^{(j)} - \xi_j) e^{-\alpha t} + \sum_{T_{k}^{(j)} < t} c_{j,j} e^{-\alpha(t-T_{k}^{(j)})} + \sum_{i \neq j} \sum_{T_{k}^{(i)} < t} c_{i,j} e^{-\alpha(t-T_{k}^{(i)})}.
\]

The initial conditions \(X_0, \lambda_0\) should be simulated according to the invariant distribution (and \(\lambda_0^{(j)}\) should be larger than \(\xi_j\)). This measure of probability is not explicit. Thus we choose: \(\lambda_0^{(j)} = \xi_j\) and \(X_0 = 0\). Also, the exogenous intensities \(\xi_j\) are chosen equal to 0.5 for \(j = 1, 2\). The weight matrix \(c\) is chosen as:

\[
c = \begin{pmatrix} 0.2 & 0.3 \\ 0.5 & 0.4 \end{pmatrix}
\]

and \(\alpha = 5\) (then the spectral radius of \(H\) is nearly 0.02 < 1). Figure 1 shows one simulation of the Hawkes process for \(T = 10\). The graph on the left represents the jump times for the two components and the graph on the right represents the sum process of the intensities \(\lambda_t^{(1)} + \lambda_t^{(2)}\).

![Figure 1](image.png)
Then we simulate \((X_Δ, \ldots, X_{(n+1)Δ})\) from an Euler scheme. Indeed because of the additional term (when \(a \neq 0\)) to the best of our knowledge, it is not possible to use classical more sophisticated scheme. Finally we follow the following steps:

1. Simulate \(N^{(1)}, N^{(2)}\) on \([0, T]\) and deduce the jump times from the thinning method.
2. Create a large vector containing the grid of the \(n\) times with time step \(Δ\) and the jump times, sorted.
3. Simulate the process at these new times through an Euler scheme:
   \[
   X_{tk} = (t_k - t_{k-1})b(X_{tk}) + \sqrt{t_k - t_{k-1}}σ(X_{tk})W + a(X_{tk})1_{tk=jump time}.
   \]
4. Keep only in the grid the points corresponding to the regular grid of time step \(Δ\) without considering the jump times.

A simulation algorithm is also detailed in Dion et al. (2019) Section 2.3.

We choose to show only relevant examples of processes with different types of drift function. There are four:

1. Model 1: (Ornstein Uhlenbeck diffusion) \(b(x) = -2x, σ(x) = 1\), with \(a(x) = x1_{[-5,5]}(x)\).
2. Model 2: \(b(x) = -(x - 1/4)^3 - (x + 1/4)^3, σ(x) = 1\), \(a(x) = x1_{[-5,5]}(x)\).
3. Model 3: \(b(x) = -2x + \sin(3x), σ(x) = \sqrt{3x^2 + 1^2}, a(x) = x1_{[-5,5]}(x)\).
4. Model 4: \(b(x) = -2x, σ(x) = 1\), \(a(x) = 0.2x\).

Model 1 is the simplest one. The second model do not satisfy the assumptions because in particular \(b\) is not Lipschitz. The process is not ergodic and is really unstable. In order to obtain a result in this case we kept only the trajectories that do not explode in finite time. (This may change the distribution of the process thus this is only an illustration). The third model has drift function similar to model 1, but with a larger diffusion coefficient. Finally, model 4 illustrate a case where the function \(a\) is not positive and change sign. This last case allows an unbounded coefficient \(a\).

Assumption 2.1.2. is particularly strong and not verified in practice. Nevertheless it is true on the compact of estimation.

### 4.2 Numerical results for the estimation of \(b\)

The results are illustrated for the trigonometric basis on the compact \(A = [-1, 1]\). (Let us precise that the simulated process take values in a larger interval and we bring the values back over the interval \([-1, 1]\) from the random interval \([\min(X), \max(X)]\)). Then, the regression is done on the random variables \(U_{kΔ} = Y_{kΔ} - T_{M,kΔ}\). This requires to compute the term \(T_{M,kΔ}\) that we remind the reader is given by

\[
T_{M,kΔ} = \frac{1}{Δ} \int_{kΔ}^{(k+1)Δ} a(X_{s-}) \sum_{j=1}^{M} dN_s^{(j)}.
\]

This term is thus approximated by

\[
\sum_{j=1}^{M} (N_{(k+1)Δ}^{(j)} - N_{kΔ}^{(j)}) \quad \text{(as} a \text{ is assumed to be known and)}
\]

\[
N_{(k+1)Δ}^{(j)} - N_{kΔ}^{(j)} \text{ is the number of jumps of component number} j \text{ on interval} [kΔ, (k + 1)Δ]\]

\((which \ is \ observed)\). Figure 2 illustrates the gain considering variables \(U_{kΔ} = Y_{kΔ} - T_{M,kΔ}\) instead of \(Y_{kΔ}\). For model 1 and model 3 (top and bottom) the left graph is \(Y_{kΔ}\) versus \(X_{kΔ}\) together with the drift function \(b\) in plain line, and on the right \(U_{kΔ}\) versus \(X_{kΔ}\). As attended, one can notice that the right graphs are more suited for regression, the points follows the curve of the true function \(b\).
Fig. 2 At the left, $Y_{k,\Delta}$ versus $X_{k,\Delta}$, at the right, $U_{k,\Delta}$ versus $X_{k,\Delta}$. On the top, model 1, at the bottom, model 3. The true function $b$ is in plain line. Simulations for $n = 1000$ and $\Delta = 1/10$.

Then, we have implemented the least squares estimator $\hat{b}_m$ with the trigonometric basis. The dimension $D_m = 2m + 1$ with $m \in \{1, \ldots, m_{\text{max}}\}$ with $m_{\text{max}} = 20$ on simulations. The adaptive procedure given in formula (3.7) and (3.8) is implemented with the true value $\sigma_1$ (the bound of function $\sigma$ is known). The calibration constant $\rho$ in the penalty function given in Eq. (3.8) $\text{pen}(m) = \rho \sigma^2 D_m/(n\Delta)$ is calibrated on a large preliminary simulation study where we investigate various models with known parameters and let $\rho$ vary. We show for example on Fig. 3 one of the models on which the calibration is done. Here $b(x) = 0.1(-\sin(2\pi x) + \cos(2\pi x) + 16\sin(3\pi x) - 5\cos(3\pi x))$, $\sigma(x) = 1$, $a(x) = x \mathbb{1}_{[-5,5]}(x)$, and $\rho$ evolves in the grid $\{0.1, 1, 2, 3, 4, 5, 6\}$. We have represented for each couple $n, \Delta$ the graph of the Monte-Carlo estimator of the empirical risks $\mathbb{E}[\|\hat{b}_m - b\|_n^2]$ as a function of $\rho$. We see that $\rho = 3$ is the best constant in most cases. This constant is chosen equal to 3 in the following. Figure 4 shows for model 2 a collection of estimators $\hat{b}_m$ in (green) dotted line, together with the true function $b$ in dark plain line and the chosen estimator in red bold line. Finally, we have computed a Monte-Carlo approximation of the empirical risks from 1000 simulations. We show the results for $\Delta \in \{1/10, 1/100\}$ with $n \in \{1000, 10000\}$. On Table 1 one can see the influence of the parameters $n$ and $\Delta$ on the empirical risk, and also the impact of the model. When $\Delta$ is multiplied by 10 and $n$ is fixed the risk is also multiplied by 10.
Fig. 3 Calibration of $\rho$: example with $b(x) = 0.1(-\sin(2\pi x) + \cos(2\pi x) + 16\sin(3\pi x) - 5\cos(3\pi x))$, $\sigma(x) = 1$, $a(x) = x\mathbb{1}_{[-5,5]}(x)$, and $\rho$ evolves in the grid $\{0.1, 1, 2, 3, 4, 5, 6\}$.

Fig. 4 Collection of estimators $\hat{b}_m$ in (green) dotted line, true function $b$ in dark plain line and final estimator $\hat{b}_m$ in red bold line for Model 2. (Color figure online)
Table 1  Empirical risk $\mathbb{E}[\|\hat{b}_m - \hat{b}\|^2_n]$ from 1000 repetitions of Monte-Carlo on $[-1, 1]$  

| $n, \Delta$ | $n = 1000$ | $n = 10000$ | $n = 1000$ | $n = 10000$ |
|-------------|-------------|-------------|-------------|-------------|
| $\Delta = 1/10$ | $0.068$ | $0.011$ | $1.917$ | $0.061$ |
| $\Delta = 1/100$ | $0.080$ | $0.007$ | $2.948$ | $0.104$ |
| $\Delta = 1/1000$ | $0.161$ | $0.014$ | $4.528$ | $0.154$ |
| $\Delta = 1/10000$ | $0.058$ | $0.005$ | $4.398$ | $0.065$ |

thus as one can attend, when $\Delta$ decreases the error decreases. But if $T = n\Delta$ is unchanged the influence of $\Delta$ is less clear. It is in line with the result of Theorem 3.5.

The most difficult case is $T = 10$. Of course in this case the process may not be stationary yet as we do not start from the invariant distribution. Model 1 shows the best results. Indeed the model is simple and satisfies all the assumptions. Nevertheless the other results are not so far. One can compare for example with the results obtain for the estimation of the drift in the classical jump-diffusion in Schmisser (2014a), the results are very close.

5 Discussion

In this paper, we provide a nonparametric estimator of the drift function in a diffusion process with jumps driven by a Hawkes process. Section 3 is devoted to theoretical guarantees for the non-adaptive and adaptive estimator. Section 4 deals with the implementation of the process and of the estimator.

In future works, we plan to generalize the obtained results. There are many remaining questions of interest. The case of an unknown jump coefficient $a(\cdot)$ is obviously a crucial issue. Indeed, the knowledge of $a(\cdot)$ is used for the computation of $T_M$. Also, the presented adaptive estimator requires the knowledge of $\sigma_1^2$ (see Eq. (3.8)). The relation

$$
\frac{(X_{(k+1)\Delta} - X_{k\Delta})^2}{\Delta} = \sigma^2(X_{k\Delta}) + a^2(X_{k\Delta}) \left( \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} \sum_{j=1}^{M} \phi_j(x) ds \right) + \text{terms}
$$

could lead to some estimation of $\sigma^2$, $a$, $\sigma^2 + a$ (or at least in the parametric case). Nevertheless it is not a simple issue because the additional terms here are not centered and are hard to control. It will be investigated in future works.

One may also think of generalizing the initial model. For instance, the case of unknown kernel functions $h$ for the Hawkes process is of interest. Indeed, as several nonparametric estimators of $\lambda$ are available (mainly Bacry et al. 2012; Hansen et al. 2015) an estimator of $\lambda$ could be plug-in the procedure of the estimation of the coefficients.

Finally, this estimation procedure will be tested on a neuronal dataset and put in competition with classical methods of machine learning. This is a work in progress.

6 Proofs

We drive the reader attention on the fact that through the proofs the constants are generically noted $C, C'$... for a strictly positive, real-valued constant, not necessarily the same.
Sometimes, the notation $C(k)$ is used to emphasize that the constant depends on $k$ in some unspecified manner.

### 6.1 Preliminary result

Let us remind the reader a result obtained in Bacry et al. (2013), named Lemma 4 and Lemma 6.

**Lemma 6.1** For all $\delta > 0$,

1. There exists a positive constant $C$ such that
   \[
   \sum_{j=1}^{M} \mathbb{E}[(N_{t+\delta}^{(j)} - N_{t}^{(j)})^2] \leq C(\delta + \delta^2), \tag{6.1}
   \]

2. and there exists a positive constant $C$ such that
   \[
   \sum_{j=1}^{M} \mathbb{E}[(N_{t+\delta}^{(j)} - N_{t}^{(j)})^4] \leq C(\delta + \delta^2 + \delta^4). \tag{6.2}
   \]

**Proof of Lemma 6.1** (i) Following Bacry and Muzy (2014), denoting $A_t^{(j)} := N_t^{(j)} - \mathbb{E}[N_t^{(j)}]$, it comes

\[
\sum_{j=1}^{M} \mathbb{E}[(N_{t+\delta}^{(j)} - N_{t}^{(j)})^2] \leq 2 \sum_{j=1}^{M} \mathbb{E}[(A_{t+\delta}^{(j)} - A_{t}^{(j)})^2] + (\mathbb{E}[(N_{t+\delta}^{(j)} - N_{t}^{(j)})^2]
\]

\[
\leq 2\|\mathbb{E}[N_{t+\delta} - N_t]\|^2 + 2\mathbb{E}[\|A_{t+\delta} - A_t\|^2],
\]

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^M$. The first term is controlled by Eq. (2.3): $\mathbb{E}[N_{t+\delta} - N_t] = \delta(t_H - H)^{-1}\xi$ where $\xi$ is the M-vector of $\xi_j$'s and $H = (c_{j,k}/\alpha)_{j,k \in \{1, \ldots, M\}}$, it is of order $\delta^2$. The second term is controlled by Lemma 6 of Bacry et al. (2013) and is of order $\delta$. There are constants that depend on $c_{j,k}, \alpha, \xi_j$.

(ii) The same notations as in Bacry and Muzy (2014) are used:

\[
A_t^{(j)} = N_t^{(j)} - \mathbb{E}[N_t^{(j)}] = \tilde{N}_t^{(j)} + \int_0^t \sum_{i=1}^{M} \psi_{ij} \tilde{N}_s^{(i)} ds,
\]

where $\tilde{N}_t^{(j)} = N_t^{(j)} - \int_0^t \lambda_t^{(j)}$ is a local martingale, $\psi_{ij} = \sum_{n \geq 1} h_{i n}^{(j)}$ and $h_{i n}^{(j)}$ is such that $h_{n+1}^{(j)}(t) = \int_0^t h_{ij}(t-s) h_{s}^{(j)}(s) ds$ and $h_{1}^{(j)} = h_{ij}$. With these notations, we have for a constant $C$ that may change from a line to another

\[
\mathbb{E}[(N_{t+\delta}^{(j)} - N_{t}^{(j)})^4] \leq C\mathbb{E}[(A_{t+\delta}^{(j)} - A_{t}^{(j)})^4] + C\mathbb{E}[(N_{t+\delta}^{(j)} - N_{t}^{(j)})^4].
\]

From Eq. (2.3), we have $\mathbb{E}[(N_{t+\delta}^{(j)} - N_{t}^{(j)})^4] \leq C\delta^4$. The second term in the previous inequality can be controlled as follows

\[
\mathbb{E}[(A_{t+\delta}^{(j)} - A_{t}^{(j)})^4] \leq C\mathbb{E}[(\tilde{N}_{t+\delta}^{(j)} - \tilde{N}_{t}^{(j)})^4] + C\mathbb{E}[(\int_0^{+\infty} \sum_{i=1}^{M} \psi_{ij}(u)(\tilde{N}_{t+\delta}^{(j)} - \tilde{N}_{t}^{(j)}) du)^4].
\]
From Bacry and Muzy (2014), Lemma 9, we have

$$\mathbb{E}\left[\sup_{t \leq s \leq t+\delta} (\tilde{N}_s^{(j)} - \tilde{N}_t^{(j)})^4 | \mathcal{F}_t \right] \leq C(\delta + \delta^2).$$

We deduce that

$$\mathbb{E}[(A_{t+\delta}^{(j)} - A_t^{(j)})^4] \leq C(\delta + \delta^2) + C \int_0^{+\infty} \left( \sum_{i=1}^M \psi_{ij}(u) \right)^4 \mathbb{E}[|N_{t+\delta-u}^{(j)} - N_{t-u}^{(j)}|^4] du \leq C(\delta + \delta^2).$$

The last inequality holds due to the fact that following Gloter (2000) Proposition 5.1, we first show that for $6.2$ Proof of Proposition 2.5

Following Gloter (2000) Proposition 5.1, we first show that for $p = 2$ or $p = 4$, $\mathbb{E}\left[\sup_{s \in [t, t+\delta]} |X_s|^p | \mathcal{F}_t \right] \leq c(p)(1 + |X_t|^p)$. Choose $s \in [t, t+\delta]$ with $\delta < 1$. Then,

$$X_s - X_t = \int_t^s b(X_u) du + \int_t^s \sigma(X_u) dW_u + \sum_{j=1}^M \int_t^s a(X_{u-}) dN_u^{(j)}.$$ 

Using Burkholder–Davis–Gundy inequality, the fact that $\sigma$ and $b$ are Lipschitz and that $a$ is bounded, we have for $p = 2$ or $p = 4$

$$\mathbb{E}\left[\sup_{s \in [t, t+\delta]} |X_s|^p | \mathcal{F}_t \right] \leq c(p)|X_t|^p + c(p)\mathbb{E}\left[\left( \int_t^s \sigma^2(X_u) du \right)^{p/2} | \mathcal{F}_t \right]$$

$$+ c(p)\mathbb{E}\left[\left( \int_t^s |b(X_u)| du \right)^p | \mathcal{F}_t \right] + c(p)((\sum_{j=1}^M \int_t^s a(X_{u-}) dN_u^{(j)})^p | \mathcal{F}_t \right)$$

$$\leq c(p)|X_t|^p + c(p)\mathbb{E}\left[\left( \int_t^s (|\sigma(X_u)|^p + |b(X_u)|^p) du \right)^p | \mathcal{F}_t \right]$$

$$+ c(p, M) a_1^p \sum_{j=1}^M \mathbb{E}[(S_t^{(j)} - N_t^{(j)})^p | \mathcal{F}_t]$$

$$\leq c(p)|X_t|^p + c(p)\mathbb{E}[1 + |X_t|^p]$$

$$+ c(p, M) a_1^p \sum_{j=1}^M \mathbb{E}[(N_t^{(j)} - S_t)^p | \mathcal{F}_t]$$

$$+ c(p, M) a_1^p \sum_{j=1}^M \mathbb{E}[(N_t^{(j)} - N_t^{(j)})^p | \mathcal{F}_t]$$

$$\leq c(p)|X_t|^p + c(p)\mathbb{E}[1 + |X_t|^p]$$

$$+ c(p, M) a_1^p \sum_{j=1}^M \mathbb{E}[(N_t^{(j)} - N_t^{(j)})^p | \mathcal{F}_t].$$
where the constants may change from a line to another. We then introduce the following notation: \( \phi(s) \leq \sup_{u \in [t,s]} \mathbb{E}[|X_u|^p|\mathcal{F}_t] \). We have with this notation

\[
\phi(s) \leq c(p, M)(1 + |X_t|^p + a_1^p \sum_{j=1}^{M} \mathbb{E}[(N_s^{(j)} - N_t^{(j)})^p|\mathcal{F}_t]) + c(p) \int_t^s \phi(u)du.
\]

From Assumption 2.1.1, \( \phi(s) \) is almost surely finite. We can apply Gronwall’s Lemma (see in “Appendix” Lemma 7.1) to obtain

\[
\phi(s) \leq c(p, M) \left(1 + |X_t|^p + a_1^p \sum_{j=1}^{M} \mathbb{E}[(N_s^{(j)} - N_t^{(j)})^p|\mathcal{F}_t]\right) + c(p)(1 + |X_t|^p)(s - t) + c(p, M)a_1^p \int_t^s \sum_{j=1}^{M} \mathbb{E}[(N_u^{(j)} - N_t^{(j)})^p|\mathcal{F}_t]e^{c(p)(t-u)}du.
\]

Since \( s \in [t, t + \delta] \) and \( \delta < 1 \), applying Lemma 6.1, we obtain

\[
\sup_{u \in [t, t + \delta]} \mathbb{E}[|X_u|^p|\mathcal{F}_t] \leq c(p)(1 + |X_t|^p). \tag{6.3}
\]

Now, we consider \( \Delta_{t, t + \delta} := \sup_{s \in [t, t + \delta]} |X_s - X_t| \). As previously, we have

\[
\mathbb{E}[(\Delta_{t, t + \delta})^p|\mathcal{F}_t] \leq c(p) \mathbb{E} \left[ \left( \int_t^{t + \delta} \sigma^2(X_u)du \right)^{p/2} \big| \mathcal{F}_t \right] + 2 \mathbb{E} \left[ \left( \int_t^{t + \delta} |b(X_u)|du \right)^p \big| \mathcal{F}_t \right] + c(p, M, a_1) \sum_{j=1}^{M} \mathbb{E} \left[ (N_{t + \delta}^{(j)} - N_t^{(j)})^p \big| \mathcal{F}_t \right]
\]

\[
\leq c(p) \mathbb{E} \left[ \delta^{p/2} \sup_{t \leq s \leq t + \delta} \sigma^p(X_s) + \delta^p \sup_{t \leq \tilde{s} \leq t + \delta} |b(X_{\tilde{s}})| \big| \mathcal{F}_t \right] + c(p, M, a_1) \sum_{j=1}^{M} \mathbb{E} \left[ (N_{t + \delta}^{(j)} - N_t^{(j)})^p \big| \mathcal{F}_t \right]
\]

and

\[
\mathbb{E}[(\Delta_{t, t + \delta})^p|\mathcal{F}_t] \leq c(p) \delta^{p/2} \mathbb{E} \left[ \sup_{t \leq \tilde{s} \leq t + \delta} (1 + |X_{\tilde{s}}|^p) \big| \mathcal{F}_t \right] + c(p, M, a_1) \sum_{j=1}^{M} \mathbb{E} \left[ (N_{t + \delta}^{(j)} - N_t^{(j)})^p \big| \mathcal{F}_t \right].
\]

Finally, according to Eqs. (6.1), (6.2) and (6.3), since \( p = 2 \) or \( p = 4 \), there exists a constant \( C(M, a_1) \), such that

\[
\mathbb{E}[(\Delta_{t, t + \delta})^p] \leq C(M, a_1)(1 + \mathbb{E}[|X_t|^p]).
\]

Finally since \( \mathbb{E}[X_t^p] < \infty \), we deduce the assertion. \( \square \)
6.3 Additional results

Lemma 6.2 For $p = 2$ or $p = 4$, it yields

$$E\left[\left(\int_{k\Delta}^{(k+1)\Delta} b(X_{k\Delta}) - b(X_s) ds\right)^p\right] \leq C(M, a_1)\Delta^{3p/2}.$$  

Proof of Lemma 6.2 According to Cauchy–Schwarz and Jensen’s inequalities, we have

$$E\left[\left(\int_{k\Delta}^{(k+1)\Delta} b(X_{k\Delta}) - b(X_s) ds\right)^p\right] = \Delta^p E\left[\left(\int_{k\Delta}^{(k+1)\Delta} b(X_{k\Delta}) - b(X_s) ds\right)^p\right] \leq \Delta^p E\left[\int_{k\Delta}^{(k+1)\Delta} (b(X_{k\Delta}) - b(X_s))^p ds\right] \leq \Delta^{p-1} \int_{k\Delta}^{(k+1)\Delta} (b(X_{k\Delta}) - b(X_s))^p ds.$$ 

Applying Proposition 2.5, gives

$$E\left[\left(\int_{k\Delta}^{(k+1)\Delta} b(X_{k\Delta}) - b(X_s) ds\right)^p\right] \leq \Delta^{p-1} E\left[\int_{k\Delta}^{(k+1)\Delta} (b(X_{k\Delta}) - b(X_s))^p ds\right] \leq C(M, a_1)\Delta^{p+1}.$$ 

\[\square\]

Lemma 6.3 For some constant $C(M, a_1) > 0$ (depending on $M$ and $a_1$) we have

$$E[I_{k\Delta}^2] \leq C(M, a_1)\Delta, \quad E[I_{k\Delta}^4] \leq C(M, a_1)\Delta.$$ 

There exists $C > 0$ such that

$$E[Z_{k\Delta}^4|\mathcal{F}_{k\Delta}] \leq C \frac{\sigma_1^4}{\Delta^2}.$$ 

Proof of Lemma 6.3 The first point comes from Lemma 6.2 with $p = 2$, the second point can be deduced when $p = 4$. Finally, with Burkholder–Davis–Gundy inequality we find,

$$E\left[Z_{k\Delta}^4|\mathcal{F}_{k\Delta}\right] \leq C \frac{\sigma_1^4}{\Delta^4} \left[\int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds\right]^2|\mathcal{F}_{k\Delta}] \leq C \frac{\sigma_1^4}{\Delta^4} \sigma^4 \Delta^2 = C \frac{\sigma_1^4}{\Delta^2}.$$ 

\[\square\]

6.4 Proof of Proposition 3.3

From the definition of the contrast function it comes for $t \in \mathbb{L}^2(A),$

$$\gamma_{n,M}(t) - \gamma_{n,M}(b) = \|t - b\|^2_n + 2\sum_{k=1}^n (b - t)(X_{k\Delta})I_{k\Delta} + 2\sum_{k=1}^n (b - t)(X_{k\Delta})Z_{k\Delta}.$$ 

Let us define

$$\mu_n(t) := \frac{1}{n} \sum_{k=1}^n t(X_{k\Delta})Z_{k\Delta}.$$ 

(6.4)
By definition of \( \hat{b}_m \), we have
\[
\gamma_{n,M}(\hat{b}_m) - \gamma_{n,M}(b) \leq \gamma_{n,M}(b_m) - \gamma_{n,M}(b),
\]
where \( b_m \in S_m \), for example \( b_m \) is the orthogonal projection of \( b \) on \( S_m \). This implies
\[
\| \hat{b}_m - b \|_n^2 \leq \| b_m - b \|_n^2 + \frac{2}{n} \sum_{k=1}^{n} (\hat{b}_m - b_m) (X_{k\Delta}) I_{k\Delta} + 2 \mu_n (\hat{b}_m - b_m)
\]
(where \( b \) denotes the restriction of \( b \) to \( A \)). Let us denote:
\[
\mathcal{B}_{m,\pi} = \{ t \in S_m, \| t \|_{\pi^X} = 1 \}.
\]
If \( t \) is a deterministic function, \( \mathbb{E}[\| t \|_n^2] = \| t \|_{\pi^X}^2 \).
Using Cauchy–Schwarz inequality and the relation \( 2xy \leq x^2/d + dy^2 \) for \( x, \ y, d > 0 \), it comes
\[
\begin{align*}
\| \hat{b}_m - b \|_n^2 & \leq \| b_m - b \|_n^2 + 2 \| \hat{b}_m - b_m \|_n \left( \frac{1}{n} \sum_{k=1}^{n} I_{k\Delta}^2 \right)^{1/2} \\
& \quad + 2 \| \hat{b}_m - b_m \|_{\pi^X} \sup_{t \in \mathcal{B}_{m,\pi^X}} |\mu_n(t)| \\
& \leq \| b_m - b \|_n^2 + \frac{1}{d} \| \hat{b}_m - b_m \|_{\pi^X}^2 + d \sup_{t \in \mathcal{B}_{m,\pi^X}} |\mu_n(t)|^2 \\
& \quad + \frac{1}{d} \| \hat{b}_m - b_m \|_{\pi^X}^2 + \frac{n}{d} \sum_{k=1}^{n} I_{k\Delta}^2.
\end{align*}
\]
(6.5)

First, let us consider the following set where the \( \pi^X \)-norm and the empirical norm are equivalent:
\[
\Omega_n := \{ \omega, \forall t \in S_n \setminus \{0\}, \left| \frac{\| t \|_n^2}{\| t \|_{\pi^X}^2} - 1 \right| \leq 1/2 \}.
\]
(6.6)

**Bound of the risk on** \( \Omega_n \) On \( \Omega_n \), we have \( \| \hat{b}_m - b_m \|_{\pi^X}^2 \leq 2 \| \hat{b}_m - b_m \|_n^2 \) and \( \| \hat{b}_m - b_m \|_n^2 \leq 2(\| \hat{b}_m - b \|_n^2 + \| b_m - b \|_n^2) \) thus it comes from Eq. (6.5) with \( d = 7 \)
\[
\| \hat{b}_m - b \|_n^2 \leq 13 \| b_m - b \|_n^2 + 49 \sum_{k=1}^{n} I_{k\Delta}^2 + 49 \sup_{t \in \mathcal{B}_{m,\pi^X}} |\mu_n(t)|^2
\]
(6.7)

Denote \( (\psi_\ell)_\ell \) an orthogonal basis of \( S_m \) for the \( L^2_{\pi^X} \)-norm (thus \( \int \psi_\ell^2(x) \pi^X(x) dx = 1 \)). It comes from Cauchy–Schwarz inequality that,
\[
\sup_{t \in \mathcal{B}_{m,\pi^X}} |\mu_n(t)|^2 = \sup_{\sum_{\ell} \alpha_\ell^2 \leq 1} \left( \sum_{\ell} \alpha_\ell \psi_\ell \right)^2 \leq \sup_{\sum_{\ell} \alpha_\ell^2 \leq 1} \left( \sum_{\ell} \alpha_\ell^2 \right) \left( \sum_{\ell} \mu_n^2(\psi_\ell) \right)
\]
\[
= \sum_{\ell=1}^{D_m} \mu_n^2(\psi_\ell).
\]
(6.8)
Moreover,

\[
E \left[ \sup_{t \in B_{m,n}X} \mu_n^2(t) \right] \leq \sum_{\ell=1}^{D_m} E \left[ \mu_n^2(\psi_{\ell}) \right] \\
\leq \frac{2}{n^2 \Delta^2} \sum_{k=1}^{n} \left[ \sum_{\ell=1}^{D_m} \psi_{\ell}^2(X_{k,\Delta}) \right] E \left[ \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s \right]^2 |F_{k\Delta}| \\
\leq \frac{2}{n^2 \Delta^2} \sum_{k=1}^{n} \left[ \sum_{\ell=1}^{D_m} \psi_{\ell}^2(X_{k,\Delta}) \right] E \left[ \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds |F_{k\Delta}| \right] \\
\leq \frac{2D_m \sigma_1^4}{n\Delta} 
\]  

(6.9)

(if \(\sigma\) is not bounded one can use the stationarity of the process). With Lemma 6.3 the obtained result is

\[
E \left[ \| \hat{b}_m - b \|^2_n \mathbb{I}_{\Omega_n^c} \right] \leq C_1 E \left[ |b - b_m|^2_n \right] + C_2 \frac{D_m}{n\Delta} + C_3 \Delta 
\]  

(6.10)

with \(C_1 = 13, C_2 = 98\sigma_1^4, C_3\) depending on \(a_1, M, \sigma_1\).

**Bound of the risk on \(\Omega_n^c\)**

Let us set \(e = (e_1, \ldots, e_n)\), where \(e_k = U_k - b(X_{k,\Delta}) = I_{k,\Delta} + Z_{k,\Delta}\) and \(\Pi_m U = \Pi_m(U_1, \ldots, U_n) = (\hat{b}_m(X_{\Delta}), \ldots, \hat{b}_m(X_{n,\Delta}))\) where \(\Pi_m\) is the Euclidean orthogonal projection over \(S_m\). Then according to the projection definition,

\[
\| \hat{b}_m - b \|^2_n = \| \Pi_m U - b \|^2_n = \| \Pi_m b - b \|^2_n + \| \Pi_m U - \Pi_m b \|^2_n \\
\leq \| b \|^2_n + \| U - b \|^2_n = \| b \|^2_n + \| e \|^2_n.
\]

Cauchy–Schwarz inequality implies,

\[
E[\| e \|^4_n] = E \left[ \left( \frac{1}{n} \sum_{k=1}^{n} (I_{k,\Delta} + Z_{k,\Delta})^2 \right) \right] \\
\leq \frac{1}{n} \sum_{k=1}^{n} 2^3 E \left[ I_{k,\Delta}^4 + Z_{k,\Delta}^4 \right].
\]

With the controls given in Lemma 6.3, the additional following result will end the proof.

**Lemma 6.4** If \(n\Delta / \ln^2(n) \to \infty\) if \(D_n \leq O((n\Delta)/(\ln^2(n)))\), the probability of the event \(\Omega_n^c\) is bounded as follows:

\[
P(\Omega_n^c) \leq \frac{c_0}{n^2}.
\]

This Lemma is proven in Sect. 6.6.

From Lemma 6.3, Lemma 6.4, Cauchy–Schwarz inequality, we finally obtain

\[
E[\| e \|^2_n \mathbb{I}_{\Omega_n^c}] \leq \left( E[\| e \|^4_n] \right)^{1/2} \left( P(\Omega_n^c) \right)^{1/2} \leq \frac{C}{n\Delta}.
\]
Using that $\|\hat{b}_n - b\|_{n}^2 \leq \|b\|_{n}^2 + \|e\|_{n}^2$ and Cauchy–Schwarz inequality, we deduce under the stationarity assumption, that

$$
\mathbb{E}[\|\hat{b}_n - b\|_{n}^2 1_{\Omega_n^c}] \leq \mathbb{E}\left[ \left( \frac{1}{n} \sum_{k=1}^{n} b(X_k \Delta)^2 \right)^{1/2} \right] \mathbb{P}(\Omega_n^c)^{1/2} + \mathbb{E}\left[ \left( \frac{1}{n} \sum_{k=1}^{n} (e_k \Delta)^2 \right)^{2} \right] \mathbb{P}(\Omega_n^c)^{1/2}
$$

$$
\leq \left( \mathbb{E}[b^4(X_0)]^{1/2} + \mathbb{E}[e^4_\Delta]^{1/2} \right) \mathbb{P}(\Omega_n^c)^{1/2}.
$$

From the Lipschitz property of $b$, we get $\mathbb{E}[b^4(X_0)] \leq C(1 + \mathbb{E}[X^4_0]) < \infty$, and from the Lemma 6.3 $\mathbb{E}[e^4_\Delta]^{1/2} \leq C'/\Delta$, thus

$$
\mathbb{E}[\|\hat{b}_n - b\|_{n}^2 1_{\Omega_n^c}] \leq \frac{C''}{n \Delta}.
$$

(6.11)

Gathering Eqs. (6.10) and (6.11) gives the attended result. □

6.5 Proof of Theorem 3.5

On $\Omega_n^c$, the proof can be lead as in the proof of Proposition 3.3.

On $\Omega_n$, by definition of $\hat{m}$, we have $\gamma_n(\hat{b}_n) + \text{pen}(\hat{m}) \leq \gamma_n(\hat{b}_n) + \text{pen}(m)$. We deduce with the same computation as previously leading to (6.10) that we have for all $m \in \mathcal{M}_n$,

$$
\mathbb{E}[\|\hat{b}_n - b\|_{n}^2 1_{\Omega_n}] \leq 13\|b_n - b\|_{n}^2 + 49C \Delta + 49\mathbb{E}\left[ \sup_{t \in B_{m,\hat{m}},\pi} |\mu_n(t)|^2 1_{\Omega_n} \right]
$$

$$
+ 7\text{pen}(m) - 7\mathbb{E}[\text{pen}(\hat{m})],
$$

where $B_{m,m',\pi} = \{h \in S_m + S_{m'} : \|h\|_{\pi} \leq 1\}$ and $\mu_n$ given in Eq. (6.4). The challenge here is to compute the expectation of the supremum on a random ball. Let us introduce the notation

$$
G_m(m') := \sup_{t \in B_{m,m',\pi}} |\mu_n(t)|.
$$

Then,

$$
\mathbb{E}[\|\hat{b}_n - b\|_{n}^2 1_{\Omega_n}] \leq 13\mathbb{E}[\|b_n - b\|_{n}^2] + 49C \Delta + 49\mathbb{E}\left[ G^2_m(\hat{m}) 1_{\Omega_n} \right]
$$

$$
+ 7\text{pen}(m) - 7\mathbb{E}[\text{pen}(\hat{m})]
$$

$$
\leq 13\mathbb{E}[\|b_n - b\|_{n}^2] + 49C \Delta + 49 \sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[ (G^2_m(m') - p(m,m')) 1_{\Omega_n} \right]_{+}
$$

$$
+ 49 p(m,\hat{m}) + 7\text{pen}(m) - 7\mathbb{E}[\text{pen}(\hat{m})]
$$

where $\text{pen}(m)$ must satisfy $7p(m,m') \leq \text{pen}(m) + \text{pen}(m')$.

To control the term $\mathbb{E}\left[ G^2_m(m') - p(m,m') 1_{\Omega_n} \right]_{+}$ we have to prove a Bernstein-type inequality. Here we follow Baraud et al. (2001b). The idea is presented in Comte et al. (2007) Lemma 2 page 533. We have the following control.

Lemma 6.5 For any positive numbers $\varepsilon$; $\nu$, we have

$$
\mathbb{P}\left( \sum_{k=1}^{n} t(X_k \Delta) Z_k \geq \varepsilon \|t\|_{n}^2 \leq \nu^2 \right) \leq \exp\left( -\frac{n \varepsilon^2 \Delta}{2 \nu^2 \sigma^2_1} \right).
$$
The proof is based on the martingale $\int_0^t H_u dW_u$, $H_u = \sum_{k=1}^n 1_{[k,\Delta, (k+1),\Delta]}(u) t(X_k) \sigma(X_u)$ (for the filtration $\sigma(X_u, u \leq n \Delta)$), and follows exactly the one detailed in Comte et al. (2007) and is omitted here.

Then, the chaining argument of Baraud et al. (2001b) Proposition 6.1 page 45 can be applied here. We finally obtain:

$$\mathbb{E} \left[ (G_{\infty}(m') - p(m, m')) 1_{\Omega_n} \right] + \leq c \sigma_1^2 e^{-D_{m'}} \frac{n \Delta}{n \Delta}$$

and $p(m, m') = \rho_0^2 \frac{D_{m'} + D_{m}}{n \Delta}$, thus $\text{pen}(m) \geq 7 \rho \sigma_1^2 D_{m}/(n \Delta)$. Finally, we have shown that

$$\mathbb{E}[\|\hat{b}_m - b\|_n^2 1_{\Omega_n}] \leq 13 \mathbb{E}[\|b_m - b\|_n^2] + 49 \Delta + 49 \sigma_1^2 \sum_{m' \in \mathcal{M}_n} \frac{e^{-D_{m'}}}{n \Delta} + 7 \text{pen}(m),$$

which ends the proof.

\[\square\]

### 6.6 Proof of Lemma 6.4

The proof of this result follows the proof of Lemma 1 of Comte et al. (2007). The difficulty occurs here in the fact that the projection of the invariant measure $\pi$ onto the $X$–coordinate $\pi^X$ is not bounded from above on the compact $A$. Then, the proof must be adapted from Baraud et al. (2001a).

Under the exponential $\beta$–mixing control, similarly to Comte et al. (2007) using Berbee’s coupling method (for the random variables $X_{k,\Delta}$), we obtain the decomposition:

$$\mathbb{P}(\Omega_n^\pi) \leq \mathbb{P}(\Omega_n^\pi \cap \Omega^*) + \mathbb{P}(\Omega^*)$$

and

$$\mathbb{P}(\Omega^*) \leq n \beta_X(q_n \Delta)$$

where $q_n$ is an integer such that $q_n < n$. The notations are those used in Comte et al. (2007), in particular the notation $^*$ referring to the coupling variables is not detailed here.

Let us control the first term now. Let $(\varphi_j)_j = 1, \ldots, D_n$ be an $L^2(A)$-orthonormal basis of $S_n$ and define the matrices:

$$V_{\pi} = \left[ \left( \int_A \varphi_j^2(x) \varphi_{j'}^2(x) \pi^X(x) dx \right)^{1/2} \right]_{j,j' \in \{1, \ldots, D_n\}}, \quad B = \left( \|\varphi_j \varphi_{j'}\|_{\infty} \right)_{j,j' \in \{1, \ldots, D_n\}}^2.$$

Besides, denote $L_n(\varphi) := \max(\rho^2(V_{\pi} \varphi, \rho(B))$, where for a symmetric matrix $A$:

$$\rho(A) := \sum_{a_j \neq 0} \sum_{a_{j'} \neq 0} |a_j| |a_{j'}| |A_{j,j'}|.$$

We note $\mathbb{P}^* = \mathbb{P}(\cdot \cap \Omega^*)$. Let us consider $v_n(t) := (1/n) \sum_{k=1}^n t(X_k) - \mathbb{E}[t(X_k)]$, $B_{\pi} = \{t \in S_n, \|t\|_{\infty} \leq 1\}$ and $B(0,1) = \{t \in S_n, \|t\| \leq 1\}$, where $\|\cdot\|$ is the $L^2$-norm. As on $A$, $\pi_0 \leq \pi^X(x)$, we have:

$$\mathbb{E} \left[ |v_n(t^2)| \right] = \sup_{t \in B_{\pi}(0,1)} \left| \frac{\|t\|^2_{\pi^X}}{\|t\|^2_{\pi^X} - 1} \right| \leq \pi_0^{-1} \sup_{t \in \mathcal{B}(0,1)} |v_n(t^2)|.$$
Thus,
\[ \mathbb{P}^* \left( \sup_{t \in B_n x (0,1)} |v_n(t^2)| \geq \rho_0 \right) \leq \mathbb{P}^* \left( \sup_{t \in B(0,1)} |v_n(t^2)| \geq \pi_0 \rho_0 \right) \leq \mathbb{P}^* \left( \sup_{\sum_{j=1}^{D_n} a_j^2 \leq 1} \sum_{j, j'} |a_j a_{j'}^*| v_n(\varphi_j \varphi_{j'}) \geq \pi_0 \rho_0 \right). \]

On the set \( \{ (j, j') \in \{ 1, \ldots, D_n \}^2, \ |v_n(\varphi_j \varphi_{j'})| \leq 2V_{\pi x, j, j'}(2x)^{1/2} + 3B_{j, j'} x \} \) we have
\[ \sup_{\sum_{j=1}^{D_n} a_j^2 \leq 1} \sum_{j, j'} |a_j a_{j'}^*| v_n(\varphi_j \varphi_{j'}) \leq 2 \rho(V_{\pi x})(2x)^{1/2} + 3 \rho(B)x. \]

Choosing \( x = (\rho_0 \pi_0)^2/(2L_n(\varphi)) \) with \( \rho_0 = 4(1 + \sqrt{1 + 3\pi_0})/(3\pi_0) \) it comes
\[ \sup_{\sum_{j=1}^{D_n} a_j^2 \leq 1} \sum_{j, j'} |a_j a_{j'}^*| v_n(\varphi_j \varphi_{j'}) \leq \pi_0/2. \]

Thus we have obtained:
\[ \mathbb{P}^*(\Omega^C_n) = \mathbb{P}^* \left( \sup_{t \in B_n x (0,1)} |v_n(t^2)| \geq 1/2 \right) \leq \mathbb{P}^* \left( \{ (j, j') \in \{ 1, \ldots, D_n \}^2, \ |v_n(\varphi_j \varphi_{j'})| \leq 2V_{\pi x, j, j'}(2x)^{1/2} + 3B_{j, j'} x \} \right). \]

Then, we have to bound this last probability term. This is done in Baraud et al. (2001a) Claim 6 of Proposition 7 using a Bernstein’s inequality. Let us denote as in Baraud et al. (2001a) \( Z^k_{f, j} \) the random variables to which the inequality is applied. As the projection of the invariant measure \( \pi \) onto the \( X \)-coordinate \( \pi^X \) is not bounded from above, we have only
\[ E\left[(Z^k_{f, j})^2 \right] \leq \int_A \varphi_j^2(x) \varphi_j^2(x) \pi^X(x) dx = V_{\pi x, j, j'} \]
and still \( ||Z^k_{f, j}||_\infty \leq B_{j, j'} \). Then, it comes,
\[ \mathbb{P}^* \left( \{ (j, j') \in \{ 1, \ldots, D_n \}^2, \ |v_n(\varphi_j \varphi_{j'})| \leq 2V_{\pi x, j, j'}(2x)^{1/2} + 3B_{j, j'} x \} \right) \leq 2D_n^2 \exp \left( -\frac{n x}{q_n} \right). \]

The chosen value for \( x \) implies that there is a constant \( C > 0 \) such that
\[ \mathbb{P}^* \left( \{ (j, j') \in \{ 1, \ldots, D_n \}^2, \ |v_n(\varphi_j \varphi_{j'})| \leq 2V_{\pi x, j, j'}(2x)^{1/2} + 3B_{j, j'} x \} \right) \leq 2D_n^2 \exp \left( -C \frac{n}{q_n} \right). \]

It remains to control \( L_n(\varphi) \). Following Lemma 2 of Baraud et al. (2001a) we have for the trigonometric basis according to Cauchy–Schwarz:
\[ \rho^2(V_{\pi x}) \leq \sum_{j, j' \in \{ 1, \ldots, D_n \}^2} \int_A \varphi_j^2(x) \varphi_j^2(x) \pi^X(x) dx \leq \left\| \sum_{j=1}^{D_n} \varphi_j^2 \right\|_\infty \sum_{j'=1}^{D_n} \int_A \varphi_j^2(x) \pi^X(x) dx \leq \phi_{\pi x} \phi_1^2 D_n^2. \]
under Assumption 3.2.1 and 3.2.3, \( \rho(B) \leq \phi_1^2 D_n \). Thus, \( L_n(\varphi) \leq \max(1, \phi_1^2 \Delta) D_n^2 \). Now, following Eq. (2.4), \( \beta_X(q_n \Delta) \leq e^{-b q_n \Delta} \), decomposition (6.12) leads to

\[
\mathbb{P}(\Omega_n^c) \leq \frac{1}{n^4} + 2n^2 \exp \left( -\frac{C' n \Delta}{\ln(n) D_n} \right),
\]

with \( q_n = \lfloor \frac{5 \ln(n)}{\theta \Delta} \rfloor + 1 \). Finally, for the trigonometric basis, if \( D_n^2 \leq C' n \Delta \frac{6 \ln^2(n)}{\theta} \), thus \( D_n \leq C' \sqrt{\frac{\Delta}{\ln(n)}} \) and \( \mathbb{P}(\Omega_n^c) \leq c / n^4 \), if \( n \Delta / \ln^2(n) \to \infty \).

\( \square \)

Acknowledgements The authors are very greatfull to CNRS for the financial support PEPS. They also thank Gilles Pages for the fruitful discussions. The authors wish to thank Patricia Reynaud Bouret for her advice and her listening. Finally, the authors are very grateful to Eva Löcherbach who supported the project and helped to improve the paper.

7 Appendix: Theoretical results

7.1 Gronwall’s lemma

Lemma 7.1 Let \( \phi, \psi \) and \( y \) be three non negative continuous functions on \([a, b]\), that satisfy

\[
\forall t \in [a, b], \quad y(t) \leq \phi(t) + \int_a^t \psi(s) y(s) ds.
\]

Then

\[
\forall t \in [a, b], \quad y(t) \leq \phi(t) + \int_a^t \phi(s) \psi(s) \exp \left( \int_s^t \psi(u) du \right) ds.
\]

7.2 Talagrand’s inequality

The following result follows from the Talagrand concentration inequality (see Klein and Rio 2005).

Theorem 7.2 Consider \( n \in \mathbb{N}^*, \mathcal{F} \) a class at most countable of measurable functions, and \((X_i)_{i \in \{1, \ldots, n\}}\) a family of real independent random variables. One defines, for all \( f \in \mathcal{F} \),

\[
\nu_n(f) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]).
\]

Supposing there are three positive constants \( M, H \) and \( v \) such that \( \sup_{f \in \mathcal{F}} \|f\|_\infty \leq M \), \( \mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_n(f)|] \leq H \), and \( \sup_{f \in \mathcal{F}} (1/n) \sum_{i=1}^n \text{Var}(f(X_i)) \leq v \), then for all \( \alpha > 0 \),

\[
\mathbb{E} \left[ \left( \sup_{f \in \mathcal{F}} |\nu_n(f)|^2 - 2(1 + 2\alpha)H^2 \right)^+ \right] \leq \frac{4}{b} \left( \frac{v}{n} \exp \left( -b\alpha \frac{nH^2}{v} \right) + \frac{49 M^2}{bC^2(\alpha)n^2} \exp \left( -\frac{\sqrt{2bC(\alpha)\sqrt{\alpha} nH}}{7} \right) \right)
\]

with \( C(\alpha) = (\sqrt{1 + \sqrt{\alpha}} - 1) \wedge 1 \), and \( b = \frac{1}{6} \).
References

Aït-Sahalia Y, Cacho-Diaz J, Laeven RJ (2015) Modeling financial contagion using mutually exciting jump processes. J Financ Econ 117(3):585–606

Amorino C, Gloter A (2018) Contrast function estimation for the drift parameter of ergodic jump diffusion process. arXiv preprint arXiv:1807.08965

Bacry E, Muzy JF (2014) Second order statistics characterization of Hawkes processes and non-parametric estimation. arXiv preprint arXiv:1401.0903

Bacry E, Dayri K, Muzy J (2012) Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. Eur Phys J 85:157

Bacry E, Delattre S, Hoffmann M, Muzy JF (2013) Some limit theorems for Hawkes processes and application to financial statistics. Stoch Process Their Appl 123(7):2475–2499

Bacry E, Mastromatteo I, Muzy JF (2015) Hawkes processes in finance. Mark Microstruct Liq 1(01):1550005

Baraud Y (2002) Model selection for regression on a random design. ESAIM Probab Stat 6:127–146

Baraud Y, Comte F, Viennet G (2001a) Adaptive estimation in autoregression or –mixing regression via model selection. Ann Stat 29(3):839–875. https://doi.org/10.1214/aos/1009210692

Baraud Y, Comte F, Viennet G (2001b) Model selection for (auto-)regression with dependent data. ESAIM Probab Stat 5:33–49

Bibby BM, Sørensen M (1995) Martingale estimation functions for discretely observed diffusion processes. Bernoulli 1:17–39

Bonnet A, Rivoirard V, Picard F (2018) Modeling spatial genomic interactions with the Hawkes model. bioRxiv p 241874

Brémaud P, Massoulié L (1996) Stability of nonlinear Hawkes processes. Ann Probab 24:1563–1588

Carmona R, Fouque JP, Sun LH (2013) Mean field games and systemic risk. Available SSRN

Comte F, Genon-Catalot V (2019) Nonparametric drift estimation for iid paths of stochastic differential equations. hal-02083474

Comte F, Genon-Catalot V, Rozenholc Y (2007) Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli 13(2):514–543

Daley D, Vere-Jones D (2007) An introduction to the theory of point processes: volume II: general theory and structure. Springer, Berlin

Dassios A, Zhao H (2013) Exact simulation of Hawkes process with exponentially decaying intensity. Electron Commun Probab 18(62):1–13

Delattre S, Fournier N, Hoffmann M (2016) Hawkes processes on large networks. Ann Appl Probab 26(1):216–261

DeVore R, Lorentz G (1993) Constructive approximation, vol 303. Springer, Berlin

Dion C (2014) New adaptive strategies for nonparametric estimation in linear mixed models. J Stat Plan Inference 150:30–48

Dion C, Lemler S, Löcherbach E (2019) Exponential ergodicity for diffusions with jumps driven by a Hawkes process. arXiv preprint arXiv:1904.06051

Ditlevsen S, Löcherbach E (2017) Multi-class oscillating systems of interacting neurons. Stoch Process Their Appl 127(6):1840–1869

Donnet S, Samson A (2013) A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. Adv Drug Deliv Rev 65(7):929–939

Duarte A, Galves A, Löcherbach E, Ost G (2016) Estimating the interaction graph of stochastic neural dynamics. arXiv preprint arXiv:1604.00419

El Karoui N, Peng S, Quenez MC (1997) Backward stochastic differential equations in finance. Math Finance 7(1):1–71

Gloiter A (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. ESAIM Probab Stat 4:205–227

Gobet E (2002) Lan property for ergodic diffusions with discrete observations. Ann Inst H Poincaré Probab Stat 38(5):711–737

Gobet E, Matulewicz G (2016) Parameter estimation of Ornstein-Uhlenbeck process generating a stochastic graph. Stat Inference Stoch Process 20:1–25

Gobet E, Hoffmann M, Reiß M (2004) Nonparametric estimation of scalar diffusions based on low frequency data. Ann Stat 32(5):2223–2253

Györfi L, Kohler M, Krzyzak A, Walk H (2006) A distribution-free theory of nonparametric regression. Springer, Berlin

Hansen N, Reynaud-Bouret P, Rivoirard V (2015) Lasso and probabilistic inequalities for multivariate point processes. Bernoulli 21(1):83–143. https://doi.org/10.3150/13-BEJ562

Has’minskii R (1980) Stochastic stability of differential equations. Sijthoff & Noordhoff, New York
Hawkes A (1971) Spectra of some self-exciting and mutually exciting point processes. Biometrika 58(1):83–90
Hawkes A, Oakes D (1974) A cluster process representation of a self-exciting process. J Appl Probab 11(3):493–503
Hoffmann M (1999) Adaptive estimation in diffusion processes. Stoch Process Their Appl 79(1):135–163
Höpfner R (2007) On a set of data for the membrane potential in a neuron. Math Biosci 207(2):275–301
Höpfner R, Löcherbach E, Thieullen M (2016) Ergodicity and limit theorems for degenerate diffusions with
 time period drift. Application to a stochastic Hodgkin–Huxley model. ESAIM Probab Stat 20:527–554
Jahn P, Berg R, Hougaard J, Ditllevsen S (2011) Motoneuron membrane potentials follow a time inhomoge-
 neous jump diffusion process. J Comput Neurosci 31(3):563–579
Jaisson T, Rosenbaum M (2015) Limit theorems for nearly unstable Hawkes processes. Ann Appl Probab
25(2):600–631
Kessler M, Sørensen M (1999) Estimating equations based on eigenfunctions for a discretely observed diffusion
 process. Bernoulli 5(2):299–314
Kirchner M (2017) An estimation procedure for the Hawkes process. Quant Finance 17(4):571–595
Klein T, Rio E (2005) Concentration around the mean for maxima of empirical processes. Ann Probab
33(3):1060–1077
Le Gall J (2010) Calcul stochastique et processus de markov. Notes de cours
Lemonnier R, Vaayatis N (2014) Nonparametric Markovian learning of triggering kernels for mutually excit-
 ing and mutually inhibiting multivariate Hawkes processes. In: Joint European conference on machine
 learning and knowledge discovery in databases. Springer, Berlin, pp 161–176
Lukasik M, Srijith P, Vu D, Bontcheva K, Zubiaga A, Cohn T (2016) Hawkes processes for continuous time
 sequence classification: an application to rumour stance classification in twitter. In: Proceedings of the
 54th annual meeting of the association for computational linguistics (volume 2: short papers), vol 2, pp 393–398
Mancini C (2009) Non-parametric threshold estimation for models with stochastic diffusion coefficient and
 jumps. Scand J Stat 36(2):270–296
Masuda H (2007) Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps.
Stoch Process Their Appl 117(1):35–56
Parisi G, Sourlas N (1992) Supersymmetric field theories and stochastic differential equations. World Scientific,
Singapore
Rambaldi M, Pennesi P, Lillo F (2015) Modeling foreign exchange market activity around macroeconomic
news: Hawkes-process approach. Phys Rev E 91:012819
Renault E, Touzi N (1996) Option hedging and implied volatilities in a stochastic volatility model 1. Math
Finance 6(3):279–302
Reynaud-Bouret P, Roy E (2007) Some non asymptotic tail estimates for Hawkes processes. Bull Belgian
Math Soc Simon Stevin 13(5):883–896
Reynaud-Bouret P, Rivoirard V, Tuleau-Malot C (2013) Inference of functional connectivity in neurosciences
via Hawkes processes. In: 2013 IEEE global conference on signal and information processing (Global-
SIP). IEEE, pp 317–320
Schmisser E (2014a) Nonparametric adaptive estimation of the drift for a jump diffusion process. Stoch Process
Their Appl 124:883–914
Schmisser E (2014b) Nonparametric estimation of coefficients of a diffusion with jumps. Hal
Shimizu Y, Yoshida N (2006) Estimation of parameters for diffusion processes with jumps from discrete
observations. Stat Inference Stoch Process 9(3):227–277
Tankov P, Voltchkova E (2009) Jump-diffusion models: a practitioner’s guide. Banque et Marchés 99(1):24
Vasicek O (1977) An equilibrium characterization of the term structure. J Financ Econ 5(2):177–188
Vere-Jones D, Ozaki T (1982) Some examples of statistical estimation applied to earthquake data. Ann Inst
Stat Math 34(1):189–207
Veretennikov A (1997) On polynomial mixing bounds for stochastic differential equations. Stoch Process
Their Appl 70(1):115–127

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and
institutional affiliations.