A PRACTICAL INFERENCE FOR DISCRETELY OBSERVED JUMP-DIFFUSIONS FROM FINITE SAMPLES

Yasutaka Shimizu*

In the inference for jump-diffusion processes, we often need to get the information of the jump part and of the continuous part separately from the data. Although some asymptotic theories have been studied on this issue, a practical interest is the inference from finitely many discrete samples. In this paper we propose a numerical procedure to construct a filter to judge whether or not a jump occurred from finite samples. The paper includes a discussion about the validity of the procedure.

Key words and phrases: Finite samples, Jump-diffusions, jump-discriminant filter, mathematical validity, plug-in method, threshold selection.

1. Introduction

On a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$, we consider the following $d$-dimensional stochastic differential equation with jumps:

\begin{equation}
\begin{aligned}
dX_t &= a(X_t)dt + b(X_t)dw_t + dz_t, \\
X_0 &= x,
\end{aligned}
\end{equation}

where $x$ is an $\mathbb{R}^d$-valued random variable, $a$ and $b$ are respectively $\mathbb{R}^d$ and $\mathbb{R}^d \otimes \mathbb{R}^m$-valued measurable functions defined on $\mathbb{R}^d$, $w$ is an $m$-dimensional Wiener process, $z$ is an $\mathbb{R}^d$-valued compound Poisson process of the form $z_t = \sum_{i=1}^{N_t} \varepsilon_i$, $N$ is a Poisson process with intensity parameter $\lambda (\geq 0)$, $\varepsilon_i$'s ($i \in \mathbb{N}$) are $\mathbb{R}^d$-valued random variables satisfying $P\{\varepsilon_i \in A\} = \int_A F(z)dz$ for any $A \subset \mathbb{R}^d$, and we put $f(z) = \lambda F(z)$. If $z$ has a jump $\varepsilon_i$ at the time $\tau_i$ then $|\Delta X_{\tau_i}| > 0$ a.s., where $\Delta X_t := X_t - X_{t-}$ for any $t \geq 0$. We assume that $w$, $N$ and $\varepsilon_i$'s are independent to each other, and that (1.1) has a solution.

The statistical inference for jump-diffusion models is recently important in practice since the dynamics of asset prices in finance and insurance are modeled by diffusions with jumps. Particularly, the inference in the case where the process is observed at discrete time points would be of major interest for practitioners, since the real data are always discrete while the assumed model is a continuous-time process. If we can observe the complete path of the process; $X^T = \{X_t\}_{t \in [0,T]}$ then the inference is relatively easy since we can observe the exact jump size $\varepsilon_i$ and its exact jump time $\tau_i$. We can write the likelihood ratio of $X^T$ in an explicit form; see Sørensen (1989, 1999) where the maximum likelihood method for jump-diffusions was discussed. However if we can obtain only

Received November 1, 2006. Revised October 9, 2007. Accepted April 22, 2008.

*Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan.
discrete samples $X^n = \{X^n_t\}_{t=0}^n$ for $0 = t^n_0 < t^n_1 < \cdots < t^n_n = T$ then it grows too complex to write down the likelihood ratio of $X^n$ explicitly.

Shimizu and Yoshida (2006) tackled that problem and proposed an approximation of the log-likelihood of $X^n$. They proposed an asymptotic filter to judge whether or not a jump had occurred in an observational interval $(t^n_{i-1}, t^n_i]$ of the form

\[
\mathcal{H}^n_i = \{ \omega \in \Omega; |\Delta^n_i X| > L h^n \rho \},
\]

for an arbitrarily given constant $L > 0$ and $\rho \in (0, 1/2)$, where $\Delta^n_i X = X^n_{t^n_i} - X^n_{t^n_{i-1}}$ and $t^n_i = ih^n$; see also Shimizu (2006). They showed that we could judge whether a single jump had occurred if $|\Delta^n_i X| > L h^n \rho$ in the interval $(t^n_{i-1}, t^n_i]$, and that the real jump size could be approximated by $\Delta^n_i X$ under the asymptotics that $h^n \to 0$ and $t^n \to \infty$ as $n \to \infty$. Using this judgement, they approximated the likelihood of $X^n$ as its natural discretization.

It is intuitively easy to understand the condition that $\rho \in (0, 1/2)$ in (1.2) since the expected variation of the continuous part of $X$ on $(t^n_{i-1}, t^n_i]$ is of order $\sqrt{h^n}$. That is, we expect that $|\Delta^n_i X|$ is at most of order $\sqrt{h^n}$ if no jump occurs in $(t^n_{i-1}, t^n_i]$. Therefore it would be natural that the threshold $L h^n \rho$ is asymptotically independent of how to select $L > 0$ and $\rho < 1/2$ under the asymptotics that $h^n \to 0$.

The asymptotic theory discussed in Shimizu and Yoshida (2006) or Shimizu (2006) does not indicate how to select the constants $L$ and $\rho$ in (1.2). However the most important problem in practice is to determine these constants since, for example, if we choose $L$ as being too large then we might detect no jump in every interval, on the other hand, if we choose $L$ as being too small, then we might always judge that a jump occurred in every interval. Moreover there is the same problem in selecting $\rho$. Indeed Shimizu (2006) showed some examples that the asymptotic theory does not work under the inappropriate selection of $L$ and $\rho$ even for an enormously large sample size, e.g. $n = 15000$. The same kind of examples are also displayed in Section 2. We must be very careful when we deal with discretely observed stochastic processes, without blindly believing in the philosophy of the asymptotic theory.

In the case of real data, the sample size $n$ is fixed. We would like to choose a “good” threshold $L h^n \rho$ in some sense according to the sample size $n$ and the model. In this paper, we propose a method to select the suitable threshold which improves the performance of the estimators for some unknowns in the model (1.1). For that purpose, we simply consider a filter of the form

\[
\mathcal{H}^n_i(r_n) = \{ \omega \in \Omega; |\Delta^n_i X| > r_n \},
\]

and determine the suitable threshold $r_n$ according to the model (1.1) and the sample size $n$.

Our plan of this paper is as follows. In Section 2 we recall the asymptotic filter and its practical problem discussed in Shimizu (2006) with some simulations. In Sections 3 and 4, we discuss what kind of $r_n$ should be selected, and in
Section 5, we propose a numerical method to determine the desired $r_n$ without a theoretical discussion. Some simulation results will show the usefulness of our method. Section 6 is devoted to the theoretical discussion. Some mathematical validities for our numerical method are shown in a restricted situation. Section 7 gives a remark to apply our method to an unrestricted situation and notes some problems for the future.

2. The asymptotic filter and its problem

In this paper we suppose that process $X$ is observed at discrete time points $\{t^n_i\}_{i=0}^n$ and that $h_n := t^n_i - t^n_{i-1}$ does not depend on $i = 0, \ldots, n$.

The essential idea in Shimizu and Yoshida (2006) or Shimizu (2006) is to use the size of $|\Delta^n X|$ in order to judge the existence of a jump in an interval $(t^n_{i-1}, t^n_i]$, and they used the jump-discriminant filter of the form (1.2). The key result to show the mathematical validity of using such a filter was the following: under some regularities on $a$ and $b$, it follows for any $\rho \in [0, 1/2)$, any $L > 0$ and any $p \geq 1$ that

$$P \left\{ \sup_{t \in [t^n_{i-1}, t^n_i]} |X_t - X_{t^n_{i-1}}| > L h^n_\rho, J^n_i = 0 \right\} \leq C_p h^p_n,$$

where $J^n_i := \# \{ t \in (t^n_{i-1}, t^n_i); |\Delta X_t| > 0 \}$, and $C_p$ is a positive constant depending on $p$; see Lemma 2.1 in Shimizu and Yoshida (2006). The condition that $\rho < 1/2$ is essential as described in Section 1. According to this result, the increment of $X$ with continuous transition could hardly exceed the threshold $L h^n_\rho$. Therefore we could judge a jump occurred if $|\Delta^n X| > L h^n_\rho$, and consequently, $\Delta^n X$ can be an approximation of the jump size as $h_n \to 0$ as $n \to \infty$.

However, as pointed out in Section 4.2 of Shimizu (2006), the accuracy of the judgment of jumps for given $L$ and $\rho$ depends on each model, especially the diffusion coefficient and the distribution of jumps. That is, when we fix constant $L$ and $\rho$ independent of the sample size $n$ then the filter does not work well in some models.

Let us show some simulations where the filter does not work well. Consider the 1-dimensional case of (1.1) with $a(x) = -\mu x$, $b(x) = \sqrt{\sigma}$, $F(z) = (2\pi\theta_2)^{-1/2}e^{-(z-\theta_1)^2/(2\theta_2)}$ and the Poisson intensity $\lambda$:

$$dX_t = -\mu X_t dt + \sqrt{\sigma} dw_t + z_t^{(\lambda, \theta)}.$$

Set the true value of the parameter as $(\mu, \theta_1, \theta_2, \lambda) = (0.3, 0.5, 0.1, 3.0)$, and consider the two models as $\sigma = 0.1$ and $\sigma = 0.3$. We estimate the parameter $(\mu, \sigma, \theta_1, \theta_2, \lambda)$ jointly from the discrete observation $X^n$ with $t^n_i = ih_n$ via the method of Shimizu and Yoshida (2006) with a slight extension to the case where the Lévy density is bounded around the origin. We set $h_n = n^{-0.8}$ and $\rho = 0.49$.

In the case where $\sigma = 0.1$, we set $L = 1.0$ and obtain Table 1. From this result, one may think that the filter works well. This is because the diffusion parameter $\sigma$ is relatively small compared with the jump size. However, when
Table 1. The mean and the standard deviation (s.d.) of estimators over 500 iterations. \( \sigma = 0.1, L = 1.0 \).

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>50</th>
<th>500</th>
<th>3000</th>
<th>TRUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma = 0.1 )</td>
<td>( \hat{\mu} )</td>
<td>0.2702</td>
<td>0.3009</td>
<td>0.3013</td>
<td>0.3</td>
</tr>
<tr>
<td>( L = 1.0 )</td>
<td>s.d.</td>
<td>0.1474</td>
<td>0.0477</td>
<td>0.0277</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\sigma} )</td>
<td>0.1169</td>
<td>0.1003</td>
<td>0.0998</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.0295</td>
<td>0.0066</td>
<td>0.0025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\theta}_1 )</td>
<td>0.6450</td>
<td>0.5408</td>
<td>0.5135</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.1294</td>
<td>0.0745</td>
<td>0.0558</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\theta}_2 )</td>
<td>0.0919</td>
<td>0.0924</td>
<td>0.0956</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.0886</td>
<td>0.0495</td>
<td>0.0269</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\lambda} )</td>
<td>2.0659</td>
<td>2.6893</td>
<td>2.9067</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.7266</td>
<td>0.6514</td>
<td>0.5076</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. The mean and the standard deviation (s.d.) of estimators over 500 iterations. \( \sigma = 0.3, L = 1.0 \).

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>50</th>
<th>500</th>
<th>3000</th>
<th>TRUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma = 0.3 )</td>
<td>( \hat{\mu} )</td>
<td>0.2358</td>
<td>0.2528</td>
<td>0.2536</td>
<td>0.3</td>
</tr>
<tr>
<td>( L = 1.0 )</td>
<td>s.d.</td>
<td>0.2366</td>
<td>0.0809</td>
<td>0.0428</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\sigma} )</td>
<td>0.2310</td>
<td>0.2381</td>
<td>0.2471</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.0433</td>
<td>0.0133</td>
<td>0.0054</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\theta}_1 )</td>
<td>0.4505</td>
<td>0.2230</td>
<td>0.1108</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.0151</td>
<td>0.0522</td>
<td>0.0162</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\theta}_2 )</td>
<td>0.2131</td>
<td>0.1445</td>
<td>0.0839</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.1178</td>
<td>0.0331</td>
<td>0.0162</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\lambda} )</td>
<td>2.7419</td>
<td>5.7280</td>
<td>11.453</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.8759</td>
<td>0.9882</td>
<td>1.0682</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.7266</td>
<td>0.6514</td>
<td>0.5076</td>
<td></td>
</tr>
</tbody>
</table>

\( \sigma = 0.3 \), we obtain the result as in Table 2. The estimator \( \hat{\theta}_1 \) behaves strange and \( \hat{\lambda} \) is overestimated as the standard deviation increases. This is because the filter misjudges the jumps, especially it overestimates the number of jumps. This is the situation where the asymptotic theory does not work yet. In order to improve the performance of the estimation, we change the value \( L = 1.0 \) to a suitable number. Here we chose \( L = 1.8 \) as in Table 3. Then the result were dramatically improved.

These results indicate that we need to select \( L \) depending on the model and the sample size \( n \). See Section 4.2 in Shimizu (2006) for other examples.

Shimizu (2006) left the constant \( L \) in the threshold to describe some intuitive procedures for selecting the filter. However there is no theoretical reason to separate \( L \) and \( h_n^\rho \) in general. Therefore, in this paper, we rewrite the filter simply as in (1.3) and consider the selection problem of not \( L \), but \( r_n \) itself, which is the most important parameter in applications. In the next section, we discuss what kind of \( r_n \) is suitable to improve the performance for fixed \( n \).
Table 3. The mean and the standard deviation (s.d.) of estimators over 500 iterations. \( \sigma = 0.1, L = 1.8 \).

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>50</th>
<th>500</th>
<th>3000</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma = 0.3 )</td>
<td>( \hat{\mu} )</td>
<td>0.2001</td>
<td>0.2977</td>
<td>0.3010</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.2366</td>
<td>0.0866</td>
<td>0.0428</td>
<td></td>
</tr>
<tr>
<td>( L = 1.8 )</td>
<td>( \hat{\sigma} )</td>
<td>0.3902</td>
<td>0.3044</td>
<td>0.2978</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.0901</td>
<td>0.0191</td>
<td>0.0074</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\theta}_1 )</td>
<td>0.7517</td>
<td>0.5750</td>
<td>0.5147</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.1879</td>
<td>0.0800</td>
<td>0.0552</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\theta}_2 )</td>
<td>0.2014</td>
<td>0.0965</td>
<td>0.1002</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.2054</td>
<td>0.0531</td>
<td>0.0279</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\lambda} )</td>
<td>1.5156</td>
<td>2.4717</td>
<td>2.8938</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.6450</td>
<td>0.5881</td>
<td>0.4682</td>
<td></td>
</tr>
</tbody>
</table>

3. Criterion for threshold selection

Notice that in a time interval \((s_1, s_2]\) where no jump occurs, \( X \) is a solution to the following stochastic integral equation:

\[
\alpha_t^{(s_1)} = X_{s_1} + \int_{s_1}^{t} a(\alpha_u^{(s_1)}) du + \int_{s_1}^{t} b(\alpha_u^{(s_1)}) dw_u.
\]

(3.1)

The statistic

\[
\hat{\lambda}_n(r_n) = \frac{1}{nh_n} \sum_{i=1}^{n} I(\mathcal{H}_i^n(r_n))
\]

is an estimator of the Poisson intensity of jumps over the \([0, 1]\) time interval. It is shown in Shimizu (2007) that this estimator was consistent for \( \lambda \) unless \( X \) is ergodic, and it is easily found that (3.2) is asymptotically unbiased, that is, \( E[\hat{\lambda}_n(r_n)] \to \lambda \) as \( n \) tends to infinity. Therefore it is desired that at least the bias of the estimator \( \hat{\lambda}_n(r_n) \) for large values of fixed \( n \) is as small as possible.

Let us estimate the following expectation analytically:

\[
E[\hat{\lambda}_n(r_n)] = \frac{1}{nh_n} \sum_{i=1}^{n} P(\mathcal{H}_i^n(r_n)) = \frac{1}{nh_n} \sum_{i=1}^{n} P(\{|\Delta_i^n X| > r_n\}).
\]

In the sequel we use the following notations:

\[
\epsilon(r_n) := \frac{1}{n} \sum_{i=1}^{n} P(|\Delta_i^n \alpha(t_{i-1}^n)| > r_n),
\]

(3.3)

\[
\delta(r_n, A) := \frac{1}{n} \sum_{i=1}^{n} P(|S_i^n(\tau)| > r_n \mid \{J_i^n = 1\} \cap A)
\]

(3.4)

for \( A \in \mathcal{F} \), where \( \Delta_i^n \alpha(t_{i-1}^n) = \alpha(t_{i-1}^n) - \alpha(t_{i-1}^n) \) and \( S_i^n(\tau) = (\alpha(t_{i-1}^n) - \alpha(t_{i-1}^n) + (\alpha(t_{i-1}^n) - \alpha(t_{i-1}^n)) \), \( \tau_i^n = \inf \{t \in (t_{i-1}^n, t_i^n); |\Delta X_t| > 0\} \) and \( J_i^n := \# \{t \in (t_{i-1}^n, t_i^n); |\Delta X_t| > 0\} \).
Lemma 3.1. For any integer $n$, it follows that

\[
E[\lambda_n(r_n)] = h_n^{-1}e(r_n)e^{-\lambda h_n} + (1 - \delta(r_n, |\Delta z_{\tau_i^n}| > 2 r_n))e^{-\lambda h_n} \int_{|z| > 2 r_n} f(z) dz + T_n + U_n + e_n,
\]

where $e_n$, $T_n$ and $U_n$ are sequences satisfying the following inequalities, respectively:

\[
0 \leq e_n \leq \lambda^2 h_n,
\]

\[
0 \leq e^{\lambda h_n} T_n \leq \int_{|z| > 2 r_n} f(z) dz \cdot \delta(r_n, |\Delta z_{\tau_i^n}| > 2 r_n),
\]

\[
0 \leq e^{\lambda h_n} U_n \leq \int_{\frac{r_n}{2} < |z| \leq 2 r_n} f(z) dz + \delta \left( \frac{r_n}{2}, |\Delta z_{\tau_i^n}| \leq \frac{r_n}{2} \right) \int_{|z| \leq \frac{r_n}{2}} f(z) dz.
\]

Proof. Notice the following decomposition:

\[
E[\lambda_n(r_n)] = \frac{1}{nh_n} \sum_{i=1}^{n} \left[ P\{\mathcal{E}_i^n(r_n) \cap \{ J_{i}^n = 0 \} \} + P\{\mathcal{E}_i^n(r_n) \cap \{ J_{i}^n = 1 \} \} + P\{\mathcal{E}_i^n(r_n) \cap \{ J_{i}^n \geq 2 \} \} \right].
\]

(3.5)

Let the last term be $e_n$ then it follows from the Poissonian property that

\[
e_n := \frac{1}{nh_n} \sum_{i=1}^{n} P\{\mathcal{E}_i^n(r_n) \cap \{ J_{i}^n \geq 2 \} \} \leq \frac{1}{nh_n} \sum_{i=1}^{n} P\{ J_{i}^n \geq 2 \} \leq \lambda^2 h_n.
\]

(3.6)

Since $dX_t = d\alpha_t^{(\tau_i^n)}$ in $(t_{i-1}^n, t_i^n]$ on the set $\{ J_{i}^n = 0 \}$, we see for the first term of (3.5) that

\[
P\{\mathcal{E}_i^n(r_n) \cap \{ J_{i}^n = 0 \} \} = e^{-\lambda h_n} P\{ |\Delta_i^n \alpha(t_{i-1}^n) | > r_n \}.
\]

The task remains to estimate the second term of (3.5). Noticing that

\[
S_{i-1}^n(\tau) = (X_{t_{i-1}^n} - X_{\tau_i^n}) + (X_{\tau_i^n} - X_{t_{i-1}^n})
\]
on $\{ J_{i}^n = 1 \}$, we obtain the following decomposition.

\[
P\{\mathcal{E}_i^n(r_n) \cap \{ J_{i}^n = 1 \} \} = P\{|S_i^n(\tau) + |\Delta z_{\tau_i^n}| > r_n, |\Delta z_{\tau_i^n}| > 2 r_n, J_{i}^n = 1 \} + P\{|S_i^n(\tau) + |\Delta z_{\tau_i^n}| > r_n, |\Delta z_{\tau_i^n}| \leq 2 r_n, J_{i}^n = 1 \} \leq P\{|S_i^n(\tau) | \leq r_n, |\Delta z_{\tau_i^n}| > 2 r_n, J_{i}^n = 1 \} + P\{|S_i^n(\tau) + |\Delta z_{\tau_i^n}| > r_n, |S_i^n(\tau) | > r_n, |\Delta z_{\tau_i^n}| > 2 r_n, J_{i}^n = 1 \}.
\]

(3.7)

(3.8)
Let (3.7) – (3.10) be

\[ A_i^n = \lambda h_n e^{-\lambda h_n} \int_{|z| > 2r_n} F(z) dz \cdot P\{|S_i^n(\tau) - \hat{\lambda} h_n| \leq 2r_n, J_i^n = 1\}, \]

\[ B_i^n = \lambda h_n e^{-\lambda h_n} \int_{|z| > 2r_n} F(z) dz \cdot P\{|S_i^n(\tau) - \hat{\lambda} h_n| > r_n, |\Delta z_i^n| = 1\}, \]

\[ C_i^n = \lambda h_n e^{-\lambda h_n} \int_{2r_n < |z| \leq 2r_n} F(z) dz, \]

\[ D_i^n = \lambda h_n e^{-\lambda h_n} \int_{|z| \leq r_n/2} F(z) dz \cdot P\{|S_i^n(\tau) - \hat{\lambda} h_n| > r_n, |\Delta z_i^n| \leq \frac{r_n}{2}, J_i^n = 1\}. \]

Then we obtain the consequence by setting the following as:

\[ T_n = \frac{1}{nh_n} \sum_{i=1}^{n} B_i^n, \]

\[ U_n = \frac{1}{nh_n} \sum_{i=1}^{n} (C_i^n + D_i^n). \]

This completes the proof. \( \square \)

4. A bias correction

Let \( b(r_n) \) be the bias of \( \hat{\lambda}_n(r_n) \), that is,

\[ b(r_n) := E[\hat{\lambda}_n(r_n)] - \lambda. \]

Our final goal is to select an \( r_n \) which minimizes the absolute bias \( |b(r_n)| \). However it would be difficult to estimate the bias directly, and what we can show is the upper and lower bounds of the bias. We easily obtain the following theorem from the previous lemma and the fact that \( \lambda = \int f(z) dz \).

**Theorem 4.1.** Define \( \tilde{\ell}_n \) and \( \ell_n \) as follows.

\[ \tilde{\ell}_n := h_n^{-1} \epsilon(r_n) - \int_{|z| < 2r_n} f(z) dz, \]

\[ \ell_n := \tilde{\ell}_n + \int_{r_n/2 < |z| \leq 2r_n} f(z) dz. \]

Then it follows that

\[ \tilde{\ell}_n + O(\tilde{\delta}_n \vee h_n) \leq e^{\lambda h_n} b(r_n) \leq \ell_n + O(\delta_n \vee h_n) \]

as \( h_n \to 0 \), where \( \tilde{\delta}_n = \delta(r_n, |\Delta z_i^n| > 2r_n) \vee \delta(r_n/2, |\Delta z_i^n| \leq r_n/2) \).
Roughly speaking, we can regard $\ell_n$ and $\tilde{\ell}_n$ as the first order terms of the upper and lower bounds, respectively since $h_n \vee \delta_n \to 0$ as $n \to \infty$ under $r_n \sim h_n$ for $\rho < 1/2$, which follows from (2.1). For these bounds, it is desired that the amplitude of the bias $|\tilde{\ell}_n - \ell_n|$, which corresponds to a kind of variance, is as small as possible from the aspect of stability of the estimated bias. Therefore the threshold $r_n$ should be selected so that $|\tilde{\ell}_n - \ell_n| \approx 0$. On the other hand, if we concentrate only on minimizing the distance $|\tilde{\ell}_n - \ell_n|$ then $\tilde{\ell}_n$ might be strictly positive, or $\ell_n$ might be strictly negative, which induce strictly biased estimators. Therefore the center of the interval $[\tilde{\ell}_n, \ell_n]$ should be nearly to zero in order to make the maximum of the absolute bias as small as possible; the aspect of unbiasedness. Therefore $r_n$ should also be selected so that $(\tilde{\ell}_n + \ell_n) \approx 0$. From these points of view, it would be natural to select $r_n$ which minimizes the following quantity:

$$(1 - u)|\tilde{\ell}_n + \ell_n| + u|\tilde{\ell}_n - \ell_n| \quad (0 \leq u < 1).$$

The weight $u$ should not be 1 since $r_n = 0$ or $\infty$ is clearly selected in this case and each of them does not play a role as the filter. Therefore it is convenient to rewrite it as follows.

$$L_{n,w}(r_n) := |\tilde{\ell}_n + \ell_n| + w|\tilde{\ell}_n - \ell_n| = |L(r_n)| + w \int_{r_n/2 < |z| \leq 2r_n} f(z)dz$$

for $w \geq 0$, where $L(r_n) = 2h_n^{-1}\epsilon(r_n) - J(r_n)$, and $J(r_n) = \int_{|z| \leq r_n/2} f(z)dz + \int_{|z| < 2r_n} f(z)dz$. We would like to select $r_n$ which minimizes the function $L_{n,w}(r_n)$. Although we must note that it still has unknown functions such as $\epsilon$ and $f$, the target for $r_n$ would be defined as follows.

**Definition 4.1.** We denote by $r_{opt}^{(n,w)}$ a minimizer of the function $L_{n,w}(r)$:

$$r_{opt}^{(n,w)} := \arg \min_{r \geq 0} L_{n,w}(r).$$

It is easy to see that $r_{opt}^{(n,w)}$ is well defined for any $n$ and $w$ from the form of $L_{n,w}(r)$. A constant $w$ is the weight on the amplitude of the bias; $|\tilde{\ell}_n - \ell_n|$. Setting $w$ as being large, one puts weight not on the unbiased estimation but the stable estimation. However too large a value of $w$ can induce the definitely positive or negative bias. Since we do not obtain the exact bias but only less strict bounds from both sides, the aspect of unbiasedness should be given more weight rather than compulsory minimization of the bias range. Moreover, from the technical point of view, it is not necessarily that $r_{opt}^{(n,w)}$ is determined uniquely if $w > 1$. Actually, if $w \to \infty$, then the selected threshold would tend to zero or infinity as already pointed out. Therefore it would not be suitable to choose large $w$. If we have no prior information about the true bias then $w = 1$ would be appropriate.
Hence, both the function $\epsilon$ and $L_{n,w}$ is uniquely determined and is independent of $w$ as the next lemma shows. The lemma implies that, if $0 \leq w \leq 1$, minimizing $L_{n,w}$ is equivalent to finding an unique root of $L(r) = 0$.

**Lemma 4.1.** If $0 \leq w \leq 1$ then $r_{\text{opt}}^{(n,w)}$ is the unique solution to $L(r) = 0$.

**Proof.** Note that

$$L_{n,w}(r) = \begin{cases} 2h_n^{-1}\epsilon(r) - J_w(r) & \text{if } L(r) \geq 0 \\ J_w(r) - 2h_n^{-1}\epsilon(r) & \text{if } L(r) \leq 0 \end{cases},$$

where $J_w(r) = J(r) - w \int_{r/2 < |z| \leq 2r} f(z)dz$. We see that the function $J_{\tilde{w}}(r)$ is increasing in $r$ for each $\tilde{w} \in [-1,1]$ since we can rewrite $J_{\tilde{w}}$ as follows.

$$J_{\tilde{w}}(r) = (1 + \tilde{w}) \int_{|z| \leq r/2} f(z)dz + (1 - \tilde{w}) \int_{|z| < 2r} f(z)dz.$$

Hence, both the function $l_1(r) := 2h_n^{-1}\epsilon(r) - J_w(r)$ and the function $l_2(r) := 2h_n^{-1}\epsilon(r) - J_{-w}(r)$ are decreasing in $r \geq 0$ for each $w \in [0,1]$. Similarly, the function $L(r)$ is also decreasing. Moreover the equation $L(r) = 0$ has the unique root since $\epsilon(r)$ is decreasing in $r \geq 0$ and $\epsilon(0) = 1$, $\lim_{r \to \infty} \epsilon(r) = 0$, and $J(r)$ is increasing in $r \geq 0$ with $J(0) = 0$. From these facts, it follows that $l_1$ is minimized at $r_0 := \max\{r \geq 0; L(r) \geq 0\}$ on the closed set $\{r \geq 0; L(r) \geq 0\} \subset [0, \infty)$, and also

$$r_0 = \arg\min L(r) \quad \text{on} \quad \{r \geq 0; L(r) \geq 0\},$$

that is, $L(r_0) = 0$. Similarly, we can show that $l_2$ is maximized at $r'_0$ satisfying $L(r'_0) = 0$. Consequently we obtain that $r_0 = r'_0 = r_{\text{opt}}^{(n,w)}$. \qed

This lemma says that $r_{\text{opt}}^{(n,w)}$ is independent of $w$ while $w \in [0,1]$. In the sequel we consider the case where $w \in [0,1]$. Under this assumption we put $r_{\text{opt}}^{(n)} := r_{\text{opt}}^{(n,w)}$ for simplicity. Then our interest is to find the solution $r_{\text{opt}}^{(n)}$ to the equation

$$L(r) = 2h_n^{-1}\epsilon(r) - J(r) = 0. \quad (4.3)$$

Equation (4.3) still includes the unknown quantities $\epsilon$ and $f$. Therefore we must substitute some suitable estimators for $\epsilon$ and $f$, and construct an estimator of $r_{\text{opt}}^{(n)}$. However $\epsilon$ and $f$ must be constructed by the filter which should be selected in our goal, so it goes back and forth! In the next section, we propose a plug-in method in order to avoid this dilemma, and we show the performance of the method in some simulations.
5. Direct plug-in method

5.1. Plug-in rule

The goal of this section is to estimate the threshold $r_{\text{opt}}^{(n)}$ from a finitely fixed $n$. For that purpose, we have to construct an estimator of $L(r)$. As a general notation, we denote by $\hat{G}_n(x; r_n)$ an estimator of a function $G(x)$ constructed by the data $\{X_{t_i}^{(n)}\}_{i=0}^{n}$ and the filter $\mathcal{H}_i^n(r_n)$. Using this notation, the natural estimator of $J$ is written as follows.

\begin{equation}
\hat{J}_n(r; r_n) = \int_{|z| \leq r/2} \hat{f}_n(z; r_n) dz + \int_{|z| \leq 2r} \hat{f}_n(z; r_n) dz.
\end{equation}

Let us consider a 1-dimensional case of $X$ for simplicity. In order to calculate the above integrals easily, it is convenient to use

\begin{equation}
\hat{f}_n(z; r_n) = \frac{1}{nh_n^{1+\delta}} \sum_{i=1}^{n} \phi \left( \frac{z - \Delta_i^n X}{h_n^\delta} \right) I(\mathcal{H}_i^n(r_n))
\end{equation}

proposed in Shimizu (2006), where $\delta \in (0, 1/2)$ is a constant, and $\phi$ is the standard normal kernel, i.e. $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. When we assume a parametric model for $f$, we can use a parametric estimator for $J$. On the other hand, the natural approximation of $\epsilon(r_n) = \frac{1}{n} \sum_{i=1}^{n} P\{|\Delta_i^n \alpha(t_{i-1}^n)| > r_n\}$ for diffusion (3.1) is a local-Gauss approximation of the transition probability of $\alpha(t_{i-1}^n)$, that is, for $a_{i-1} := a(X_{t_{i-1}^n})$ and $\beta_{i-1} := b^2(X_{t_{i-1}^n})$,

\begin{equation}
\epsilon(r_n) \approx \frac{1}{n} \sum_{i=1}^{n} \int_{|y| > r_n} \frac{1}{\sqrt{2\pi} |\beta_{i-1}| h_n} \exp \left( -\frac{1}{2h_n \beta_{i-1}} (y - h_n a_{i-1})^2 \right) dy
\end{equation}

Substituting $\beta$ by the estimator $\hat{\beta}_n(X_{t_{i-1}^n}^n; r_n)$, we have

\begin{equation}
\hat{\epsilon}_n(r; r_n) = \frac{1}{n} \sum_{i=1}^{n} \int_{|y| > r_n} \frac{1}{\sqrt{2\pi} |\hat{\beta}_n(X_{t_{i-1}^n}^n; r_n)| h_n} \times \exp \left( -\frac{y^2}{2h_n \hat{\beta}_n(X_{t_{i-1}^n}^n; r_n)} \right) dy.
\end{equation}

We note that the above procedure can be easily extended to the multidimensional case.

Now let us proceed to the algorithm to find the approximator of $r_{\text{opt}}^{(n)}$. The following **Plug-in method** is executable:

**Plug-in Algorithm:**
- **Step 0.** Choose a pilot threshold $r_n^{(0)} > 0$ arbitrarily.
• **Step $k (\geq 1)$**. Solve the equation

$$
\hat{L}_n(r; r_n^{(k-1)}) := 2h_n^{-1}\hat{\epsilon}_n(r; r_n^{(k-1)}) - \hat{J}_n(r; r_n^{(k-1)}) = 0 \quad \cdots \quad (*)
$$

and define the root as $r = r_n^{(k)}$.

• **Iterate Step $k$ ($k = 1, 2, \ldots$)** until the sequence $\{r_n^{(k)}\}_{k \in \mathbb{N}}$ converges.

We call the $k$th solution to the equation ($*$) the $k$-stage threshold and call the function $\hat{L}_n(r; r_n^{(k-1)})$ the $k$-stage threshold selector. By the same argument as in Lemma 4.1, we see that the equation $\hat{L}_n(r; r_n^{(k-1)}) = 0$ has the unique solution $r_n^{(k)}$ for any $k, n \in \mathbb{N}$.

We expect that $\lim_{k \to \infty} r_n^{(k)}$ exists and that the limit is near the value of $r_{\text{opt}}^{(n)}$ in some sense. Before the theoretical study, we try this algorithm with simulations.

### 5.2. Simulation results

Let us show the performance of our threshold selector $\hat{L}_n$ using (5.2) and (5.4).

We consider a 1-dimensional data generating process as follows:

$$
dX_t = -\mu X_t dt + \sigma dw_t + dz_t,
$$

where $z$ is the compound Poisson process with the Lévy density $f(z) = \frac{\lambda}{\sqrt{2\pi}} e^{-z^2/2}$, and the true parameter is $(\mu, \sigma, \lambda) = (0.03, 0.3, 15.0)$.

For fixed $n$, the experiment is done as follows:

1. Choose the pilot threshold as $r_n^{(0)} = 1.0$.
2. Observe one path of $X$ at the time $t_n^i = ih_n$, where $h_n = n^{-0.8}$.
3. Calculate 1 to 7-stage thresholds solving $\hat{L}_n(r; r_n^{(k-1)}) = 0$ ($k = 0, 1, \ldots, 6$) and the estimator of $(\mu, \sigma, \lambda)$ in each stage as follows; see Shimizu and Yoshida (2006):

$$
\hat{\mu}_n^{(k)} = \frac{\sum_{i=1}^n \Delta_{i-1}^n X_i X_{i-1} \mathbf{1}_{\{|\Delta_i^n X| \leq r_n^{(k)}\}}}{\sum_{i=1}^n X_{i-1}^2 h_n \mathbf{1}_{\{|\Delta_i^n X| \leq r_n^{(k)}\}}},
$$

$$
\hat{\sigma}_n^{(k)} = \left\{ \frac{\sum_{i=1}^n (\Delta_i^n X - \hat{\mu}_n X_{i-1} h_n)^2 \mathbf{1}_{\{|\Delta_i^n X| \leq r_n^{(k)}\}}}{h_n \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i^n X| \leq r_n^{(k)}\}}} \right\}^{1/2},
$$

$$
\hat{\lambda}_n^{(k)} = \frac{1}{nh_n} \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i^n X| > r_n^{(k)}\}}.
$$

The experiment (1)–(3) is iterated 300 times. Tables 4 and 5 in the sequel are the sample means and the sample standard deviations (s.d.) in each stage throughout 300 iterations. The values in the last line are $r_{\text{opt}}^{(n)}$ and the true value for each parameter.
Table 4. The 1st–7th thresholds with a sample size of \( n = 1000 \). Each estimator is the mean over 300 iterations, and the s.d. is their standard deviation.

<table>
<thead>
<tr>
<th>( n = 1000 )</th>
<th>( r_n^{(k)} )</th>
<th>( \lambda_n^{(k)} )</th>
<th>( \hat{\mu}_n^{(k)} )</th>
<th>( \hat{\sigma}_n^{(k)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-stage</td>
<td>1.0</td>
<td>4.68</td>
<td>−0.11718</td>
<td>1.74590</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0</td>
<td>1.082</td>
<td>0.319</td>
<td>0.186</td>
</tr>
<tr>
<td>1-stage</td>
<td>0.36071</td>
<td>10.52</td>
<td>−0.04073</td>
<td>0.51620</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0336</td>
<td>1.471</td>
<td>0.079</td>
<td>0.077</td>
</tr>
<tr>
<td>2-stage</td>
<td>0.09943</td>
<td>13.40</td>
<td>−0.03253</td>
<td>0.30624</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0136</td>
<td>1.663</td>
<td>0.050</td>
<td>0.009</td>
</tr>
<tr>
<td>3-stage</td>
<td>0.05836</td>
<td>14.38</td>
<td>−0.03089</td>
<td>0.29766</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0020</td>
<td>1.835</td>
<td>0.049</td>
<td>0.007</td>
</tr>
<tr>
<td>4-stage</td>
<td>0.05599</td>
<td>14.66</td>
<td>−0.03052</td>
<td>0.29619</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0020</td>
<td>1.946</td>
<td>0.048</td>
<td>0.007</td>
</tr>
<tr>
<td>5-stage</td>
<td>0.05551</td>
<td>14.72</td>
<td>−0.03043</td>
<td>0.29589</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0019</td>
<td>1.986</td>
<td>0.048</td>
<td>0.007</td>
</tr>
<tr>
<td>6-stage</td>
<td>0.05540</td>
<td>14.74</td>
<td>−0.03042</td>
<td>0.29582</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0019</td>
<td>2.004</td>
<td>0.048</td>
<td>0.007</td>
</tr>
<tr>
<td>7-stage</td>
<td>0.05540</td>
<td>14.74</td>
<td>−0.03037</td>
<td>0.29582</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0019</td>
<td>2.004</td>
<td>0.048</td>
<td>0.007</td>
</tr>
<tr>
<td>( r_n^{(n)} / \text{True} )</td>
<td>0.05560</td>
<td>15.0</td>
<td>−0.03</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 5. The 1st–7th thresholds with a sample size of \( n = 3000 \). Each estimator is the mean over 300 iterations, and the s.d. is their standard deviation.

<table>
<thead>
<tr>
<th>( n = 3000 )</th>
<th>( r_n^{(k)} )</th>
<th>( \lambda_n^{(k)} )</th>
<th>( \hat{\mu}_n^{(k)} )</th>
<th>( \hat{\sigma}_n^{(k)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-stage</td>
<td>1.0</td>
<td>4.77</td>
<td>−0.09615</td>
<td>1.74004</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0</td>
<td>0.995</td>
<td>0.257</td>
<td>0.173</td>
</tr>
<tr>
<td>1-stage</td>
<td>0.26503</td>
<td>11.73</td>
<td>−0.03510</td>
<td>0.40342</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0238</td>
<td>1.436</td>
<td>0.048</td>
<td>0.038</td>
</tr>
<tr>
<td>2-stage</td>
<td>0.05592</td>
<td>14.16</td>
<td>−0.03331</td>
<td>0.30113</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0048</td>
<td>1.702</td>
<td>0.037</td>
<td>0.004</td>
</tr>
<tr>
<td>3-stage</td>
<td>0.04098</td>
<td>14.80</td>
<td>−0.03315</td>
<td>0.29901</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0008</td>
<td>1.851</td>
<td>0.037</td>
<td>0.004</td>
</tr>
<tr>
<td>4-stage</td>
<td>0.04036</td>
<td>14.90</td>
<td>−0.03291</td>
<td>0.29874</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0008</td>
<td>1.883</td>
<td>0.037</td>
<td>0.004</td>
</tr>
<tr>
<td>5-stage</td>
<td>0.04028</td>
<td>14.92</td>
<td>−0.03288</td>
<td>0.29870</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0008</td>
<td>1.887</td>
<td>0.037</td>
<td>0.004</td>
</tr>
<tr>
<td>6-stage</td>
<td>0.04027</td>
<td>14.92</td>
<td>−0.03288</td>
<td>0.29870</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0008</td>
<td>1.888</td>
<td>0.037</td>
<td>0.004</td>
</tr>
<tr>
<td>7-stage</td>
<td>0.04027</td>
<td>14.92</td>
<td>−0.03288</td>
<td>0.29870</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0008</td>
<td>1.888</td>
<td>0.037</td>
<td>0.004</td>
</tr>
<tr>
<td>( r_n^{(n)} / \text{True} )</td>
<td>0.04016</td>
<td>15.0</td>
<td>−0.03</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Table 6. The 1st–7th thresholds with a sample size of $n = 10000$. Each estimator is the mean over 300 iterations, and the s.d. is their standard deviation.

<table>
<thead>
<tr>
<th>$n = 10000$</th>
<th>$r_n^{(k)}$</th>
<th>$\lambda_n^{(k)}$</th>
<th>$\mu_n^{(k)}$</th>
<th>$\sigma_n^{(k)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-stage</td>
<td>1.0</td>
<td>4.75</td>
<td>-0.07670</td>
<td>1.75522</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0</td>
<td>0.877</td>
<td>0.179</td>
<td>0.152</td>
</tr>
<tr>
<td>1-stage</td>
<td>0.18738</td>
<td>12.70</td>
<td>-0.03042</td>
<td>0.34009</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0150</td>
<td>1.283</td>
<td>0.034</td>
<td>0.016</td>
</tr>
<tr>
<td>2-stage</td>
<td>0.03205</td>
<td>14.57</td>
<td>-0.03005</td>
<td>0.30008</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0014</td>
<td>1.442</td>
<td>0.029</td>
<td>0.002</td>
</tr>
<tr>
<td>3-stage</td>
<td>0.02777</td>
<td>14.91</td>
<td>-0.02997</td>
<td>0.29957</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0003</td>
<td>1.495</td>
<td>0.029</td>
<td>0.002</td>
</tr>
<tr>
<td>4-stage</td>
<td>0.02760</td>
<td>14.95</td>
<td>-0.02995</td>
<td>0.29953</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0003</td>
<td>1.503</td>
<td>0.029</td>
<td>0.002</td>
</tr>
<tr>
<td>5-stage</td>
<td>0.02759</td>
<td>14.95</td>
<td>-0.02995</td>
<td>0.29952</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0003</td>
<td>1.504</td>
<td>0.029</td>
<td>0.002</td>
</tr>
<tr>
<td>6-stage</td>
<td>0.02759</td>
<td>14.95</td>
<td>-0.02995</td>
<td>0.29952</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0003</td>
<td>1.504</td>
<td>0.029</td>
<td>0.002</td>
</tr>
<tr>
<td>7-stage</td>
<td>0.02759</td>
<td>14.95</td>
<td>-0.02995</td>
<td>0.29952</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.0003</td>
<td>1.504</td>
<td>0.029</td>
<td>0.002</td>
</tr>
<tr>
<td>$r_{opt}^{(n)}/$True</td>
<td>0.02752</td>
<td>15.0</td>
<td>-0.03</td>
<td>0.3</td>
</tr>
</tbody>
</table>

These results show that our threshold selector can find $r_{opt}^{(n)}$ approximately at the limit of the $k$-stage threshold, and as a result, the parameters are estimated well. Although we do not know yet if the $k$-stage threshold can theoretically converge to a positive constant, we can easily imagine that $r_n^{(k)}$ stops absolutely after several stages in applications where only one sample path of $X$ is available and the sample size $n$ is fixed. Because, if $r_n^{(k)}$ goes near $r_{opt}^{(n)}$, then the difference $|r_n^{(k)} - r_{n}^{(k+1)}|$ is getting small. Consequently, the estimated jump’s number $I_k := \#\{i; |\Delta_i^n X| > r_n^{(k)}\}$ is not updated, that is, $I_k = I_{k+1}^{n}$ for sufficiently large $k$. Furthermore the estimators $\hat{\epsilon}_n$ and $\hat{f}_n$ are not updated either. This yields that $r_n^{(k)} = r_n^{(k+1)}$. Actually the variance of the last stage threshold is sufficiently small, and it indicates that we can always select the good threshold.

In this simulation, we chose the pilot threshold as $r_n^{(0)} = 1.0$, but we can check that any other choice of pilot threshold can also lead to similar results.

6. Theoretical discussion

6.1. What is the validity?

In this section, we investigate the asymptotic behavior of $r_n^{(k)}$ as $k \to \infty$ with fixed $n \in \mathbb{N}$, and as $n \to \infty$ after $k \to \infty$.

According to the numerical studies in the previous section, the sequence $\{r_n^{(k)}\}_{k \in \mathbb{N}}$ for fixed $n$ seems to converge to a positive constant, which is near $r_{opt}^{(n)}$. Consequently, the limit of $r_n^{(k)}$ as $k \to \infty$ becomes a good threshold, and the corresponding filter shows a high-performance. Therefore we at least expect that
the following convergences hold true: there exists a positive constant $\gamma_n$ for each $n \in \mathbb{N}$ such that

\begin{align}
\lim_{k \to \infty} r_n^{(k)} &= \gamma_n \quad \text{a.s.,} \\
\lim_{n \to \infty} \left( \gamma_n + \frac{\sqrt{h_n}}{\gamma_n} \right) &= 0.
\end{align}

On the first property (6.1), though our greatest hope is $\gamma_n = r_{opt}^{(n)}$ for each $n \in \mathbb{N}$, it may be impossible while $n$ is finite. Therefore it will be desirable that

\begin{equation}
\Delta_n := |\gamma_n - r_{opt}^{(n)}| \quad \text{is sufficiently small for large } n.
\end{equation}

The condition (6.3) is described in Theorem 6.1, (6.14) more clearly.

The second property (6.2) is a necessary condition for the asymptotic filter. The $\sqrt{h_n}$-order is the order of the variation of continuous part of process $X$. Therefore it is desired that the speed of the convergence of the threshold is slower than $\sqrt{h_n}$ in order to detect the variation of discontinuous part. In Shimizu and Yoshida (2006), they used the asymptotic filter as $\{ |\Delta_n X| > L h_n^\rho \}$ for a constant $L > 0$ and $\rho \in (0, \frac{1}{2})$, and this threshold certainly satisfies condition (6.2). Furthermore Mancini (2004) proposed a similar type of filter $\{ |\Delta_n X| > L \sqrt{h_n \log h_n^{-1}} \}$ for a constant $L > 0$, which also satisfied the above conditions. Similarly we demand those conditions on $\gamma_n$.

\subsection*{6.2. Mathematical validity}

In order to show the mathematical validities (6.1)–(6.3) of our plug-in method, we first make the following assumption.

A 1. The coefficients $a(x)$ and $b(x)$ of the stochastic differential equation (1.1) are known.

This assumption implies that the function $\epsilon(r)$ is implicitly known. First we suppose that $\epsilon(r)$ is known. Therefore we consider the case where

\begin{equation}
\hat{L}_n(r; s) = 2h_n^{-1} \epsilon(r) - \hat{J}_n(r; s).
\end{equation}

We make some remarks later on the case where $\epsilon(r)$ is unknown.

Let $\hat{I}_n(r; s)$ be an estimator of the integral $\int_{|z| \leq r} f(z) \, dz$ constructed in some way using the filter $\{ |\Delta_n X| > s \}$. We assume the following.

A 2. For any $r > 0$ and any $n \in \mathbb{N}$,

\begin{align}
\inf_{k \in \mathbb{N}} \hat{I}_n(r; r_n^{(k-1)}) &= 0 \quad \text{a.s.,} \\
\sup_{s_2 \leq s_1} \hat{I}_n(r; s_1) &\leq \hat{I}_n(r; s_2) \quad \text{a.s.}
\end{align}
Remark 6.1. The family of such estimators $\hat{L}_n(r; s)$ with above conditions is not empty. Actually, the density estimator (5.1) with (5.2) satisfies (6.5) if
\[
0 < r_n^{(k-1)} < \max_{1 \leq i \leq n} |\Delta_i^n X|
\]
for any $k \in \mathbb{N}$, and (6.6) clearly holds true.

Although the following results are clear from the definition of the function $L(r)$ and $\hat{L}_n(r; r_n^{(k-1)})$, we present them as a lemma since those results will be used repeatedly below.

**Lemma 6.1.** Suppose Condition A1. Then two functions $L(r)$ and $\hat{L}_n(r; r_n^{(k-1)})$ are strictly decreasing in $r$ for any $n, k \in \mathbb{N}$. Moreover the equations
\[
L(r) = 0 \quad \text{and} \quad \hat{L}_n(r; r_n^{(k-1)}) = 0
\]
have the unique roots $r^{(n)}_{\text{opt}}$ and $r_n^{(k)}$, respectively. Therefore, in particular,
\[
r^{(n)}_{\text{opt}} \leq r \Leftrightarrow L(r) \leq 0,
\]
\[
r_n^{(k)} \leq r \quad \text{a.s.} \Leftrightarrow \hat{L}_n(r; r_n^{(k-1)}) \leq 0 \quad \text{a.s.}
\]

The following theorem shows the monotonicity of the sequence $\{r_n^{(k)}\}_{k \in \mathbb{N}}$.

**Lemma 6.2.** Let $k \in \mathbb{N}$. Suppose Conditions A1 and A2. Then it follows that $r_n^{(k)} \geq r_n^{(k+1)}$ for all $\omega \in \{\omega \in \Omega; r_n^{(k-1)} \geq r_n^{(k)}\}$. Moreover it follows that $r_n^{(k)} \leq r_n^{(k+1)}$ for all $\omega \in \{\omega \in \Omega; r_n^{(k-1)} \leq r_n^{(k)}\}$.

**Proof.** Fix an arbitrary $\omega \in \{\omega \in \Omega; r_n^{(k-1)} \geq r_n^{(k)}\}$. From the definition of $r_n^{(k)}$,
\[
2h_n^{-1} \epsilon(r_n^{(k)}) = \hat{J}_n(r_n^{(k)}; r_n^{(k-1)}).
\]
Noticing that $\hat{J}_n(r; r_n^{(k)}) = \hat{J}_n(r/2; r_n^{(k)}) + \hat{J}_n(2r; r_n^{(k)})$, we see from the conditions (6.5) and (6.6) that
\[
\frac{2h_n^{-1} \epsilon(r_n^{(k)})}{\hat{J}_n(r_n^{(k)}; r_n^{(k)})} = \frac{\hat{J}_n(r_n^{(k)}; r_n^{(k-1)})}{\hat{J}_n(r_n^{(k)}; r_n^{(k)})} \leq 1.
\]
Hence
\[
\hat{L}_n(r_n^{(k)}; r_n^{(k)}) \leq 0.
\]
Lemma 6.1 yields that $r_n^{(k+1)} \leq r_n^{(k)}$. The last half of the statement follows by the same argument as above. □
We make further assumptions.

A 3. For any \( s \geq 0 \), the process \( \alpha^{(s)} \) satisfies for any \( p \geq 1 \) and any \( t \geq 0 \) with \( |t - s| < 1 \) that
\[
E[|\alpha_t^{(s)} - \alpha_s^{(s)}|^p] \leq C_p |t - s|^{p/2},
\]
where \( C_p \) is a positive constant depending on \( p \).

This assumption holds true if, for example, the coefficients \( a \) and \( b \) are bounded, or if \( X \) satisfies the conditions presented in Shimizu and Yoshida (2006).

A 4. For any \( c \in (0, 1/2) \), there exists a constant \( \delta > 0 \) such that
\[
0 < r_{\text{opt}}^{(n)} \leq r_n^{(k)} < \kappa h_c^n
\]
for each \( k, n \in \mathbb{N} \).

**Lemma 6.3.** Suppose Conditions A 1–A 4, and that \( h_n < 1 \) for any \( n \in \mathbb{N} \). Then, for any \( c \in (0, 1/2) \), there exists a constant \( \kappa > 0 \) which is independent of \( n \) such that
\[
0 < r_{\text{opt}}^{(n)} \leq r_n^{(k)} < \kappa h_c^n
\]
for each \( k, n \in \mathbb{N} \).

**Proof.** Under A 4, for any \( \delta > 0 \), \( c \in (0, 1/2) \) and large enough \( p \geq 1 \), there exists a constant \( \kappa_1 > 0 \) which is independent of \( n \) such that, for each \( n \in \mathbb{N} \),
\[
\kappa_1 \geq \left( \sup_n \frac{h_n^\delta}{J(k_1 h_c^n)} \right)^{1/p} (2C_p)^{1/p} \geq \left( \frac{2C_p h_n^\delta}{J(k_1 h_c^n)} \right)^{1/p}
\]
since \( J(k_1 h_c^n) \) is increasing in \( k_1 \) for each \( n \), where \( C_p \) is a constant given in A 3. For the constants \( \kappa_1 \) and \( c \) in (6.12), it follows from Chebysev’s inequality and A 3 that
\[
\frac{2h_n^{-1} \epsilon(k_1 h_c^n)}{J(k_1 h_c^n)} \leq \frac{2C_p}{h_n J(k_1 h_c^n)} \left( \frac{\sqrt{h_n}}{k_1 h_c^n} \right)^p \leq h_n^{p(1/2-c)-1-\delta}.
\]
Therefore, taking $p$ large enough such that $\delta < p(1/2 - c) - 1$, we find that the last term is less than 1. This implies by Lemma 6.1 that $r_{(n)}^{(n)} \leq h_n^c$. Similarly we also see that $\hat{L}_n(\kappa h_n^c; r_{(n)}^{(k-1)}) \leq 0$ for a constant $\kappa_2 > 0$. Hence the statement holds for $\kappa = \kappa_1 \lor \kappa_2$. □

**Remark 6.2.** Although the statement says for any $c \in (0, 1/2)$, we can not take $c = 1/2$ since it may be that $r_{(n)}^{(n)} \sim |\log h_n| \sqrt{h_n}$. Such a sequence satisfies $r_{(n)}^{(n)} \leq \kappa h_n^c$ for sufficiently large $n$ and any $c \in (0, 1/2)$. However $r_{(n)}^{(n)} > \kappa h_n^{1/2}$ for sufficiently large $n$.

**Lemma 6.4.** Suppose Conditions A 1–A 3, and fix any $c \in (0, 1/2)$ and $k, n \in \mathbb{N}$. Then it follows that

$$r_{(n)}^{(k)} < r_{(n)}^{(k-1)}$$

for all $\omega \in \{\omega \in \Omega; r_{(n)}^{(k-1)} > \kappa h_n^c\}$, where $\kappa$ is given in Lemma 6.3.

**Proof.** By the similar argument as in the proof of Lemma 6.3 and the constant $\kappa$ given in the lemma, we see that

$$\frac{2h_n^{-1} \epsilon(r_{(n)}^{(k-1)})}{\hat{J}_n(r_{(n)}^{(k-1)}; r_{(n)}^{(k-1)})} = 2(nh_n \hat{J}_n(r_{(n)}^{(k-1)}; r_{(n)}^{(k-1)}))^{-1} \sum_{i=1}^{n} P\{|\Delta_i^n X| > r_{(n)}^{(k-1)}, J_i^n = 0\}$$

$$\leq 2C_p(h_n \hat{J}_n(\kappa h_n^c; r_{(n)}^{(k-1)}))^{-1} \left(\frac{\sqrt{h_n}}{\kappa h_n^c}\right)^p < 1$$

for any $p \geq 1$. Hence Lemma 6.1 yields the consequence. □

Lemma 6.2 and 6.4 indicate a way to choose the pilot threshold. If we take $r_{(n)}^{(0)} > \kappa h_n^c$ then the sequence $\{r_{(n)}^{(k)}\}_{k \in \mathbb{N}}$ is almost surely decreasing. Since it is bounded from the bottom, it converges to a limit $\gamma_n$. Moreover, even if one chooses the pilot threshold as being too large, Lemma 6.4 ensures an improvement in the threshold. Indeed, Lemma 6.3 implies that $r_{(n)}^{(1)} \leq \kappa h_n^c$, which would be nearer to $r_{(n)}^{(0)}$ than $r_{(n)}^{(0)}$ since $r_{(opt)}^{(n)} \leq \kappa h_n^c$.

The validity (6.1) and the first half of the validity (6.2) is obtained by the following theorem.

**Theorem 6.1.** Suppose Conditions A 1–A 3. For arbitrary $r_{(n)}^{(0)}$ with $\hat{I}_n(r; r_{(n)}^{(0)}) > 0$ for any $r > 0$, there exists a positive constant $\gamma_n$ such that

$$\lim_{k \to \infty} r_{(n)}^{(k)} = \gamma_n \quad a.s.$$  

(6.13)

for any fixed $n \in \mathbb{N}$. Moreover

$$\Delta_n := |\gamma_n - r_{(opt)}^{(n)}| \leq \kappa h_n^c$$  

(6.14)
for any \( c \in (0, 1/2) \) and \( \kappa \) given in Lemma 6.3.

**Proof.** By Lemma 6.2 and 6.3, we can see that the sequence \( \{r^{(k)}_n\}_{k \in \mathbb{N}} \) is monotone and bounded. Therefore \( r^{(k)}_n \) converges to a limit \( \gamma_n \geq 0 \). For this \( \gamma_n \), we have \( \hat{L}_n(\gamma_n; \gamma_n) = 0 \). If \( \gamma_n = 0 \) then it must be \( \hat{L}_n(0; 0) = 0 \). However it contradicts that \( \hat{L}_n(r; 0) = 2h_n^{-1} > 0 \) for any \( r \geq 0 \). Hence \( \gamma_n > 0 \). The inequality \( \gamma_n \leq \kappa h_n^c \) is clear by Lemma 6.3. Therefore we obtain (6.14). \( \square \)

The result (6.14) was also one of the validities stated in (6.3). Therefore we find that \( \gamma_n \) is close to \( r_{opt}^{(n)} \) if the sample size \( n \) is sufficiently large, and we can check this phenomenon in the simulation results displayed in Section 5.

Let us consider the following quantity:

\[
\mathcal{D}_\alpha^{(n)}(r_1, r_2) = \frac{1}{n} \sum_{i=1}^{n} P\{r_1 \wedge r_2 \leq |\Delta_n^i \alpha(t_{i-1}^n)| \leq r_1 \vee r_2\},
\]

where \( \alpha \) is a solution process to (3.1). If the distribution function of \(|\alpha_t^s - \alpha_s^s|\) for any \( t, s \geq 0 \) is strictly increasing, or if the support of the probability density of \(|\alpha_t^s - \alpha_s^s|\) is \( \mathbb{R}_+ \), then \( \mathcal{D}_\alpha^{(n)}(r_1, r_2) \) for fixed \( n \) can be a distance between \( r_1 \) and \( r_2 \).

Let us consider an estimator of the Lévy density \( f \), and make a natural estimator \( \hat{I}_n \):

\[
\hat{I}_n(r; s) := \int_{|z| \leq r} \hat{f}_n(z; s) dz.
\]

In this case, we can estimate \( \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \) as in the next theorem. Although this estimate says nothing about the direct estimate of the error \( \Delta_n \), this gives us indirectly the order conditions for \( \gamma_n \) and \( r_{opt}^{(n)} \) which should be satisfied asymptotically; Condition (6.2).

**Theorem 6.2.** Suppose Conditions A 1–A 3, and let \( \hat{f}_n(z; r) := \hat{f}_n(z; r) - f(z) \). Then

\[
\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \leq \sqrt{\frac{33}{2} \kappa h_n^{1+c} \sup_{z \in \mathbb{R}_d} \|\hat{f}_n(z; \gamma_n)\|_{L^2(P)}}
\]

is valid for any \( c \in (0, 1/2) \) and \( \kappa \) given in Lemma 6.3.

**Proof.** First we suppose that \( r_{opt}^{(n)} \geq \gamma_n \). By the definition of \( r_{opt}^{(n)} \) and \( \gamma_n \), we have

\[
J(r_{opt}^{(n)}) = 2h_n^{-1} \epsilon(r_{opt}^{(n)}),
\]

\[
\hat{J}_n(\gamma_n; \gamma_n) = 2h_n^{-1} \epsilon(\gamma_n).
\]

Hence we obtain that

\[
\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) = \epsilon(\gamma_n) - \epsilon(r_{opt}^{(n)})
\]
\[
\frac{1}{2} h_n \left[ \hat{J}_n (\gamma_n ; \gamma_n) - J (r_{opt}^{(n)}) \right] \\
\leq \frac{1}{2} h_n \left[ \hat{J}_n (r_{opt}^{(n)} ; \gamma_n) - J (r_{opt}^{(n)}) \right].
\]

According to Jensen’s inequality, it follows that
\[
\left\{ \mathcal{D}_\alpha^{(n)} (\gamma_n, r_{opt}^{(n)}) \right\}^2 \leq \frac{h_n^2 r_{opt}^{(n)} }{2} \left[ \left( \frac{1}{r_{opt}^{(n)}} \int_{|z| \leq r_{opt}^{(n)} / 2} \bar{f}_n (z ; \gamma_n) \, dz \right)^2 + 16 \left( \frac{1}{4 r_{opt}^{(n)}} \int_{|z| \leq 2 r_{opt}^{(n)}} \bar{f}_n (z ; \gamma_n) \, dz \right)^2 \right]
\leq \frac{h_n^2 r_{opt}^{(n)} }{2} \left[ \int_{|z| \leq r_{opt}^{(n)} / 2} \bar{f}_n^2 (z ; \gamma_n) \, dz + 4 \int_{|z| \leq 2 r_{opt}^{(n)}} \bar{f}_n^2 (z ; \gamma_n) \, dz \right].
\]

Taking the expectation in both sides, we obtain that
\[
\mathcal{D}_\alpha^{(n)} (\gamma_n, r_{opt}^{(n)}) \leq \sqrt{\frac{33}{2} h_n r_{opt}^{(n)} } \sup_{z \in E} \| \bar{f}_n (z ; \gamma_n) \|_{L^2 (P)},
\]
and Lemma 6.3 yields the consequence.

When \( r_{opt}^{(n)} < \gamma_n \), the same argument as above holds since
\[
\mathcal{D}_\alpha^{(n)} (\gamma_n, r_{opt}^{(n)}) = \epsilon (r_{opt}^{(n)}) - \epsilon (\gamma_n) \\
\leq \frac{h_n}{2} \left[ J (r_{opt}^{(n)}) - \hat{J}_n (r_{opt}^{(n)} ; \gamma_n) \right].
\]

This completes the proof. \( \square \)

Let us use the kernel density estimator proposed in Shimizu (2006) as \( \hat{f}_n \):
\[
(6.16) \quad \hat{f}_n (z) = \frac{1}{n h_n^{1+\delta} } \sum_{i=1}^{n} K \left( \frac{z - \Delta_i^n X}{\delta_n} \right) I (\mathcal{H}_i^n (r_n)),
\]
where \( \delta_n \) satisfies \( \delta_n^d = h_n^\delta \) for a constant \( \delta \in (0, 1/2) \), and \( K \) is a bounded kernel with some conditions described in Shimizu (2006). If the true density \( f \) is bounded, then we find by rough estimate that \( \sup_{z \in \mathbb{R}^d} \| \bar{f}_n (z ; \gamma_n) \|_{L^2 (P)} = O (h_n^{-(1+\delta)}) \) for any sequence \( \gamma_n \). Therefore we obtain from Theorem 6.2 that
\[
(6.17) \quad \mathcal{D}_\alpha^{(n)} (\gamma_n, r_{opt}^{(n)}) \leq O (h_n^{c-\delta}).
\]
Noticing that \( c \in (0, 1/2) \) can be taken arbitrarily, we have
\[
\mathcal{D}_\alpha^{(n)} (\gamma_n, r_{opt}^{(n)}) \to 0 \quad (n \to \infty)
\]
by taking \( c > \delta \). We state this fact as a corollary.
Corollary 6.1. Suppose Conditions A 1–A 3, and that f is bounded. Let \( \hat{f}_n \) be the kernel density estimator (6.16). Then

\[
\limsup_{n \to \infty} \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) = 0.
\]

The last half of the validities (6.2) remain to be studied. First, let us consider the simplest case where \( \alpha^{(s)} \) is a one-dimensional Brownian motion for each \( s > 0 \): \( \alpha^{(s)}_t = X_s + \mu t + \sigma w_t \) for constants \( \mu \in \mathbb{R}, \sigma > 0 \). Then we can show that \( \sqrt{h_n} r_{opt}^{(n)} = o_p(1) \). Actually, considering a sequence \( r_n = M_n \sqrt{h_n} + o_p(\sqrt{h_n}) \) with \( M_n \to M \geq 0, \) we have

\[
L(r_n) = 2h_n^{-1} \epsilon(r_n) - \mathcal{J}(r_n)
\]

\[
\geq 2h_n^{-1} \left[ \int_{r_n}^{\infty} \frac{1}{\sqrt{2\pi \sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du + \int_{-\infty}^{-r_n} \frac{1}{\sqrt{2\pi \sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du \right] - 2\lambda_0
\]

\[
\sim 2h_n^{-1} \left[ \int_{M_n}^{\infty} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du + \int_{-\infty}^{-M_n} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du \right] - 2\lambda_0.
\]

Therefore \( L(r_n) > 0 \) for sufficiently large \( n \), and Lemma 6.1 yields \( r_{opt}^{(n)} > r_n \) a.s. for sufficiently large \( n \). Since we can take \( M_n \) as an arbitrary sequence while \( M_n \to M \geq 0 \), it follows that \( r_{opt}^{(n)}/\sqrt{h_n} \to \infty \) a.s. by \( M \to \infty \).

In this situation, we can show that \( \sqrt{h_n} \gamma_n^{-1} = o(1) \) from Corollary 6.1. For example, assume that \( \gamma_n \sim M \sqrt{h_n} \) for a constant \( M \geq 0 \). For sufficiently large \( n \), we have

\[
\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) = E \left[ \int_{\gamma_n}^{r_{opt}^{(n)}} \frac{1}{\sqrt{2\pi \sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du + \int_{-\gamma_n}^{-r_{opt}^{(n)}} \frac{1}{\sqrt{2\pi \sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du \right]
\]

\[
\sim \int_{\sqrt{M}}^{r_{opt}^{(n)}/\sqrt{h_n}} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{v^2}{2\sigma^2}} dv + \int_{-r_{opt}^{(n)}/\sqrt{h_n}}^{-\sqrt{M}} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{v^2}{2\sigma^2}} dv
\]

\[
\geq \Delta(M),
\]

where \( \Delta(M) \) is a positive constant. This contradicts Corollary 6.1. These facts are proved more generally in the next theorem.

In the general case where \( \alpha^{(s)} \) is a diffusion process (3.1), we suppose the following condition.
A 5. For any $t \geq 0$, the process $\alpha^{(t)}$ has the transition density $p(h, x, y)$: $P\{\alpha^{(t)}_{t+h} \in A \mid X_t = x\} = \int_A p(h, x, y)dy$ for any $h > 0$, such that

$$p(h, x, y) \geq \frac{K}{\sqrt{h}} \exp \left( c h |x|^2 - \frac{|x - y|^2}{ch} \right)$$

for constants $c > 1$ and $K > 1$.

Although one might think that this condition is not easy to be checked, we can replace this condition with a more concrete one using the coefficients of the SDE (3.1); see Gobet (2002) which gives us a sufficient condition for A 5 under the elliptic diffusion case by using the coefficient of the SDE, and it can be easily checked. However we need only the fact (6.19) for our purpose, so we dare to impose this condition directly.

We obtain the following theorem, which was one of the validities presented in (6.2).

**Theorem 6.3.** Suppose Conditions A 1–A 5 and that $f$ is bounded. Then

$$\lim_{n \to \infty} \sqrt{h_n} (\gamma_n^{-1} + (r_{opt}^{(n)})^{-1}) = 0 \quad \text{a.s.}$$

**Proof.** The process $\alpha$ follows the equation

$$\alpha^{(t_n)}_{t_{i-1}} = X^{(t_n)}_{t_{i-1}} + \int_{t_{i-1}}^{t_n} a(\alpha^{(t_n)}_{s})ds + \int_{t_{i-1}}^{t_n} b(\alpha^{(t_n)}_{s})dw_s,$$

for $t \in (t_{i-1}, t_i)$, and suppose that $\alpha^{(t_n)}_{t_{i-1}}$ has a transition density satisfying (6.19).

It follows from the lower bound (6.19) that

$$\begin{align*}
P\{r_n \leq \Delta_n^{(n)} \leq R_n\} &= E[P\{r_n + X^{(n)}_{t_{i-1}} \leq \alpha^{(n)}_{t_{i-1}} \leq R_n + X^{(n)}_{t_{i-1}} \mid X^{(n)}_{t_{i-1}}\}] \\
&\geq E \left[ \int_{r_n + X^{(n)}_{t_{i-1}}}^{R_n + X^{(n)}_{t_{i-1}}} \frac{K}{\sqrt{h_n}} \exp \left( ch_n |X^{(n)}_{t_{i-1}}|^2 - \frac{|y - X^{(n)}_{t_{i-1}}|^2}{ch_n} \right) dy \right] \\
&= E \left[ \exp(ch_n |X^{(n)}_{t_{i-1}}|^2) \int_{r_n/\sqrt{h_n}}^{R_n/\sqrt{h_n}} Ke^{-y^2/c}dy \right] \\
&\geq E \left[ \int_{r_n/\sqrt{h_n}}^{R_n/\sqrt{h_n}} Ke^{-y^2/c}dy \right].
\end{align*}$$

Putting $r_n = M_n \sqrt{h_n} + o(\sqrt{h_n})$, where $M_n$ is an arbitrary sequence satisfying $M_n \to M$ $(0 \leq M < \infty)$, and $R_n = \infty$, then we obtain that

$$L(r_n) \geq 2h_n^{-1} \int_{M_n + o(1)}^{\infty} Ke^{-y^2/c}dy - 2\lambda_0 > 0$$
for sufficiently large $n$. Hence we obtain that $r_{opt}^{(n)} > r_n$ a.s. by Lemma 6.1. This implies that $\sqrt{h_n} (r_{opt}^{(n)})^{-1} = o_p(1)$.

Let $r_n = \gamma_n \wedge r_{opt}^{(n)}$ and $R_n = \gamma_n \vee r_{opt}^{(n)}$. Then

$$\mathcal{D}_\alpha^{(n)}(r_n, R_n) = P\{r_n \leq |\Delta_t^n \alpha(t^n_{i-1})| \leq R_n\} = P\{r_n \leq \Delta_t^n \alpha(t^n_{i-1}) \leq R_n\} + P\{-R_n \leq \Delta_t^n \alpha(t^n_{i-1}) \leq -r_n\} \geq \frac{1}{E} \left[ \int_{\gamma_n \vee \sqrt{h_n}}^{\gamma_n \wedge \sqrt{h_n}} Ke^{-y^2/c} dy + \int_{\gamma_n \wedge \sqrt{h_n}}^{-\gamma_n \vee \sqrt{h_n}} Ke^{-y^2/c} dy \right].$$

If we suppose that $\gamma_n \leq r_{opt}^{(n)}$ a.s. for sufficiently large $n$, we have

$$\liminf_{n \to \infty} \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \geq \frac{1}{E} \left[ \liminf_{n \to \infty} \left\{ \int_{\gamma_n \vee \sqrt{h_n}}^{\gamma_n \wedge \sqrt{h_n}} Ke^{-y^2/c} dy + \int_{\gamma_n \wedge \sqrt{h_n}}^{-\gamma_n \vee \sqrt{h_n}} Ke^{-y^2/c} dy \right\} \right] \geq 0.$$  

This implies that $\sqrt{h_n} \gamma_n^{-1} + \sqrt{h_n} (r_{opt}^{(n)})^{-1} = o(1)$ since $\sqrt{h_n} (r_{opt}^{(n)})^{-1} = o(1)$, and the right hand side of the last inequality must be zero from Corollary 6.1. When $\gamma_n \geq r_{opt}^{(n)}$ for sufficiently large $n$, the same argument holds. This completes the proof. □

7. To a practical approach

So far, we have discussed a kind of validity of the plug-in method in the case where the function $\epsilon(r)$ is known, and we obtained sufficient conditions for (6.1)–(6.3).

(i) The properties (6.1) and (6.3) hold true under A 1–A 4.

(ii) The property (6.2) holds true under A 1–A 5 and if $f$ is bounded.

Of course we are actually interested in the case where $\epsilon(r)$ is unknown; $\epsilon(r)$ is replaced by $\tilde{\epsilon}_n(r; s)$. In Section 5, we tried some simulations in such a case and we obtained some preferable results. However we can not show yet rigorous validity when $\epsilon(r)$ is unknown. This is the critical problem for the future.

Nevertheless this method is often useful. Let

$$\tilde{L}_n(r; r_n^{(k-1)}) := 2h_n^{-1} \tilde{\epsilon}_n(r; r_n^{(k-1)}) - \hat{J}_n(r; r_n^{(k-1)}),$$

and note that

$$\tilde{L}_n(r; r_n^{(k-1)}) = \hat{L}_n(r; r_n^{(k-1)}) + 2h_n^{-1} \Delta \epsilon_n(r; r_n^{(k-1)}),$$

where $\Delta \epsilon_n(r; r_n^{(k-1)}) := \tilde{\epsilon}_n(r; r_n^{(k-1)}) - \epsilon(r)$ and $\hat{L}_n$ is given in (6.4).

If we choose the pilot threshold $r_n^{(0)}$ to be relatively small, and (6.7), then many simulations show that the estimator $\hat{b}_n(x; r_n^{(0)})$ becomes relatively robust;
see also Shimizu (2006), and \( \hat{\epsilon}_n(r; r_n^{(k-1)}) \) gives a good approximation of \( \epsilon(r) \). Consequently, \( \Delta \epsilon \) becomes almost zero, and this implies that finding the root of \( \hat{L}_n(r; r_n^{(0)}) = 0 \) is similar to finding the root of \( \tilde{L}_n(r; r_n^{(0)}) = 0 \) in the 1st-stage. The same argument is possible after that stage and the discussion in the previous section can be approximately applied to the case where \( \epsilon(r) \) is unknown.

These are the empirical and intuitive explanations why our plug-in rule with unknown \( \epsilon(r) \) works well. We need more rigorous study for the case where \( \epsilon(r) \) is unknown. Moreover the case where the jump part is \( c(X_{t-})dZ_t \), or more general cases should be also studied in the future.

**Acknowledgements**

The author expresses many thank to the associate editor and the anonymous referees for their helpful comments and exact corrections. He is also grateful to Professor Masanobu Taniguchi for his help. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 19740049, 2007–2008.

**References**


