J. Appl. Prob. **50,** 931–942 (2013)
Printed in England
© *Applied Probability Trust* 2013

NONPARAMETRIC ESTIMATION FOR A CLASS OF PIECEWISE-DETERMINISTIC MARKOV PROCESSES

TAKAYUKI FUJII,* *Osaka University*

Abstract

In this paper we study nonparametric estimation problems for a class of piecewise-deterministic Markov processes (PDMPs). Borovkov and Last (2008) proved a version of Rice's formula for PDMPs, which explains the relation between the stationary density and the level crossing intensity. From a statistical point of view, their result suggests a methodology for estimating the stationary density from observations of a sample path of PDMPs. First, we introduce the local time related to the level crossings and construct the local-time estimator for the stationary density, which is unbiased and uniformly consistent. Secondly, we investigate other estimation problems for the jump intensity and the conditional jump size distribution.

Keywords: Piecewise-deterministic Markov process; local time; stationary density; uniform consistency; nonparametric estimation

2010 Mathematics Subject Classification: Primary 62G20
Secondary 60J55

1. Introduction

Let us consider a piecewise-deterministic Markov process (PDMP) driven by a drift coefficient μ, a jump intensity λ, and a jump size distribution J. We suppose that the sample path of the PDMP strictly increases continuously between each jump and decreases only by jumps. Furthermore, the PDMP is assumed to be stationary and ergodic with stationary density f. This type of PDMP includes well-known stochastic models, for instance, risk processes in mathematical insurance, stress release processes in seismology, and many others. For general information about the PDMP, we refer the reader to Davis (1993).

Our main purpose in this paper is to study the problems of nonparametric estimation of these characteristic functions f, λ, and J by observation of a sample path of the PDMP. One of the motivations for this work comes from Borovkov and Last's (2008) proof of the following version of Rice's formula:

$$\nu(x) = \mu(x) f(x), \qquad x \in \mathbb{R}.$$

Here $\nu(x)$ is the continuous crossing intensity of the level x. Rice's formula is classical in stochastic analysis and mainly applied in engineering fields; see, e.g. Rice (1944) and Leadbetter (1966). This theoretically important result suggests that the estimation of the stationary density $f(x)$ is associated with the number of crossings of level x. In a related work, the level crossing method was developed to study the stationary distribution of queues; see Brill and Posner (1977).

Received 19 January 2012; revision received 8 February 2013.
* Current address: Faculty of Economics, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan. Email address: takayuki-fujii@biwako.shiga-u.ac.jp

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

In the present paper we first introduce the local time, which is related to the number of upcrossings, and construct the local-time estimator (LTE) for the stationary density. We note that the LTE is familiar in the context of density estimation for stationary processes; see Bosq (1998) and Bosq and Davydov (1999). Detailed studies on the LTE have been carried out in the case of diffusion processes (see Kutoyants (1998), (2004) and van Zanten (2000)) since the properties of the local time of diffusion processes are well known (see, e.g. Karatzas and Shreve (1991)). Recently, Fujii and Nishiyama (2010) obtained some properties of the LTE of the basic stress release process. The present study is a generalization of this work.

As a preliminary step, the stochastic integral representation of the local time of a PDMP is presented. This formula is derived by noting the simple relation that the number of upcrossings at level x is almost equal to that of the downcrossings for our considered PDMPs, where the number of the downcrossings is given by the random counting measure. The stochastic integral form plays an important role in studying the properties (e.g. uniform consistency) of the LTE. By using the LTE, we also address estimation problems for the jump intensity and the conditional jump size distribution of PDMPs.

2. Model and assumptions

Let us first introduce the model treated in this paper. The following notation is based on that used in Borovkov and Last (2008). Consider a marked point process $\Phi = \{(T_n, Z_n)\}_{n\in\mathbb{N}}$ on $[0, \infty)$ with $0 < T_1 < T_2 < \dots$ and nonnegative random variables (marks) Z_n. We define the PDMP X_t by

$$X_t = X_0 + \int_0^t \mu(X_{s-})\,\mathrm{d}s - \sum_{n=1}^{N_t} Z_n, \qquad t \in [0, T], \tag{1}$$

where X_0 is an initial random variable,

$$N_t = \#\{n \in \mathbb{N} : T_n \le t\}, \qquad t \ge 0$$

($\#A$ denotes the number of elements in a set A), and μ is a continuous function on $\mathbb{R}$ satisfying $\inf_{x\in\mathbb{R}} \mu(x) > 0$. We assume that, for any $x \in \mathbb{R}$, there exists a unique continuous function $q_x(\cdot)\colon [0, \infty) \to \mathbb{R}$ satisfying the integral equation

$$q_x(t) = x + \int_0^t \mu(q_x(s))\,\mathrm{d}s, \qquad t \ge 0.$$

Therefore, between each jump, the PDMP (1) is represented by

$$X_t = q_{X_{T_n}}(t - T_n) \quad \text{on } t \in [T_n, T_{n+1}) \text{ for } n \ge 0,$$

where we set $T_0 = 0$. We note that the drift function $\mu(x)$ can always be assumed to be known, since it is obtained by taking the right limit

$$\lim_{\delta\to 0} \frac{X_{t+\delta} - X_t}{\delta} = \lim_{\delta\to 0} \frac{1}{\delta} \int_t^{t+\delta} \mu(X_{s-})\,\mathrm{d}s = \mu(X_t).$$

This is a reason for the name 'piecewise deterministic'.

Let us introduce some assumptions on the jump intensity and the jump size distribution for PDMP (1). The occurrence of jumps is governed by the stochastic jump intensity $\lambda(X_t)$, which

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

is assumed to be measurable and locally bounded, and to satisfy

$$\int_0^\infty \lambda(q_x(t))\,\mathrm{d}t = \infty, \qquad x \in \mathbb{R}. \tag{2}$$

For $n \geq 0$, the conditional distributions of $T_{n+1} - T_n$ given X_{T_n} are specified by

$$\mathbb{P}(T_{n+1} - T_n \leq t \mid X_{T_n}) = 1 - \exp\left\{-\int_0^t \lambda(q_{X_{T_n}}(s))\,\mathrm{d}s\right\} \quad \mathbb{P}\text{-a.s.} \tag{3}$$

So assumption (2) implies that the jump epochs T_n are all finite almost surely (a.s.). The conditional distribution functions of the jump size are given by

$$\mathbb{P}(Z_n \in A \mid X_{T_n-}) = J(X_{T_n-}, A) \quad \mathbb{P}\text{-a.s.},\ n \geq 1, \tag{4}$$

where A is a measurable subset on $[0, \infty)$. We also suppose that $J(x, \{0\}) = 0$ for all $x \in \mathbb{R}$, that is, $Z_n \neq 0$ for all $n \geq 1$.

In general, the marked point process is characterized by a random counting measure given by

$$\Phi(t, A) = \#\{n \in \mathbb{N} \mid T_n \leq t,\ Z_n \in A\},$$

where $t \geq 0$ and A is a measurable subset on $[0, \infty)$. By using this counting measure, PDMP (1) can be rewritten as the stochastic integral

$$X_t = X_0 + \int_0^t \mu(X_{s-})\,\mathrm{d}s - \int_0^t \int_0^\infty z\Phi(\mathrm{d}s, \mathrm{d}z), \qquad t \in [0, T].$$

We define $\mathcal{F}_t$ to be the filtration generated by X. Then, from (3) and (4), we have

$$\mathbb{E}\left[\int_0^\infty \int_0^\infty h(s, z)\Phi(\mathrm{d}s, \mathrm{d}z)\right] = \mathbb{E}\left[\int_0^\infty \int_0^\infty h(s, z)J(X_{s-}, \mathrm{d}z)\lambda(X_{s-})\,\mathrm{d}s\right] \tag{5}$$

for any predictable process $h(s, z)$, and, furthermore, if the predictable process $h(s, z)$ satisfies

$$\mathbb{E}\left[\int_0^t \int_0^\infty |h(s, z)|J(X_{s-}, \mathrm{d}z)\lambda(X_{s-})\,\mathrm{d}s\right] < \infty, \qquad t \geq 0,$$

then the process

$$\int_0^t \int_0^\infty h(s, z)\{\Phi(\mathrm{d}s, \mathrm{d}z) - J(X_{s-}, \mathrm{d}z)\lambda(X_{s-})\,\mathrm{d}s\}, \qquad t \geq 0,$$

becomes an $\mathcal{F}_t$-martingale; see Chapter 8 of Brémaud (1981) and Chapter 4 of Last and Brandt (1995).

In the following, we assume that PDMP (1) is ergodic with invariant density f. Also, PDMP (1) is assumed to be stationary, that is, the initial random variable X_0 is governed by f. We note that the condition

$$\liminf_{x \to -\infty}(\mu(x) - \lambda(x)m(x)) > 0 > \limsup_{x \to \infty}(\mu(x) - \lambda(x)m(x)),$$

where

$$m(x) = \int_0^\infty zJ(x, \mathrm{d}z),$$

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

is well known and reasonable for the ergodicity. We now further assume that

$$\int_{\mathbb{R}} \lambda(x) f(x)\,\mathrm{d}x < \infty. \tag{6}$$

Precise sufficient conditions for these two assumptions are discussed in Appendix A of Borovkov and Last (2008); see also Last (2004).

3. Local time of the PDMP

In this section we provide some auxiliary results for the local time $\Lambda_T(x)$ of PDMP (1).

Definition 1. The local time $\Lambda_T(x)$ of PDMP (1) is defined by

$$\Lambda_T(x) = \frac{1}{\mu(x)} \#\{t\colon X_t = x,\ 0 \le t \le T\}. \tag{7}$$

The local time $\Lambda_T(x)$ is regarded as the occupation density, since, for any Borel measurable function $g(x)$, we have

$$\begin{aligned}
\int_0^T g(X_t)\,\mathrm{d}t &= \sum_{n=1}^{\infty} \int_{[T_{n-1}\wedge T, T_n \wedge T)} g(X_t)\,\mathrm{d}t \\
&= \sum_{n=1}^{\infty} \int_{[T_{n-1}\wedge T, T_n \wedge T)} g\biggl(X_{T_{n-1}} + \int_{T_{n-1}}^{t} \mu(X_s)\,\mathrm{d}s\biggr)\,\mathrm{d}t \\
&= \sum_{n=1}^{\infty} \int_{[X_{T_{n-1}\wedge T}, X_{T_n \wedge T})} g(x) \frac{1}{\mu(x)}\,\mathrm{d}x \\
&= \int_{\mathbb{R}} g(x) \Lambda_T(x)\,\mathrm{d}x.
\end{aligned}$$

For all $T > 0$, the local time (7) is right continuous with a left limit in x and can be used as the occupation density of its left limit process since $\Lambda_T(x) = \tilde{\Lambda}_T(x)$ for almost all x, where $\tilde{\Lambda}_T(x)$ is the local time for the left limit process X_{t-}, defined similarly as in (7).

Note that the number of upcrossings at level x is almost equal to that of the downcrossings, which can be represented by

$$\int_0^T \int_0^{\infty} \mathbf{1}_{\{x < X_{t-} \le x+z\}}\, \Phi(\mathrm{d}t, \mathrm{d}z),$$

where $\mathbf{1}_A$ is the indicator function of the set A. Then the first jump overcounts for upcrossings if $X_0 > x$. On the other hand, the last upcrossing cannot be counted by the jump if $X_T > x$; see Figure 1 for an illustration. Hence, we provide the stochastic integral representation of the local time $\Lambda_T(x)$ as

$$\Lambda_T(x) = \frac{1}{\mu(x)} \biggl\{ \mathbf{1}_{\{X_T > x\}} - \mathbf{1}_{\{X_0 > x\}} + \int_0^T \int_0^{\infty} \mathbf{1}_{\{x < X_{t-} \le x+z\}}\, \Phi(\mathrm{d}t, \mathrm{d}z) \biggr\}. \tag{8}$$

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

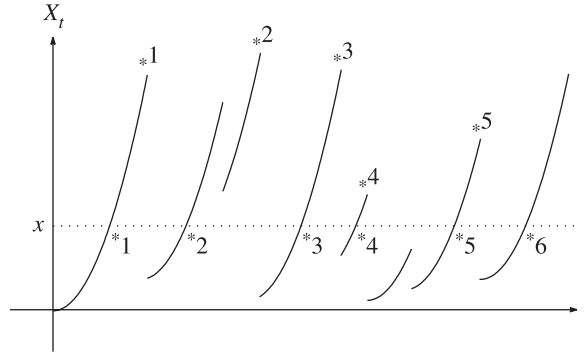

FIGURE 1: Relation between the upcrossing at level x and the jump.

4. Results

First, we consider the nonparametric estimation of the stationary density f of PDMP (1) in the situation where the conditional intensity $\lambda(\cdot)$ and the jump size distribution $J(\cdot,\cdot)$ are unknown. We denote the LTE $f_T^\circ(x)$ for the stationary density of PDMP (1) by

$$f_T^\circ(x) = \frac{\Lambda_T(x)}{T}. \tag{9}$$

As mentioned in Bosq and Davydov (1999), the LTE $f_T^\circ(x)$ is an unbiased estimator of $f(x)$, i.e. $f(x) = \mathbb{E}[f_T^\circ(x)]$ for any T. Therefore, by (5) and (8), we obtain the relation

$$\begin{aligned} f(x) &= \frac{1}{T}\frac{1}{\mu(x)}\mathbb{E}\biggl[\int_0^T\int_0^\infty \mathbf{1}_{\{x<X_{t-}\le x+z\}}\,\Phi(\mathrm{d}t,\mathrm{d}z)\biggr] \\ &= \frac{1}{\mu(x)}\int_{\mathbb{R}}\int_0^\infty \mathbf{1}_{\{x<y\le x+z\}}\,J(y,\mathrm{d}z)\lambda(y)f(y)\,\mathrm{d}y. \end{aligned} \tag{10}$$

We now show the uniform consistency of LTE (9). For this proof, we recall the definition of bracketing numbers of a seminormed space $(\Psi, \|\cdot\|)$. Given two functions ℓ and u, each of which may or may not belong to Ψ, the bracket $[\ell, u]$ is the set of all functions $\psi \in \Psi$ with $\ell \le \psi \le u$. An ε-bracket is a bracket $[\ell, u]$ such that $\|\ell - u\| < \varepsilon$. The bracketing number $N_{[\,]}(\varepsilon, \Psi, \|\cdot\|)$ is the minimum number of ε-brackets needed to cover Ψ. For more information of this method, we refer the reader to van der Vaart and Wellner (1996). In the following, we always consider the limit $T \to \infty$.

Theorem 1. *The LTE f_T° is uniformly consistent, i.e.*

$$\sup_{x\in\mathbb{R}} |f_T^\circ(x) - f(x)| \xrightarrow{\mathbb{P}} 0.$$

Proof. For the proof, it is sufficient to show that

$$\sup_{x\in\mathbb{R}} |\mu(x)f_T^\circ(x) - \mu(x)f(x)| \xrightarrow{\mathbb{P}} 0,$$

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

since $\inf_{x\in\mathbb{R}} \mu(x) > 0$. We note that

$$\begin{aligned}
&|\mu(x) f_T^\circ(x) - \mu(x) f(x)| \\
&\quad \le \frac{1}{T} |\, \mathbf{1}_{\{X_T > x\}} - \mathbf{1}_{\{X_0 > 0\}} \,| \\
&\qquad + \left| \frac{1}{T} \int_0^T \int_0^\infty \mathbf{1}_{\{x < X_{t-} \le x+z\}} \, J(X_{t-}, \mathrm{d}z) \lambda(X_{t-}) \, \mathrm{d}t - \mu(x) f(x) \right| \\
&\qquad + \left| \frac{1}{T} \int_0^T \int_0^\infty \mathbf{1}_{\{x < X_{t-} \le x+z\}} \{\Phi(\mathrm{d}t, \mathrm{d}z) - J(X_{t-}, \mathrm{d}z) \lambda(X_{t-}) \, \mathrm{d}t\} \right|
\end{aligned} \tag{11}$$

and take the supremum over all $x \in \mathbb{R}$. The first term on the right-hand side of (11) obviously converges to 0. In order to evaluate the other two terms, we introduce the class

$$\mathcal{H} = \{h(y, z) = \mathbf{1}_{\{x < y \le x+z\}}, \ x \in \mathbb{R}\}.$$

By Theorem 2.7.5 of van der Vaart and Wellner (1996), for any $\varepsilon > 0$, the class $\mathcal{H}$ satisfies $N(\varepsilon) = N_{[\,]}(\varepsilon, \mathcal{H}, \|\cdot\|) < \infty$, where the seminorm is given by

$$\|h\| = \int_{\mathbb{R}} \int_0^\infty h(y, z) J(y, \mathrm{d}z) \lambda(y) f(y) \, \mathrm{d}y.$$

Then there exist $N(\varepsilon)$-brackets $[l_k, u_k]$, $k = 1, \dots, N(\varepsilon)$, such that

$$\varepsilon_k = \int_{\mathbb{R}} \int_0^\infty |u_k(y, z) - l_k(y, z)| J(y, \mathrm{d}z) \lambda(y) f(y) \, \mathrm{d}y < \varepsilon.$$

For the second term in (11), it follows from (10) that, for any $h \in [l_k, u_k]$,

$$\begin{aligned}
&\frac{1}{T} \int_0^T \int_0^\infty h(X_{t-}, z) J(X_{t-}, \mathrm{d}z) \lambda(X_{t-}) \, \mathrm{d}t - \mu(x) f(x) \\
&\qquad \le \frac{1}{T} \int_0^T \int_0^\infty u_k(X_{t-}, z) J(X_{t-}, \mathrm{d}z) \lambda(X_{t-}) \, \mathrm{d}t \\
&\qquad\quad - \int_{\mathbb{R}} \int_0^\infty l_k(y, z) J(y, \mathrm{d}z) \lambda(y) f(y) \, \mathrm{d}y.
\end{aligned}$$

Therefore, the ergodicity of X_{t-} leads to

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \left\{ \frac{1}{T} \int_0^T \int_0^\infty h(X_{t-}, z) J(X_{t-}, \mathrm{d}z) \lambda(X_{t-}) \, \mathrm{d}t - \mu(x) f(x) \right\} \\
&\quad \le \max_{1 \le k \le N(\varepsilon)} \left\{ \frac{1}{T} \int_0^T \int_0^\infty u_k(X_{t-}, z) J(X_{t-}, \mathrm{d}z) \lambda(X_{t-}) \, \mathrm{d}t \right. \\
&\qquad\qquad \left. - \int_{\mathbb{R}} \int_0^\infty l_k(y, z) J(y, \mathrm{d}z) \lambda(y) f(y) \, \mathrm{d}y \right\} \\
&\quad \to \max_{1 \le k \le N(\varepsilon)} \left\{ \int_{\mathbb{R}} \int_0^\infty u_k(y, z) J(y, \mathrm{d}z) \lambda(y) f(y) \, \mathrm{d}y \right. \\
&\qquad\qquad \left. - \int_{\mathbb{R}} \int_0^\infty l_k(y, z) J(y, \mathrm{d}z) \lambda(y) f(y) \, \mathrm{d}y \right\} \\
&\quad < \varepsilon.
\end{aligned}$$

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

By an analogous argument for the lower bound

$$\begin{aligned}\inf_{x\in\mathbb{R}}&\left\{\frac{1}{T}\int_0^T\int_0^\infty h(X_{t-},z)J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t-\mu(x)f(x)\right\}\\&\geq\min_{1\leq k\leq N(\varepsilon)}\left\{\frac{1}{T}\int_0^T\int_0^\infty l_k(X_{t-},z)J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t\right.\\&\qquad\left.-\int_{\mathbb{R}}\int_0^\infty u_k(y,z)J(y,\mathrm{d}z)\lambda(y)f(y)\,\mathrm{d}y\right\},\end{aligned}$$

we obtain

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_0^T\int_0^\infty \mathbf{1}_{\{x<X_{t-}\leq x+z\}}\,J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t-\mu(x)f(x)\right|\xrightarrow{\mathbb{P}}0,$$

since the choice of ε is arbitrary.

The proof is completed by showing that

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_0^T\int_0^\infty \mathbf{1}_{\{x<X_{t-}\leq x+z\}}\{\Phi(\mathrm{d}t,\mathrm{d}z)-J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t\}\right|\xrightarrow{\mathbb{P}}0,$$

which can be done similarly. That is,

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left\{\frac{1}{T}\int_0^T\int_0^\infty \mathbf{1}_{\{x<X_{t-}\leq x+z\}}\{\Phi(\mathrm{d}t,\mathrm{d}z)-J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t\}\right\}\\&\quad\leq\max_{1\leq k\leq N(\varepsilon)}\left\{\frac{1}{T}\int_0^T\int_0^\infty u_k(X_{t-},z)\{\Phi(\mathrm{d}t,\mathrm{d}z)-J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t\}\right\}\\&\qquad+\max_{1\leq k\leq N(\varepsilon)}\left\{\frac{1}{T}\int_0^T\int_0^\infty \{u_k(X_{t-},z)-l_k(X_{t-},z)\}J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t\right\}\\&\quad\xrightarrow{\mathbb{P}}0+\max_{1\leq k\leq N(\varepsilon)}\varepsilon_k\\&\quad<\varepsilon,\end{aligned}$$

where the convergence of the martingale part is obtained by using Doob's inequality (see, e.g. Theorem 1.9.1 of Liptser and Shiryaev (1989)), i.e. for each k and $\eta>0$,

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{T}\sup_{t\in[0,T]}\left|\int_0^t\int_0^\infty u_k(X_{t-},z)\{\Phi(\mathrm{d}t,\mathrm{d}z)-J(X_{t-},\mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t\}\right|>\eta\right)\\&\quad\leq\frac{1}{T}\frac{1}{\eta^2}\int_{\mathbb{R}}\int_0^\infty u_k(y,z)^2J(y,\mathrm{d}z)\lambda(y)f(y)\,\mathrm{d}y\\&\quad\to 0,\end{aligned}$$

since $0\leq u_k(\cdot,\cdot)\leq 1$.

Based on the results of the LTE, we consider nonparametric estimation problems for the jump intensity $\lambda(x)$ and the jump size distribution function $J(x,y)$ of PDMP (1), where we use the shortened notation $J(x,y)$ for the conditional jump size distribution $J(x,[0,y])$. In the following, we assume that the intensity function $\lambda(x)$ and the jump size distribution $J(x,y)$ are continuous on any compact interval on $\mathbb{R}$ and $\mathbb{R}\times[0,\infty)$, respectively.

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

As an estimator for the jump intensity function $\lambda(x)$, we define

$$\hat{\lambda}_T(x) = \frac{\tilde{A}_T(x)}{f_T^\circ(x)},$$

where $f_T^\circ(x)$ is the LTE for the stationary density and

$$\tilde{A}_T(x) = \frac{1}{T}\int_0^T \int_0^\infty \frac{1}{b_T} K\left(\frac{X_{t-} - x}{b_T}\right) \Phi(\mathrm{d}t, \mathrm{d}z). \tag{12}$$

Here kernel K is a continuous function of bounded variation on $I = [\alpha, \beta]$ which vanishes outside I and has integral 1, and b_T is the bandwidth satisfying

$$b_T \to 0 \quad \text{and} \quad T b_T^2 \to \infty \quad \text{as } T \to \infty.$$

To prove the uniform consistency of $\hat{\lambda}_T(x)$, we need the following lemma.

Lemma 1. *For any compact interval* $[\gamma_1, \gamma_2]$,

$$\sup_{x \in [\gamma_1, \gamma_2]} |\tilde{A}_T(x) - \lambda(x) f(x)| \xrightarrow{\mathbb{P}} 0.$$

Proof. With obvious notation, we write

$$\begin{aligned}
\tilde{A}_T(x) &= \frac{1}{T}\int_0^T \int_0^\infty \frac{1}{b_T} K\left(\frac{X_{t-} - x}{b_T}\right) J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t \\
&\quad + \frac{1}{T}\int_0^T \int_0^\infty \frac{1}{b_T} K\left(\frac{X_{t-} - x}{b_T}\right) \{\Phi(\mathrm{d}t, \mathrm{d}z) - J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t\} \\
&= \frac{1}{T} U_T + \frac{1}{T} V_T.
\end{aligned}$$

It then follows from Theorem 1 that

$$\begin{aligned}
\frac{1}{T} U_T &= \frac{1}{T}\int_0^T \frac{1}{b_T} K\left(\frac{X_{t-} - x}{b_T}\right) \lambda(X_{t-})\,\mathrm{d}t \\
&= \frac{1}{T}\int_{\mathbb{R}} \frac{1}{b_T} K\left(\frac{y - x}{b_T}\right) \lambda(y) \Lambda_T(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}} \frac{1}{b_T} K\left(\frac{y - x}{b_T}\right) \lambda(y) f(y)\,\mathrm{d}y + \int_{\mathbb{R}} \frac{1}{b_T} K\left(\frac{y - x}{b_T}\right) \lambda(y) \left(\frac{\Lambda_T(y)}{T} - f(y)\right) \mathrm{d}y \\
&\xrightarrow{\mathbb{P}} \lambda(x) f(x) \quad \text{uniformly in } x \in [\gamma_1, \gamma_2],
\end{aligned}$$

since both λ and f are uniformly continuous on any compact interval.

Next we show that V_T/T uniformly converges to 0 by checking the conditions in Theorem 3.2 of Nishiyama (2000). For condition [C2] of Nishiyama (2000), by the assumption that $T b_T^2 \to \infty$ and (6), we have

$$\begin{aligned}
\left\langle \frac{1}{T} V_\cdot \right\rangle_T &= \frac{1}{T^2}\int_0^T \frac{1}{b_T^2} K\left(\frac{X_{t-} - x}{b_T}\right)^2 \lambda(X_{t-})\,\mathrm{d}t \\
&\le \frac{L^2}{T b_T^2} \frac{1}{T}\int_0^T \lambda(X_{t-})\,\mathrm{d}t \\
&\xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

where $L = \sup_{x\in[\alpha,\beta]} |K(x)| < \infty$. Analogously, condition [L2] of Nishiyama (2000),

$$\int_0^T |\bar{W}_t|^2 \mathbf{1}_{\{\bar{W}_t > \varepsilon\}} \lambda(X_{t-})\,\mathrm{d}t \xrightarrow{\mathbb{P}} 0,$$

is easily obtained, where

$$\bar{W}_t = \sup_{x\in[\alpha,\beta]} \left\{ \frac{1}{T}\frac{1}{b_T} K\left(\frac{X_{t-}-x}{b_T}\right)\right\} = \frac{L}{Tb_T}.$$

Finally, we check condition [PE] of Nishiyama (2000). For any $\varepsilon > 0$, choose some finite points $\gamma_1 = x_0 < x_1 < \cdots < x_{N(\varepsilon)} = \gamma_2$ such that $x_k - x_{k-1} \le \varepsilon^2$ with $N(\varepsilon) \le (\text{constant})\varepsilon^{-2}$. Then the entropy condition $\int_0^1 \sqrt{\log N(\varepsilon)}\,\mathrm{d}\varepsilon < \infty$ also holds. The proof is completed by showing that

$$\begin{aligned}\sup_{\varepsilon>0} \max_{1\le k\le N(\varepsilon)} \frac{1}{\varepsilon^2}\frac{1}{(Tb_T)^2}\int_0^T \sup_{u,v\in[x_{k-1},x_k]} \left| K\left(\frac{X_{t-}-u}{b_T}\right) - K\left(\frac{X_{t-}-v}{b_T}\right)\right|^2 \lambda(X_{t-})\,\mathrm{d}t \\ = O_{\mathbb{P}}(1). \end{aligned} \tag{13}$$

Since the kernel function K has bounded variation on $I = [\alpha,\beta]$, the Lebesgue–Stieltjes integral yields the representation

$$K(x) - K(\alpha) = \int_I \mathbf{1}_{\{z<x\}}\,\mathrm{d}K(z).$$

Furthermore, the kernel function can be decomposed as $K = K_1 - K_2$, where K_1 and K_2 are nondecreasing functions having bounded variation on I. Therefore, we obtain

$$\begin{aligned}&\left| K\left(\frac{X_{t-}-u}{b_T}\right) - K\left(\frac{X_{t-}-v}{b_T}\right)\right| \\ &\qquad \le \int_I |\mathbf{1}_{\{z<(X_{t-}-u)/b_T\}} - \mathbf{1}_{\{z<(X_{t-}-v)/b_T\}}|\,\mathrm{d}(K_1(z)+K_2(z)).\end{aligned}$$

By Hölder's inequality, it follows that

$$\begin{aligned}&\sup_{u,v\in[x_{k-1},x_k]} \left| K\left(\frac{X_{t-}-u}{b_T}\right) - K\left(\frac{X_{t-}-v}{b_T}\right)\right|^2 \\ &\qquad \le |K|_{\mathrm{TV}} \int_I |\mathbf{1}_{\{z<(X_{t-}-x_{k-1})/b_T\}} - \mathbf{1}_{\{z<(X_{t-}-x_k)/b_T\}}|\,\mathrm{d}(K_1(z)+K_2(z)),\end{aligned}$$

where $|K|_{\mathrm{TV}}$ means the total variation of K. Thus, we have

$$\begin{aligned}&\max_{1\le k\le N(\varepsilon)} \frac{1}{T}\int_0^T \sup_{u,v\in[x_{k-1},x_k]} \left| K\left(\frac{X_{t-}-u}{b_T}\right) - K\left(\frac{X_{t-}-v}{b_T}\right)\right|^2 \lambda(X_{t-})\,\mathrm{d}t \\ &\qquad \le |K|_{\mathrm{TV}} \max_{1\le k\le N(\varepsilon)} \int_I\int_{\mathbb{R}} \mathbf{1}_{\{x_{k-1}+zb_T<x\le x_k+zb_T\}} \lambda(x)\frac{\Lambda_T(x)}{T}\,\mathrm{d}x\,\mathrm{d}(K_1(z)+K_2(z)) \\ &\qquad \le |K|_{\mathrm{TV}}^2 \sup_{x\in[\gamma_1+\alpha b_T,\gamma_2+\beta b_T]} \left\{\lambda(x)\frac{\Lambda_T(x)}{T}\right\}\varepsilon^2.\end{aligned}$$

Hence, (13) holds since $\sup_z\{\Lambda_T(z)/T\}$ is stochastically bounded.

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

As a direct consequence of this lemma and Theorem 1, we have the following result.

Theorem 2. *For any* $\varepsilon > 0$,

$$\sup_{x \in I_\varepsilon} |\hat{\lambda}_T(x) - \lambda(x)| \xrightarrow{\mathbb{P}} 0,$$

where $I_\varepsilon = \{x : f(x) \geq \varepsilon\}$.

Next, for the conditional jump size distribution $J(x, y)$, we propose the estimator

$$\hat{J}_T(x, y) = \frac{\tilde{B}_T(x, y)}{\tilde{A}_T(x)},$$

where $\tilde{A}_T$ is given by (12) and

$$\tilde{B}_T(x, y) = \frac{1}{T} \int_0^T \int_0^y \frac{1}{b_T} K\left(\frac{X_{t-} - x}{b_T}\right) \Phi(\mathrm{d}t, \mathrm{d}z).$$

Here the kernel function K and bandwidth b_T are selected similarly as for $\tilde{A}_T$. Then the following result holds.

Theorem 3. *For any* $\varepsilon > 0$,

$$\sup_{(x,y) \in H_\varepsilon \times \mathbb{R}_+} |\hat{J}_T(x, y) - J(x, y)| \xrightarrow{\mathbb{P}} 0,$$

where $H_\varepsilon = \{x : \lambda(x) f(x) \geq \varepsilon\}$ *and* $\mathbb{R}_+ = [0, \infty)$.

Proof. As with the case for the intensity estimation, we show that, for any compact interval $[\gamma_1, \gamma_2]$,

$$\sup_{(x,y) \in [\gamma_1, \gamma_2] \times \mathbb{R}_+} \{\tilde{B}_T(x, y) - J(x, y)\lambda(x) f(x)\} \xrightarrow{\mathbb{P}} 0. \tag{14}$$

We write

$$\begin{aligned}\tilde{B}_T(x, y) = {} & \frac{1}{T} \int_0^T \int_0^y \frac{1}{b_T} K\left(\frac{X_{t-} - x}{b_T}\right) J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\, \mathrm{d}t \\ & + \frac{1}{T} \int_0^T \int_0^y \frac{1}{b_T} K\left(\frac{X_{t-} - x}{b_T}\right) \{\Phi(\mathrm{d}t, \mathrm{d}z) - J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\, \mathrm{d}t\}.\end{aligned}$$

It follows, from the uniform continuity of λ, f, and J, that the first term converges in probability to $J(x, y)\lambda(x) f(x)$ uniformly on $(x, y) \in [\gamma_1, \gamma_2] \times \mathbb{R}_+$. As above, we prove that the martingale part converges to 0 in probability. Conditions [C2] and [L2] in Theorem 3.2 of Nishiyama (2000) can be checked analogously. For condition [PE] of Nishiyama (2000), we use the following partitioning. For any $\varepsilon > 0$, choose some finite point $\gamma_1 = x_0 < x_1 < \cdots < x_{N_1(\varepsilon)} = \gamma_2$ such that $x_k - x_{k-1} \leq \varepsilon^2$ with $N_1(\varepsilon) \leq (\text{constant})\varepsilon^{-2}$ and, for the same ε, choose $0 = z_0 < z_1 < \cdots < z_{N_2(\varepsilon)} = \infty$ such that

$$\sup_{y \in \mathbb{R}} \int_{z_{j-1}}^{z_j} J(y, \mathrm{d}z) < \varepsilon^2,$$

with $N_2(\varepsilon) \leq (\text{constant})\varepsilon^{-2}$. Then the total number of partitions $\mathcal{N}(\varepsilon)$ for $[\gamma_1, \gamma_2] \times \mathbb{R}_+$ satisfies $\mathcal{N}(\varepsilon) = N_1(\varepsilon) \times N_2(\varepsilon) \leq (\text{constant})\varepsilon^{-4}$. Therefore, it holds that

$$\int_0^1 \sqrt{\log \mathcal{N}(\varepsilon)}\, \mathrm{d}\varepsilon < \infty.$$

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

Set $\Delta_{k,j} = (x_{k-1}, x_k] \times (z_{j-1}, z_j]$ for $1 \le k \le N_1(\varepsilon)$ and $1 \le j \le N_2(\varepsilon)$. We show that

$$\begin{aligned}\sup_{\varepsilon>0} \max_{k,j} \frac{1}{\varepsilon^2} \frac{1}{(Tb_T)^2} \int_0^T \int_0^\infty \sup_{(u,\xi),(v,\eta)\in\Delta_{k,j}} \left| \mathbf{1}_{\{z\le\xi\}} K\left(\frac{X_{t-}-u}{b_T}\right) - \mathbf{1}_{\{z\le\eta\}} K\left(\frac{X_{t-}-v}{b_T}\right)\right|^2 \\ \times J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t \\ = O_{\mathbb{P}}(1).\end{aligned}$$

It is obvious that, for any $(u, \xi), (v, \eta) \in \Delta_{k,j}$,

$$\begin{aligned}&\left| \mathbf{1}_{\{z\le\xi\}} K\left(\frac{x-u}{b_T}\right) - \mathbf{1}_{\{z\le\eta\}} K\left(\frac{x-v}{b_T}\right)\right| \\ &\quad \le \left| K\left(\frac{x-u}{b_T}\right) - K\left(\frac{x-v}{b_T}\right)\right| + |\,\mathbf{1}_{\{z\le\xi\}} - \mathbf{1}_{\{z\le\eta\}}\,| K\left(\frac{x-u}{b_T}\right).\end{aligned}$$

Note that

$$\begin{aligned}&\max_{k,j} \frac{1}{T} \int_0^T \int_0^\infty \sup_{\substack{\xi,\eta\in[z_{j-1},z_j]\\ u\in[x_{k-1},x_k]}} |\,\mathbf{1}_{\{z\le\xi\}} - \mathbf{1}_{\{z\le\eta\}}\,|^2 K\left(\frac{X_{t-}-u}{b_T}\right)^2 J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t \\ &\quad \le \sup_{x\in[\gamma_1+\alpha b_T, \gamma_2+\beta b_T]} \left\{\lambda(x)\frac{\Lambda_T(x)}{T}\right\} L^2\varepsilon^2,\end{aligned}$$

so, by taking account of (13), we have

$$\begin{aligned}&\sup_{\varepsilon>0} \max_{k,j} \frac{1}{\varepsilon^2} \frac{1}{(Tb_T)^2} \int_0^T \int_0^\infty \sup_{(u,\xi),(v,\eta)\in\Delta_{k,j}} \left| \mathbf{1}_{\{z\le\xi\}} K\left(\frac{X_{t-}-u}{b_T}\right) - \mathbf{1}_{\{z\le\eta\}} K\left(\frac{X_{t-}-v}{b_T}\right)\right|^2 \\ &\qquad \times J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t \\ &\quad \le \sup_{\varepsilon>0} \max_{k,j} \frac{2^2}{\varepsilon^2} \frac{1}{Tb_T^2} \left\{ \frac{1}{T} \int_0^T \sup_{u,v\in[x_{k-1},x_k]} \left| K\left(\frac{X_{t-}-u}{b_T}\right) - K\left(\frac{X_{t-}-v}{b_T}\right)\right|^2 \lambda(X_{t-})\,\mathrm{d}t \right. \\ &\qquad + \frac{1}{T} \int_0^T \int_0^\infty \sup_{\substack{\xi,\eta\in[z_{j-1},z_j]\\ u\in[x_{k-1},x_k]}} |\,\mathbf{1}_{\{z\le\xi\}} - \mathbf{1}_{\{z\le\eta\}}\,|^2 K\left(\frac{X_{t-}-u}{b_T}\right)^2 \\ &\qquad \left. \times J(X_{t-}, \mathrm{d}z)\lambda(X_{t-})\,\mathrm{d}t \right\} \\ &\quad = O_{\mathbb{P}}(1).\end{aligned}$$

Hence, we obtain the uniform convergence (14). Together with Lemma 1, this completes the proof.

Acknowledgements

The author would like to express his gratitude to the anonymous referee for valuable comments which helped to improve the first version of this paper. This study was financially supported by the CREST (Core Research for Evolutional Science and Technology) program of the Japan Science and Technology Cooperation.

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press

References

Borovkov, K. and Last, G. (2008). On level crossings for a general class of piecewise-deterministic Markov processes. *Adv. Appl. Prob.* **40,** 815–834.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (Lecture Notes Statist. **110**), 2nd edn. Springer, New York.

Bosq, D. and Davydov, Y. (1999). Local time and density estimation in continuous time. *Math. Meth. Statist.* **8,** 22–45.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer, New York.

Brill, P. H. and Posner, M. J. M. (1977). Level crossings in point processes applied to queues: single-server case. *Operat. Res.* **25,** 662–674.

Davis, M. H. A. (1993). *Markov Models and Optimization.* Chapman and Hall, London.

Fujii, T. and Nishiyama, Y. (2012). Some problems in nonparametric inference for the stress release process related to the local time. *Ann. Inst. Statist. Math.* **64,** 991–1007.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd edn. Springer, New York.

Kutoyants, Y. A. (1998). Efficient density estimation for ergodic diffusion processes. *Statist. Infer. Stoch. Process.* **1,** 131–155.

Kutoyants, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, London.

Last, G. (2004). Ergodicity properties of stress release, repairable system and workload models. *Adv. Appl. Prob.* **36,** 471–498.

Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line: The Dynamic Approach*. Springer, New York.

Leadbetter, M. R. (1966). On crossings of levels and curves by a wide class of stochastic processes. *Ann. Math. Statist.* **37,** 260–267.

Liptser, R. S. and Shiryaev, A. N. (1989). *Theory of Martingales*. Kluwer, Dordrecht.

Nishiyama, Y. (2000). Weak convergence of some classes of martingales with jumps. *Ann. Prob.* **28,** 685–712.

Rice, S. O. (1944). Mathematical analysis of random noise. *Bell System Tech. J.* **23,** 282–332.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Van Zanten, J. H. (2001). On the uniform convergence of the empirical density of an ergodic diffusion. *Statist. Infer. Stoch. Process.* **3,** 251–262.

https://doi.org/10.1239/jap/1389370091 Published online by Cambridge University Press