ELSEVIER

Statistics & Probability Letters 44 (1999) 97–106

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

On the asymptotic variance of the continuous-time kernel density estimator

Martin Sköld *, Ola Hössjer

Centre for Mathematical Sciences, Division of Mathematical Statistics, Lund University, Box 118, S-221 00 Lund, Sweden

Received April 1998; accepted September 1998

Abstract

We reformulate the conditions of Blanke and Bosq (1997) for achieving the $(\log T)/T$-rate of convergence of the kernel density estimator for a smooth process and give under slightly stronger assumptions the exact asymptotic form of the variance giving an expression for the asymptotic optimal bandwidth. Conditions for the full T^{-1} and discrete-time rates are also considered.

Keywords: Density estimation; Kernel estimation; Stationary processes

1. Introduction

For a kernel density estimator of a discrete-time stationary ergodic process, X_i; $i \in \{1,\dots,n\}$, the optimal rate of convergence of the mean-squared error to zero is typically $n^{-2m/(2m+1)}$ if the density f has m continuous derivatives (see Wahba, 1975). The interest in kernel density estimation given a continuous-time sample, X_t; $t \in [0,T]$, started with the paper of Castellana and Leadbetter (1986) where they gave conditions under which the faster T^{-1}-rate of convergence could be achieved. Given a differentiable process however this result is usually not possible and we will get the $(\log T)/T$-rate as examined by Blanke and Bosq (1997). The improvement from discrete time is perhaps not surprising since in continuous time the observation of complete sample paths gives us a chance to actually observe the event $\{X_\tau = u\}$ for some $\tau \in [0,T]$, this allows us to construct unbiased estimators in the form of occupation-time densities (OTD) and local times, see e.g. Geman and Horowitz (1960) and Kutoyants (1996), which is not possible in discrete time. We will show that the existence of an OTD is closely related to achieving a faster rate than in discrete time, also the slower rate for differentiable processes is due to the OTD having an infinite variance. We will reformulate the conditions for achieving the different rates and under slightly stronger assumptions give the exact form of the asymptotic variance, a result which is important when selecting the optimal bandwidth parameter. Another difference from discrete time is that the possibility to localize our estimators to $\{\tau; X_\tau = u\}$ makes smoothing

* E-mail: martins@maths.lth.se.

0167-7152/99/$ – see front matter
PII: S0167-7152(98)00297-1

unnecessary and the convergence rate will not depend on m above as long as the density is Hölder continuous. Finally we will give example of a non-parametric class of processes for which the kernel estimator is not rate optimal.

2. The kernel density estimator

The kernel density estimator for a stationary process observed over an interval $[0,T]$ is defined as:

$$f_h(u)=\frac{1}{T}\int_0^T K_h(X_t-u)\,\mathrm{d}t, \tag{2.1}$$

where $K_h(\cdot)=h^{-1}K(\cdot/h)$ for a *kernel function* K integrating to 1 scaled with a *bandwidth* h. We will further assume K to be compactly supported and bounded. Properties of this estimator have been examined by e.g. Castellana and Leadbetter (1986), Bosq (1996), Blanke and Bosq (1997) and in the case of an ergodic diffusion in Kutoyants (1996), differentiable processes are considered in Blanke and Bosq (1997) and Sköld (1996) but so far an explicit expression for the asymptotic variance has to our knowledge not appeared in literature, this is important e.g. when developing methods for automatic bandwidth selection. Castellana and Leadbetter (1986) derived the following form for the asymptotic variance of f_h:

$$\lim_{\substack{T\to\infty\\ h\to 0}} T\,\mathrm{Var}(f_h(u))=\lim_{\substack{T\to\infty\\ h\to 0}} 2\int_0^T(1-\tau/T)\alpha_h(\tau,u)\,\mathrm{d}\tau=2\int_0^\infty (f_{X_0,X_\tau}(u,u)-f^2(u))\,\mathrm{d}\tau,$$

where

$$\alpha_h(\tau,u)=\int\int K_h(s-u)K_h(v-u)(f_{X_0,X_\tau}(s,v)-f(s)f(v))\,\mathrm{d}s\,\mathrm{d}v\to f_{X_0,X_\tau}(u,u)-f^2(u),$$

as $h\to 0$. For processes which behave locally as Brownian motion the right-hand side is usually finite since $f_{X_0,X_\tau}(u,u)=\mathrm{O}(\tau^{-1/2})$ for small τ, given a differentiable process however $f_{X_0,X_\tau}(u,u)=\mathrm{O}(\tau^{-1})$, as will be apparent in the proof of Theorem 2.1, and the integral will diverge.

Theorem 2.1. *Let X_t be an ergodic stationary stochastic process with marginal density function f continuous in a neighborhood of u. Define $Y_\tau=(X_\tau-X_0)/\tau$ and $Y_0:=X_0'$. Then if*

1. $\lim_{h\to 0}\int_\delta^\infty|\alpha_h(\tau,u)|\,\mathrm{d}\tau<\infty$ *for all* $\delta>0$.
2. *There is a constant* $f_{X_0,X_0'}(u,0)$ *such that* $\lim_{\varepsilon\to 0}\sup_{(x,y,\tau)\in B_\varepsilon\times[0,\varepsilon)}|f_{X_0,Y_\tau}(x,y)-f_{X_0,X_0'}(u,0)|=0$ *where* $B_\varepsilon=\{(x,y);(x-u)^2+y^2<\varepsilon^2\}$.

we have that

$$\mathrm{Var}(f_h(u))=2f_{X_0,X_0'}(u,0)\log(h^{-1})T^{-1}+\mathrm{o}(\log(h^{-1})T^{-1}), \tag{2.2}$$

as $h\to 0$ and $T\to\infty$.

Proof. We have, assuming a uniform kernel $K_h(\cdot)=(2h)^{-1}\mathbf{1}_{\{|\cdot|<h\}}$, where $\mathbf{1}_A$ is the indicator of the set A, that $\alpha_h(\tau,u)=(2h)^{-2}(P(|X_0-u|<h,|X_\tau-u|<h)-P^2(|X_0-u|<h))$ and

$$\begin{aligned} T\,\mathrm{Var}(f_h(u))&=\frac{1}{4h^2}\int_0^T\int_0^T \mathrm{Cov}(\mathbf{1}_{\{|X_t-u|<h\}},\mathbf{1}_{\{|X_s-u|<h\}})\,\mathrm{d}t\,\mathrm{d}s\\ &=\frac{1}{2h^2}\int_0^T(1-\tau/T)\mathrm{Cov}(\mathbf{1}_{\{|X_0-u|<h\}},\mathbf{1}_{\{|X_\tau-u|<h\}})\,\mathrm{d}\tau\\ &=2\int_0^T(1-\tau/T)\alpha_h(\tau,u)\,\mathrm{d}\tau. \end{aligned}$$

To study this integral we divide it into three segments:

$$\int_0^{\gamma h}(1-\tau/T)\alpha_h(\tau,u)\,\mathrm{d}\tau+\int_{\gamma h}^{\delta}(1-\tau/T)\alpha_h(\tau,u)\,\mathrm{d}\tau+\int_{\delta}^{T}(1-\tau/T)\alpha_h(\tau,u)\,\mathrm{d}\tau, \tag{2.3}$$

where we have introduced monotonous functions $\delta=\delta(h)$ and $\gamma=\gamma(h)$ such that $\lim_{h\to 0}\delta(h)=0$, $\lim_{h\to 0}\delta(h)/h=\infty$ and $\lim_{h\to 0}\int_{\delta(h)}^{\infty}|\alpha_h(\tau,u)|\,\mathrm{d}\tau/\log(h^{-1})=0$ which is possible by assumption 1, further we select $\gamma(h)$ such that $\gamma(h)h<\delta(h)$, $\lim_{h\to 0}\gamma(h)=\infty$ and $\lim_{h\to 0}\gamma(h)/\log(h^{-1})=0$. The first integral is bounded by $\gamma f(u)=\mathrm{o}(\log(h^{-1}))$ since

$$|(1-\tau/T)\alpha_h(\tau)|\leqslant\frac{1}{4h^2}P(|X_0-u|<h)\leqslant h^{-1}(1/2+\varepsilon')f(u)\leqslant h^{-1}f(u)$$

for h sufficiently small, some $\varepsilon'>0$ and $f(x)$ continuous when $|x-u|<h$. For the second integral we have with $\mathrm{O}(1)$ meaning that the term is bounded as $T\to\infty$ that

$$\begin{aligned}\int_{\gamma h}^{\delta}(1-\tau/T)\alpha_h(\tau,u)\,\mathrm{d}\tau&=\frac{1}{4h^2}\int_{\gamma h}^{\delta}(1-\tau/T)P(|X_0-u|<h,|\tau Y_\tau+(X_0-u)|<h)\,\mathrm{d}\tau+\mathrm{O}(1)\\&=\frac{1}{4h^2}\int_{\gamma h}^{\delta}(1-\tau/T)\int\int_{\substack{|s-u|<h\\|\tau v'+(s-u)|<h}}f_{X_0,Y_\tau}(s,v')\,\mathrm{d}s\,\mathrm{d}v'\,\mathrm{d}\tau+\mathrm{O}(1)\\&=\frac{1}{4h^2}\int_{\gamma h}^{\delta}\tau^{-1}\int\int_{\substack{|s-u|<h\\|v+(s-u)|<h}}f_{X_0,Y_\tau}(s,v/\tau)\,\mathrm{d}s\,\mathrm{d}v\,\mathrm{d}\tau+\mathrm{O}(1)\\&=\int_{\gamma h}^{\delta}\tau^{-1}(f_{X_0,X_0'}(u,0)+R_\tau(h))\,\mathrm{d}\tau+\mathrm{O}(1)\\&=f_{X_0,X_0'}(u,0)(\log(\delta)-\log(\gamma h))+\mathrm{o}(\log(h^{-1})),\end{aligned}$$

since in the area of integration $v/\tau\leqslant 2h/\tau\leqslant 2/\gamma$ and with $\varepsilon_h=\max\{h,2/\gamma,\delta\}\to 0$ as $h\to 0$ we have by assumption 2 that

$$\int_{\gamma h}^{\delta}\tau^{-1}R_\tau(h)\,\mathrm{d}\tau\leqslant\int_{\gamma h}^{\delta}\tau^{-1}\,\mathrm{d}\tau\sup_{(x,y,\tau)\in B_{\varepsilon_h}\times[0,\varepsilon_h)}|f_{X_0,Y_\tau}(x,y)-f_{X_0,X_0'}(u,0)|=\mathrm{o}(\log(h^{-1})).$$

The third integral is $\mathrm{o}(\log(h^{-1}))$ by the assumptions on $\delta(h)$ and the theorem follows.

For a general compactly supported kernel function the proof is only notationally more complicated, note that the size of the support does not matter since a rescaling $h'=ah$ only affects (2.2) with a factor of $\mathrm{O}(T^{-1})$. □

Remark 2.1. Assumption 1, to ensure the variance is not affected by long-range dependence, is true under more standard integrability of mixing-coefficients conditions, see e.g. Bosq (1996). It can be easily seen from the proof that the somewhat weaker $\int_{\delta}^{\infty}|\alpha_h(\tau,u)|\,\mathrm{d}\tau=\mathrm{o}(\log(h^{-1}))$ as $h\to 0$ is sufficient.

Theorem 2.2. *Assume f is continuous and bounded, if further*
1. $\lim_{h\to 0}\int\int_{\delta}^{\infty}|\alpha_h(\tau,u)|\,\mathrm{d}\tau\,\mathrm{d}u<\infty$
2. *There is a constant* $f_{X_0'}(0)$ *such that* $\lim_{\varepsilon\to 0}\sup_{(y,\tau)\in(-\varepsilon,\varepsilon)\times[0,\varepsilon)}|f_{Y_\tau}(y)-f_{X_0'}(0)|=0.$
3. $f^2(u)<g(u)$, *where g is an integrable function monotonous when $|u|>M$ for some constant $M<\infty$.*

then

$$\int\mathrm{Var}(f_h(u))\,\mathrm{d}u=2f_{X_0'}(0)\log(h^{-1})T^{-1}+\mathrm{o}(\log(h^{-1})T^{-1}), \tag{2.4}$$

as $h \to 0$ *and* $T \to \infty$.

Proof. For fixed $h > 0$, f_h and thus also its variance is bounded and we can change order of integration;

$$T \int \mathrm{Var}(f_h(u))\,\mathrm{d}u = 2\int_0^T \int (1 - \tau/T)\alpha_h(\tau, u)\,\mathrm{d}u\,\mathrm{d}\tau,$$

and decompose the outer integral into three terms as in (2.3). The first is bounded by $\gamma(h) = \mathrm{o}(\log h^{-1})$ by the inequality

$$\left| \int (1 - \tau/T)\alpha_h(\tau, u)\,\mathrm{d}u \right| \leqslant \frac{1}{4h^2} \int E(\mathbf{1}_{\{|X_0 - u| < h\}})\,\mathrm{d}u = \frac{1}{2h}.$$

The bound of the third follows from assumption 1 as in Theorem 2.1 and for the second we have that

$$\begin{aligned} I_2 &:= 2\int_{\gamma h}^{\delta} (1 - \tau/T) \int \alpha_h(\tau, u)\,\mathrm{d}u\,\mathrm{d}\tau \\ &= \frac{1}{2h^2} \int_{\gamma h}^{\delta} (1 - \tau/T) \int \mathrm{Cov}(\mathbf{1}_{\{|X_0 - u| < h\}} \mathbf{1}_{\{|X_\tau - u| < h\}})\,\mathrm{d}u\,\mathrm{d}\tau \\ &= \frac{1}{2h^2} \int_{\gamma h}^{\delta} E\left(\int \mathbf{1}_{\{|X_0 - u| < h\}} \mathbf{1}_{\{|X_\tau - u| < h\}}\,\mathrm{d}u \right) - \frac{1}{2h^2} \int P^2(|X_0 - u| < h)\,\mathrm{d}u\,\mathrm{d}\tau + \mathrm{O}(1). \end{aligned}$$

Assumption 3 now gives us a bound for the last term above since

$$\frac{1}{2h^2} \int P^2(|X_0 - u| < h)\,\mathrm{d}u = 2\int f^2(s_u)\,\mathrm{d}u \leqslant 2\int g(u)\,\mathrm{d}u + 4M \sup(f^2),$$

where $|s_u - u| < 2h$. Continuing with the first we have that

$$\begin{aligned} I_2 &= \frac{1}{2h^2} \int_{\gamma h}^{\delta} E(\mathbf{1}_{\{|Y_\tau| < 2h/\tau\}}(2h - \tau|Y_\tau|))\,\mathrm{d}\tau + \mathrm{O}(1) \\ &= \frac{f_{X_0'}(0)}{2h^2} \int_{\gamma h}^{\delta} \int_{-2h/\tau}^{2h/\tau} (2h - \tau|s|)\,\mathrm{d}s\,\mathrm{d}\tau + \int_{\gamma h}^{\delta} R_\tau(h)\,\mathrm{d}\tau + \mathrm{O}(1) \\ &= f_{X_0'}(0)(4 - 2)\log(h^{-1}) + \mathrm{o}(\log(h^{-1})), \end{aligned}$$

by using assumption 2 as in the proof of Theorem 2.1. □

Example 2.1. For a zero mean unit variance Gaussian process X_t satisfying assumption 1 we have since Y_τ is also zero mean Gaussian with $E(Y_\tau^2) = (2 - 2r_\tau)\tau^{-2}$ where $r_\tau = E(X_0 X_\tau)$ is the covariance function of the process that

$$f_{Y_\tau}(s) = C\tau(1 - r_\tau)^{-1/2} \exp(-(s\tau)^2/(4 - 4r_\tau)).$$

which satisfies assumptions 2 and 3 of Theorem 2.2 provided $0 < \lim_{\tau \to 0}(1 - r_\tau)\tau^{-2} = \mathrm{Var}(X_0') < \infty$. For assumption 1, assume there is $d < 1$ and $\delta > 0$ such that $r(\tau) < d$ when $\tau > \delta$ and write ϕ_r for the standard bivariate normal density with correlation r. By the mean-value theorem

$$\int |\alpha_h(\tau,u)|\,\mathrm{d}u = \int \left| \int\int K_h(s-u)K_h(t-u)(\phi_{r(\tau)} - (s,t)\phi_0(s,t))\,\mathrm{d}s\,\mathrm{d}t \right| \mathrm{d}u$$

$$= r(\tau)\int \left| \frac{\mathrm{d}}{\mathrm{d}r}\phi_r(s_u,t_u)|_{r=s(\tau)} \right| \mathrm{d}u$$

$$= r(\tau)D(h,\tau),$$

where $|s_u - u| \leqslant h$, $|t_u - u| \leqslant h$ and $0 \leqslant s(\tau) \leqslant r(\tau)$. Now $\lim_{h\to 0} D(h,\tau)$ is bounded for $\tau > \delta$ by standard properties of the bivariate normal density and a sufficient additional condition is $\int_0^\infty |r(\tau)|\,\mathrm{d}\tau < \infty$.

Corollary 2.1. *Under the conditions in the theorem above using a positive symmetric kernel K with compact support and if f is twice continuously differentiable, the bandwidth minimizing the expression for the asymptotic integrated mean-square error* (AIMSE) *is given by*

$h(T) = (CT)^{-\alpha}$ *with* $\alpha = 1/4$ *and*

$$C = \left(\int x^2K(x)\,\mathrm{d}x\right)^2 \int f''(u)^2\,\mathrm{d}u/(2f_{X_0'}(0)), \tag{2.5}$$

which gives the rate of convergence:

$$\lim_{T\to\infty} T(\log T)^{-1}\int_{-\infty}^{\infty} E(f_{h(T)}(u) - f(u))^2\,\mathrm{d}u = 2\alpha f_{X_0'}(0). \tag{2.6}$$

Moreover (2.6) *is true for all* $\alpha \geqslant \frac{1}{4}$ *independently of the choice of* $C > 0$.

Proof. The bias of $f_h(u)$ is found by standard Taylor-expansion arguments to equal

$$E(f_h(u)) - f(u) = E(K_h(X_0 - u)) - f(u)$$

$$= \int K(x)f(u - xh)\,\mathrm{d}x - f(u)$$

$$= h^2 f''(u)\int x^2K(x)\,\mathrm{d}x/2 + \mathrm{o}(h^2),$$

since $\int xK(x)\,\mathrm{d}x = 0$ for a symmetric kernel K. By writing the mean-square error as the sum of variance and squared bias and neglecting the lower order terms we get

$$\mathrm{AIMSE}(h) = h^4\left(\int x^2K(x)\,\mathrm{d}x\right)^2\int f''(u)^2\,\mathrm{d}u/4 + 2\log(h^{-1})T^{-1}f_{X_0'}(0), \tag{2.7}$$

which is minimized by $h = (CT)^{-\alpha}$ as in (2.5). If further $\alpha \geqslant \frac{1}{4}$, the bias will be of $\mathrm{O}(T^{-1}) = \mathrm{o}((\log T)/T)$ and thus asymptotically negligible compared to the variance which is asymptotically independent of C. □

Remark 2.2. Note that since the bias with an uniform kernel equals $(2h)^{-1}\int_{u-h}^{u+h}(f(t) - f(u))\,\mathrm{d}t$ it is sufficient for f to be Hölder-continuous in a neighborhood of u, i.e. $|f(u) - f(u+h)| < C|h|^\beta$, $\beta > 0$ for sufficiently small $|h|$, together with the assumptions of Theorem 2.2 to achieve the $(\log T)/T$-rate, though with possibly larger α.

3. Rates

In the following section we will assume f to be continuous in a neighborhood of u, $h=T^{-\alpha}$, $\alpha \in (0,1)$ and assumption 1 of Theorem 2.1 to hold.

Theorem 3.1. *Assume there is an open neighborhood* Ω *of* $(u,0)$, $\varepsilon>0$ *and a constant* $M<\infty$ *such that* $(s,v,\tau)\mapsto f_{X_0,Y_\tau}(s,v)<M$ *in* $\Omega\times[0,\varepsilon)$, *then*

$$\limsup_{T\to\infty} T(\log T)^{-1}\mathrm{Var}(f_{T^{-\alpha}}(u))<\infty. \tag{3.1}$$

If f_{X_0,Y_τ} *is also bounded from below this is the exact rate, i.e. if* $0<m<f_{X_0,Y_\tau}(s,v)$ *in* $\Omega\times[0,\varepsilon)$,

$$\liminf_{T\to\infty} T(\log T)^{-1}\mathrm{Var}(f_{T^{-\alpha}}(u))>0. \tag{3.2}$$

If on the other hand there exists $\beta>0$ *such that* $f_{X_0,Y_\tau}(s,v)\leqslant M(\tau^\beta+|s-u|^\beta+v^\beta\tau^\beta)$ *when* $(s,v,\tau)\in\Omega\times[0,\varepsilon)$ *for some* $M<\infty$, *then*

$$\limsup_{T\to\infty} T\,\mathrm{Var}(f_{T^{-\alpha}}(u))<\infty. \tag{3.3}$$

Let $\lambda_t=\lambda(\{s\in[t,t+1];X_s=X_t\})$ *where* λ *is the Lebesgue measure. If there is* $\varepsilon>0$ *and* $\delta>0$ *such that* $x\mapsto P(\lambda_0>\delta|X_0=x)>\delta$ *when* $|x-u|<\varepsilon$ *(the process can spend positive length of time at level* x*), we get the discrete-time rate*:

$$\liminf_{T\to\infty} T^{1-\alpha}\mathrm{Var}(f_{T^{-\alpha}}(u))>0. \tag{3.4}$$

Proof. (3.1)–(3.3) are easy consequences of the proof of Theorem 2.1, we show (3.1) which follows from the inequality

$$\limsup_{T\to\infty} T(\log T)^{-1}\mathrm{Var}(f_{T^{-\alpha}}(u))\leqslant \lim_{T\to\infty} 2M(\log T)^{-1}\int_{T^{-\alpha}}^{\varepsilon}\tau^{-1}\,\mathrm{d}\tau=2M\alpha<\infty.$$

For (3.4) we have

$$T\,\mathrm{Var}(f_{T^{-\alpha}}(u))=2\int_0^1(1-\tau/T)\alpha_{T^{-\alpha}}(\tau,u)\,\mathrm{d}\tau+2\int_1^T(1-\tau/T)\alpha_{T^{-\alpha}}(\tau,u)\,\mathrm{d}\tau,$$

where the last term is bounded as $T\to\infty$ by assumption 1 of Theorem 2.1. For the first we have

$$\begin{aligned}4\int_0^1(1-\tau/T)\alpha_{T^{-\alpha}}(\tau,u)\,\mathrm{d}\tau &= T^{2\alpha}E\left(\int_0^1(1-\tau/T)\mathbf{1}_{\{|X_0-u|<T^{-\alpha}\}}\mathbf{1}_{\{|X_\tau-u|<T^{-\alpha}\}}\,\mathrm{d}\tau\right)\\ &\quad -T^{2\alpha}\int_0^1(1-\tau/T)E^2(\mathbf{1}_{\{|X_0-u|<T^{-\alpha}\}})\,\mathrm{d}\tau\\ &\geqslant T^{2\alpha}(1-1/T)E(\lambda_0\mathbf{1}_{\{|X_0-u|<T^{-\alpha}\}})+\mathrm{O}(1)\\ &\geqslant T^{2\alpha}\int_{u-T^{-\alpha}}^{u+T^{-\alpha}}E(\lambda_0|X_0=x)f(x)\,\mathrm{d}x+\mathrm{O}(1)\\ &\geqslant T^{\alpha}\delta^2 f(u)+\mathrm{O}(1)\end{aligned}$$

if $T^{-\alpha}<\varepsilon$ is sufficiently small. □

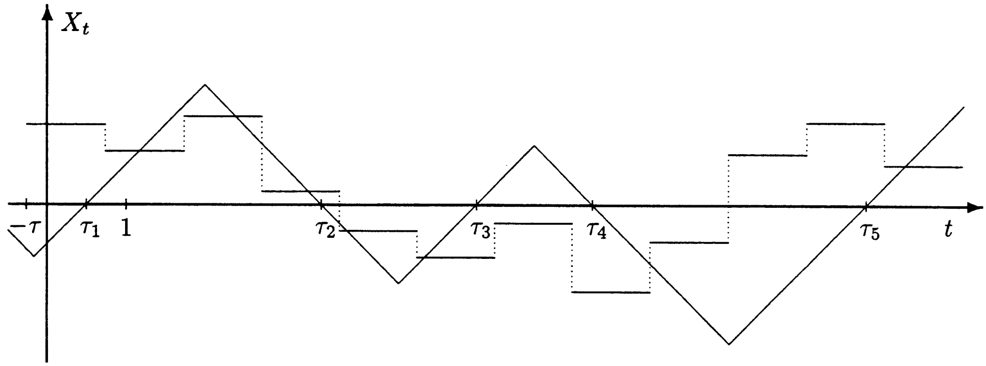

Fig. 1. The processes of examples 3.2 and 3.3.

Remark 3.1. The condition $P(\lambda_0 > \delta | X_0 = x) > \delta$ for the discrete-time rate should be compared with Geman and Horowitz's sufficient condition $P(X_0' = 0) = 0$ for the existence of an OTD provided f exists. Thus when the sample paths satisfy $\lambda(t; X_t' = 0) > 0 \Rightarrow \exists x; \lambda(t; X_t = x) > 0$ with positive probability and there is no OTD, we cannot improve the discrete-time rate.

Example 3.1. Processes where the full T^{-1}-rate is achieved includes e.g. processes which behave locally as Brownian motion since for standard Brownian motion we have

$$f_{X_1,(X_{1+\tau}-X_1)/\tau}(s,v) = C\tau^{1/2}\exp(-(s^2+\tau v^2)/2) \leqslant C\tau^{1/2},$$

which satisfies the conditions of (3.3) with $M = C$ and $\beta = \frac{1}{2}$.

Example 3.2. A second example to show that the sample paths need not be very "irregular" to achieve the full T^{-1}-rate are processes where the absolute derivative is bounded from below by some constant $\delta > 0$ when X_t is close to u, e.g. the saw-tooth process obtained from a stationary point process τ_i on the real line with interarrival times $\tau_{i+1} - \tau_i > 1$ and independent by

$$X_t = \min(t - \tau_i, \tau_{i+1} - t)(-1)^{i+Z},$$

$t \in [\tau_i, \tau_{i+1})$, where $P(Z=0) = P(Z=1) = \frac{1}{2}$ (see Fig. 1). Here the conditions of (3.3) are satisfied for $u = 0$ since $f_{X_0,Y_\tau}(s,v) = 0$ when $(s,v,\tau) \in (-0.25, 0.25) \times (-1,1) \times [0, 0.25)$.

Example 3.3. The discrete-time case can be seen as a special case of (3.4) by viewing a discrete sample, $\{X_i; i = 1,\dots,n\}$, as a piecewise constant process (see Fig. 1 and Theorem 4.3 in Bosq, 1996):

$$\begin{aligned} f_h(u) &= \frac{1}{n}\sum_1^n K_h(X_i - u) \\ &= \frac{1}{n}\int_0^{n-1} K_h(X_{\lceil t+\tau \rceil} - u)\,\mathrm{d}t + O_P(n^{-1}), \end{aligned}$$

where $\tau \in \mathrm{Unif}(0,1)$ is a random shift of time to make the process stationary. Clearly the conditions of (3.4) are satisfied since $P(\lambda_0 > \delta | X_0 = x) = P(1 - \tau > \delta) > \delta$ if $0 < \delta < \frac{1}{2}$.

4. Beating the rate

If X_t is a differentiable process and $N_T(u)$ the number of level-crossings of u, $t \in [0,T]$, we have under general smoothness conditions (Geman and Horowitz, 1960) the occupation-time density

$$\beta_T(u) = 2T \lim_{h\to 0} f_h(u) = \sum_{i=1}^{N_T(u)} |X'_{\tau_i}|^{-1}, \tag{4.1}$$

where τ_i, $i=1,\dots,N_T(u)$ are the times of level-crossings of u. This is, scaled by $(2T)^{-1}$, an unbiased estimator of the density but in practise not very useful since it will often have an infinite variance and has a singularity at every level u where a crossing of zero of the derivative is observed (a local extrema). Closely related to this is *Rice's formula* for the mean number of level crossings in the unit interval,

$$2f(u) = E(N_1(u))/E(|X'_0||X_0 = u) = E(N_1(u))E(|X'_{\tau_1}|^{-1}), \tag{4.2}$$

see e.g. Marcus (1977), Geman and Horowitz (1960) and in a kernel estimation context Sköld (1996). To motivate why $\beta_T(u)$ has an infinite variance, note that the crossing-derivatives $|X'_{\tau_i}|$ will have density $g_u(x) = Cxf_{X_0,X'_0}(u,x)$ due to a length-bias phenomena (the process spends more time close to u if the crossing derivative is close to zero) and if $f_{X_0,X'_0}(u,0) > 0$ the inverse second moment will not exist. This is why $f_{X_0,X'_0}(u,0)$ appears in main term of the asymptotic variance in (2.2). Note that if $N_T(u)$ is well behaved, i.e. it is close to $TE(N_1(u))$ for large T, we can construct a new $(\log T)/T$-estimator by using the truncated crossing derivatives $|X'_{\tau_i}|\mathbf{1}_{\{|X'_{\tau_i}|>T^{-\alpha}\}}$ in (4.1).

Now as indicated the slower rate of convergence $(\log T)/T$ comes from the difficulty in estimating $E(|X'_{\tau_1}|^{-1})$. If however the mean derivative is independent of the level; $E(|X'_0||X_0 = u) = E(|X'_0|)$, which e.g. is the case for Gaussian processes, we can use the estimator given by

$$\hat{f}(u) = N_T(u) \Big/ \left(2\int_0^T |X'_t|\,\mathrm{d}t + 2\right), \tag{4.3}$$

which under general conditions in this class of processes achieves the full T^{-1}-rate and thus beating the $(\log T)/T$-rate of the kernel estimator.

Theorem 4.1. *Assume that X_t is a stationary mean-square differentiable process, write $S_T = N_T(u) - E(N_T(u)) + 2(TE|X'_0| - \int_0^T |X'_t|\,\mathrm{d}t)f(u)$ and $S'_T = \int_0^T \min(1,|X'_t|)\,\mathrm{d}t - TE(\min(1,|X'_t|))$, then if there is $M < \infty$ and $\delta > 0$ such that with $p = 1 + \delta/2$ and $q = (2+\delta)/\delta$*
1. *$E(N_1(u)) = 2f(u)E|X'_0|$ (Rice's formula when X_0 and X'_0 independent),*
2. *$\limsup_{T\to\infty} E|T^{-1/2}S_T|^{2p} < M$,*
3. *$\limsup_{T\to\infty} E|T^{-1/2}S'_T|^{4q} < M$,*

we have

$$\limsup_{T\to\infty} TE(\hat{f}(u) - f(u))^2 < \infty.$$

Proof. Write $\hat{I}_T = \int_0^T |X'_t|\,\mathrm{d}t/T$ and $\tilde{I}_T = \int_0^T \min(1,|X'_t|)\,\mathrm{d}t/T$. Under assumption 1, $S_T = N_T(u) - 2T\hat{I}_T f(u)$ and using that $(a+b)^2 \leqslant 2a^2 + 2b^2$ and Hölder's inequality we have

$$TE(\hat{f}(u) - f(u))^2 = TE\left[\frac{N_T(u)/T - 2\hat{I}_T f(u) + 2f(u)/T}{2\hat{I}_T + 2/T}\right]^2$$

$$\leqslant 2E\left(\frac{T^{-1/2}S_T}{2\hat{I}_T+2/T}\right)^2 + 2f^2(u)T^{-1}E(2\hat{I}_T+2/T)^{-2}$$

$$\leqslant 2(E|T^{-1/2}S_T|^{2p})^{1/p}(E(2\hat{I}_T+2/T)^{-2q})^{1/q} + 2f^2(u)T^{-1}E(2\hat{I}_T+2/T)^{-2},$$

and by assumption 2 it remains to show that $E(\hat{I}_T+1/T)^{-2q}$ and thus also $E(\hat{I}_T+1/T)^{-2}$ is bounded. Now, by Markov's inequality for $z<E(\tilde{I}_T)/2$ and assumption 3 we have for T large enough that

$$\begin{aligned} F_{\tilde{I}_T}(z) &\leqslant P(|\tilde{I}_T - E(\tilde{I}_T)| \geqslant E(\tilde{I}_T)-z) \\ &\leqslant \frac{E|\tilde{I}_T-E(\tilde{I}_T)|^{4q}}{|E(\tilde{I}_T)-z|^{4q}} = \frac{E|T^{-1/2}S'_T|^{4q}}{T^{2q}|E(\tilde{I}_T)-z|^{4q}} \\ &\leqslant M(E(\tilde{I}_T)/2)^{-4q}T^{-2q}, \end{aligned}$$

and further that

$$\begin{aligned} E(\hat{I}_T+1/T)^{-2q} &\leqslant E(\tilde{I}_T+1/T)^{-2q} \\ &= \int_0^\infty (z+1/T)^{-2q} f_{\tilde{I}_T}(z)\,\mathrm{d}z \\ &= 2q\int_0^\infty (z+1/T)^{-2q-1}F_{\tilde{I}_T}(z)\,\mathrm{d}z \\ &\leqslant M(E(\tilde{I}_T)/2)^{-4q}T^{-2q}2q\int_0^{E(\tilde{I}_T)/2}(z+1/T)^{-2q-1}\,\mathrm{d}z + 2q\int_{E(\tilde{I}_T)/2}^\infty z^{-2q-1}\,\mathrm{d}z \\ &< M' < \infty \end{aligned}$$

for all T large enough.

Remark 4.1. Assumption 1 is satisfied for processes for which Rice's formula is valid (see e.g. Marcus, 1977) and X_0 and X'_0 are independent, e.g. mean-square differentiable Gaussian processes. Assumptions 2 and 3 are satisfied if the stationary zero-mean sequences $\xi_i = S_i - S_{i-1}$ and $\zeta_i = S'_i - S'_{i-1}$, respectively are φ-mixing with $\sum \varphi_i^{1/2} < \infty$ and $E|\xi_1|^{2+\delta}<\infty$ (Yokoyama, 1980), $E|\zeta_1|^{4q}$ is bounded for any $q>0$ since $|\zeta_1|\leqslant 1$.

Remark 4.2. $\hat{f}$ can be generalized by introducing a function $\lambda(T)=\mathrm{o}(T^{1/2})\uparrow\infty$ as $T\to\infty$ and using $\hat{f}_\lambda(u) = N_T(u)/(2\int_0^T |X'_t|\,\mathrm{d}t + \lambda(T))$, allowing us to relax the moment conditions in assumption 3 of Theorem 4.1 at the price of a possibly increased (but still $\mathrm{O}(T^{-1/2})$) bias.

References

Blanke, D., Bosq, D., 1997. Accurate rates of density estimators for continuous-time processes. Statist. Probab. Lett. 33, 185–191.

Bosq, D., 1996. Nonparametric statistics for stochastic processes: estimation and Prediction. Springer Lecture Notes in Statistics vol. 110. Springer, Berlin.

Castellana, J.V., Leadbetter, M.R., 1986. On smoothed probability density estimation for stationary processes. Stochastic Process. Appl. 21, 179–193.

Geman, D., Horowitz, J., 1960. Occupation times for smooth stationary processes. Ann. Probab. 1, 131–137.

Kutoyants, Yu., 1996. Efficient density estimation for ergodic diffusion. Research Memo. 607, Institute of Statistical Mathematics, Tokyo.

Marcus, M.B., 1977. Level crossings of a stochastic process with absolutely continuous sample paths. Ann. Probab. 5, 52–71.

Sköld, M., 1996. Kernel intensity estimation for marks and crossings of differentiable stochastic processes. Theory Stochastic Process. 2 (1–2), 273–284.

Wahba, G., 1975. Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation. Ann. Statist. 3, 15–29.

Yokoyama, R., 1980. Moment bounds for stationary mixing sequences. Z. Wahrsch. Gebiete 52, 45–57.