

Efficiency of the empirical distribution for ergodic diffusion

YURY A. KUTOYANTS

*Laboratoire de Statistique et Processus, Université du Maine, 72017 Le Mans, France
e-mail: kutoyants@univ-lemans.fr*

We consider the problem of stationary distribution function estimation at a given point by the observations of an ergodic diffusion process on the interval $[0, T]$ as $T \rightarrow \infty$. First we introduce a lower (minimax) bound on the risk of all estimators and then we prove that the empirical distribution function attains this bound. Hence this estimator is asymptotically efficient in the sense of the given bound.

Keywords: diffusion process; minimax bound; nonparametric estimation.

1. Introduction

We consider the problem of estimation of a one-dimensional distribution function $F(x)$ by the observation of a diffusion process $\{X_t, 0 \leq t \leq T\}$ as $T \rightarrow \infty$. We suppose that the process $X_t, t \geq 0$, possesses ergodic properties with invariant measure \mathbf{P}^* and $F(x) = \mathbf{P}^*((-\infty, x])$. We introduce a lower (minimax) bound on the risks of all estimators, then we define the asymptotically efficient estimators as estimators attaining this bound, and finally we show that the empirical distribution function is asymptotically efficient in this problem. The same program was already realized for several other models of observations. For independent identically distributed random variables this was done by Dvoretzky *et al.* (1956) see also Millar 1983, section VIII, and references therein); then Penev (1991) proved the asymptotic efficiency of the empirical distribution function for exponentially ergodic Markov chains with state space $[0, 1]$ (the more general case was treated by van der Vaart and Wellner 1990) and further generalizations were given by Greenwood and Wefelmeyer (1995). The difference between these results lies in the types of model, the regularity conditions and the choice of the definitions of asymptotic optimality, and common to all of them is the possibility of $n^{1/2}$ -consistent estimation of the underlying distribution. An exhaustive description of such situations has been given in Bickel *et al.* (1993). Our statement of the problem can be termed *semi-parametric* because we estimate the one-dimensional parameter $\tau = F(x)$, i.e., the value of the unknown function at one point x only (Bickel 1993, p. 59).

The diffusion process $X_t, t \geq 0$, is supposed to be a solution of the stochastic differential equation

$$dX_t = S(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T, \quad (1)$$

where $\{W_t, t \geq 0\}$ is the standard Wiener process. The trend coefficient $S(\cdot)$ is unknown to the observer and the diffusion coefficient $\sigma(\cdot)^2$ is a known positive function. Recall that the diffusion coefficient can be estimated without error by the observation of (1) (see, for example, Genon-Catalot and Jacod 1994).

We suppose that the functions $S(\cdot)$ and $\sigma(\cdot)$ satisfy the global Lipschitz condition

$$|S(x) - S(y)| + |\sigma(x) - \sigma(y)| \leq L|x - y|;$$

so (1) has a unique strong solution (see, for example, Liptser and Shirayev (1977, Theorem 4.6). This condition will not be used directly; so any other condition providing that property of the solution can replace the given one. We assume for simplicity of exposition that the functions $S(\cdot)$ and $\sigma(\cdot)$ are continuous.

We suppose that

$$\int_0^y \frac{S(v)}{\sigma(v)^2} dv \rightarrow -\infty, \quad \text{as } |y| \rightarrow \infty,$$

and the condition

$$G(S) \equiv G = \int_{-\infty}^{\infty} \sigma(y)^{-2} \exp\left(2 \int_0^y \frac{S(v)}{\sigma(v)^2} dv\right) dy < \infty$$

are fulfilled. The class of such functions $S(\cdot)$ we denote by Θ . These conditions provide the existence of a stationary distribution

$$F(x) = G(S)^{-1} \int_{-\infty}^x \sigma(y)^{-2} \exp\left(2 \int_0^y \frac{S(v)}{\sigma(v)^2} dv\right) dy$$

for the diffusion process (1) (Mandl 1968). Under this condition the random process X_t , $t \geq 0$, has ergodic properties, i.e., for any measurable function $g(\cdot)$ with $\mathbf{E}|g(\xi)| < \infty$ (ξ has $F(\cdot)$ as distribution function)

$$\mathbf{P}\left(\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(X_t) dt = \mathbf{E}g(\xi)\right) = 1$$

(Mandl 1968, p. 93).

Below we are interested in the problem of nonparametric ($S(\cdot)$ is unknown) estimation of the function $F(x)$. The empirical distribution function (EDF)

$$\hat{F}_T(x) = \frac{1}{T} \int_0^T \chi_{\{X_t < x\}} dt \tag{2}$$

is a *natural* estimator of the function $F(x)$. Recall that the EDF

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^n \chi_{\{X_j < x\}}$$

in the case of independent identically distributed observations $\{X_1, \dots, X_n\}$ is uniformly consistent, asymptotically normal and asymptotically efficient. In this paper it is shown that for the ergodic diffusion process (1) the EDF (2) has similar properties.

The density function will be denoted as

$$f(y) = G(S)^{-1} \sigma(y)^{-2} \exp \left\{ 2 \int_0^y \frac{S(v)}{\sigma(v)^2} dv \right\}. \tag{3}$$

2. Lower bound

To derive the lower minimax bound we shall follow the same approach which was applied in the density estimation problem for this model of observations by Kutoyants (1995). The idea of approximating this nonparametric bound by a family of bounds for parametric families and then choosing the worst parametric family was suggested by Stein (1956) and was realized by Levit (1973), and Koshevnik and Levit (1976) (see also the monographs by Ibragimov and Khasminskii 1981, Chapter 4 and Bickel *et al.* 1993, Chapter 3). In the present work we take the same approach.

So as in the work of Kutoyants (1995) we fix some function $S(\cdot) \in \Theta$ and suppose that the observed process is

$$dX_t = [S(X_t) + H(X_t)] dt + \sigma(X_t) dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T, \tag{4}$$

where $H(\cdot)$ is such that $S(\cdot) + H(\cdot) \in \Theta$. Introduce the set

$$U_\delta \equiv \{H(\cdot): \sup_{x \in \mathbb{R}} |H(x)| \leq \delta, S(\cdot) + H(\cdot) \in \Theta\}$$

and denote by $\{\mathbf{P}_H^{(T)}, H(\cdot) \in U_\delta\}$ the corresponding family of measures $\mathbf{P}_H^{(T)}$ induced by the process (4) in the space $\mathcal{C}[0, T]$ of continuous functions on $[0, T]$. The mathematical expectation with respect to this measure will be denoted by \mathbf{E}_H . Further consider the distribution function

$$F_H(x) = G(S + H)^{-1} \int_{-\infty}^x \sigma(y)^{-2} \exp \left(2 \int_0^y \frac{S(v) + H(v)}{\sigma(v)^2} dv \right) dy,$$

with corresponding normalizing constant $G(S + H)$ and put

$$I_* = \left\{ 4 \mathbf{E} \left(\frac{F(\xi \wedge x) \{1 - F(\xi \vee x)\}}{\sigma(\xi) f(\xi)} \right)^2 \right\}^{-1}, \tag{5}$$

where $\xi \wedge x = \min(\xi, x)$ and $\xi \vee x = \max(\xi, x)$. Introduce the set

$$\Theta_* = \{S(\cdot): \sup_{H(\cdot) \in U_\delta} G(S + H) < \infty, S(\cdot) \in \Theta\}.$$

The set Θ_* is not empty. For example, if $\sigma(\cdot) \equiv 1$ and $S(y) \operatorname{sgn} y \leq -\gamma_*$ for all large values of $|y|$ with some $\gamma_* > 0$ then $S(\cdot) \in \Theta_*$ for any $\delta < \gamma_*$.

We suppose that the loss function $\mathcal{L}(\cdot)$ has the following ‘usual’ properties: $\mathcal{L}(y, z) = \mathcal{L}(y - z)$, $y, z \in \mathbb{R}$; $\mathcal{L}(\cdot)$ is non-negative on \mathbb{R} ; $\mathcal{L}(0) = 0$ is continuous at $z = 0$ but is not identically 0; $\mathcal{L}(\cdot)$ is symmetric and non-decreasing on \mathbb{R}_+ ; finally the function $\mathcal{L}(z)$ grows as $z \rightarrow \infty$ more slowly than any one of the functions $\exp(\varepsilon z^2)$, $\varepsilon > 0$ (see, for example, Ibragimov and Khasminskii 1981).

Theorem 1. Let $S(\cdot) \in \Theta_*$ and $I_* > 0$; then

$$\lim_{\delta \rightarrow 0} \liminf_{T \rightarrow \infty} \inf_{F_T} \sup_{H(\cdot) \in U_\delta} \mathbf{E}_H \mathcal{L} [T^{1/2} \{F_T(x) - F_H(x)\}] \geq \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \mathcal{L}(xI_*^{-1/2}) e^{-x^2/2} dx, \quad (6)$$

where \inf is taken over all possible estimators $F_T(x)$.

Proof. The supremum on U_δ is estimated from below by supremum on some parametric family with a special parametrization passing through the model with $S(\cdot)$. For this parametric model we apply the Hajek–Le Cam inequality, then maximize the right-hand side of the last inequality and find the *worst parametric family* (with minimal Fisher information). This last quantity (risk) is just the right-hand side of (6).

Let us introduce the parametric family of functions

$$S_h(x) = S(x) + (h - \tau)\psi(x)\sigma(x)^2,$$

where $h \in (\tau - \gamma, \tau + \gamma)$, $\gamma > 0$, and the function $\psi(\cdot)$ has a compact support, $S_h(\cdot) \in \Theta$. Then, for small γ ,

$$(h - \tau)\psi(x)\sigma(x)^2 \in U_\delta.$$

The corresponding family of stochastic differential equations is

$$dX_t = [S(X_t) + (h - \tau)\psi(X_t)\sigma(X_t)^2] dt + \sigma(X_t) dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T, \quad (7)$$

with $h \in (\tau - \gamma, \tau + \gamma)$ and the corresponding family of measures is $\{\mathbf{P}_h^{(T)}, h \in (\tau - \gamma, \tau + \gamma)\}$. We can consider the problem of the estimation of h based on observation of (7). In our assumptions the stochastic process (7) has also the ergodic properties if $h = \tau$; hence the family of measures $\{\mathbf{P}_h^{(T)}, h \in (\tau - \gamma, \tau + \gamma)\}$ is *locally asymptotically normal* at the point $h = \tau$, i.e., the likelihood ratio admits the representation (see, for example, Kutoyants 1984, Theorem 3.3.8)

$$\frac{d\mathbf{P}_{\tau+T^{-1/2}u}^{(T)}}{d\mathbf{P}_\tau^{(T)}}(X) = \exp \left(u\Delta_T - \frac{u^2}{2} I_\psi + r_T \right),$$

where $u \in R$,

$$I_\psi = \int_{-\infty}^{\infty} \psi(x)^2 \sigma(x)^2 f(x) dx, \quad \mathbf{P}_\tau - \lim_{T \rightarrow \infty} r_T = 0,$$

$$\Delta_T = T^{-1/2} \int_0^T \psi(X_t)\sigma(X_t) dW_t, \quad \mathcal{L}_\tau\{\Delta_T\} \Rightarrow \mathcal{N}(0, I_\psi).$$

Thus by the Hajek–Le Cam inequality

$$\lim_{\delta \rightarrow 0} \liminf_{T \rightarrow \infty} \sup_{h_T \mid |h - \tau| < \delta} \mathbf{E}_h \mathcal{L} \{T^{1/2}(h_T - h)\} \geq \mathbf{E} \mathcal{L}(\xi I_\psi^{-1/2}), \quad (8)$$

with $\xi \sim \mathcal{N}(0, 1)$ and, as was noted by Ibragimov and Khasminskii (1981, p. 217), the difference $h_T - h$ can be replaced by $h_T - h + o(|h - \tau|)$ without changing the right-hand side of (8).

We put

$$F_h(x) = G_h^{-1} \int_{-\infty}^x \sigma(y)^{-2} \exp \left(2 \int_0^y \frac{S(v)}{\sigma(v)^2} dv + 2(h - \tau) \int_0^y \psi(v) dv \right) dy.$$

The function $\psi(\cdot)$ has a compact support; hence we can expand $F_h(\cdot)$ in powers of $h - \tau$ in the vicinity of the point τ and obtain

$$\begin{aligned} F_h(x) &= F(x) + 2(h - \tau) \left(\int_{-\infty}^x \int_0^y \psi(v) dv f(y) dy - F(x) \int_{-\infty}^{\infty} \int_0^y \psi(v) dv f(y) dy \right) + o(|h - \tau|) \\ &= F(x) + 2(h - \tau) \mathbf{E}\{[\chi_{\{\xi < x\}} - F(x)]\Psi(\xi)\} + o(|h - \tau|), \end{aligned}$$

where ξ as before has the distribution function $F(\cdot)$ and

$$\Psi(\xi) = \int_0^{\xi} \psi(y) dy.$$

Let us put $\tau = F(x)$, $\tau_h = F_h(x)$ and introduce the class \mathcal{H} of functions $\psi(\cdot)$ satisfying the equality

$$\mathbf{E}\{[\chi_{\{\xi < x\}} - F(x)]\Psi(\xi)\} = \frac{1}{2}.$$

Then for $\psi(\cdot) \in \mathcal{H}$ we have the expansion

$$\begin{aligned} F_h(x) &= F(x) + 2(h - \tau) \mathbf{E}\{[\chi_{\{\xi < x\}} - F(x)]\Psi(\xi)\} + o(|h - \tau|) \\ &= h + o(|h - \tau|) \end{aligned}$$

or $\tau_h = h + o(|h - \tau|)$. Let the function $\psi(\cdot) \in \mathcal{H}$ then

$$\begin{aligned} \sup_{H(\cdot) \in U_\delta} \mathbf{E}_H \ell [T^{1/2} \{F_T(x) - F_H(x)\}] &\geq \sup_{|h - \tau| < \gamma} \mathbf{E}_h \ell [T^{1/2} \{F_T(x) - F_h(x)\}] \\ &= \sup_{|h - \tau| < \gamma} \mathbf{E}_h \ell [T^{1/2} \{h_T - h - o(|h - \tau|)\}], \end{aligned}$$

where $h_T = F_T(x)$ is an arbitrary estimator of h . So we can choose $\gamma = \gamma(\delta) \rightarrow 0$ as $\delta \rightarrow 0$ in such a way that for any estimator $F_H(x)$

$$\begin{aligned} \lim_{\delta \rightarrow 0} \liminf_{T \rightarrow \infty} \sup_{H(\cdot) \in U_\delta} \mathbf{E}_H \ell [T^{1/2} \{F_T(x) - F_H(x)\}] &\geq \lim_{\gamma \rightarrow 0} \liminf_{T \rightarrow \infty} \sup_{|h - \tau| < \gamma} \mathbf{E}_h \ell [T^{1/2} \{h_T - h - o(|h - \tau|)\}] \\ &\geq \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \ell(xI_\psi^{-1/2}) e^{-x^2/2} dx. \end{aligned}$$

The last integral is a monotonically decreasing function of $I_\psi = \mathbf{E}\psi(\xi)^2 \sigma(\xi)^2$; hence to find the worst parametric family (defined by $\psi(\cdot)$) we have to minimize I_ψ on the class \mathcal{H} .

Fubini's theorem allows us to write

$$\begin{aligned}
 \mathbf{E}\{\chi\{\xi < x\} - F(x)\}\Psi(\xi) &= \int_{-\infty}^x \{1 - F(x)\}\Psi(y)f(y) \, dy - F(x)\int_x^{\infty} \Psi(y)f(y) \, dy \\
 &= \{F(x) - 1\} \int_{-\infty}^0 \int_{-\infty}^0 \chi_{\{v \geq y\}} \psi(v)f(y) \, dv \, dy \\
 &\quad + \{1 - F(x)\} \int_0^x \int_0^x \chi_{\{v < y\}} \psi(v)f(y) \, dv \, dy \\
 &\quad - F(x)\{1 - F(x)\}\Psi(x) - F(x)\int_x^{\infty} \int_x^{\infty} \chi_{\{v < y\}} \psi(v)f(y) \, dv \, dy \\
 &= \{F(x) - 1\} \int_{-\infty}^x \psi(v)F(v) \, dv - F(x)\int_x^{\infty} \psi(v)\{1 - F(v)\} \, dv \\
 &= \int_{-\infty}^{\infty} \psi(v)[\{F(x) - 1\}F(v)\chi_{\{v < x\}} + \{F(v) - 1\}F(x)\chi_{\{v \geq x\}}] \, dv.
 \end{aligned}$$

The function $\psi(\cdot)$ belongs to the class \mathcal{K} hence by the Cauchy-Schwarz inequality

$$\begin{aligned}
 \frac{1}{4} &= [\mathbf{E}\{\chi_{\{\xi < x\}} - F(x)\}\Psi(\xi)]^2 \\
 &= \left(\int_{-\infty}^{\infty} \psi(v)\{F(v \vee x) - 1\}F(v \wedge x) \, dv \right)^2 \\
 &\leq \int_{-\infty}^{\infty} \psi(v)^2 \sigma(v)^2 f(v) \, dv \int_{-\infty}^{\infty} \frac{[\{F(v \vee x) - 1\}F(v \wedge x)]^2}{\sigma(v)^2 f(v)} \, dv \\
 &= \mathbf{E}\psi(\xi)^2 \sigma(\xi)^2 \mathbf{E}\left(\frac{\{F(\xi \vee x) - 1\}F(\xi \wedge x)}{\sigma(\xi)f(\xi)} \right)^2.
 \end{aligned}$$

Therefore

$$I_{\psi} \geq \left\{ 4\mathbf{E}\left(\frac{\{1 - F(\xi \vee x)\}F(\xi \wedge x)}{\sigma(\xi)f(\xi)} \right)^2 \right\}^{-1} = I_*$$

for all $\psi(\cdot) \in \mathcal{K}$ and we obtain the equality for the function

$$\psi_*(v) = \frac{C}{\sigma(v)^2 f(v)} \{1 - F(v \vee x)\}F(v \wedge x),$$

with some constant $C > 0$. This function does not have compact support but we can introduce a sequence $\psi_*^N(v) = \psi_*(v)\chi_{\{|v| < N\}}$, $N \rightarrow \infty$ to obtain (6). \square

The inequality (6) suggests that we introduce the following definition of asymptotically efficient estimator.

Definition. Let the conditions of Theorem 1 be fulfilled then we say that the estimator $F_T^*(x)$ is locally asymptotically minimax (LAM) for the loss function $\mathcal{L}(\cdot)$ if for any $x \in R$

$$\lim_{\delta \rightarrow 0} \lim_{T \rightarrow \infty} \sup_{H(\cdot) \in U_\delta} \mathbf{E}_{H \mathcal{L}} [T^{1/2} \{F_T^*(x) - F_H(x)\}] = \mathbf{E} \mathcal{L}(\zeta I_*^{-1/2})$$

where $\mathcal{L}(\zeta) = \mathcal{N}(0, 1)$.

We shall show below that for the polynomial loss functions $\mathcal{L}(\cdot)$ the EDF is LAM.

3. Empirical distribution function

By the law of large numbers the EDF is a consistent estimator of the value $F(x)$ and even uniformly consistent.

Theorem (Glivenko–Cantelli).

$$\mathbf{P} \left\{ \lim_{T \rightarrow \infty} \sup_x |\hat{F}_T(x) - F(x)| = 0 \right\} = 1.$$

Proof. The proof is well known and coincides with the proof of this theorem in the i.i.d. case. □

The normed difference

$$W_T(x) = T^{1/2} [\hat{F}_T(x) - F(x)] = \frac{1}{T^{1/2}} \int_0^T \{\mathcal{X}_{\{X_t < x\}} - F(x)\} dt$$

is asymptotically normal by the following *central limit theorem*.

Lemma 1. Let the integrals

$$\mathbf{E} h(\xi) = \int_{-\infty}^{\infty} h(y) f(y) dy = 0$$

and

$$0 < 4G \int_{-\infty}^{\infty} h(y) \int_y^s h(z) f(z) p(s) f(y) dz ds dy \equiv D(h, h) < \infty,$$

where

$$p(s) = \exp \left(-2 \int_0^s \frac{S(v)}{\sigma(v)^2} dv \right)$$

converge absolutely; then

$$\mathcal{L} \left\{ \frac{1}{T^{1/2}} \int_0^T h(X_t) dt \right\} \Rightarrow \mathcal{N}(0, D(h, h)). \tag{9}$$

Proof. For the proof see Mandl (1968, pp. 93 and 94) (see also Lanska 1979 for this formulation). □

Hence the EDF is a uniformly consistent and asymptotically normal estimator of the one-dimensional distribution function $F(x)$. Below we calculate its limit variance and compare with the bound (6).

By Lemma 1 the limit variance of $W_T(x)$ is $D(h_x, h_x)$ with

$$h_x(y) = \chi_{\{y < x\}} - F(x)$$

This quantity is calculated in the Appendix and is equal to

$$D(h_x, h_x) = 4\mathbf{E} \left(\frac{\{1 - F(\xi \vee x)\} F(\xi \wedge x)}{\sigma(\xi) f(\xi)} \right)^2 = I_*^{-1}.$$

Therefore, if the convergence is uniform, then the EDF is LAM for the bounded loss functions.

Remark. The central limit theorem (9) was proved by Mandl (1968) with the help of the central limit theorem for sums of independent identically distributed random variables and the conditions of uniform asymptotic normality for such sums can be found in Ibragimov and Khasminskii (1981, Appendix 1).

The asymptotic efficiency of the EDF can be also proved for the polynomial loss functions $\mathcal{L}(u) = |u|^p$ with $p \geq 2$. To show this we shall use another form of Lemma 1 and we have to strengthen the conditions.

Let us introduce the functions

$$I_*(H) = \left\{ 4f_H(x)^2 \mathbf{E}_H \left(\frac{F_H(\xi \wedge x) - F_H(x)F_H(\xi)}{\sigma(\xi)f_H(\xi)} \right)^2 \right\}^{-1}$$

and

$$g(y) = 2 \int_{-\infty}^y \frac{F_H(v \wedge x) - F_H(v)F_H(x)}{\sigma(v)^2 f_H(v)} dv$$

and introduce the following condition.

(C₁) There exists a number $p_* > 2$ such that

$$\begin{aligned} & \sup_{H(\cdot) \in U_\delta} \mathbf{E}_H |g(\xi)|^{p_*} < \infty, \\ & \sup_{H(\cdot) \in H_\delta} \mathbf{E}_H \left(\frac{F_H(\xi \wedge x) - F_H(\xi)F_H(x)}{\sigma(\xi)f_H(\xi)} \right)^{p_*} < \infty \end{aligned}$$

and the law of large numbers

$$\mathbf{P}_H - \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left(\frac{F_H(X_t \wedge x) - F_H(x)F_H(X_t)}{\sigma(X_t)f_H(X_t)} \right)^2 dt = I_*(H)^{-1}$$

is uniform on $H(\cdot) \in U_\delta$.

Theorem 2. *Let the condition (C₁) be fulfilled, and let $S(\cdot) \in \Theta_*$, $I_* > 0$ and $I_*(H)$ be continuous at the point $H(\cdot) = 0$; then the EDF $\hat{F}_T(x)$ is LAM for the loss functions $\ell(u) = |u|^p$ with $p < p_*$.*

Proof. By the Itô formula we have

$$\frac{1}{T^{1/2}} \int_0^T [\mathcal{X}_{\{X_t < x\}} - F_H(x)] dt = \frac{g(X_T) - g(X_0)}{T^{1/2}} - \frac{2}{T^{1/2}} \int_0^T \frac{F_H(X_t \wedge x) - F_H(X_t)F_H(x)}{\sigma(X_t)f_H(X_t)} dW_t.$$

The last stochastic integral is, by condition (C₁), asymptotically normal, uniformly on $H(\cdot) \in U_\delta$:

$$\frac{2}{T^{1/2}} \int_0^T \frac{F_H(X_t \wedge x) - F_H(X_t)F_H(x)}{\sigma(X_t)f_H(X_t)} dW_t \Rightarrow \mathcal{N}(0, I_*(H)^{-1})$$

(Kutoyants 1984, Theorem 3.3.3) and the random variables $\eta_T(x) = \ell[T^{1/2}\{\hat{F}_T(x) - F_H(x)\}]$ are uniformly integrable; for any $p < p_*$,

$$\begin{aligned} \sup_{T, H(\cdot) \in U_\delta} \mathbf{E}_H |\eta_T(x)|^{p^*/p} &\leq C_1 T^{-p^*/2} \mathbf{E}_H \{g(X_T) + g(X_0)\}^{p^*} \\ &+ C_2 T^{-1} \mathbf{E}_H \int_0^T \left(\frac{F_H(X_t \wedge x) - F_H(X_t)F_H(x)}{\sigma(X_t)f_H(X_t)} \right)^{p^*} dt \leq C. \end{aligned}$$

Therefore

$$\sup_{H(\cdot) \in U_\delta} \mathbf{E}_H \ell[T^{1/2}\{\hat{F}_T(x) - F_H(x)\}] \rightarrow \sup_{H(\cdot) \in U_\delta} \mathbf{E} \ell\{\xi I_*(H)^{-1/2}\},$$

and the LAM of EDF now follows from the continuity of $I_*(H)$. □

4. Concluding remarks

It is interesting to have the similar lower bound for loss functions such as $\ell(\sup_x T^{1/2}\{\hat{F}_T(x) - F(x)\})$ and to prove the LAM of EDF in this situation as was done in i.i.d. case.

Another problem closely related with this model of observations is the density function $f(x)$ estimation. As was shown by Castellana and Leadbetter (1986) the rate of convergence

of kernel-type estimators is $T^{1/2}$. It can be shown that the lower bound is similar to that given above in (6) but with

$$I_* = \left\{ 4f(x)^2 \mathbf{E} \left(\frac{\chi_{\{\xi > x\}} - F(\xi)}{\sigma(\xi)f(\xi)} \right)^2 \right\}^{-1}$$

and the kernel-type estimators as well as an unbiased estimator

$$f_T^*(x) = \frac{2}{T\sigma(x)^2} \int_0^T \chi_{\{X_t < x\}} dX_t$$

of the density are LAM in this problem (Kutoyants 1995).

All these results, including those given in the present paper, were given (without detailed proofs) in Kutoyants (1996).

Appendix

Below we calculate the limit variance $D(h_x, h_x)$ of EDF $\hat{F}_T(x)$ as follows. First we have

$$\int_{-\infty}^s h_x(z)f(z) dz = F(s \wedge x) - F(s)F(x).$$

Then

$$\begin{aligned} D(h_x, h_x) = & 4 \int_{-\infty}^x \{1 - F(x)\}^2 \int_y^0 \frac{F(s)}{\sigma(s)^2 f(s)} ds f(y) dy \\ & - 4F(x) \int_x^\infty f(y) \int_y^0 \frac{F(s \wedge x) - F(s)F(x)}{\sigma(s)^2 f(s)} ds dy, \end{aligned}$$

because

$$Gp(s) = \sigma(s)^{-2} f(s)^{-1}$$

and for all $y < x$ in the first integral we have $s \wedge x = s$.

Exchanging the order of integration we obtain

$$\begin{aligned} \int_{-\infty}^x f(y) \int_y^0 \frac{F(s)}{\sigma(s)^2 f(s)} ds dy &= \int_{-\infty}^0 f(y) \int_{-\infty}^0 \chi_{\{s \geq y\}} \frac{F(s)}{\sigma(s)^2 f(s)} ds dy \\ &\quad - \int_0^x f(y) \int_0^x \chi_{\{s < y\}} \frac{F(s)}{\sigma(s)^2 f(s)} ds dy \\ &= \int_{-\infty}^0 \frac{F(s)^2}{\sigma(s)^2 f(s)} ds - \int_0^x \frac{F(s)\{F(x) - F(s)\}}{\sigma(s)^2 f(s)} ds \end{aligned}$$

and

$$\begin{aligned} & \int_x^\infty f(y) \int_y^0 \frac{F(s)F(x) - F(s \wedge x)}{\sigma(s)^2 f(s)} ds dy \\ &= \int_x^\infty f(y) \left(\int_0^x \frac{F(s)\{1 - F(x)\}}{\sigma(s)^2 f(s)} ds + \int_x^y \frac{F(x)\{1 - F(s)\}}{\sigma(s)^2 f(s)} ds \right) dy \\ &= \{1 - F(x)\}^2 \int_0^x \frac{F(s)}{\sigma(s)^2 f(s)} ds + F(x) \int_x^\infty \frac{\{1 - F(s)\}^2}{\sigma(s)^2 f(s)} ds. \end{aligned}$$

Therefore

$$\begin{aligned} D(h_x, h_x) &= 4\{1 - F(x)\}^2 \int_{-\infty}^x \frac{F(s)^2}{\sigma(s)^2 f(s)} ds + 4F(x)^2 \int_x^\infty \frac{\{1 - F(s)\}^2}{\sigma(s)^2 f(s)} ds \\ &= 4 \int_{-\infty}^\infty \frac{F(s \wedge x)^2 \{1 - F(s \vee x)\}^2}{\sigma(s)^2 f(s)} ds \\ &= 4\mathbf{E} \left(\frac{\{1 - F(\xi \vee x)\} F(\xi \wedge x)}{\sigma(\xi) f(\xi)} \right)^2 = I_*^{-1}. \end{aligned}$$

Acknowledgements

The author is indebted to the referee for pointing out several references and for advice improving the presentation of the paper.

References

- Bickel, P.J. (1993) Estimation in semiparametric models. In C.R. Rao (ed.), *Multivariate Analysis, Future Directions*, pp. 55–73. Amsterdam: Elsevier.
- Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: John Hopkins University Press.
- Castellana, J.V. and Leadbetter, M.R. (1986) On smoothed density estimation for stationary processes. *Stochastic Proc. Appl.*, **21**, 179–193.
- Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. *Ann. Statist.*, **27**, 642–669.
- Genon-Catalot, V. and Jacod, J. (1994) On the estimation of diffusion coefficient for diffusion processes. *Scand. J. Statist.*, **21**, 193–221.
- Greenwood, P.E. and Wefelmeyer, W. (1995) Efficiency of empirical estimators for Markov chains. *Ann. Statist.*, **23**, 132–143.
- Ibragimov, I.A. and Khraminskii, R.Z. (1981) *Statistical Estimation. Asymptotic Theory*. New York: Springer-Verlag.

- Koshevnik, Ya.A. and Levit, B.Ya. (1976) On a non-parametric analog of the information matrix. *Theor. Probab. Appl.*, **21**, 738–753.
- Kutoyants, Yu.A. (1984) *Parameter Estimation for Stochastic Processes*. Berlin: Heldermann.
- Kutoyants, Yu.A. (1995) On density estimation by the observations of ergodic diffusion process. Preprint 95-8, January, Université du Maine, Le Mans.
- Kutoyants, Yu.A. (1997) Some problems of nonparametric estimation by the observations of ergodic diffusion process. *Statist. Probab. Lett.* To appear.
- Lanska, V. (1979) Minimum contrast estimation in diffusion processes. *J. Appl. Probab.*, **16**, 65–75.
- Levit, B.Ya (1973) On optimality of some statistical estimates. *Proceedings of the Prague Symposium on Asymptotic Statistics*, Vol. 2, pp. 215–238.
- Liptser, R.S. and Shiriyayev, A.N. (1977) *Statistics of Random Processes*, Vol. 1. New York: Springer-Verlag.
- Mandl, P. (1968) *Analytical Treatment of One-Dimensional Markov Processes*. Prague: Academia. Berlin: Springer.
- Millar, P.W. (1983) The minimax principle in asymptotic statistical theory. In P.L. Hennequin, (ed.), *Ecole d'Été de Probabilités de Saint Flour XI 1981*, pp. 75–266. Lect. Notes in Math. 976. Berlin: Springer.
- Penev, S. (1991) Efficient estimation of the stationary distribution for exponentially ergodic Markov chains. *J. Statist. Plan. Inference*, **27**, 105–123.
- Stein, C. (1956) Efficient nonparametric testing and estimation. *Proceedings of the Third Berkeley Symposium*, Vol. 1, pp. 187–195.
- van der Vaart, A.W. and Wellner, J.A. (1990) Prohorov and continuous mapping theorems in the Hoffmann–Jørgensen weak convergence theory with applications to convolution and asymptotic minimax theorems. Technical Report 157, Department of Statistics, University of Washington, Seattle, WA.

Received March 1995 and revised November 1996