



UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI[°]

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series “*Mathematics and Statistics*”

n. 10/MS – 2008

*On optimality of the empirical distribution function
for the estimation of the invariant distribution
function of a diffusion process*

by

Ilia Negri

COMITATO DI REDAZIONE[§]

Series Information Technology (IT): Stefano Paraboschi
Series Mathematics and Statistics (MS): Luca Brandolini, Ilia Negri

[§] L'accesso alle *Series* è approvato dal Comitato di Redazione. I *Working Papers* della Collana dei Quaderni del Dipartimento di Ingegneria dell'Informazione e Metodi Matematici costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

On optimality of the empirical distribution function for the estimation of the invariant distribution function of a diffusion process *

Ilia Negri[†]

Abstract

In this work we present some results on the optimality of the empirical distribution function as estimator of the invariant distribution function of an ergodic diffusion process. The result presented here were obtained in different previous works under conditions that we have uniformed in this work. It is well known that the empirical distribution function it is an unbiased and uniformly consistent estimator for the invariant distribution function of an ergodic diffusion process. It is also an efficient estimator in the sense that the risk of this estimator attain an asymptotic minimax lower bound. In this paper we review some results on the problem of the efficiency of the empirical distribution considering three type of risk functions. The first one is in a semi-parametric set-up. The second one is the integrated mean square error and the third is based on the *sup* norm.

key words: invariant distribution function, efficiency, lower bound, efficient estimator.

2000 MSC: 60G35, 62M20.

*This work has been supported by the local Grant sponsored by University of Bergamo: *Theoretical and computational problems in statistics for continuously and discretely observed diffusion processes*

[†]Department of Information Technology and Mathematical Methods, Viale Marconi 5, I-24044 Dalmine (BG), e-mail: ilia.negri@unibg.it

1 Introduction

In this paper we review some results on the optimality, in a minimax sense, of the empirical distribution function as estimator of the invariant distribution function $F(x)$, $x \in \mathbb{R}$ by the continuous observation of a diffusion process $\{X_t : 0 \leq t \leq T\}$. Let X_t , $t \geq 0$ be an ergodic diffusion process with invariant measure μ and $F(x) = \mu\{(-\infty, x]\}$. The empirical distribution function, denoted by $\hat{F}_T(x)$, $x \in \mathbb{R}$, is a natural estimator for $F(x)$. It is well known that the empirical distribution function is an unbiased and uniformly consistent estimator, by the Glivenko–Cantelli theorem, for the invariant distribution function of an ergodic diffusion process. It is also an efficient estimator in the sense that the risk of this estimator attain an asymptotic minimax lower bound. In this paper we review some results on the problem of the efficiency of the empirical distribution considering three type of risk functions. The first one is in a semi-parametric set-up. The second one is the integrated mean square error and the third is based on the *sup* norm. These results where obtained in previous works (Kutoyants, 1997, Kutoyants and Negri, 2002 and Negri 1998) under different conditions. Here we have rewritten the conditions in a form that they are comparable. In this way we can see how changing the definition of risk, also the conditions have to be strength in some natural directions. For example passing to the risk defined when the estimation is in one fixed point to the risk based on the sup norm the conditions required involve the supremum over every $x \in \mathbb{R}$. In a similar way when it is considered the integrated mean square error risk the required conditions involves integral of the same type as in the definition of the risk. Synthetically in the sup norm we need conditions that assure the weakly uniform convergence of the empirical process. To prove the asymptotically efficiency for the integrated mean square error the conditions needed are less restrictive, essentially because to reach the lower bound we do not need the weak convergence of the empirical process. To prove the asymptotic efficiency the procedure is the same in the three different set-up. First it is derived a local asymptotic minimax lower bound for the risk of any estimator of the invariant distribution function, then the efficient estimator is defined as the estimator that attains the lower bound, and finally it is showed that the empirical distribution function attains that bound.

In the next section we present the statement of the problem and the assumptions. In section 3 we present the statistical problem and the conditions used through the text. In section 4 we present the principal properties of the empirical distribution function. Section 5 is divided in three sub-section where in each of them we present an asymptotic minimax bound for three different risk function. Finally in the conclusive section 6 we give an example.

2 Preliminaries

In this section we introduce the model and his principal properties. Let us consider a one dimensional diffusion process

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

where $\{W_t : t \geq 0\}$ is a standard Wiener process, and the initial value X_0 is independent of W_t , $t \geq 0$. Let us introduce some conditions on the pair (S, σ) .

Condition \mathcal{A}_1 . The function S is locally bounded, the function σ^2 is positive and continuous and for some $A > 0$ the condition

$$xS(x) + \sigma(x)^2 \leq A(1 + x^2), \quad x \in \mathbb{R}$$

holds.

If condition \mathcal{A}_1 holds true the equation (1) has an unique weak solution (see for example Kutoyants, 2004 p. 27). Note that every condition that assures the existence of a weak solution can be assumed instead of \mathcal{A}_1 . Although a strong solution of (1) is not required also conditions that assure such kind of solution are sometimes required for simplicity. For example in Kutoyants and Negri (2002), the Lipschitz conditions were required on the coefficient S and σ .

Condition \mathcal{A}_2 . The functions S and σ are such that:

$$V_S(x) = \int_0^x \exp \left\{ 2 \int_0^y \frac{S(v)}{\sigma(v)^2} dv \right\} dy \rightarrow \pm\infty, \quad as \quad x \rightarrow \pm\infty \quad (2)$$

and

$$G(S) = \int_{-\infty}^{+\infty} \frac{1}{\sigma(x)^2} \exp \left\{ 2 \int_0^x \frac{S(v)}{\sigma(v)^2} dv \right\} dx < +\infty. \quad (3)$$

If condition \mathcal{A}_2 is satisfied then the weak solution of (1), X_t , $t \geq 0$, has the ergodic property (see for example Gikhman and Skorohod, 1972 or Durrett, 1996), that is, there exists an invariant probability measure μ such that for every measurable function g such that $\mathbf{E}|g(\xi)| < \infty$, where ξ has the invariant measure as distribution, and \mathbf{E} denote the expectation with respect the invariant measure, we have, with probability one,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(X_t) dt = \int_{\mathbb{R}} g(z) f(z) dz$$

where f is the invariant density of the invariant measure μ given by

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

The invariant distribution function is given by

$$F(x) = \int_{-\infty}^x \frac{1}{G(S)} \frac{1}{\sigma(y)^2} \exp \left\{ 2 \int_0^y \frac{S(v)}{\sigma(v)^2} dv \right\} dy. \quad (4)$$

From now on we suppose that the initial value X_0 has the invariant distribution, so the process (1) is strictly stationary.

Suppose we observe different diffusion processes $\{X_t : 0 \leq t \leq T\}$ given by equation (1) with drift coefficients respectively given by S_1 , S_2 and $S_0 = 0$ and initial value respectively X_0^1 , X_0^2 and X_0^0 . Let us introduce the following condition.

\mathcal{A}_3 . *The functions S_1 , S_2 and σ satisfy condition \mathcal{A}_1 and the densities (with respect to the Lebesgue measure) of the corresponding initial values X_0^1 , X_0^2 and X_0^0 have the same support (if the initial value is nonrandom, then we suppose that it takes the same value for all processes).*

If condition \mathcal{A}_3 holds true, all the measure P_S^T induced in the space \mathcal{C}_T , the space of all the continuous function on $[0, T]$ with uniform metric and Borel σ -algebra $\mathcal{B}(\mathcal{C}_T)$, by the processes $\{X_t : 0 \leq t \leq T\}$ given by equation (1) are equivalent and the corresponding Radon-Nikodym derivative or likelihood ratio,

$$L(S, S_1, X^T) = \frac{dP_S^T}{dP_{S_1}^T}(X^T)$$

are given by

$$\begin{aligned} L(S, S_1, X^T) &= \frac{G(S_1)}{G(S)} \exp \left\{ 2 \int_0^{X_0} \frac{S(v) - S_1(v)}{\sigma(v)^2} dv + \int_0^T \frac{S(X_t) - S_1(X_t)}{\sigma(X_t)^2} dX_t \right\} \\ &\cdot \exp \left\{ -\frac{1}{2} \int_0^T \frac{S(X_t)^2 - S_1(X_t)^2}{\sigma(X_t)^2} dt \right\}. \end{aligned}$$

Let us denote by \mathbf{E}_S^T the expected value with respect to the measure P_S^T . Here and in the sequel we use a subscript S (for example we have denoted with f_S and F_S the invariant density and the invariant distribution function respectively) to emphasize the dependence of a certain quantity from the unknown function S . So we denote by \mathbf{E}_S the expected value with respect to the invariant density depending on different S .

3 Statement of the statistical problem and conditions

Given the diffusion process

$$dX_t = S(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad X_0,$$

we suppose that conditions \mathcal{A}_1 and \mathcal{A}_2 are satisfied, and that X_0 has the invariant distribution, so the process X_t , $t \geq 0$ is ergodic and is strictly stationary. We are interested in the estimation of the invariant distribution function

$$F_S(x) = \int_{-\infty}^x \frac{1}{G(S)} \frac{1}{\sigma(y)^2} \exp \left\{ 2 \int_0^y \frac{S(v)}{\sigma(v)^2} dv \right\} dy \quad (5)$$

by the full continuous observation $X^T = \{X_t : 0 \leq t \leq T\}$ solution of (1) when σ is known and S is unknown. The fact that σ it is supposed known it is not a limitation. Indeed it is known that the measure induced on \mathcal{C}_T endowed by the Borel σ -algebra, by diffusion processes with different diffusion coefficients are singular, so the statistical inference can be trivial. For a fixed function σ let us introduce the classes

$$\mathcal{S}_\sigma = \{S : \text{conditions } \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3 \text{ are fulfilled}\}$$

and $\mathcal{S}_\sigma^* \subset \mathcal{S}_\sigma$ such that for every S_* in \mathcal{S}_σ^* there exist a $\delta > 0$, and a vicinity

$$V_\delta = \left\{ S : \sup_{x \in \mathbb{R}} |S_*(x) - S(x)| < \delta, \quad S \in \mathcal{S}_\sigma^* \right\},$$

such that

$$\sup_{S \in V_\delta} G(S) < +\infty.$$

This means that for every function $S \in \mathcal{S}_\sigma^*$ there exist a vicinity of function in this class for which the solution of the corresponding stochastic differential equation has the ergodic

property with invariant distribution function given by (5). For x and y in \mathbb{R} we denote by $x \wedge y$ and by $x \vee y$ respectively the minimum and the maximum between x and y .

Let us introduce some functions and some conditions on them that will be used through the sequel in the text.

$$R_S(x, y) = 4 \int_{-\infty}^{+\infty} \frac{F_S(v \wedge x)(1 - F_S(v \vee x))F_S(v \wedge y)(1 - F_S(v \vee y))}{\sigma(v)^2 f_S(v)} dv \quad (6)$$

Note that for $x = y$ (6) can be rewritten as

$$R_S(x, x) = 4 \mathbf{E}_S \left(\frac{F_S(\xi \wedge x) - F_S(x)F_S(\xi)}{\sigma(\xi)f_S(\xi)} \right)^2 \quad (7)$$

$$H_{x,S}(y) = 2 \int_0^y \frac{F_S(v \wedge x) - F_S(v)F_S(x)}{\sigma(v)^2 f_S(v)} dv, \quad (8)$$

Let ν be a finite measure on \mathbb{R} . The quantity $\rho_*(S)$ is defined as

$$\rho_*(S) = \int_{\mathbb{R}} 4 \mathbf{E}_S \left(\frac{F_S(x \wedge \xi)(1 - F_S(x \vee \xi))}{\sigma(\xi)f_S(\xi)} \right)^2 \nu(dx) = \int_{\mathbb{R}} R_S(x, x) \nu(dx)$$

Condition \mathcal{Q}_1 . The function $S_* \in \mathcal{S}_\sigma^*$ and for some $\delta > 0$

$$\sup_{S \in V_\delta} \rho_*(S) = \sup_{S \in V_\delta} \int_{\mathbb{R}} 4 \mathbf{E}_S \left(\frac{F_S(x)F_S(\xi) - F_S(\xi \wedge x)}{\sigma(\xi)f_S(\xi)} \right)^2 \nu(dx) < +\infty. \quad (9)$$

Condition \mathcal{Q}_2 . The function $S_* \in \mathcal{S}_\sigma^*$ and for some $\delta > 0$

$$\sup_{S \in V_\delta} \int_{\mathbb{R}} \mathbf{E}_S H_{x,S}(\xi)^2 \nu(dx) < +\infty.$$

Condition \mathcal{S}_1 . The function $S_* \in \mathcal{S}_\sigma^*$ and for some $\delta > 0$

$$\sup_{S \in V_\delta} \sup_{x \in \mathbb{R}} \mathbf{E}_S \left(\frac{F_S(\xi \wedge x)(1 - F_S(\xi \vee x))}{\sigma(\xi)f_S(\xi)} \right)^2 < +\infty \quad (10)$$

Condition \mathcal{S}_2 . The function $S_* \in \mathcal{S}_\sigma^*$ and for some $\delta > 0$

$$\sup_{S \in V_\delta} \sup_{x \in \mathbb{R}} \mathbf{E}_S H_{x,S}(\xi)^2 < +\infty.$$

Moreover let us introduce the following conditions

Condition \mathcal{S}_3 . Let $R > 0$. The function $S_* \in \mathcal{S}_\sigma^*$ and for some $\delta > 0$, for $x, y \in [-R, R]$

$$\sup_{S \in V_\delta} \mathbf{E}_S |H_{x,S}(\xi) - H_{y,S}(\xi)|^2 < P(R)|x - y|^2$$

where $P(R)$ a polynomial function depending only on R .

For $x > L > 0$

$$\sup_{S \in V_\delta} \mathbf{E}_S |H_{x,S}(\xi)|^2 \leq C e^{-\alpha|x|}$$

where $C > 0$ and $\alpha > 0$ are constant.

Condition \mathcal{U}_1 . With probability 1 the convergence

$$\lim_{T \rightarrow +\infty} \frac{4}{T} \int_0^T \left(\frac{F_S(X_t \wedge x) - F_S(x)F_S(X_t)}{\sigma(X_t)f_S(X_t)} \right)^2 dt = R_S(x, x)$$

is uniformly on $S \in V_\delta$.

Let $\bar{F}_T(x)$ be any estimator of $F_S(x)$, $x \in \mathbb{R}$. If x is fixed the risk associated at any estimator can be defined as

$$\rho_T^p(\bar{F}_T, F_S) = E_S^T \left(\ell \left(\sqrt{T}(\bar{F}_T(x^*) - F_S(x^*)) \right) \right)$$

where $\ell \in \mathcal{W}_p$ is a loss function with the usual properties having a polynomial majorant of order p . If we are interested in the risk defined with loss function that involve the estimation procedure in all the point x where F is defined we can introduce other types of risk function. More precisely here we consider two types of global risk functions. We define the integrated mean square error as

$$\rho_T^{L^2}(\bar{F}_T, F_S) = T \mathbf{E}_S^T \int_{\mathbb{R}} |\bar{F}_T(x) - F_S(x)|^2 \nu(dx)$$

where ν is a finite measure on \mathbb{R} . The second type of global risk considered is the risk based on the *sup* norm. Let us consider $\mathcal{C}_0(\mathbb{R})$ the space of all the continuous function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ vanishing at infinity, endowed with the sup norm $\|\varphi\| = \sup_{x \in \mathbb{R}} |\varphi(x)|$. We denote \mathcal{B}_0 the corresponding Borel σ -algebra. Let us introduce the following loss function $\ell(\varphi) = g(\sup_{x \in \mathbb{R}} |\varphi(x)|)$ with $\varphi \in \mathcal{C}_0(\mathbb{R})$ where $g : [0, +\infty) \rightarrow \mathbb{R}$ is non negative, non decreasing and $g(0) = 0$. The function ℓ is a sub-convex function. For any estimators $\bar{F}_T(x)$ and for a fixed S the risk is defined as

$$\rho_T^{\text{sup}}(\bar{F}_T, F_S) = \mathbf{E}_S^T \left(g(\sup_{x \in \mathbb{R}} \sqrt{T} |\bar{F}_T(x) - F_S(x)|) \right)$$

A natural estimator of $F_S(x)$ is the empirical distribution function defined as follows

$$\hat{F}_T(x) = \frac{1}{T} \int_0^T \chi_{(-\infty, x]}(X_t) dt,$$

where χ_A is the indicator function of a measurable set A . This estimator satisfies all the nice conditions of the *first order asymptotic statistics*, that are uniform consistency, asymptotic normality and asymptotic efficiency as we will show in the next sections.

4 Empirical distribution function

A natural estimator for the stationary distribution function is the empirical distribution function (EDF)

$$\hat{F}_T(x) = \frac{1}{T} \int_0^T \chi_{(-\infty, x]}(X_t) dt.$$

This is a natural estimator because in virtue of the strong law of large number we have that for every $x \in \mathbb{R}$ the relation

$$\frac{1}{T} \int_0^T \chi_{(-\infty, x]}(X_t) dt \rightarrow \int_{\mathbb{R}} \chi_{(-\infty, x]}(y) dF(y) = F(x)$$

holds with probability one.

This estimator is uniformly consistent with respect to $x \in \mathbb{R}$ by the Glivenko–Cantelli theorem (see Kutoyants, 1997).

On $(\mathcal{C}_0(\mathbb{R}), \mathcal{B}_0)$ we define the Gaussian process

$$\eta_S = \{\eta_S(x), x \in \mathbb{R}\}$$

with mean 0 for every $x \in \mathbb{R}$ and with covariance function $R_S(x, y)$. We denote with P_S the law on $(\mathcal{C}_0(\mathbb{R}), \mathcal{B}_0)$ of the process η_S . For a fixed S , consider the empirical process $\{\eta_T^S(x), x \in \mathbb{R}\}$ where

$$\eta_T^S(x) = \sqrt{T}(\hat{F}_T(x) - F_S(x)) = \frac{1}{T} \int_0^T (\chi_{(-\infty, x]}(X_t) - F_S(x)) dt. \quad (11)$$

We are interested in the weak convergence of the empirical process.

The uniform weak convergence of the empirical process to the gaussian process η_S uniformly with respect to $S \in V_\delta$ was proved in Negri, (1998). The result is the following.

Theorem 1. *Let conditions $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ and \mathcal{U}_1 hold. Then the empirical process $\{\eta_T^S(x), x \in \mathbb{R}\}$, weakly converge to the process $\{\eta_S(x), x \in \mathbb{R}\}$ uniformly on V_δ .*

Condition \mathcal{S}_1 assures the weak convergence of finite dimensional laws. Conditions \mathcal{S}_2 and \mathcal{S}_3 are used to prove the tightness of the family of process. Condition \mathcal{U}_1 assures the uniformity of the result. The convergence of $\eta_T^S(x)$ to $\eta^S(x)$ for a fixed x was proved in Kutoyants, (1995). See also Kutoyants, (2004). To prove the convergence of $\eta_T(x) = \sqrt{T}(\hat{F}_T(x) - F_S(x))$ to $\eta_S = \{\eta_S(x) : x \in R\}$ for a fixed S all the conditions can be relaxed avoiding to require uniformity with respect to $S \in V_\delta$.

Regarding the weak convergence of the empirical process η_T^S , for a fixed S , Van der Vaart and Van Zanten, (2005) have established the following general result. The process η_T^S weak converges to the process η_S in $\ell^\infty(J)$ for every compact $J \subset \mathbb{R}$ if and only if $\int_{\mathbb{R}} F_S^2(1 - F_S)^2 dp < \infty$, where p is the scale function of the diffusion. Here $\ell^\infty(J)$ denotes the space of the continuous function $g : J \rightarrow \mathbb{R}$ endowed with the sup norm. Moreover we have the convergence in $\ell^\infty(\mathbb{R})$ if and only if the limit process η_S lies in $C_0(\mathbb{R})$. The condition on $\int_{\mathbb{R}} F_S^2(1 - F_S)^2 dp$ is equivalent to the existence of the covariance function of the process η_S , that is $R_S(x, x) < \infty$ for every $x \in \mathbb{R}$. Their result is based on the relation $\frac{1}{t} \sup_{x \in I} l_t^X(x) = O_{\mathbf{P}}(1)$, where l_t^X denote the diffusion local time of the diffusion X . See also Van Zanten, 2003.

5 Asymptotic efficiency of the empirical distribution function

In this section we review some results on the efficiency of the empirical distribution function as estimator of the invariant distribution function when the efficiency is evaluated with respect to different risk functions. The EDF is asymptotically efficient in a minimax sense. That is, first an asymptotical minimax lower bound for the risk of all the estimators is founded, then an efficient estimator is defined as the estimator whose risk is equal to the asymptotical lower bound and finally it is shown that the empirical distribution function attains this bound.

5.1 An asymptotic global bound in a semiparametric set-up

For the problem of the estimation of $F_S(x^*)$ in a fixed x^* let us consider the risk defined as

$$\rho_T^p(\bar{F}_T, F_S) = \mathbf{E}_S^T \left(\ell \left(\sqrt{T}(\bar{F}_T(x^*) - F_S(x^*)) \right) \right)$$

where $\ell \in \mathcal{W}_p$ is a loss function with the usual properties having a polynomial majorant of order p . Kutoyants, (1997) has proven the following result.

Theorem 2. *Let $S_* \in \mathcal{S}_\sigma^*$ and $R_S(x^*, x^*) < +\infty$. Then for any loss function $\ell \in \mathcal{W}_p$*

$$\lim_{\delta \rightarrow 0} \lim_{T \rightarrow \infty} \inf_{\bar{F}_T} \sup_{S \in V_\delta} \mathbf{E}_S^T \left(\ell \left(\sqrt{T}(\bar{F}_T(x^*) - F_S(x^*)) \right) \right) \geq \mathbf{E} \ell \left(\zeta R_{S_*}(x^*, x^*)^{-1/2} \right)$$

where $\zeta \sim \mathcal{N}(0, 1)$

Moreover suppose that condition \mathcal{U}_1 is satisfied and the following conditions hold for some $p^* \geq 2$

$$\sup_{S \in V_\delta} 4\mathbf{E}_S \left(\frac{F_S(x^*)F_S(\xi) - F_S(\xi \wedge x^*)}{\sigma(\xi)f_S(\xi)} \right)^{p^*} < +\infty \quad \sup_{S \in V_\delta} \mathbf{E}_S |H_{x^*, S}(\xi)|^{p^*} < +\infty; \quad (12)$$

Then $\hat{F}_T(x^*)$ is asymptotically efficient for loss function $\ell \in \mathcal{W}_p$ for every $p < p^*$, provided that $R_S(x^*, x^*)$ is continuous in S_* . See also Kutoyants, (2004). The proof is essentially based on the construction of a parametric vicinity of the non parametric model given by the fixed S_* . In this sense the problem is called semi-parametric. The asymptotically minimax risk bound by the Hájek-Le Cam inequality follows from the local asymptotic normality (LAN) property of the parametric model. Note that the conditions (12) required to prove the efficiency of the EDF are essentially conditions \mathcal{S}_1 and \mathcal{S}_2 without asking the supremum with respect to x and for $p^* = 2$.

5.2 An asymptotic global bound based on the integrated mean square error

A natural extension of the work of Kutoyants was to try to extend the efficiency of the empirical distribution function in a global sense. Suppose that instead in one fixed point you are interested in the risk of the estimator in all the points where the invariant distribution function is defined. This statement of the problem can be called nonparametric

(see Bikel *et al.*, 1993 and van der Vaart, 1995) because not only the underlying model is a nonparametric one (due to the fact the drift coefficient is supposed to be an unknown function) but also because in the risk considered we measure the difference between the estimator and the *true value* F in all the points where they are defined. For the risk defined as the integrated mean square error and the model of independent and identical distributed random variables such result have been established earlier by Levit (1978) and Millar (1979), using the theory of local asymptotic normality. Gill and Levit (1995), obtained the same result using a different approach based on a multidimensional version of the van Trees inequality.

The empirical distribution function achieves a local asymptotic minimax lower bound for the integrated mean square error of an arbitrary estimator.

Theorem 3. *Let $S_* \in \mathcal{S}_\sigma^*$ and condition \mathcal{Q}_1 be fulfilled. Then*

$$\lim_{\delta \rightarrow 0} \liminf_{T \rightarrow \infty} \inf_{\bar{F}_T} \sup_{S \in V_\delta} \rho_T^{L^2}(\bar{F}_T, F_S) \geq \rho_*(S^*) \quad (13)$$

where the inf is taken over any estimator \bar{F}_T of F_S .

The argument used in Kutoyants and Negri (2002) to prove this result is essentially based on the reduction of the problem to a semiparametric one. It is considered a nonparametric vicinity of a fixed model and then it is constructed a parametric family of ergodic processes that belongs to this nonparametric vicinity. But instead of the Hájek-Le Cam inequality, as in Kutoyants (1998), the multidimensional van Trees inequality (see Gill and Levit, 1995) is applied to this parametric model and the lower bound is obtained. This technique has also been used by Iacus (1998) for the problem of state estimation of a diffusion process with small dispersion and can be applied in the same manner in the stationary density estimation problem. Let us introduce the following

Definition 1. *Let condition \mathcal{Q}_1 be fulfilled. Then an estimator \hat{F}_T is called asymptotically efficient if for every $S_* \in \mathcal{S}_\sigma^*$ we have*

$$\lim_{\delta \rightarrow 0} \liminf_{T \rightarrow \infty} \sup_{S \in V_\delta} \rho_T^{L^2}(\hat{F}_T, F_S) = \rho_*(S^*)$$

The efficiency of the EDF is given by the following result.

Theorem 4. *Let conditions \mathcal{Q}_1 , \mathcal{Q}_2 and \mathcal{U}_1 hold and $\rho_*(S)$ be continuous in the uniform topology at the point S^* , then the empirical distribution function is asymptotically efficient.*

Note that conditions \mathcal{Q}_1 and \mathcal{Q}_2 are the integrated version with respect the measure ν of conditions (12) with $p^* = 2$.

5.3 An asymptotic global bound with the risk based on the sup norm

For this last kind of risk the problem of asymptotic efficiency of the empirical distribution function was already studied for many models. Dvoretzky, Kiefer and Wolfowitz (1956) have shown that the EDF is asymptotically minimax in the case of independent and identically distributed observations. See also the works of Koshevnik and Levit (1976) and Millar (1983) and the references therein for further generalizations on this model. For the model of dependent observations Penev (1991) has shown that the EDF is efficient in the problem of the invariant distribution estimation for an exponentially ergodic Markov chain when the state space is $[0, 1]$. A more general case can be found in Greenwood and Wefelmeyer (1995). Further reference for the same kind of problem for other model can be found in Negri (1998).

In this context, where the risk function is based on the sup norm, we are not able to reduce our model to a parametric one, and so we cannot apply the same technique used when the risk is the integrated mean square error to find a lower minimax bound. However also in this context the empirical distribution function is efficient but to prove it we have to follow a different approach. We can state the result for the asymptotic lower bound for the risk based on the sup norm of all the estimator. This result is a bit different from that in Negri (1998) because here we give the theorem under the introduced conditions, so we give also the proof.

Theorem 5. *Let $S_* \in \mathcal{S}_\sigma^*$ and condition \mathcal{S}_1 be fulfilled. Then*

$$\lim_{\delta \rightarrow 0} \liminf_{T \rightarrow +\infty} \inf_{\tilde{F}_T} \sup_{S \in V_\delta} \mathbf{E}_S^T \left(g(\sup_{x \in \mathbb{R}} \sqrt{T} |\tilde{F}_T(x) - F_S(x)|) \right) \geq \int \ell(w) P_*(dw)$$

where P_* is the law of the process $\{\eta_{S^*}(x), x \in \mathbb{R}\}$

Proof. First of all the perturbed model as to be constructed in a non parametric way. Let us consider the Hilbert space

$$\mathcal{H} = \left\{ H : \|H\|_{\mathcal{H}}^2 := \int_{\mathbb{R}} H(x)^2 f_{S^*}(x) dx < +\infty \right\},$$

and the subspaces of \mathcal{H}

$$\mathcal{H}_m = \left\{ H \in \mathcal{H} : \sup_{x \in \mathbb{R}} |H(x)| \leq m, \text{Supp } H \subseteq [-m, m], H \in \mathcal{C}^1(\mathbb{R}) \right\}$$

We can prove (see Negri, 1998) that the space $\mathcal{H}_0 := \bigcup_{m=1}^{+\infty} \mathcal{H}_m$ is dense in \mathcal{H} . We define the perturbed model by means of the function of these subspaces

$$S_H(x) = S^*(x) + \frac{1}{\sqrt{T}} H(x) \sigma(x), \quad H \in \mathcal{H}_m$$

We suppose that the observation is given by the process

$$dX_t^H = S_H(X_t^H) dt + \sigma(X_t^H) dW_t \quad X_0^H \quad 0 \leq t \leq T,$$

Let $F_H(x)$ be the distribution function of the perturbed model. Consider the following sequence of experiments

$$E^T = \{P_H^T, H \in \mathcal{H}_m\}$$

where the P_H^T is the distribution of the solution of the perturbed model. They converge (see Negri, 1998) to the following experiment

$$E = \{P_H, H \in \mathcal{H}_m\}$$

with

$$\frac{dP_H}{dP_0} = \exp \left\{ L_H - \frac{1}{2} \|H\|^2 \right\},$$

where L_H is $\mathcal{N}(0, \|H\|^2)$ and $(L_{H_1}, \dots, L_{H_k})$ is $\mathcal{N}(0, \Sigma)$, $\Sigma_{i,j} = \langle H_i, H_j \rangle$. It turns out that the experiment E is the Gaussian shift experiment of an appropriate Wiener space. To prove this fact we follow the same approach of Millar, (1983) applied to the problem of the efficiency of the empirical distribution function for the case of i.i.d observations. Let us introduce in \mathcal{H} the linear operator

$$\tau H(x) = \int_{-\infty}^{+\infty} 2H(v) \frac{F(x \wedge v)(F(x \vee v) - 1)}{\sigma(v)} dv$$

It can be proved that τ satisfy the following conditions: $\tau H(x)$ is well defined for every $H \in \mathcal{H}$, it is continuous, the image is such that $\tau H \in \mathcal{C}_0(\mathbb{R})$ and $\tau\mathcal{H}$ is dense in $\mathcal{C}_0(\mathbb{R})$. Moreover we can prove that if $H \in \mathcal{H}_m$ for some m then for $T \rightarrow +\infty$ we have

$$\sqrt{T}(F_H(x) - F(x)) = \tau H(x) + o_P(1).$$

Let Q be the Gaussian canonical cylindric measure on \mathcal{H} (see Millar, 1983), then the image of Q under τ is P_* . All this allows us to say that $(\tau, \mathcal{H}, C_0(\mathbb{R}))$ is a Wiener space and its Gaussian shift experiment is exactly $E = \{P_H, H \in \mathcal{H}\}$. Now we can write

$$\sup_{S \in V_\delta} \mathbf{E}_S^T \left(g(\sup_{x \in \mathbb{R}} \sqrt{T} |\bar{F}_T(x) - F_S(x)|) \right) \geq \sup_{H \in \mathcal{H}} \int_{\mathcal{C}([0, T])} \ell(\sqrt{T}(\tilde{F}_T - F_H)) dP_H^T$$

and, denoting with \tilde{F}'_T an appropriate procedure (a generalization of the concept of estimator, see Millar, 1983) we have the following inequalities

$$\begin{aligned} & \inf_{\tilde{F}_T} \sup_{H \in \mathcal{H}} \int_{\mathcal{C}([0, T])} \ell(\sqrt{T}(\tilde{F}_T - F_H)) dP_H^T \\ & \geq \inf_{\tilde{F}_T} \sup_{H \in \mathcal{H}_m} \int_{\mathcal{C}([0, T])} \ell(\sqrt{T}(\tilde{F}_T - F_H)) dP_H^T \\ & = \inf_{\tilde{F}'_T} \sup_{H \in \mathcal{H}_m} \int_{\mathcal{C}([0, T])} \ell(\tilde{F}'_T - \tau H) dP_H^T \\ & = \inf_{\tilde{F}_T} \sup_{H \in \mathcal{H}_m} \int_{\mathcal{C}([0, T])} \ell(\tilde{F}_T - \tau H) dP_H^T \\ & \geq \inf_{b \in \mathcal{K}} \sup_{H \in \mathcal{H}_m} \int_{\mathcal{C}([0, T])} \int_{\mathcal{C}_0(\mathbb{R})} \ell(y - \tau H) b(w, dy) P_H^T(dw) \end{aligned}$$

By the Hajek–Le Cam theorem we have

$$\begin{aligned} & \liminf_{T \rightarrow +\infty} \inf_{\tilde{F}_T} \sup_{H \in \mathcal{H}_m} \int_{\mathcal{C}([0, T])} \ell(\sqrt{T}(\tilde{F}_T - F_H)) dP_H^T \\ & \geq \inf_{b \in \mathcal{K}} \sup_{H \in \mathcal{H}_m} \int_{\mathcal{C}_0(\mathbb{R})} \int_{\mathcal{C}_0(\mathbb{R})} \ell(y - \tau H) b(w, dy) P_H(dw) \\ & = \inf_{b \in \mathcal{K}} \sup_{H \in \mathcal{H}_m} \rho(b, H), \end{aligned}$$

where is given the definition of $\rho(b, H)$. Now it can be proved that (see Millar, 1993 and Negri, 1998)

$$\inf_{b \in \mathcal{K}} \sup_{H \in \mathcal{H}_m} \rho(b, H) = \sup_{\nu \in \mathcal{M}_f(\mathcal{H}_m)} \inf_{b \in \mathcal{K}} \rho(b, \nu).$$

Then

$$\begin{aligned} \lim_{m \rightarrow +\infty} \sup_{\nu \in \mathcal{M}_f(\mathcal{H}_m)} \inf_{b \in \mathcal{K}} \rho(b, \nu) &= \\ &= \sup_{\nu \in \mathcal{M}_f(\mathcal{H})} \inf_{b \in \mathcal{K}} \rho(b, \nu) \end{aligned}$$

By the minimax theorem for Wiener spaces

$$\inf_{b \in \mathcal{K}} \sup_{H \in \mathcal{H}} \rho(b, H) = \int_{\mathcal{C}_0(\mathbb{R})} \ell(z) P_*(dz)$$

And this complete the proof. \square

Now we turn to the problem of find an asymptotically efficient estimator. We want to prove that the empirical distribution function is efficient. Let us introduce to following

Definition 2. Let $S_* \in \mathcal{S}_\sigma^*$ and condition \mathcal{S}_1 be fulfilled. The estimator F_T^* is asymptotically efficient if

$$\lim_{\delta \rightarrow 0} \liminf_{T \rightarrow +\infty} \sup_{S \in V_\delta} \mathbf{E}_S^T \left(g(\sup_{x \in \mathbb{R}} \sqrt{T} |F_T^*(x) - F_S(x)|) \right) = \int \ell(w) P_S^*(dw)$$

The efficiency of the EDF is given by the following result.

Theorem 6. Let conditions \mathcal{S}_1 , \mathcal{S}_2 , \mathcal{S}_3 and \mathcal{U}_1 hold and $\int \ell(w) P_S^*(dw)$ be continuous in the uniform topology at the point S^* . Moreover suppose that the function g is bounded, then the empirical distribution function is asymptotically efficient.

Proof. Under conditions \mathcal{S}_1 , \mathcal{S}_2 , \mathcal{S}_3 and \mathcal{U}_1 the process

$$\eta_T^S(x) = \sqrt{T}(\hat{F}_T^S(x) - F_S(x)) \quad x \in \mathbb{R}, \quad S \in V_\delta$$

weakly converge to the process $\eta_S(x)$, $x \in \mathbb{R}$ uniformly with respect to the functions $S \in V_\delta$. Then the efficiency of the empirical distribution function follows from the continuous mapping theorem when g is bounded. \square

Unfortunately we cannot relax the assumption of boundness on g and prove the efficiency of the EDF also for loss functions with a polynomial majorant as in the semiparametric set-up. Note that the assumptions we need to prove the efficiency of the empirical distribution function when the risk function is based on $\sup_{x \in \mathbb{R}} \sqrt{T} |\bar{F}_T(x) - F(x)|$ are more restrictive that the ones we need to prove the efficiency of the EDF for the mean

square integrated error and this is essentially due to the structure of the integrated mean square error and on the fact that we don't need the weak convergence of the process in the space of real continuous function vanishing at infinity in this case.

6 Example

Example 1. Let us consider a class of functions satisfying the condition \mathcal{A}_1 such that $S(y)\text{sgn}(y) \leq -\gamma$ for $|y| > M$, and S is continuous. We denote such class of function with $\mathcal{S}(\gamma, M)$. Let the observed process be

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

where S is an unknown function that belongs to $\mathcal{S}(\gamma, M)$ and there exist two constants k_1 and k_2 such that $0 < k_1 < \sigma(x) < k_2 < +\infty$ for all $x \in \mathbb{R}$. In Negri (1998) the belonging of the drift coefficient to the class $\mathcal{S}(\gamma, M)$ and the condition required on σ are fundamental assumptions to prove the efficiency of the EDF. Moreover it is proved that with such kind of S and σ the conditions $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{U}_1$ and \mathcal{A}_2 are satisfied. Also conditions \mathcal{Q}_1 and \mathcal{Q}_2 are satisfied (see Kutoyants, 1998) and the continuity of $R_S(x, x)$ in the uniform topology can also be proved. So we have a huge class of processes for which the empirical distribution function is an efficient estimator for the invariant distribution function in the three different set-up described previously.

Acknowledgements

My warmest thanks go to Yury A. Kutoyants for introducing me to the statistics for diffusion processes and for his helpful advice and comments in the last years.

References

- Bickel, P.J. (1993). Estimation in semiparametric models, *Multivariate analysis: future directions*, C.R. Rao ed., Elsevier.
- Bickel, P.J., Klaassen, C.A.J., Ritov, Y. & Wellner, J.A. (1993). *Efficient and adaptive estimation for semiparametric models*, Baltimore, John Hopkins University Press.

- Gill, R.D. & Levit, B.Y. (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound, *Bernoulli*, **1+2**, 59-79.
- Greenwood, P.E., Wefelmeyer, W., (1995), Efficiency of empirical estimators for Markov Chains, *Ann. Statist.*, **23**, 132–143.
- Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956), Asymptotic minimax character of the sample distribution function and the classical multinomial estimator, *Ann. Statist.*, **27**, 642–669.
- Durrett, R. (1996). *Stochastic Calculus: A Practical Introduction*, CRC Press, Boca Raton.
- Gikhman, I. I., Skorokhod, A. V., (1972), *Stochastic Differential Equations*, Springer-Verlag, New York.
- Ibragimov, I.A., Hasminski, R.Z., (1981), *Statistical Estimation: Asymptotic Theory*, Springer Verlag, New York.
- Iacus, S.M. (1998). The van Trees inequality for diffusion processes with small diffusion coefficient and asymptotic efficiency in a semiparametric problem, *submitted*.
- Koshevnik, Yu.A., Levit, B.Ya., (1976), On a non-parametric analogue of the information matrix, *Theory Probab. Applic.*, **21**, 738–753.
- Kutoyants, Yu.A. (1984). *Parameter estimation for stochastic processes*, Berlin, Heldermann Verlag.
- Kutoyants, Yu.A. (1997). Efficiency of the empirical distribution for ergodic diffusion, *Bernoulli*, **3**(4), 445-456.
- Kutoyants, Yu.A. (2004). *Statistical Inference for Ergodic Diffusion Processes*, Springer.
- Kutoyants Yu.A., Negri I., (2002), On L_2 -efficiency of empiric distribution for diffusion process, *Theory of Probability and its Applications*, v.46, No 1, p.164-169.
- Levit, B.Ya. (1974). On the optimality of some statistical estimates, *Proceedings of the Prague Symposium on Asymptotic Statistics*, Prague, Czechoslovakia Vol 2. (J. Hájek, ed.), 215-238, Charles University Press.
- Levit, B.Ya. (1978). Infinite-dimensional information inequalities, *Theory Prob Appl*, **23**, 371-377.
- Liptser, R.S & Shiriyayev, A.N. (1977). *Statistics of random processes*, 1, New York,

- Springer.
- Mandl, P. (1968). *Analytical Treatment of One-Dimensional Markov Processes*, Berlin, Springer.
- Millar, P.W. (1979). Asymptotic minimax theorems for the sample distribution function, *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **48**, 233-252.
- Millar, P. W., (1983), THE MINIMAX PRINCIPLE IN ASYMPTOTIC STATISTICAL THEORY, in *Ecole d'Eté de Probabilités de Saint Flour XI 1981*, Lect. Notes, Math., Springer, Berlin.
- Negri, I. (1998a), Stationary distribution function estimation for ergodic diffusion process, *C.R. Acad. Sci. Paris*, t. 326, Série I, 879-884, Elsevier, Paris.
- Negri, I. (1998), Stationary distribution function estimation for ergodic diffusion process, *Statistical Inference for Stochastic Processes*, **1**, 61–84.
- Penev, S., (1991), Efficient estimation of the stationary distribution for exponentially ergodic Markov chains, *J. Stat. Plan. and Infer.*, **27**, 105–123.
- van der Vaart, A.W. (1995). Semiparametric models: an evaluation, *Statistica Neerlandica*, v 49, **1**, 111-125.
- Van der Vaart, A.W., Van Zanten, H. (2005). Donsker theorems for diffusion: necessary and sufficient conditions, *The Annals of Probability*, **33**, No. 4, 1422-1451.
- Van Zanten, H. (2003). On uniform laws of large numbers for ergodic diffusions and consistency of estimators. *Stat. Inference Stoch. Process.*, **6**, no. 2, 199–213.