A Diffusion Representation of the Nonparametric IID Experiment on an Interval

Valentine Genon-Catalot, Catherine Larédo and Michael Nussbaum

June 1998

Abstract

We consider a diffusion model of small variance type with positive drift function varying in a nonparametric set. We investigate discrete versions of this continuous model with respect to statistical equivalence, in the sense of the asymptotic theory of experiments. It is shown that the collection of level crossing times for a uniform grid of levels is asymptotically equivalent to the continuous model in the sense of Le Cam's deficiency distance, when the discretization step decreases with the noise intensity ε. It follows that in the continuous diffusion model, the statistic of level crossing times is asymptotically sufficient. Since the level crossing times obey a nonparametric regression model with independent data, a further asymptotic equivalence can be established, leading to a simple Gaussian signal-in-white noise problem. When the drift density f is also a probability density, this in turn is asymptotically equivalent to i.i.d data with density f on the unit interval.

1 Introduction

Comparison of statistical experiments by means of Le Cam's notion of deficiency distance has recently proved feasible in nonparametric settings (Brown and Low (1996), Nussbaum (1996), Grama and Nussbaum (1997)).

When two families of experiments are asymptotically equivalent in the sense that their Le Cam deficiency distance goes to 0, then it is also generally possible to prove that

---

1990 *Mathematics Subject Classification.* Primary 62 B 15; Secondary 62M05, 62 G 07
Key words and phrases. Nonparametric experiments, asymptotic equivalence, diffusion process, discretization, level crossing, inverse Gaussian regression, signal in white noise, iid model, density estimation.

the minimax risks are the same for both families. This allows major simplifications of proofs when studying the minimax risk of the simplest family of experiments (see e. g. Pinsker (1980), Korostelev (1993), Lepskii and Spokoiny (1995) for minimax risks in Gaussian models; Nussbaum (1996) for equivalence of Gaussian with other models).

In particular, Milstein and Nussbaum (1996) considered the problem of estimating the drift function f from an observed diffusion process $(Y_t\ ,\ t \in [0,1])$, defined by the stochastic differential equation

$$dY_t = f(Y_t)dt + \varepsilon dW_t \quad ; \quad t \geq 0,\ Y_0 = 0. \tag{1}$$

Here (W_t) is a Wiener process defined on a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\geq 0}, \mathbb{P})$ and ε a small parameter. The function f belongs to a nonparametric set of functions $\mathcal{F}$ satisfying appropriate conditions. These authors consider the Euler scheme associated with f and with the sampling interval $\frac{1}{n}$, i. e.

$$y_i = y_{i-1} + \frac{1}{n} f(y_{i-1}) + \frac{\varepsilon}{\sqrt{n}}\xi_i \quad ; \quad y_0 = 0 \tag{2}$$

where $(\xi_i, i = 1, \ldots, n)$ is a n-sample of i. i. d. standard normal variables.
They prove that the deficiency distance of these experiments tends to 0 as ε goes to 0 if $n = n_\varepsilon$ tends to infinity in such a way that $\varepsilon n_\varepsilon \to \infty$. An important consequence is that the statistic $(Y_{t_1}, ..., Y_{t_n})$, where $t_i = \frac{i}{n}$, is asymptotically sufficient in the first experiment. This extends the result obtained in the corresponding parametric estimation problem (see Larédo (1990), Genon–Catalot (1990)).

In this paper, we address a closely related problem. We consider the problem of estimating the function f on $[0,1]$ from the diffusion (Y_t) defined in (1) when it is observed up to its first hitting time $T_1(Y)$ of the level 1. The function f belongs to the set $\mathcal{F} = \mathcal{F}_{K,m}$ associated with two positive constants K and m :

$$\mathcal{F} = \left\{ \begin{array}{ll} f : \mathbb{R} \to \mathbb{R} : & f(x) \geq m, \qquad f(0) \leq K, \\ & |f(x) - f(y)| \leq K|x-y|, \forall x, y \in \mathbb{R} \end{array} \right\}. \tag{3}$$

This will be our first experiment $\mathcal{E}_0^\varepsilon$ with parameter set $\mathcal{F}$. Let us stress that the condition that f is positive implies that $T_1(Y)$ is almost surely finite.

The second experiment consists in a triangular array of n independent random variables (X_n^i) distributed according to an inverse Gaussian law $IG\left(n^{-1}(f(\frac{i-1}{n}))^{-1}, n^{-2}\varepsilon^{-2}\right)$. Recall that the inverse Gaussian distribution $IG(\mu, \lambda)$ can be defined as the distribution of the hitting time of level $\sqrt{\lambda}$ by the process $X_t = \lambda^{1/2}\mu^{-1}t + W_t\ (\mu, \lambda > 0)$. It has a density

$$h_{\mu,\lambda}(t) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right) \mathbf{1}_{t>0} \tag{4}$$

(see e.g. Chikara and Folks (1989)). Denote by $\mathcal{G}^{n,\varepsilon}$ this experiment indexed by the same nonparametric set for f, i. e. $\mathcal{F}$.

Our first result (Theorem 1) states that, as ε goes to 0, the deficiency distance of these two families of experiments tends to 0 if $n = n_\varepsilon$ goes to infinity in such a way that $\varepsilon n_\varepsilon \to \infty$. As an important consequence, we obtain that the statistic consisting of the hitting times of levels $i/n,\ i = 1, \ldots, n$ of the diffusion $(Y_t, t \geq 0)$, i. e. the statistic $(T_{i/n}(Y), \quad i = 1, \ldots, n)$ is asymptotically sufficient (Corollary 2). Here again, these results extend those obtained in the parametric drift estimation problem for diffusion hitting times (Genon–Catalot and Larédo (1987), Larédo (1990)).

The experiment $\mathcal{G}^{n,\varepsilon}$ can be seen as a nonparametric regression model with independent data. Using results of Grama and Nussbaum (1997) for such models, we arrive at a Gaussian approximation for our diffusion experiment (1).

Indeed, consider a family of experiments given by an observed diffusion process

$$dZ_u = f^{1/2}(u)du + \frac{\varepsilon}{2} dW_u,\ u \in [0,1],\ Z_0 = 0 \tag{5}$$

with ε tending to 0 and $f \in \mathcal{F}$. Taking $n_\varepsilon = [\varepsilon^{-2}]$, we prove that $\mathcal{G}^{n_\varepsilon,\varepsilon}$ and the signal-in-white-noise model (5) are asymptotically equivalent (Theorem 3).

Now, in the special case where the restriction of the drift function f to $[0,1]$ is a probability density, it has been proved in Nussbaum (1996) that the signal-in-white-noise model (5) is asymptotically equivalent to the experiment given by $n = [\varepsilon^{-2}]$ observed independent identically distributed variables having density f on the unit interval. We thus obtain a rather unexpected connection between the i. i. d. model and the diffusion experiment (1), in the sense of asymptotic equivalence (Corollary 4).

Kutoyants (1985) considers nonparametric estimation of the drift function f for model (1), when it is continuously observed on a fixed time interval $[0,T]$, under the assumption that f is bounded away from 0. He proves, using kernel type estimates, that the rates of convergence are identical to the ones obtained for density estimation of i.i.d. variables, for a given smoothness condition on f. The equivalences stated in Theorem 2 and Corollary 2 both clarify and explain these results.

Let us point out that, except for $\mathcal{E}_0^\varepsilon$, f need not be defined outside the interval $[0,1]$. So, the parameter can be taken to be the restriction of the function f to the interval $[0,1]$, for the last two experiments. In fact, for $\mathcal{E}_0^\varepsilon$, our results show that we could have defined f on $[0,1]$ only and take any extension of f on R satisfying the conditions of $\mathcal{F}$, as, for instance, $f(x) = f(0)$ for $\quad x \leq 0 \quad ; \quad f(x) = f(1) \quad$ for $x \geq 1$. Another way to capture what happens here is just to remark that the function f is not identifiable outside $[0,1]$ from the experiments $\mathcal{E}_0^\varepsilon$.

Section 2 contains the notations, the statement of the main results, and some recap on the Le Cam deficiency distance Δ. In Section 3 we introduce an experiment which is exactly equivalent to the triangular array $(X_n^i,\ i = 1, \ldots, n)$, but comparable to the diffusion experiment $\mathcal{E}_0^\varepsilon$, as in Milstein and Nussbaum (1996). Using this experiment, in Section 4 we compute a bound for the Δ- distance between the diffusion experiment (1) and the other ones. In Section 5 we present the argument leading on to (5), specializing the exponential family nonparametric regression model of Grama and Nussbaum (1997) to the inverse Gaussian case.

2 Notations and main results

2.1 Definition of the experiments

Let $(\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\geq 0}, \mathbb{P})$ be a probability space endowed with a filtration $(\mathcal{A}_t)$ satisfying the usual conditions, and let $(W_t,\, t \geq 0)$ be an $(\mathcal{A}_t)$- Brownian motion defined on Ω. For $f : \mathbb{R} \to \mathbb{R}$, consider the process (Y_t) defined by the stochastic differential equation (1).

The parameter ε is here assumed to be known. The function f varies in a set $\mathcal{F} = \mathcal{F}_{K,m}$ associated with two positive constants K, m, and which is defined by the following conditions

$$
\begin{array}{ll}
(C1) & f(x) \geq m, \ \forall x \in \mathbb{R} \\
(C2) & f(0) \leq K, \ |f(x) - f(y)| \leq K|x - y| \qquad \forall x, y \in \mathbb{R}.
\end{array}
$$

It follows from $(C1), (C2)$ that any function $f \in \mathcal{F}$ satisfies the linear growth condition $0 < f(x) \leq K(1 + |x|)$. Hence the stochastic differential equation (1) has a unique strong solution $(Y_t, t \geq 0)$. Let $T_1(Y)$ be the first hitting time of level 1 by the sample path $(Y_t, t \geq 0)$. Condition C1 implies that $T_1(Y)$ is finite almost surely. The first experiment considered here is associated with the observation $(Y_t, t \in [0, T_1(Y)])$.

We may now construct the canonical experiment. Let $C(\mathbb{R}^+, \mathbb{R})$ be the space of continuous real functions defined on $\mathbb{R}^+$, let $(X_t,\, t \geq 0)$ be the canonical process of $C(\mathbb{R}^+, \mathbb{R})$, $\quad \mathcal{C}_t^0 = \sigma(X_s, s \leq t)$, $\quad \mathcal{C}_t = \bigcap_{s>t} \mathcal{C}_s^0$ and $\mathcal{C} = \bigvee_{t\geq 0} \mathcal{C}_t$.

Denote by P_f^ε the distribution of $(Y_t,\, t \geq 0)$ defined by (1) on $(C(\mathbb{R}^+, \mathbb{R}), \mathcal{C})$. Now, for $x \in C(\mathbb{R}^+, \mathbb{R})$ and $a \in \mathbb{R}$, let

$$
T_a(x) = \inf\{t \geq 0 : \ x(t) = a\}. \tag{6}
$$

Define $T = T_1(X)$ the hitting time of level 1 by the canonical process $(X_t,\, t \geq 0)$ and let

$$
P_f^{T,\varepsilon} = P_f^\varepsilon \,|\mathcal{C}_T \tag{7}
$$

be the restriction of P_f^ε to the σ- algebra $\mathcal{C}_T$. The first experiment is now described by

$$
\mathcal{E}_0^\varepsilon = \Big(C(\mathbb{R}^+, \mathbb{R}), \mathcal{C}_T, (P_f^{T,\varepsilon},\, f \in \mathcal{F}) \Big). \tag{8}
$$

Let us now present the second experiment. For $(\mu, \lambda) \in (\mathbb{R}^+)^2$, let us denote by $IG(\mu, \lambda)$ the Inverse Gaussian distribution with density given in (4). The mean of this distribution is μ and the variance is μ^3/λ . Consider now a triangular array of n independent random variables (X_n^i) such that

$$X_n^i \sim IG\left(n^{-1}(f(\frac{i-1}{n}))^{-1}, n^{-2}\varepsilon^{-2}\right). \tag{9}$$

The realization of such a triangular array can be obtained in the following way. Let $B^1, \dots, B^n$ be n independent Brownian motions and set

$$\begin{aligned} X_n^i &= \inf\{t \geq 0 : \quad \frac{i-1}{n} + f(\frac{i-1}{n})t + \varepsilon B_t^i = \frac{i}{n}\} \\ &= \inf\{t \geq 0 : \quad \frac{1}{\varepsilon} f(\frac{i-1}{n})t + B_t^i = \frac{1}{n\varepsilon}\}. \end{aligned} \tag{10}$$

Denote by

$$P_{n,f}^{\varepsilon} \doteq \text{the distribution of} (X_n^1, \dots, X_n^n). \tag{11}$$

This second family of experiments is described by

$$\mathcal{G}^{n,\varepsilon} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{n,f}^{\varepsilon}, \quad f \in \mathcal{F})\right), \tag{12}$$

where $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel σ- algebra of $\mathbb{R}^n$.

The third family of experiments is defined by

$$\mathcal{E}_1^{\varepsilon} = \left(C([0,1], \mathbb{R}), \mathcal{B}, (Q_f^{\varepsilon}, f \in \mathcal{F})\right) \tag{13}$$

where $\mathcal{B}$ is the Borel sigma algebra of $C([0,1], \mathbb{R})$, and Q_f^{ε} denotes the distribution of the process $(Z_u)_{u \in [0,1]}$ given by

$$dZ_u = f^{1/2}(u)du + \frac{\varepsilon}{2} dW_u, \ u \in [0,1], \ Z_0 = 0. \tag{14}$$

Finally, consider the case where $\int_{[0,1]} f(u)du = 1$. Let $\tilde{Q}_f^n$ be the joint distribution of n i. i. d. random variables having density ($f(u), u \in [0,1]$) on the unit interval. Consider the experiment

$$\mathcal{E}^n = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\tilde{Q}_f^n, f \in \mathcal{F})\right). \tag{15}$$

Our aim is to compare these experiments which are indexed by the same parameter set $\mathcal{F} = \mathcal{F}_{K,m}$, but which are defined on different observation spaces.

2.2 Statement of results

We can now state the results contained in this paper.

Theorem 1 *If, as* $\varepsilon \to 0$, $n = n_\varepsilon \to +\infty$ *in such a way that* $\varepsilon n_\varepsilon \to \infty$ *then the experiments* $\mathcal{E}_0^\varepsilon$ *and* $\mathcal{G}^{n,\varepsilon}$ *are asymptotically equivalent : i.e. for the Le Cam deficiency distance* Δ*, we have*

$$\Delta(\mathcal{E}_0^\varepsilon\,, \mathcal{G}^{n,\varepsilon}\,) \longrightarrow 0 \qquad as \quad \varepsilon \to 0.$$

Noting that the mapping $a \to T_a(Y)$ is increasing from $[0,1]$ to $[0, T_1(Y)]$, the statistic $(T_{\frac{i}{n}}(Y), i = 1...n)$ is well defined. An important consequence of Theorem 1 is the following.

Corollary 2 *Under the conditions of Theorem 1, for the diffusion model* $(Y_t,\, t \geq 0)$ *observed up to* $T_1(Y)$ *the statistic* $\left(T_{\frac{i}{n}}(Y),\, i = 1, ..., n\right)$ *defined by the hitting times of levels* $\frac{i}{n}$, $i = 1, \ldots, n$ *is asymptotically sufficient as* $\varepsilon \to 0$.

Remark 1 *The assumption that* f *is bounded from below by a positive constant* m *is quite natural. This assumption is required if one wants to obtain a nonparametric estimator of the drift function as* $\varepsilon \to 0$ *(see e. g. Kutoyants (1985)). One has to ensure that the diffusion passes all points* x *between* 0 *and* $x_f(T)$*, where* $x_f(t)$ *is the solution of* $\frac{dx_f(t)}{dt} = f(x(t))$, $x_f(0) = 0$. *It can then be shown that for the uniform positive minorant* m *of* f *the relation* $\bigcap_{f \in \mathcal{F}}[0, x_f(T)] = [0, mT]$ *holds, so that consistent estimation of* f *on* $[0, mT]$ *is possible. This coincides with the fact that, here,* f *is identifiable on* $[0,1]$ *only.*

The next result concerns the equivalence between the signal-in-white-noise model $\mathcal{E}_1^\varepsilon$ defined in (13) and the diffusion experiment $\mathcal{E}_0^\varepsilon$.

Theorem 3 *Under the conditions of Theorem 1 we have*

$$\Delta(\mathcal{E}_0^\varepsilon\,, \mathcal{E}_1^\varepsilon\,) \longrightarrow 0 \qquad as \quad \varepsilon \to 0.$$

Corollary 4 *Suppose that* $n = n_\varepsilon = [\varepsilon^{-2}]$*, and that* f *restricted to* $[0,1]$ *is a probability density. Then, denoting by* $\mathcal{E}^n = \mathcal{E}^{n_\varepsilon}$*,*

$$\Delta(\mathcal{E}_0^\varepsilon\,, \mathcal{E}^n\,) \longrightarrow 0 \qquad as \quad \varepsilon \to 0.$$

According to this last result, we have a new asymptotic diffusion representation for the experiment given by i. i. d. random variables on the unit interval.

If the density g of these variables satisfies a uniform Lipschitz condition on $[0,1]$ and is bounded away from 0, one may consider the extension of g to the whole of $\mathbb{R}$ by setting $f_g(x) = g(0)$, for $x < 0$ and $f_g(x) = g(1)$ for $x > 1$. In that case, the model (1) for $f = f_g$ is well defined and the diffusion experiment $\mathcal{E}_0^\varepsilon$ indexed by $f = f_g$ is also an asymptotic representation for the i. i. d. experiment $\mathcal{E}^n$ with $n = [\varepsilon^{-2}]$.

Remark 2 : **Fisher Information** *We may confirm the result of Corollary 4 by a calculation of the asymptotic Fisher information in both models. Indeed, asymptotic equivalence in the Le Cam sense for the nonparametric models entails the same for parametric submodels, and hence equality of asymptotic Fisher informations for regular cases. Consider a parametric submodel of (1), where* $f = f_\vartheta$, $\vartheta \in \Theta$ *and* Θ *is an open interval and observation (of model (1)) is between* 0 *and* $T_1(Y)$. *According to Genon-Catalot and Larédo (1987), if the model is sufficiently regular, the asymptotic Fisher information (divided by* ε^{-2}*) is*

$$I_F(\vartheta) = I(\vartheta) = \int_0^1 \left(\frac{\partial}{\partial \vartheta} f_\vartheta(x) \right)^2 f_\vartheta^{-1}(x) dx.$$

This indeed coincides with the Fisher information in an i. i. d. model with regular density f_ϑ.

In order to prove these results and for the sake of clarity, we recall below the main definitions and properties of the Le Cam deficiency distance.

2.3 The Le Cam deficiency distance

This pseudo distance is generally denoted by Δ. In what follows, all measurable spaces are supposed to be Polish metric spaces equipped with their Borel σ-algebras.
Consider two experiments with the same parameter space that we shall again denote by $\mathcal{F}$, say $\mathcal{E} = (\mathcal{X}, \mathcal{A}, (P_f,\ f \in \mathcal{F}))$ and $\mathcal{G} = (\mathcal{Y}, \mathcal{B}, (Q_f,\ f \in \mathcal{F}))$. Assume also that the two families $(P_f,\ f \in \mathcal{F})$ and $(Q_f,\ f \in \mathcal{F})$ are dominated.
Consider now a Markov kernel $M(x, dy)$ from $\mathcal{X}$ to $(\mathcal{Y}, \mathcal{B})$, i.e. an application such that for all $B \in \mathcal{B}$ the mapping $x \to M(x, B)$ is $\mathcal{A}$- measurable and, for all $x \in \mathcal{X}$, $M(x, dy)$ is a probability measure on $(\mathcal{Y}, \mathcal{B})$. Denote by MP_f the image probability measure of P_f under the kernel M, i.e.

$$MP_f(B) = \int_{\mathcal{X}} M(x, B) P_f(dx) \qquad \text{for} \quad B \in \mathcal{B}$$

The experiment $M\mathcal{E} = (\mathcal{Y}, \mathcal{B}, (MP_f,\ f \in \mathcal{F})$ is called a randomization of $\mathcal{E}$ by the kernel M. It is defined on the same measurable space $\mathcal{Y}$ that $\mathcal{G}$. Let $\mathcal{M}$ denote the set of Markov kernels from $\mathcal{X}$ into $(\mathcal{Y}, \mathcal{B})$.

Definition 1 *The deficiency of* $\mathcal{E}$ *with respect to* $\mathcal{G}$ *is given by*

$$\delta(\mathcal{E}, \mathcal{G}) = \inf_{M \in \mathcal{M}} \sup_{f \in \mathcal{F}} \|MP_f - Q_f\|_{TV}, \tag{16}$$

where $\|\cdot\|_{TV}$ *denotes the total variation norm, i.e.* $\|P - Q\|_{TV} = 2 \sup_{B \in \mathcal{B}} |P(B) - Q(B)|$.

Definition 2 *The deficiency distance* Δ *in the sense of Le Cam is given by*

$$\Delta(\mathcal{E},\mathcal{G}) = \max\{\delta(\mathcal{E},\mathcal{G}), \delta(\mathcal{G},\mathcal{E})\}. \tag{17}$$

In fact the Δ-distance is a pseudo-distance. Two experiments are said to be equivalent whenever $\Delta(\mathcal{E},\mathcal{G}) = 0$. In the sequel we shall use two basic properties of Δ.

Property 1: Let $T : (\mathcal{X},\mathcal{A}) \longrightarrow (\mathcal{Y},\mathcal{B})$ be a measurable application and let T the image experiment of $\mathcal{E}$ by the (deterministic) kernel T. Then, $\Delta(\mathcal{E}, T\mathcal{E}) = 0$ if and only if T is a sufficient statistic for the experiment $\mathcal{E}$.

Property 2: If the experiments $\mathcal{E}$ and $\mathcal{G}$ have the same measurable space of observations $((\mathcal{X},\mathcal{A}) = (\mathcal{Y},\mathcal{B}))$ then the following inequality holds:

$$\Delta(\mathcal{E},\mathcal{G}) \le \sup_{f\in\mathcal{F}} \|P_f - Q_f\|_{TV}. \tag{18}$$

3 An accompanying diffusion experiment

It is well known and clear from its definition that it is difficult to compute the Δ-distance between two experiments when they are not defined on the same measurable space.

So, following Brown and Low (1996), Milstein and Nussbaum (1996), we define another experiment $\overline{\mathcal{G}}^{n,\varepsilon}$ which has the same observation space as $\mathcal{E}_0^\varepsilon$.
Let $(t,z) \in \mathbb{R}^+ \times C(\mathbb{R}^+,\mathbb{R})$. Consider, for $i = 1,\dots,n$ the times $T_{\frac{i}{n}}(z) = \inf\{t \ge 0,\, z(t) = \frac{i}{n}\}$ and the function

$$f_n(t,z) = \sum_{i=1}^{n} f\left(\frac{i-1}{n}\right) \mathbf{1}_{\left(T_{\frac{i-1}{n}}(z), T_{\frac{i}{n}}(z)\right]}(t). \tag{19}$$

Finally, define the diffusion type process $(\overline{Y}_t,\, t \ge 0)$ on $(\Omega,\mathcal{A},(\mathcal{A}_t)_{t\ge 0},\mathbb{P})$ as the solution of the stochastic differential equation

$$\begin{cases} d\overline{Y}_t &= f_n(t,\overline{Y})dt + \varepsilon dW_t \\ \overline{Y}_0 &= 0. \end{cases} \tag{20}$$

Let $T_1(\overline{Y})$ be the first hitting time of level 1 by the path $(\overline{Y}_t)$. Again, by condition (C1), $T_1(\overline{Y})$ is finite almost surely, and we can describe the experiment associated with the observation $(\overline{Y}_t, t \in [0, T_1(\overline{Y})])$. Denote by $\bar{P}^\varepsilon_{n,f}$ the distribution of $(\overline{Y}_t,\, t \ge 0)$ on $(C(\mathbb{R}^+,\mathbb{R}), \mathcal{C})$ and set

$$\bar{P}^{T,\varepsilon}_{n,f} = \bar{P}^\varepsilon_{n,f}|\mathcal{C}_T.$$

Recall that T is the first hitting time of level 1 by the canonical process of $C(\mathbb{R}^+,\mathbb{R})$.

The accompanying experiment is defined as

$$\overline{\mathcal{G}}^{n,\varepsilon} = \Big(C(\mathbb{R}^+,\mathbb{R}), \mathcal{C}_T, (\bar{P}^{T,\varepsilon}_{n,f}, f \in \mathcal{F}) \Big) .$$

Then the following holds.

Proposition 5 *(i) The statistic* $(z \to (T_{\frac{i}{n}}(z), i = 1...n)$ *is sufficient for the experiment* $\overline{\mathcal{G}}^{n,\varepsilon}$.

(ii) The experiments $\mathcal{G}^{n,\varepsilon}$ *and* $\overline{\mathcal{G}}^{n,\varepsilon}$ *are exactly equivalent, i.e.*

$$\forall n \geq 1,\ \forall \varepsilon > 0, \qquad \Delta(\mathcal{G}^{n,\varepsilon}, \overline{\mathcal{G}}^{n,\varepsilon}) = 0 .$$

The proof of Proposition 1 is based upon a precise description of the process $(\overline{Y}_t)$. Let us define by induction a sequence of processes and stopping times as follows. Let

$$X_0(t) = f(0)t + \varepsilon W_t \text{ , and}$$

$$\tau_{\frac{1}{n}} = T_{\frac{1}{n}}(X_0) = \inf\{\, t \geq 0 : \ X_0(t) = \frac{1}{n}\}.$$

Then, by induction, for $i = 1, ..., n$

$$X_{i-1}(t) = f\left(\frac{i-1}{n}\right) t + \varepsilon \left(W_{t+\tau_{\frac{i-1}{n}}} - W_{\tau_{\frac{i-1}{n}}} \right) \tag{21}$$

$$\tau_{\frac{i}{n}} = \tau_{\frac{i-1}{n}} + T_{\frac{1}{n}}(X_{i-1}). \tag{22}$$

The process $(\overline{Y}_t)$ satisfies the property stated below.

Lemma 6 *The hitting times of levels* $(\frac{i}{n}, i = 1...n\ )$ *by the process* $(\overline{Y}_t)$ *are given by*

$$T_{\frac{i}{n}}(\overline{Y}) = \tau_{\frac{i}{n}}, \qquad i = 1...n.$$

Moreover, the n-tuple $\left\{ \left(T_{\frac{i}{n}}(\overline{Y}) - T_{\frac{i-1}{n}}(\overline{Y}) \right), i = 1, ..., n \right\}$ *has the distribution* $P^{\varepsilon}_{n,f}$ *of the triangular array defined in (11).*

Proof. For $t \in [0, \tau_{\frac{1}{n}}], \overline{Y}_t = X_0(t)$ and $T_{\frac{1}{n}}(\overline{Y}) = \tau_{\frac{1}{n}}$. Consider now $t \in [\tau_{\frac{1}{n}}, \tau_{\frac{2}{n}}]$; then

$$\overline{Y}_t = \frac{1}{n} + f(\frac{1}{n}) \left(t - \tau_{\frac{1}{n}} \right) + \varepsilon \left(W_t - W_{\tau_{\frac{1}{n}}} \right) .$$

Thus $\overline{Y}_{\tau_{\frac{1}{n}}+u} = \frac{1}{n} + X_1(u)$ for $u \in [0, \tau_{\frac{2}{n}} - \tau_{\frac{1}{n}}]$, and $T_{\frac{2}{n}}(\overline{Y}) = \tau_{\frac{2}{n}}$ since by construction, $\tau_{\frac{2}{n}} - \tau_{\frac{1}{n}} = T_{\frac{1}{n}}(X_1)$. By induction, if $t \in (\tau_{\frac{i-1}{n}}, \tau_{\frac{i}{n}}]$,

$$\overline{Y}_t = \frac{i-1}{n} + f(\frac{i-1}{n}) \left(t - \tau_{\frac{i-1}{n}} \right) + \varepsilon \left(W_t - W_{\tau_{\frac{i-1}{n}}} \right) \tag{23}$$

and $T_{\frac{i}{n}}(\overline{Y}) = \tau_{\frac{i-1}{n}} + T_{\frac{1}{n}}(X_{i-1}) = \tau_{\frac{i}{n}}$. This holds for $i = 1, ..., n$. Now, the random variables $\tau_{\frac{i}{n}}$ are stopping times of $(\mathcal{A}_t)_{t\geq 0}$. Thus, $\left(W_{u+\tau_{\frac{i-1}{n}}} - W_{\tau_{\frac{i-1}{n}}}\right)_{u\geq 0}$ is a Brownian motion independent of $\mathcal{A}_{\tau_{\frac{i-1}{n}}}$ for all $i = 1, ..., n$. Hence, the random variables $\left\{\left(\tau_{\frac{i}{n}} - \tau_{\frac{i-1}{n}}\right), i = 1, ..., n\right\}$ are independent, and by construction $\left(\tau_{\frac{i}{n}} - \tau_{\frac{i-1}{n}}\right)$ has the inverse Gaussian distribution

$$IG\left(n^{-1}(f(\frac{i-1}{n}))^{-1}, n^{-2}\varepsilon^{-2}\right).$$

■

Proof of Proposition 5. Let P^ε denote the distribution of $(\varepsilon W_t,\, t \geq 0)$ on $(C(\mathbb{R}^+, \mathbb{R}), \mathcal{C})$, and P_T^ε the restriction of P^ε to $\mathcal{C}_T$. Then by the Girsanov formula

$$\begin{aligned}\log \frac{d\bar{P}_{n,f}^{T,\varepsilon}}{dP_T^\varepsilon}(\overline{Y}) &= \frac{1}{\varepsilon^2}\int_0^T f_n(t,\overline{Y})d\overline{Y}_t - \frac{1}{2\varepsilon^2}\int_0^T f_n(t,\overline{Y})^2 dt \\ &= \frac{1}{\varepsilon^2}\sum_{i=1}^n f\left(\frac{i-1}{n}\right)\frac{1}{n} - \frac{1}{2\varepsilon^2}\sum_{i=1}^n f\left(\frac{i-1}{n}\right)^2\left(T_{\frac{i}{n}}(\overline{Y}) - T_{\frac{i-1}{n}}(\overline{Y})\right).\end{aligned}$$

Hence, $S = \left(T_{\frac{i}{n}}(\overline{Y}) - T_{\frac{i-1}{n}}(\overline{Y}),\, i = 1, ..., n\right)$ is a sufficient statistic for the experiment $\overline{\mathcal{G}}^{n,\varepsilon}$ defined by $\left(\overline{Y}_t,\, t \leq T_1(\overline{Y})\right)$. This gives (i). Since, by Lemma 1, $S\overline{\mathcal{G}}^{n,\varepsilon} = \mathcal{G}^{n,\varepsilon}$, we obtain that the two experiments are equivalent by Property 1. ■

4 A bound for the Δ-distance

In this section, we prove a proposition from which Theorem 1 can be derived. It follows from the results of Section 3 and the triangular inequality that

$$\Delta(\mathcal{E}_0^\varepsilon, \mathcal{G}^{n,\varepsilon}) = \Delta(\mathcal{E}_0^\varepsilon, \overline{\mathcal{G}}^{n,\varepsilon}).$$

Now, $\mathcal{E}_0^\varepsilon$ and $\overline{\mathcal{G}}^{n,\varepsilon}$ have the same measurable space $(C(\mathbb{R}^+, \mathbb{R}), \mathcal{C}_T)$. So applying Property 2 (see (18)) we get the bound

$$\Delta(\mathcal{E}_0^\varepsilon, \mathcal{G}^{n,\varepsilon}) \leq \sup_{f\in\mathcal{F}} \|P_f^{T,\varepsilon} - \bar{P}_{n,f}^{T,\varepsilon}\|_{TV}.$$

Proposition 7 *For $f \in \mathcal{F}$ we have*

$$\|P_f^{T,\varepsilon} - \bar{P}_{n,f}^{T,\varepsilon}\|_{TV} \leq K\,C(m)\left(\frac{1}{(n\varepsilon)^2} + \frac{1}{n} + \varepsilon^2\right)^{\frac{1}{2}}$$

uniformly on $\mathcal{F}$, where K is the constant defining $\mathcal{F}$ and $C(m)$ is a constant which depends only on m.

Proof. We use here an upper bound given in Jacod and Shiryaev (1987), §4b, Theorem 4.21, p. 279, for the total variation norm between the distributions of two diffusion type processes having the same constant diffusion coefficient. Let h^f be the Hellinger process of order $\frac{1}{2}$ between P_f^ε and $\overline{P}_{n,f}^\varepsilon$ (see e.g. Jacod and Shiryaev Chap. 4). For $z \in C(\mathbb{R}^+, \mathbb{R})$, it is given by

$$h_u^f(z) = \frac{1}{8\varepsilon^2} \int_0^u (f(z(t)) - f_n(t,z))^2 dt, \text{ for } u > 0.$$

Since the two processes (Y_t) and $(\overline{Y}_t)$ have the same initial distribution $(Y_0 = \overline{Y}_0 = 0)$, the inequality for the total variation norm $\|.\|_{TV}$ is the following,

$$\|P_f^{T,\varepsilon} - \bar{P}_{n,f}^{T,\varepsilon}\|_{TV} \leq 4\sqrt{E_{\bar{P}_{n,f}^\varepsilon}(h_T^f)}, \tag{24}$$

with $T = T_1(X)$.

It is worth noting that this inequality is not symmetric: for the right hand side of (24), we may choose to take the expectation either with respect to $\bar{P}_{n,f}^\varepsilon$, or with respect to P_f^ε. The choice $\bar{P}_{n,f}^\varepsilon$ makes the computation easier here.

Let us set $E_{\bar{P}_{n,f}^\varepsilon}(h_T^f) = E(n,\varepsilon)$. We have

$$\begin{aligned} E(n,\varepsilon) &= \frac{1}{8\varepsilon^2}\mathbb{E}\int_0^T \left((f(\overline{Y}_t) - f_n(t,\overline{Y}_t)\right)^2 dt \\ &= \frac{1}{8\varepsilon^2}\sum_{i=1}^n \mathbb{E}\int_{T_{\frac{i-1}{n}}(\overline{Y})}^{T_{\frac{i}{n}}(\overline{Y})} \left(f(\overline{Y}_t) - f(\frac{i-1}{n})\right)^2 dt. \end{aligned}$$

Now, using Lemma 6 and (21)-(23),

$$\begin{aligned} &\mathbb{E} \int_{T_{\frac{i-1}{n}}(\overline{Y})}^{T_{\frac{i}{n}}(\overline{Y})} \left(f(\overline{Y}_t) - f(\frac{i-1}{n})\right)^2 dt = \mathbb{E}\int_{\tau_{\frac{i-1}{n}}}^{\tau_{\frac{i}{n}}} \left(f(\overline{Y}_t) - f(\frac{i-1}{n})\right)^2 dt \\ &= \mathbb{E}\int_0^{\tau_{\frac{i}{n}} - \tau_{\frac{i-1}{n}}} \left(f(\overline{Y}_{u+\tau_{\frac{i-1}{n}}}) - f(\frac{i-1}{n})\right)^2 du \\ &= \mathbb{E}\int_0^{T_{\frac{1}{n}}(X_{i-1})} \left(f(\frac{i-1}{n} + X_{i-1}(u)) - f(\frac{i-1}{n})\right)^2 du \\ &\leq K^2\, \mathbb{E}\int_0^{T_{\frac{1}{n}}(X_{i-1})} X_{i-1}^2(u) du, \end{aligned}$$

where K is the Lipschitz constant of f and X_{i-1} is a Brownian motion starting from 0 with drift coefficient $f\left(\frac{i-1}{n}\right)$ and diffusion coefficient ε (see (21), (22)).
It is well known that this last expectation can be computed explicitly.

Lemma 8 *Let* $X(u) = \theta u + \varepsilon W_u$, $u \geq 0$ *be a Brownian motion with drift* $\theta > 0$. *Let* $T_a = T_a(X)$ *be the first hitting time of level* a. *Then for* $a > 0$,

$$\mathbb{E}\int_0^{Ta} X^2(u)du = \frac{a^3}{3\theta} - \frac{a^2\varepsilon^2}{2\theta^2} + \frac{a\varepsilon^4}{2\theta^3}.$$

Proof. Let $s(u) = \exp\left(-\frac{2\theta u}{\varepsilon^2}\right)$ and

$$S(x) = \int_0^x s(u)du = \frac{\varepsilon^2}{2\theta}\left(1 - \exp\left(-\frac{2\theta x}{\varepsilon^2}\right)\right)$$

be the scale function of the diffusion $X(u)$. For $b < 0 < a$, it is well known that (see e. g. Karlin and Taylor (1981), Chap. 15)

$$E\int_0^\tau X^2(u)du = \varphi_{b,a}(0) \text{ with } \quad \tau = T_a \wedge T_b,$$

where $\varphi_{b,a} = \varphi$ is the solution of the equation

$$\frac{\varepsilon^2}{2}\varphi''(x) + \theta\varphi'(x) = -x^2$$

$$\varphi(b) = \varphi(a) = 0.$$

Then $\lim_{b\to-\infty}\varphi_{b,a}(0) = \mathbb{E}\int_0^{T_a} X^2(u)du$, and we have

$$\mathbb{E}\int_0^{T_a} X^2(u)du = \frac{2}{\varepsilon^2}\left\{\int_0^a (S(a) - S(u))\frac{u^2}{s(u)}du + (S(a) - S(0))\int_{-\infty}^0 u^2\frac{du}{s(u)}\right\}.$$

Straightforward computations lead to Lemma 8. ∎

Coming back to the proof of Proposition 7, we get

$$\begin{aligned} E(n,\varepsilon) &\leq \frac{1}{8\varepsilon^2}K^2\left\{\sum_{i=1}^n \frac{1}{3n^3}\frac{1}{f(\frac{i-1}{n})} - \frac{\varepsilon^2}{2n^2}\sum_{i=1}^n \frac{1}{f^2(\frac{i-1}{n})} + \frac{\varepsilon^4}{2n}\sum_{i=1}^n \frac{1}{f^3(\frac{i-1}{n})}\right\} \\ &\leq \frac{K^2}{8}\left\{\frac{1}{3(n\varepsilon)^2}\frac{1}{m} - \frac{1}{2n}\frac{1}{m^2} + \frac{\varepsilon^2}{2m^3}\right\}. \end{aligned}$$

This completes the proof of Proposition 7. ∎

Proof of Theorem 1 Now, we have,

$$\Delta(\mathcal{E}_0^\varepsilon, \mathcal{G}^{n,\varepsilon}) \leq KC(m)\left(\frac{1}{(n\varepsilon)^2} + \frac{1}{n} + \varepsilon^2\right)^{\frac{1}{2}}$$

which tends to 0 as $\varepsilon \to 0$ if $n = n_\varepsilon \to \infty$ such that $\varepsilon\, n_\varepsilon \to \infty$. ∎

Remark 3 *Setting $\varepsilon\sqrt{n} = 1$, these three terms are equal to ε^2 . This leads to the rate of convergence $\frac{1}{\sqrt{n}}$.*

Proof of Corollary 1: By Proposition 1, the statistic $z \to (T_{\frac{i}{n}}(z),\, i = 1, \ldots, n)$ is exactly sufficient for the experiment $\overline{\mathcal{G}}^{n,\varepsilon}$. So, we have

$$\Delta(\mathcal{E}_0^\varepsilon, \mathcal{G}^{n,\varepsilon}) = \Delta(\mathcal{E}_0^\varepsilon, \overline{\mathcal{G}}^{n,\varepsilon}).$$

Therefore, $\Delta(\mathcal{E}_0^\varepsilon, \overline{\mathcal{G}}^{n,\varepsilon})$ tends to 0 as $\varepsilon \to 0$. This proves that the same statistic is asymptotically sufficient for $\mathcal{E}_0^\varepsilon$. ■

5 Exponential family regression and white noise

The experiment $\mathcal{G}^{n,\varepsilon}$ defined by the triangular array $(X_n^i,\, i = 1, \ldots, n)$ where X_n^i is distributed according to

$$IG\left(n^{-1}(f(\frac{i-1}{n}))^{-1}, n^{-2}\varepsilon^{-2}\right)$$

is equivalent to a nonparametric regression experiment. Indeed, set $Z_n^i = nX_n^i$. Then

$$Z_n^i = \frac{1}{f(\frac{i-1}{n})} + \sqrt{n}\varepsilon\, \xi_n^i$$

where $E\xi_n^i = 0$, $E(\xi_n^i)^2 = \frac{1}{f^3(\frac{i-1}{n})}$ (due to the properties of the inverse Gaussian distribution). Moreover, the inverse Gaussian distribution has the following scaling property: if $X \sim IG(\mu, \lambda)$ then $cX \sim IG(c\mu, c\lambda)$ (cf. Johnson and Kotz, 1970). Hence the variables $Z_n^i = nX_n^i$ have distribution $IG\left((f(\frac{i-1}{n}))^{-1}, n^{-1}\varepsilon^{-2}\right)$.

Let us now assume

(25) $$n = n_\varepsilon = \varepsilon^{-2} \quad .$$

Note that this choice implies that $n_\varepsilon \to +\infty$ and $\varepsilon n_\varepsilon = \frac{1}{\varepsilon} \to \infty$ as $\varepsilon \to 0$. This corresponds to the conditions required in Theorem 1. Then, $Z_n^i \sim IG\left((f(\frac{i-1}{n}))^{-1}, 1\right)$. The distribution of Z_n^i is

$$\mu_{n,f}^i(dt) = \exp\left(-\frac{t}{2}f^2(\frac{i-1}{n}) + f(\frac{i-1}{n})\right)\nu(dt)$$

with

$$\nu(dt) = \frac{1}{\sqrt{2\pi t^3}} \exp\left(-\frac{1}{2t}\right) \times \mathbf{1}_{t>0}\, dt.$$

The dominating measure is now independent of n. Let $\tilde{\mathcal{G}}^{n,\varepsilon}$ be the experiment given by observing the independent variables $(Z_n^i,\ i=1,\ldots,n)$. We have $\Delta(\mathcal{G}^{n,\varepsilon},\tilde{\mathcal{G}}^{n,\varepsilon})=0$ since the mapping $X_n^i \mapsto nX_n^i$ is one-to-one. Consider the exponential family in canonical form

$$\mu_\phi(dt) = \exp\left(\phi U(t) - V(\phi)\right)\nu(dt) \tag{26}$$

where

$$U(t) = -t,\ V(\phi) = -(2\phi)^{1/2}.$$

We note that $\mu_\phi = IG((2\phi)^{-1/2},1)$ for $\phi \in (0,\infty)$, so that μ_ϕ is defined for all $\phi \in (0,\infty)$. We thus have $\mu_{n,f}^i = \mu_\phi$ for $\phi = \frac{1}{2}f^2(\frac{i-1}{n})$.

The exponential family regression model in Grama and Nussbaum (1996) was that independent variables $(Z_n^i,\ i=1,\ldots,n)$ are observed, where the law $\mathcal{L}(Z_n^i)$ is such that $\mathcal{L}(Z_n^i) = \mu_\phi$ for $\phi = g(\frac{i}{n})$, and g is a smooth function on $[0,1]$. If we now set $g(x) = \frac{1}{2}f^2(x)$, then we are in this framework.

Proof of Theorem 3 . The conditions on f guarantee that $m \le f(x) \le 2K$ for $x \in [0,1]$, with $m>0$. Evidently g satisfies a uniform Lipschitz condition:

$$|g(x)-g(y)| = \frac{1}{2}\left|f^2(x)-f^2(y)\right| = \frac{1}{2}\left|f(x)+f(y)\right|\left|f(x)-f(y)\right| \tag{27}$$

$$\le \frac{1}{2}4K\left|f(x)-f(y)\right| \le 2K^2\left|x-y\right|,\ x,y\in[0,1] \tag{28}$$

and moreover, for $x \in [0,1]$,

$$g(x) \in [m^2/2, 2K^2]. \tag{29}$$

Let $\Sigma = \Sigma(m,K)$ be the set of all functions g satisfying (28) and (29). Thus all conditions assumed in Grama and Nussbaum (1996) are satisfied. By theorem 12 in Grama and Nussbaum (1996) we obtain a Gaussian white noise approximation in the Δ-sense, as an experiment

$$dZ_t = \Gamma(g(t))dt + n^{-1/2}dW_t,\ t\in[0,1] \tag{30}$$

with $g \in \Sigma$. The function Γ is determined by the exponential family $(\mu_\phi,\ \phi\in(0,\infty))$ as an appropriate variance stabilizing transform. Let us determine Γ.

Using the notation of section 3.3 in Grama and Nussbaum (1996), we obtain (cf. relation (3.35), (3.34) there)

$$\Gamma(\phi) = F(b(\phi))$$

where

$$F'(x) = \sqrt{a'(x)},\ b(\phi) = V'(\phi)$$

a is the inverse function to $b(\phi)$ and $V(\phi) = -(2\phi)^{-1/2}$ is the function appearing in (26). We obtain

$$b(\phi) = -\frac{d}{d\phi}(2\phi)^{1/2} = -(2\phi)^{-1/2} \text{ for } \phi > 0,$$

$$a(x) = \frac{1}{2}x^{-2},\ a'(x) = -x^{-3} \text{ for } x<0.$$

Hence

$$F'(x) = |x|^{-3/2},\ F(x) = -2|x|^{-1/2} + C \text{ for } x < 0.$$

Consequently

$$\Gamma(\phi) = F(b(\phi)) = -2(2\phi)^{1/4}$$

so that (30) becomes (up to an equivalence, given by multiplication with $\frac{1}{2}$)

$$dZ_t = (2g(t))^{1/4}dt + \frac{1}{2}n^{-1/2}dW_t,\ t \in [0,1].$$

Substituting $g(x) = \frac{1}{2}f^2(x)$ we get

$$dZ_t = f^{1/2}dt + \frac{1}{2}n^{-1/2}dW_t,\ t \in [0,1].$$

Thus the proof of Theorem 2 is complete. ∎

Proof of Corollary 2 : From Nussbaum (1996), we know that, if $n = n_\varepsilon = [\varepsilon^{-2}]$, $\Delta(\mathcal{E}_1^\varepsilon, \mathcal{E}^n) \to 0$ as $\varepsilon \to 0$. Using the triangular inequality yields

$$\Delta(\mathcal{E}_0^\varepsilon, \mathcal{E}^n) \leq \Delta(\mathcal{E}_0^\varepsilon, \mathcal{E}_1^\varepsilon) + \Delta(\mathcal{E}_1^\varepsilon, \mathcal{E}^n)$$

This implies asymptotic equivalence with density estimation, more precisely with the experiment given by n observed i.i.d. random variables, as claimed. ∎

References

[1] Brown, L. D. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**2384-2398

[2] Chikara, R. S. and Folks, J. L. (1989). *The Inverse Gaussian Distribution.* Marcel Dekker, New York

[3] Donoho, D. (1994). Asymptotic minimax risk (for sup–norm loss): solution via optimal recovery. *Probab. Theor. Rel. Fields* **99** 145-170.

[4] Genon-Catalot, V., (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics* **21 (**1) 99-116

[5] Genon-Catalot, V., Larédo, C. (1987). Limit theorems for the first hitting times process of a diffusion and statistical applications. *Scand. J. Statist.* **14** 143-160

[6] Grama, I. and Nussbaum, M. (1997). Asymptotic equivalence for nonparametric generalized models. Preprint No. 289, Weierstrass Institute, Berlin (To appear, *Prob. Theor. Rel. Fields*)

[7] Jacod, J. and Shiryaev A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, New York etc.

[8] Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions,* Vol. 1. Wiley, New York

[9] Karlin, S. and Taylor, H. M. (1981). *A Second Course in Stochastic Proceses.* Academic Press, Orlando

[10] Korostelev, A. P. (1993). Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory Probab. Appl.* **38**, no. 4, 775–782

[11] Kutoyants, Yu. A. (1985). On nonparametric estimation of trend coefficients in a diffusion process. In: *Statistics and Control of Stochastic Processes, Steklov Seminar 1984* (Krylov, N. V., Liptser, R. Sh., Novikov, A. A. Eds.), Optimization Software, Inc, New York

[12] Larédo, C. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *Ann. Statist.* **18** 1158-1171

[13] Lepskii, O., Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512-2546

[14] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New-York etc.

[15] Le Cam, L. and Yang, G. (1990). *Asymptotics in Statistics: Some Basic Concepts.* Springer-Verlag, New-York etc.

[16] Milstein, G. and Nussbaum, M. (1996). Diffusion Approximation for Nonparametric Autoregression. Unpublished manuscript (to appear, *Prob. Theor. Rel. Fields).*

[17] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399-2430

[18] Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise (in Russian). *Problemy Peredachi Informatsii* 16 (2) 52-68, English translation *Problems Inform. Transmission* (1980) 120-133

[19] Strasser, H. (1985). *Mathematical Theory of Statistics.* Walter de Gruyter. Berlin, New-York.