Genon-Catalot, Valentine; Laredo, Catherine; Nussbaum, Michael

Working Paper

Asymptotic equivalence of estimating a poisson intensity and a positive diffusion drift

SFB 373 Discussion Paper, No. 2000,75

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Genon-Catalot, Valentine; Laredo, Catherine; Nussbaum, Michael (2000) : Asymptotic equivalence of estimating a poisson intensity and a positive diffusion drift, SFB 373 Discussion Paper, No. 2000,75, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10048015

This Version is available at:
https://hdl.handle.net/10419/62238

ASYMPTOTIC EQUIVALENCE OF ESTIMATING A POISSON INTENSITY AND A POSITIVE DIFFUSION DRIFT

VALENTINE GENON-CATALOT, CATHERINE LAREDO, AND MICHAEL NUSSBAUM

Université Marne-la-Vallée, INRA Jouy-en-Josas and Cornell University

ABSTRACT. We consider a diffusion model of small variance type with positive drift density varying in a nonparametric set. We investigate Gaussian and Poisson approximations to this model in the sense of asymptotic equivalence of experiments. It is shown that observation of the diffusion process until its first hitting time of level one is a natural model for the purpose of inference on the drift density. The diffusion model can be discretized by the collection of level crossing times for a uniform grid of levels. The random time increments are asymptotically sufficient and obey a nonparametric regression model with independent data. This decoupling is then used to establish asymptotic equivalence to Gaussian signal-in-white noise and Poisson intensity models on the unit interval, and also to an i.i.d. model when the diffusion drift function f is a probability density. As an application, we find the exact asymptotic minimax constant for estimating the diffusion drift density with sup-norm loss.

1. INTRODUCTION

Diffusion processes defined by stochastic differential equations have been widely used for modeling purposes. Consider the autonomous differential equation

$$\frac{dy(t)}{dt} = f(y(t)) \quad ; \quad t \geq 0,\ y(0) = 0 \tag{1.1}$$

where f is some continuous function. When f is positive, the solution y of (1.1) is monotone increasing and can be used to model growth processes. Assume that this system is modified by a small random perturbation induced by a Wiener process W_t. This leads to a stochastic differential equation

$$dY_t = f(Y_t)dt + \varepsilon dW_t \quad ; \quad t \geq 0,\ Y_0 = 0 \tag{1.2}$$

where ε is a small parameter. Diffusion processes obtained as small random perturbations of deterministic dynamical systems have been extensively studied in probability theory (cf. e.g. Freidlin and Wentsell (1998)). We are interested in the statistical model where the process Y_t is observed and the positive function f is unknown.

Date: July 25, 2000.

1991 *Mathematics Subject Classification.* Primary 62 B 15; Secondary 62M05, 62 G 07 .

Key words and phrases. Nonparametric experiments, deficiency distance, diffusion process, discretization, inverse Gaussian regression, signal in white noise, Poisson intensity, asymptotic minimax constant.

Compared with the signal function g in the white noise model

$$dZ_u = g(u)du + \varepsilon dW_u \quad ; \quad 0 \le u \le 1,\ Z_0 = 0 \tag{1.3}$$

the drift density f has to fulfill stronger conditions, in order to ensure existence and uniqueness of non-exploding solutions of (1.2) (cf. Karatzas and Shreve (1997)). The Lipschitz condition imposed here (see condition (C2) in section 2.1 below) is essentially inherited from the ordinary differential equation (1.1). Note that in (1.2) the function f is a function of the state variable Y_t, whereas g in (1.3) is a function of time. Another feature which distinguishes the diffusion model (1.2) from the white noise model (1.3) is that Y_t in (1.2) is a continuous strong Markov process.

We are interested here in the statistical theory for Markov and diffusion processes, specifically in approximation of the respective experiments by simpler ones like Gaussian and Poisson experiments. For the diffusion model (1.2), two types of asymptotics have been considered: fixed time interval $t \in [0,T]$ with noise intensity $\varepsilon \to 0$ (see Kutoyants (1984), (1994)), and fixed noise intensity ($\varepsilon = 1$) for an expanding time interval $T \to \infty$. In the latter case Y_t is commonly assumed stationary and ergodic (cf. Kutoyants (1997) and references therein). Note that the ergodic case requires a different set of conditions on the drift function; in particular, f cannot be positive as in our model.

For both types of asymptotics, standard local likelihood theory leads to Gaussian limits of experiments which provide a satisfactory basis for efficient parametric inference. Our aim however is reduction of nonparametric experiments, when the drift function f varies in a function class $\mathcal{F}$, within the framework of asymptotic equivalence (cf. Brown and Low (1996), Nussbaum (1996)). It turns out that a straightforward reduction to a Gaussian or Poisson experiment is possible under a specific assumption: *the diffusion model (1.2) with small noise intensity $\varepsilon \to 0$ is observed until its first hitting time $T_1(Y)$ of level 1*. We shall see below (Remark 1 at the end of section 2.1) that this model is appropriate when we would like to estimate $f(u)$, $0 \le u \le 1$ and we are not interested in its values beyond the interval.

This special observation model arises in a rather natural way from the Markov character of our diffusion process. For any Gaussian approximation of a stochastic process experiment, a natural first step is to approximate the process by some collection of independent random variables ("decoupling", cf. de la Peña and Giné (1999) for a general theory). The appropriate method is *not discretization of the process in time* which results in a Markov chain; in fact discretization at equidistant time points $t_i = i/n$ was the method applied in Milstein and Nussbaum (1998) to the case of a fixed observation interval $t \in [0,T]$. That method led to some interesting statistical equivalences (see below for a more detailed discussion), but not to a Gaussian approximation. Time discretization is also applied in the seminal paper of Brown and Low (1996) to the white noise model (1.3).

In the Markov diffusion model (1.2) it turns out that *space discretization* is more natural and leads to a statistically useful decoupling. Assume that the drift density f

is nonnegative; in that case the process Y_t which starts at $Y_0 = 0$ tends to increase and will eventually reach value 1 almost surely. Even though the state space for the process Y_t is the real line, consider the interval $[0,1]$ as the area of interest here and endow it with an equidistant grid of points i/n, $i = 1, \dots, n$. Let $T_{i/n}$ be the first hitting time of level i/n by the process Y_t; then the increments $T_{(i+1)/n} - T_{i/n}$ are independent random variables. That fact is an immediate consequence of the Markov character of the diffusion process: these increments depends on the past only via the value of Y_t at $T_{i/n}$, which is i/n by definition. This value depends on i, but is nonrandom and thus independent of the past; hence $T_{(i+1)/n} - T_{i/n}$ does not depend on the past of the process. This idea of using space rather than time discretization is well known in a probabilistic context; is has been applied in algorithms of path reconstruction (Kushner and Dupuis (1992), Milstein (1998), Milstein and Tretyakov (1999)).

This reduction to independent variables takes places in the sense of the Δ-distance for experiments, for small noise asymptotics $\varepsilon \to 0$. In the sequel we use available theory for nonparametric regression models (Grama and Nussbaum (1998)) to obtain a Gaussian white noise approximation of type (1.3) for $g = f^{1/2}$ where f is the drift function in (1.2). It is obvious that conceptually, the time u in this latter models stands for the space variable in the original model (1.2); thus our arguments are related to the well-known technique of time change for stochastic processes.

The Euler type (time) discretization of Y_t in (1.2) on a fixed interval $t \in [0,1]$ has been considered by Milstein and Nussbaum (1998). It was shown that the statistic $(Y_{t_1}, ..., Y_{t_n})$, where $t_i = i/n$ is *asymptotically sufficient* for $\varepsilon \to 0$, in the experiment given by observations Y_t , $t \in [0,1]$ from (1.2), if $n = n_\varepsilon \to \infty$ in such a way that $\varepsilon n_\varepsilon \to \infty$ and f varies in a certain nonparametric set of functions $\mathcal{F}$. This extended the result obtained in the corresponding parametric problem (see Larédo (1990), Genon–Catalot (1990)). The proof was based on considering the statistical model associated with the Euler scheme:

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

where $(\xi_i, i = 1, \dots, n)$ are i. i. d. standard normal variables. It was shown that Le Cam's deficiency distance between these experiments tends to 0 as $\varepsilon \to 0$.

In this paper we consider the experiment given by the diffusion process Y_t defined in (1.2) when it is observed until its first hitting time $T_1(Y)$ of level one. The unknown function f belongs to the set $\mathcal{F} = \mathcal{F}_{K,m}$ associated with two positive constants K and m; see section 2.1 below for details. This will be our first experiment $\mathcal{E}_0^\varepsilon$ having parameter set $\mathcal{F}$.

In the model (1.2), let us consider the hitting time process

$$T_a = T_a(Y) = \inf \{t \geq 0 : Y_t = a\}$$

defined on an interval of levels $a \in [0, a_1]$ for some prescribed $a_1 > 0$. This process has some nice properties. First, it is an increasing Lévy process (a process with independent positive increments). Second, its observation is equivalent to the observation of the

record process

$$M_t = \sup_{s \le t} Y_s, \ t \in [0, T_{a_1}(Y)].$$

Indeed, the function $a \mapsto T_a(Y)$, $a \in [0, a_1]$ is the left continuous pseudo-inverse of $t \mapsto M_t$, $t \in [0, T_{a_1}(Y)]$. Records of diffusion processes occur in mathematical finance (records of stock market indices, cf. Musiela and Rutkowski (1997), chap. 9.6).

For the sake of simplicity, we assume from now on $a_1 = 1$. In this paper we address the problem of inference on the function f defined on $[0,1]$ from discrete observations of the hitting time process, i.e. of times $T_{i/n}(Y)$, $i = 1, \ldots, n$. These random times are almost surely finite under the assumption that f is positive. We ask how much information about the unknown drift function f is lost when only discrete observations on the path Y_t are available. It turns out that space discretization based on successive level crossings is feasible in the diffusion model under the assumption of a positive f, and leads to further interesting statistical results.

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

$$h_{\mu,\lambda}(t) = \left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \exp\left(-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right) \mathbf{1}_{t>0}(t) \tag{1.4}$$

(see e.g. Chikara and Folks (1989)). Denote by $\mathcal{G}^{n,\varepsilon}$ this experiment indexed again by $f \in \mathcal{F}$.

Our first result (Theorem 1) states that, as ε goes to 0, the deficiency distance of these two 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 $(Y_t, t \ge 0)$, i. e. the statistic $(T_{i/n}(Y), \quad i = 1, \ldots, n)$ is asymptotically sufficient (Corollary 1). 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 (1998) for such models, we arrive at a Gaussian approximation for our diffusion experiment (1.2). Indeed, consider an experiment given by an observed signal in white noise

$$dZ_u = f^{1/2}(u)du + \frac{\varepsilon}{2}dW_u \quad ; \quad 0 \le u \le 1, \ Z_0 = 0 \tag{1.5}$$

with ε tending to 0 and $f \in \mathcal{F}$. Taking $n_\varepsilon = [\varepsilon^{-2}]$ (where $[\varepsilon^{-2}]$ denotes the largest integer $\le \varepsilon^{-2}$), we prove that $\mathcal{G}^{n_\varepsilon,\varepsilon}$ and the signal-in-white-noise model (1.5) are asymptotically equivalent (Theorem 2).

Another approximating experiment is given by an observed inhomogeneous Poisson process on $[0,1]$ with intensity $\varepsilon^{-2}f(u)$, $u \in [0,1]$ where $f \in \mathcal{F}$. For this step we use the asymptotic equivalence of the Poisson process with (1.5) which is essentially shown in Grama and Nussbaum (1998). This approximation seems to be more convenient for some purposes than (1.5), especially when inference on f rather than on $f^{1/2}$ is desired. As an application, we derive the Korostelev constant (exact asymptotic minimax risk for sup norm loss) for the diffusion drift $f \in \mathcal{F}$ from the corresponding result in the Poisson process model.

Consider the special case where the restriction of the drift function f to $[0,1]$ is not only positive, but is also assumed to have integral one. It has been shown in Nussbaum (1996) that when $\varepsilon^{-2} = n$ takes integer values then the signal-in-white-noise model (1.5) is asymptotically equivalent to the experiment given by n observed i. i. d. 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 (1.2), in the sense of asymptotic equivalence (Corollary 2).

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 $(X_n^i,\, i = 1, \dots, n)$, but comparable to the diffusion experiment $\mathcal{E}_0^\varepsilon$, as in Milstein and Nussbaum (1998). Using this experiment, in Section 4 we compute a bound for the Δ-distance between the diffusion experiment (1.2) and the other ones. In Section 5 we present the argument leading on to the Gaussian experiment, specializing the exponential family nonparametric regression model of Grama and Nussbaum (1998) to the inverse Gaussian case. In section 6 the Poisson approximation is proved, and in section 7 we discuss asymptotic minimax estimation.

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 (1.2).

The parameter ε is assumed to be known. The function f varies in a set $\mathcal{F} = \mathcal{F}_{K,m}$ associated with two positive constants K, m, 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 (1.2) 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 are now able to 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 $C(\mathbb{R}^+,\mathbb{R})$, $\mathcal{C}_t^0=\sigma(X_s,s\leq t)$, $\mathcal{C}_t=\bigcap_{s>t}\mathcal{C}_s^0$ 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 (1.2) on $(C(\mathbb{R}^+,\mathbb{R}),\mathcal{C})$. Now, for $x\in C(\mathbb{R}^+,\mathbb{R})$ and $a\in\mathbb{R}$, let

$$T_a(x)=\inf\{t\geq 0:\ x(t)=a\}. \tag{2.1}$$

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

$$P_f^{T,\varepsilon}=P_f^\varepsilon\,|\mathcal{C}_T \tag{2.2}$$

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

$$\mathcal{E}_0^\varepsilon=\left(C(\mathbb{R}^+,\mathbb{R}),\mathcal{C}_T,\{P_f^{T,\varepsilon},\ f\in\mathcal{F}\}\right). \tag{2.3}$$

Let us now present the second experiment. For $(\mu,\lambda)\in(\mathbb{R}^+)^2$, we denote by $IG(\mu,\lambda)$ the Inverse Gaussian distribution with density given in (1.4). The mean of this distribution is μ and the variance is μ^3/λ. Consider now a triangular array of n independent random variables $(X_n^i,\ i=1,\dots,n)$ such that

$$X_n^i\sim IG\left((f(\frac{i-1}{n}))^{-1},n^{-1}\varepsilon^{-2}\right). \tag{2.4}$$

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

$$\begin{aligned}\tilde{X}_n^i &= \inf\{t\geq 0:\quad \frac{i-1}{n}+f(\frac{i-1}{n})t+\varepsilon B_t^i=\frac{i}{n}\}\\ &= \inf\{t\geq 0:\quad \frac{1}{\varepsilon}f(\frac{i-1}{n})t+B_t^i=\frac{1}{n\varepsilon}\}.\end{aligned} \tag{2.5}$$

Then $\tilde{X}_n^i\sim IG\left(n^{-1}(f(\frac{i-1}{n}))^{-1},n^{-2}\varepsilon^{-2}\right)$, and from the scaling properties of the inverse Gaussian distribution (cf. (3.7) below) it follows that for $X_n^i=n\tilde{X}_n^i$ we have (2.4). Define

$$P_{n,f}^\varepsilon:=\mathcal{L}(X_n^1,\dots,X_n^n). \tag{2.6}$$

This second experiment is now described by

$$\mathcal{G}^{n,\varepsilon}=\left(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\{P_{n,f}^\varepsilon,f\in\mathcal{F}\}\right), \tag{2.7}$$

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

The third experiment is defined by

$$\mathcal{E}_1^\varepsilon=\left(C([0,1],\mathbb{R}),\mathcal{B},\{Q_f^\varepsilon,\ f\in\mathcal{F}\}\right) \tag{2.8}$$

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

$$dZ_u = f^{1/2}(u)du + \frac{\varepsilon}{2}dW_u \quad ; \quad 0\le u\le 1,\ Z_0=0. \tag{2.9}$$

Let $\tilde{Q}_f^\varepsilon$ be the distribution of an inhomogeneous Poisson process with intensity $\varepsilon^{-2}f(u)$, $u\in[0,1]$ on the unit interval. Consider the experiment

$$\mathcal{E}_2^\varepsilon = \left(\mathbb{M},\mathcal{B}_{\mathbb{M}},\{\tilde{Q}_f^\varepsilon, f\in\mathcal{F}\}\right). \tag{2.10}$$

where $(\mathbb{M},\mathcal{B}_{\mathbb{M}})$ is the appropriate measurable space of integer valued measures. Finally, let $\mathcal{F}_1=\mathcal{F}_{1,K,m}$ be the subset of those $f\in\mathcal{F}$ which integrate to one; these are probability densities. Let $\mathcal{E}_{i,1}^\varepsilon$ be the subexperiments of $\mathcal{E}_i^\varepsilon$ $i=0,1$ with parameter space $\mathcal{F}_1$ and let $\mathcal{E}_3^n$ be the experiment given by n i. i. d. random variables with density $f\in\mathcal{F}_1$ on the unit interval. Our aim is to compare these experiments which are indexed by the same parameter set $\mathcal{F}=\mathcal{F}_{K,m}$ (or the set $\mathcal{F}_1\subset\mathcal{F}$, respectively), but which are defined on different observation spaces.

Remark 1. Let us point out that, except for $\mathcal{E}_0^\varepsilon$, f need not be defined outside the interval $[0,1]$. Thus the parameter can be taken to be the restriction of the function f to the interval $[0,1]$, for the experiments $\mathcal{G}^{n,\varepsilon}$, $\mathcal{E}_1^\varepsilon$ and $\mathcal{E}_2^n$. 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 $x\le 0$; $f(x)=f(1)$ for $x\ge 1$. Note that in model (1.2), the sample path $(Y_t, t\le T_1(Y))$ converges in P_f^ε-probability unformly as $\varepsilon\to 0$ to the solution of the ordinary differential equation (1.1) $(y(t), t\le T_1(y))$. Thus asymptotically the domain where f is observed collapses to this latter set of values which is exactly the interval $[0,1]$.

2.2. **Statement of results.** The notions of deficiency distance of experiments and asymptotic sufficiency are briefly reviewed in the next subsection.

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*

$$\Delta\left(\mathcal{E}_0^\varepsilon,\mathcal{G}^{n,\varepsilon}\right)\to 0 \text{ as } \varepsilon\to 0.$$

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

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

The next two results state asymptotic equivalence between the signal-in-white-noise model $\mathcal{E}_1^\varepsilon$ defined in (2.8), the diffusion experiment $\mathcal{E}_0^\varepsilon$ and the Poisson experiment $\mathcal{E}_2^n$.

Theorem 2. *We have*

$$\Delta\left(\mathcal{E}_0^\varepsilon\,,\mathcal{E}_1^\varepsilon\right)\to 0 \;\; as \;\; \varepsilon\to 0.$$

Theorem 3. *We have*

$$\Delta\left(\mathcal{E}_0^\varepsilon\,,\mathcal{E}_2^\varepsilon\right)\to 0 \;\; as \;\; \varepsilon\to 0.$$

Theorem 2 also implies an i. i. d. approximation. The function class $\mathcal{F}_1$ is contained in the class for which the signal-in-white-noise model and the i. i. d. model with n data on the unit interval are asymptotically equivalent if $\varepsilon = n^{-1/2}$, cf. Nussbaum (1996). It then easily follows that $\Delta\left(\mathcal{E}_{1,1}^\varepsilon, \mathcal{E}_3^n\right)\to 0.$ as $\varepsilon^{-2}\to\infty$ along integer values.

Corollary 2. *Suppose that ε takes values $\varepsilon = n^{-1/2}$ where n is integer. Then*

$$\Delta\left(\mathcal{E}_{0,1}^\varepsilon, \mathcal{E}_3^n\right)\to 0 \;\; as \;\; n\to\infty.$$

Remark 2. The result of Theorem 3 can be confirmed by calculating 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 information for regular cases. Consider a parametric submodel of (1.2) where $f = f_\vartheta$, $\vartheta\in\Theta$, Θ is an open interval and the process is observed between 0 and $T_1(Y)$. According to Genon-Catalot and Larédo (1987), if the model is sufficiently regular, then the asymptotic Fisher information (divided by ε^{-2}) is

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

This coincides with the Fisher information (divided by n) in a Poisson process model with intensity $n\ f_\vartheta$, $\vartheta\in\Theta$, or for a model of n i. i. d. observations with density f_ϑ on $[0,1]$ if all f_ϑ integrate to one.

Remark 3. Kutoyants (1985) and (1994), chap. 4.1 considers nonparametric estimation of the drift function f for model 1.2, when it is continuously observed on a fixed time interval $[0,T]$, under the assumption that f is bounded away from 0. Using kernel type estimates, he proves that the rates of convergence are identical to those of density estimation for i. i. d. variables, for a given smoothness condition on f. The equivalences stated in Theorem 2 and Corollary 2 imply such a result for our model with random stopping time $T_a(Y)$, but have much more implications (cf. section 7). In the model with fixed T, special conditions have to be imposed to ensure that the drift function $f(x)$ can be estimated on a given interval $x\in[0,a]$, in view of the fact that the solution $y(t)$ of the underlying differential equation (1.1) depends on the unknown f. Thus if f is to be estimated on a given interval $[0,a]$, it appears more natural to observe the process until $T_a(Y)$ as in our model. Comparing the diffusion experiments with fixed and random stopping times is of interest, but is beyond the scope of this paper.

2.3. **The Le Cam deficiency distance.** This pseudo distance is generally denoted by Δ. In what follows, all measurable spaces (called sample spaces) are supposed to be Polish metric spaces equipped with their Borel σ-algebras.

Consider two experiments with the same parameter space $\mathcal{F}$, say $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \{P_f,\ f \in \mathcal{F}\})$ and $\mathcal{G} = (\mathcal{Y}, \mathcal{B}, \{Q_f,\ f \in \mathcal{F}\})$, but with possibly different sample spaces. Assume 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 $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$, i. e. 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 has the same sample space as $\mathcal{G}$. Let $\mathcal{M}$ denote the set of Markov kernels from $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$.

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

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

where $\|\cdot\|_{TV}$ denotes the total variation norm for measures.

Definition 2. *The* deficiency distance Δ *is given by*

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

In fact Δ 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 mapping and let $T\mathcal{E}$ 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: Let the experiments $\mathcal{E}$ and $\mathcal{G}$ have the same sample space $((\mathcal{X}, \mathcal{A}) = (\mathcal{Y}, \mathcal{B}))$ and define

$$\Delta_0(\mathcal{E}, \mathcal{G}) = \sup_{f \in \mathcal{F}} \|P_f - Q_f\|_{TV}.$$

Then

$$\Delta(\mathcal{E}, \mathcal{G}) \leq \Delta_0(\mathcal{E}, \mathcal{G}). \tag{2.11}$$

Consider now sequences $\mathcal{E}_\varepsilon, \mathcal{G}_\varepsilon$ for $\varepsilon \to 0$; here everything except the parameter space $\mathcal{F}$ may depend on ε.

Definition 3. *(i) The sequences $\mathcal{E}_\varepsilon, \mathcal{G}_\varepsilon$ are called* asymptotically equivalent *(or* accompanying experiments*) if $\Delta(\mathcal{E}_\varepsilon, \mathcal{G}_\varepsilon) \to 0$.*
(ii) Let $\mathcal{G}_\varepsilon$, $\mathcal{E}_\varepsilon$ be two experiments having the same sample space, and let T_ε be a sufficient statistic in $\mathcal{G}_\varepsilon$ with values in an arbitrary measurable space (possibly depending on ε). The statistic T_ε is called asymptotically sufficient *for $\mathcal{E}_\varepsilon$ if $\Delta_0(\mathcal{E}_\varepsilon, \mathcal{G}_\varepsilon) \to 0$.*

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 (1998), 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 $i = 1, \ldots, n$ the times $T_{i/n}(z) = \inf\{t \geq 0,\ z(t) = i/n\}$, $T_0(z) = 0$ and the function

$$f_n(t, z) = \sum_{i=1}^{n} f\left(\frac{i-1}{n}\right) \mathbf{1}_{\left(T_{(i-1)/n}(z), T_{i/n}(z)\right]}(t). \tag{3.1}$$

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

$$\begin{cases} d\overline{Y}_t &= f_n(t, \overline{Y})dt + \varepsilon dW_t \\ \overline{Y}_0 &= 0. \end{cases} \tag{3.2}$$

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 $(\overline{Y}_t,\ t \geq 0)$ on $(C(\mathbb{R}^+, \mathbb{R}), \mathcal{C})$ and set

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

Here 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

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

Lemma 1. *(i) The statistic $z \longmapsto (T_{i/n}(z),\ i = 1, \ldots, n)$ 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.

$$\Delta\left(\mathcal{G}^{n,\varepsilon}, \overline{\mathcal{G}}^{n,\varepsilon}\right) = 0 \text{ for all } n \geq 1,\ \varepsilon > 0.$$

The proof of Lemma 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

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

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

Then, by induction, for $i = 1, ..., n$ and $\tau_0 = 0$

$$X_{i-1}(t) = f\left(\frac{i-1}{n}\right) t + \varepsilon\left(W_{t+\tau_{(i-1)/n}} - W_{\tau_{(i-1)/n}}\right) \tag{3.3}$$

$$\tau_{i/n} = \tau_{(i-1)/n} + T_{1/n}(X_{i-1}). \tag{3.4}$$

Lemma 2. *The hitting times of levels i/n by the process $\overline{Y}_t$ are given by*

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

Moreover, if

$$X_n^i = n(T_{i/n}(\overline{Y}) - T_{i/n}(\overline{Y}),\ i = 1, ..., n \tag{3.5}$$

then the triangular array $(X_n^1, \ldots, X_n^n)$ has the distribution $P_{n,f}^{\varepsilon}$ described in (2.4), (2.6)

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

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

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

$$\overline{Y}_t = \frac{i-1}{n} + f\left(\frac{i-1}{n}\right)(t - \tau_{(i-1)/n}) + \varepsilon\left(W_t - W_{\tau_{(i-1)/n}}\right) \tag{3.6}$$

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

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

i. e. the distribution of the random variables $\tilde{X}_n^i$ defined in (2.5). The following scaling property of the inverse Gaussian can easily be verified using the density formula (1.4): for $a > 0$

$$Z \sim IG(\mu, \lambda) \text{ implies } aZ \sim IG(a\mu, a\lambda). \tag{3.7}$$

Consequently $X_n^i = n\tilde{X}_n^i$ has distribution $IG\left(\left(f\left(\frac{i-1}{n}\right)\right)^{-1}, n^{-1}\varepsilon^{-2}\right)$. □

Proof of Lemma 1. 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

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

Hence, $S = \left(T_{i/n}(\overline{Y}) - T_{(i-1)/n}(\overline{Y}),\ i = 1,...,n\right)$ 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 2, $S\overline{\mathcal{G}}^{n,\varepsilon} = \mathcal{G}^{n,\varepsilon}$, 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

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

Now $\mathcal{E}_0^\varepsilon$ and $\overline{\mathcal{G}}^{n,\varepsilon}$ have the same sample space $(C(\mathbb{R}^+,\mathbb{R}),\mathcal{C}_T)$. So applying Property 2 (see (2.11)) we get the bound

$$\Delta\left(\mathcal{E}_0^\varepsilon, \mathcal{G}^{n,\varepsilon}\right) \leq \Delta_0\left(\mathcal{E}_0^\varepsilon, \mathcal{G}^{n,\varepsilon}\right) = \sup_{f\in\mathcal{F}} \| P_f^{T,\varepsilon} - \bar{P}_{n,f}^{T,\varepsilon} \|_{TV}.$$

Lemma 3. *We have*

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

uniformly over $f \in \mathcal{F}$, 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 $1/2$ between P_f^ε and $\overline{P}_{n,f}^\varepsilon$. 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 distance is

$$\| P_f^{T,\varepsilon} - \bar{P}_{n,f}^{T,\varepsilon} \|_{TV} \leq 4\sqrt{E_{\bar{P}_{n,f}^\varepsilon}(h_T^f)}, \tag{4.1}$$

with $T = T_1(X)$. It is worth noting that this inequality is not symmetric: for the right hand side of (4.1), 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 $E_{\bar{P}^{\varepsilon}_{n,f}}(h^f_T) = E(n,\varepsilon)$. We have

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

Now, using Lemma 2 and (3.3)-(3.6),

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

where K is the Lipschitz constant of f and X_{i-1} is a Brownian motion starting from 0 with drift coefficient $f((i-1)/n)$ and diffusion coefficient ε (see (3.3), (3.4)). It is well known that this last expectation can be computed explicitly.

Lemma 4. *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 $a > 0$*

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

Proof. We shall use classical properties of diffusion exit times from a bounded interval (cf. Karlin and Taylor (1981), Chap. 15, problem C, pp. 193-198). Let $s(u) = \exp(-2\theta u/\varepsilon^2)$ 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 density and the scale function of the diffusion $X(u)$, respectively. For $b < 0 < a$, the first exit time of the interval (b,a) is $\tau = T_a \wedge T_b$. It is well known (see e. g. Karlin and Taylor (1981), Chap. 15) that

$$E \int_0^{\tau} X^2(u) du = \varphi_{b,a}(0)$$

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

$$\begin{aligned} \frac{\varepsilon^2}{2}\varphi''(x) + \theta\varphi'(x) &= -x^2, \quad b \leq x \leq a \\ \varphi(b) &= \varphi(a) = 0. \end{aligned}$$

The solution is found to be

$$\varphi_{b,a}(x) = \int_b^a G(x,\xi)\xi^2 d\xi$$

where $G(x,\xi)$ is the Green function:

$$G(x,\xi) = \left\{ \begin{array}{l} 2\frac{(S(x)-S(b))(S(a)-S(\xi))}{S(a)-S(b)} \frac{1}{\varepsilon^2 s(\xi)}, \text{ for } b \le x \le \xi \le a, \\ \\ 2\frac{(S(a)-S(x))(S(\xi)-S(b))}{S(a)-S(b)} \frac{1}{\varepsilon^2 s(\xi)}, \text{ for } b \le \xi \le x \le a. \end{array} \right\}$$

Therefore $\varphi_{b,a}(0)$ can be split into two terms:

$$\begin{aligned} \varphi_{b,a}(0) &= \frac{2(S(0)-S(b))}{\varepsilon^2(S(a)-S(b))} \int_0^a (S(a)-S(\xi)) \frac{\xi^2}{s(\xi)} d\xi \\ &\quad + \frac{2(S(a)-S(0))}{\varepsilon^2(S(a)-S(b))} \int_b^0 (S(\xi)-S(b)) \frac{\xi^2}{s(\xi)} d\xi. \end{aligned}$$

By continuity of sample paths, as $b \searrow -\infty$, $T_b \nearrow \infty$ and $\tau \nearrow T_a$ almost surely. Thus

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

Furthermore, for all $\xi \in (b, 0]$ we have

$$0 \le \frac{S(\xi)-S(b)}{S(a)-S(b)} \le 1.$$

Therefore, by letting $b \to -\infty$ and noting $S(b) \to -\infty$, we get

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

A somewhat lengthy but straightforward computation of the above integrals completes the proof. □

Coming back to the proof of Proposition 3, we get

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

This completes the proof of Proposition 3. □

Proof of Theorem 1 We have,

$$\Delta(\mathcal{E}_0^\varepsilon, \mathcal{G}^{n,\varepsilon}) \le KC(m) \left(\frac{1}{(n\varepsilon)^2} + \frac{1}{n} + \varepsilon^2 \right)^{\frac{1}{2}} \tag{4.2}$$

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

Remark 4. *If n is chosen such that $\varepsilon n^{1/2} = 1 + o(1)$ then the three terms in the upper bound of (4.2) are of order ε^2. This implies a rate of convergence $n^{-1/2}$ for the Δ-distance.*

Proof of Corollary 1: By Proposition 1, the statistic $z \mapsto (T_{i/n}(z),\, i = 1, \dots, n)$ is sufficient for the experiment $\overline{\mathcal{G}}^{n,\varepsilon}$. In Proposition 3 it is shown that $\Delta_0(\mathcal{E}_0^\varepsilon, \overline{\mathcal{G}}^{n,\varepsilon})$ tends to zero. The statistic is thus asymptotically sufficient in $\mathcal{E}_0^\varepsilon$ according to Definition 3 (ii). □

5. Exponential family regression and white noise

Theorem 2 will be proved via the asymptotic equivalence of observed hitting time increments X_n^i, $i = 1, \dots, n$ (the inverse Gaussian array (3.5)) and the Gaussian white noise (2.9), formally

$$\Delta\left(\mathcal{G}^{n,\varepsilon}, \mathcal{E}_1^\varepsilon\right) \to 0. \tag{5.1}$$

The heuristic background of this result is as follows. Assume first that

$$\varepsilon^{-2} \text{ takes integer values, } n = n_\varepsilon = \varepsilon^{-2} \tag{5.2}$$

and consider the subset of our parameter space $\mathcal{F}_{K,m}$ consisting of constant functions $f(t) = \vartheta$, $t \in \mathbb{R}$ (for $\vartheta \in \Theta = [m, K]$). In this case, the X_n^i in (2.4) are i.i.d. inverse Gaussian $IG(\vartheta^{-1}, 1)$. For these, it is well known that the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_n^i$ is a sufficient statistic, and is asymptotically normal

$$n^{1/2}\left(\bar{X}_n - \vartheta^{-1}\right) \overset{\mathcal{L}}{\Rightarrow} N(0, \vartheta^{-3}). \tag{5.3}$$

This suggests an approximation of the experiment by the Gaussian family

$$\left\{N\left(\vartheta^{-1}, n^{-1}\vartheta^{-3}\right),\ \vartheta \in \Theta\right\} \tag{5.4}$$

(cf. Le Cam (1986), chap. 11 on global asymptotic normality). Let now g be a smooth one-to-one function on $(0, \infty)$; then $g(\bar{X}_n)$ is asymptotically normal with centering $g(\vartheta^{-1})$. In particular if g fulfills

$$(g'(\vartheta^{-1}))^2 \vartheta^{-3} = 1,\ \vartheta \in \Theta$$

i.e. g is a *variance stabilizing transformation* for $\bar{X}_n$ then (5.3) implies

$$n^{1/2}\left(g(\bar{X}_n) - g(\vartheta^{-1})\right) \overset{\mathcal{L}}{\Rightarrow} N(0, 1). \tag{5.5}$$

In the present case, $g(x) = 2x^{-1/2}$ is variance stabilizing. Note that $g(\bar{X}_n)$ is a sufficient statistic along with $\bar{X}_n$; then (5.5) suggests another approximation to the experiment of i.i.d. inverse Gaussian data X_n^i:

$$\left\{N\left(2\vartheta^{1/2}, n^{-1}\right),\ \vartheta \in \Theta\right\}, \tag{5.6}$$

valid simultaneously with (5.4)).

Suppose now that $f(t)$ is not constant but smooth, more precisely $f \in \mathcal{F}_{K,m}$. Then the above argument should be applicable, heuristically, *locally* around a given argument t.

Since ϑ in (5.6) then stands for $f(t)$, we are led to the signal-in-white-noise model (1.5) as an asymptotically equivalent experiment. Below we make this argument precise, on the basis of an exponential family representation of the inverse Gaussian law and the results of Grama and Nussbaum (1998) for nonparametric regression models.

In the latter paper the following model is considered. Let $\mathcal{P}$ be an exponential family on the real line in canonical form given by measures

$$\mu_\theta(dt) = \exp(\theta U(t) - V(\theta))\,\nu(dt) \tag{5.7}$$

with real parameter $\theta \in \Theta$, where Θ is an open (possibly infinite) interval in $\mathbb{R}$, ν is sigma-finite measure, $U(t)$ is a measurable function and $V(\theta)$ is given by $\exp(V(\theta)) = \int \exp(\theta U(t))\,\nu(dt)$. Let $g : [0,1] \mapsto \Theta$ be an unknown regression function, assumed to be in a nonparametric set Σ, and suppose independent observations Z_n^i, $i = 1, \ldots, n$ such that $\mathcal{L}(Z_n^i) = \mu_{g(i/n)}$, $i = 1, \ldots, n$. Under smoothness assumptions on the functions $g \in \Sigma$ it is shown that this regression experiment can be approximated, in the sense of Δ-distance, by a Gaussian white noise model

$$dZ_t = \Gamma(g(t))\,dt + n^{-1/2} dW_t, \quad t \in [0,1], \tag{5.8}$$

where the function $\Gamma(\theta) : \Theta \to \mathbb{R}$ is determined by the parametric family $\mathcal{P}$. If $I(\theta)$ is the Fisher information in $\mathcal{P}$ then

$$\Gamma(\theta) = \int_0^\theta I(u)^{1/2} du. \tag{5.9}$$

The heuristics of this result is similar to (5.3)-(5.6).

To put the present experiment $\mathcal{G}^{n,\varepsilon}$ into this framework, note that for the choice of n according to (5.2) the conditions of Theorem 1 are fulfilled (i.e. $n_\varepsilon \to +\infty$ and $\varepsilon n_\varepsilon = \varepsilon^{-1} \to \infty$ as $\varepsilon \to 0$). The distribution of X_n^i is now $IG\left((f(\frac{i-1}{n}))^{-1}, 1\right)$ which can be written (cf. (1.4))

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

with

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

In (5.7) consider the case where ν is defined as above and

$$U(t) = -t,\ V(\theta) = -(2\theta)^{1/2}. \tag{5.10}$$

Note that $\mu_\theta = IG((2\theta)^{-1/2}, 1)$ for $\theta \in (0, \infty)$, so that μ_θ is defined for all $\theta \in (0, \infty)$. We thus have $\mathcal{L}(X_n^i) = \mu_\theta$ for $\theta = f^2((i-1)/n)/2$. Setting now $g(x) = f^2(x)/2$, we arrive in the framework of Grama and Nussbaum (1998). Note that the Δ-distance between two experiments is not changed under one-to-one reparametrization, and the mapping from f to g is one-to-one under our assumptions. Thus we can invoke the Gaussian approximation (5.8) for the regression experiment indexed by g, and obtain a result for the present experiments $\mathcal{E}_1^\varepsilon$, $\mathcal{G}^{n,\varepsilon}$ indexed by f.

Proof of Theorem 2 . We argue first under assumption (5.2). 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:

$$\begin{aligned} |g(x) - g(y)| &= \frac{1}{2}\left|f^2(x) - f^2(y)\right| = \frac{1}{2}\left|f(x) + f(y)\right|\left|f(x) - f(y)\right| \\ &\le \frac{1}{2}4K\left|f(x) - f(y)\right| \le 2K^2\left|x - y\right|, \ x, y \in [0,1] \end{aligned} \tag{5.11}$$

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

$$g(x) \in [m^2/2, 2K^2]. \tag{5.12}$$

Let $\Sigma = \Sigma(m, K)$ be the set of all functions g satisfying (5.11) and (5.12). Thus all conditions assumed in Grama and Nussbaum (1998) are satisfied. By Theorem 12 in this paper we obtain a Gaussian white noise approximation in the Δ-sense, as an experiment (5.8) with $g \in \Sigma$. The function Γ is determined by the exponential family $(\mu_\theta, \ \theta \in (0,\infty))$ as an appropriate variance stabilizing transform according to (5.9). Let us determine Γ.

It is well known that in the exponential family (5.7) the Fisher information $I(\theta)$ is given by $I(\theta) = V''(\theta)$. In the inverse Gaussian case we obtain from (5.10)

$$\begin{aligned} V''(\theta) &= -\frac{d^2}{d\theta^2}(2\theta)^{1/2} = (2\theta)^{-3/2}, \\ \Gamma(\theta) &= \int_0^\theta (2u)^{-3/4} du = 2(2\theta)^{1/4}, \end{aligned}$$

so that (5.8) becomes (up to an equivalence, given by multiplication with $1/2$)

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

Substituting $g(t) = f^2(t)/2$ we get

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

We have shown (5.1); invoking Theorem 1 completes the proof under assumption (5.2).

For general values of ε, set $n = n_\varepsilon = [\varepsilon^{-2}]$ and $s_\varepsilon = n_\varepsilon^{-1}\varepsilon^{-2}$. Define $\hat{X}_n^i = s_\varepsilon^{-1}X_n^i$, $i = 1, \ldots, n$; the $\hat{X}_n^i$ represent an equivalent experiment. According to (3.7) we have $\hat{X}_n^i \sim IG\left((s_\varepsilon f(\frac{i-1}{n}))^{-1}, 1\right)$. Since $s_\varepsilon \to 1$, the functions $s_\varepsilon f$, $f \in \mathcal{F}_{K,m}$ are elements of a slightly enlarged function class $\mathcal{F}_{K',m'}$ (for $K' > K$, $0 < m' < m$ and sufficiently small ε). The previous argument applied to $\mathcal{F}_{K',m'}$ now establishes asymptotic equivalence to

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

and by multiplication with $s_\varepsilon^{-1/2}$ this is exactly equivalent to (1.5). □

6. Poisson approximation

The heuristics for the Poisson approximation is as follows. Suppose that g in (5.5) is not chosen as variance stabilizing but simply as $g(x) = x^{-1}$, i.e. we are interested in the parameter ϑ. We obtain asymptotic normality of the sufficient statistic $\bar{X}_n^{-1}$: (5.5) becomes

$$n^{1/2}\left(\bar{X}_n^{-1} - \vartheta\right) \stackrel{\mathcal{L}}{\Rightarrow} N(0, \vartheta),$$

which coincides with the asymptotic distribution of the sample mean of n independent Poisson variables with parameter ϑ. This suggests that for general smooth $f \in \mathcal{F}_{K,m}$, for the experiment $\mathcal{G}^{n,\varepsilon}$ we might also have a Poisson process approximation with intensity $n\ f(t)$, or when f has integral one, one of empirical process type with i. i. d. data having density f. In this section we discuss only the Poisson approximation; it was already argued in connection with Corollary 2 that the i. i. d. approximation is an immediate consequence of Nussbaum (1996) and Theorem 2.

Recall that the experiment $\mathcal{E}_2^\varepsilon$ was defined by an observed inhomogeneous Poisson process Π on $[0,1]$ with intensity $\varepsilon^{-2}\ f(u)$, $u \in [0,1]$ where $f \in \mathcal{F}$; the respective laws were $\tilde{Q}_f^\varepsilon$. More precisely, for nonintersecting measurable sets A_1, A_2 in $[0,1]$ the random variables $\Pi(A_i)$, $i = 1, 2$ are independent with Poisson distribution $\mathrm{Po}(\varepsilon^{-2}\int_A f)$. Define also a discretized Poisson experiment $\mathcal{E}_{2,d}^\varepsilon$ as follows: for $n = n_\varepsilon = [\varepsilon^{-2}]$ and $s_\varepsilon = n_\varepsilon^{-1}\varepsilon^{-2}$, observations are n independent r.v.'s Π^i with Poisson distribution $\mathrm{Po}(s_\varepsilon f((i-1)/n))$, $i = 1, \ldots, n$ where $f \in \mathcal{F}$. The following result is a Poisson analog of Brown and Low (1996).

Lemma 5.

$$\Delta\left(\mathcal{E}_2^\varepsilon, \mathcal{E}_{2,d}^\varepsilon\right) \to 0 \ as \ \ \varepsilon \to 0. \tag{6.1}$$

Proof. Consider a step function approximation of f

$$\bar{f}_n(u) = \sum_{i=1}^{n} f\left(\frac{i-1}{n}\right) \mathbf{1}_{((i-1)/n, i/n]}(u).$$

Let $\bar{Q}_f^\varepsilon$ be the law of the Poisson process with intensity $\varepsilon^{-2}\ \bar{f}_n$, and $H(\cdot,\cdot)$ be Hellinger distance for measures. According to a well known estimate for laws of Poisson processes (cf. Reiss (1993), Theorem 3.2.1) we have

$$H^2\left(\tilde{Q}_f^\varepsilon, \bar{Q}_f^\varepsilon\right) \le \varepsilon^{-2}\int_{[0,1]} \left(f^{1/2}(u) - \bar{f}_n^{1/2}(u)\right)^2 du.$$

For $f \in \mathcal{F}$, the function $f^{1/2}$ satisfies a uniform Lipschitz condition for some constant K', hence the right side above is

$$\begin{aligned} & \varepsilon^{-2} \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} \left(f^{1/2}(u) - f^{1/2}((i-1)/n)\right)^2 du \\ \leq \; & \varepsilon^{-2} \sum_{i=1}^{n} (K')^2 \int_{(i-1)/n}^{i/n} \left(u - ((i-1)/n)\right)^2 du \\ \leq \; & \varepsilon^{-2} n^{-2} (K')^2 = O(\varepsilon^2). \end{aligned}$$

Now if $\mathcal{G}_2^\varepsilon$ is the experiment $\left\{\bar{Q}_f^\varepsilon, f \in \mathcal{F}\right\}$ then

$$\Delta\left(\mathcal{E}_2^\varepsilon, \mathcal{G}_2^\varepsilon\right) \leq \sup_{f \in \mathcal{F}} H^2\left(\tilde{Q}_f^\varepsilon, \bar{Q}_f^\varepsilon\right) \to 0. \tag{6.2}$$

Define intervals $J_i = ((i-1)/n, i/n]$. For an observed Poisson process Π, consider the statistic $U_n(\Pi) = (\Pi(J_i), i = 1, \ldots, n)$. By the Neyman criterion applied to the density of $\bar{Q}_f^\varepsilon$ (cf. Reiss (1993), Theorem 3.1.1), it is easy to see that U_n is a sufficient statistic in $\mathcal{G}_2^\varepsilon$. Let $\tilde{Q}_{f,d}^\varepsilon = \mathcal{L}(U_n)$ under $\bar{Q}_f^\varepsilon$, then $\mathcal{E}_{2,d}^\varepsilon = \left\{\tilde{Q}_{f,d}^\varepsilon, f \in \mathcal{F}\right\}$ and hence $\Delta\left(\mathcal{G}_2^\varepsilon, \mathcal{E}_{2,d}^\varepsilon\right) = 0$. In conjunction with (6.2) this proves the lemma. □

Proof of Theorem 3. Assume (5.2); the proof for general values of ε requires a slight modification, similar to the last paragraph of the proof of Theorem 2. The discretized Poisson experiment $\mathcal{E}_{2,d}^\varepsilon$ where $\mathcal{L}(\Pi^i) = \mathrm{Po}(f((i-1)/n))$, $i = 1, \ldots, n$, $f \in \mathcal{F}$ is a special case of the nonparametric regression model of Grama and Nussbaum (1998), cf. example [1] there. As noted already in the proof of Theorem 2, $0 < m \leq f(x) \leq 2K$ for $f \in \mathcal{F}$, $x \in [0,1]$ so that all required conditions on f are satisfied. As shown in Grama and Nussbaum (1998), the discretized Poisson process $\mathcal{E}_{2,d}^\varepsilon$ is asymptotically equivalent to the white noise model (2.9), i.e. $\Delta(\mathcal{E}_{2,d}^\varepsilon, \mathcal{E}_1^\varepsilon) \to 0$ for $n = n_\varepsilon = \varepsilon^{-2}$, so that Lemma 5 and Theorem 2 yield the result. □

7. Asymptotic minimax risk for sup-norm loss

As an application of asymptotic equivalence, we derive the exact asymptotic minimax risk for nonparametric estimation of f in the sup-norm for the diffusion experiment $\mathcal{E}_0^\varepsilon$. Such an exact asymptotic minimax risk (also called an "optimal constant" for an optimal rate of convergence) has been found by Korostelev (1993) in a Gaussian nonparametric regression model. Donoho (1994) and Leonov (1997) established a connection to optimal recovery; Korostelev and Nussbaum (1999) derived an analog for nonparametric density estimation. It should be mentioned that Korostelev's result represents a sup-norm loss analog of the Pinsker constant, which pertains to L_2-loss (Pinsker (1980)).

For the continuous Poisson process model $\mathcal{E}_2^\varepsilon$ with intensity $\varepsilon^{-2} f$, the Korostelev constant for estimating $f \in \mathcal{F}$ can easily be found, based on the analogy with density

estimation for i. i. d. data. We first state this result, and then use asymptotic equivalence to the diffusion model (1.2) in a straightforward manner to carry over the risk asymptotics.

Define $\psi_\varepsilon = (\varepsilon^2 \log \varepsilon^{-2})^{1/3}$, let $\|\cdot\|_\infty$ be the sup-norm for bounded functions on $[0,1]$ and let $\mathcal{W}$ be the set of continuous nondecreasing functions $w(u)$, $u \geq 0$ such that $w(0) = 0$. In the Poisson process experiment $\mathcal{E}_2^\varepsilon$ consider the minimax risk

$$r_\varepsilon(\mathcal{E}_2^\varepsilon, w) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}_{K,m}} E_f^\varepsilon w\left(\psi_\varepsilon^{-1} \left\| \hat{f} - f \right\|_\infty\right) \tag{7.1}$$

where expectation E_f^ε is with respect to the law $\tilde{Q}_f^\varepsilon$ and the infimum is taken over all estimators $\hat{f}$ of f.

Theorem 4. *For any m, K such that $0 < m < K$ and for any bounded function $w \in \mathcal{W}$ we have*

$$\lim_{\varepsilon \to 0} r_\varepsilon(\mathcal{E}_2^\varepsilon, w) = w(K^{2/3} 2^{1/3}).$$

Proof. For a model of n i. i. d. observations with density f on the unit interval and $\varepsilon = n^{1/2}$, Korostelev and Nussbaum (1999) established the limit of $r_\varepsilon(\cdot, w)$ when the function class $\mathcal{F}$ is given by all densities which fulfill a Lipschitz condition $|f(x) - f(y)| \leq K\,|x - y|$. Set

$$B_* = \sup_{f \in \mathcal{F}} \|f\|_\infty \tag{7.2}$$

and

$$A_1 = \max\{g(0) : \|g\|_2 \leq 1, g \in \mathcal{F}_0\}$$

where $\mathcal{F}_0$ is the class functions on $\mathbb{R}$ fulfilling $|f(x) - f(y)| \leq |x - y|$ and $\|\cdot\|_2$ is the norm in $L_2(\mathbb{R})$. The limit of $r_\varepsilon(\cdot, w)$ for the density case is $w(C)$ where

$$C = C(K, B_*) = A_1 \left(\frac{2B_* K}{3}\right)^{1/3}. \tag{7.3}$$

For the Poisson process case with intensity $\varepsilon^{-2}\, f$, the proof needs only minor modifications. Moreover, the restriction that f integrates to one can be dropped in the Poisson case; one needs only to find the appropriate constant B_*. It is easy to see that if $\mathcal{F} = \mathcal{F}_{K,m}$ in (7.2) then $B_* = 2K$, and this value is attained by the function $f(x) = K + Kx$ at $x = 1$. Another look at the proof in the density case reveals that the result remains true under an additional restriction $f(x) \geq m \geq 0$ for some $m < K$. Donoho (1994), section 2.2 gives $A_1 = \left((2\beta+1)(\beta+1)/4\beta^2\right)^{\beta/(2\beta+1)}$ for $\beta = 1$, i.e. $A_1 = (3/2)^{1/3}$, thus from (7.3)

$$C = (3/2)^{1/3} \left(\frac{4K^2}{3}\right)^{1/3} = \left(2K^2\right)^{1/3}. \square$$

To obtain the analog in the diffusion model, let $r_\varepsilon(\mathcal{E}_0^\varepsilon, w)$ be the sup-norm risk corresponding to (7.1) in the diffusion experiment $\mathcal{E}_0^\varepsilon$ for the same norming sequence ψ_ε,

i.e. expectation E_f^ε is taken with respect to the law $P_f^{T,\varepsilon}$ given by (2.2). As a consequence of asymptotic equivalence (Theorem 3), we obtain $r_\varepsilon(\mathcal{E}_0^\varepsilon, w) - r_\varepsilon(\mathcal{E}_2^\varepsilon, w) \to 0$ as $\varepsilon \to 0$ (cf. Brown and Low (1996) for details on the decision theoretic meaning of the Δ-distance). An immediate consequence is the following.

Corollary 3. *For any m, K such that $0 < m < K$ and for any bounded function $w \in \mathcal{W}$ we have*

$$\lim_{\varepsilon \to 0} r_\varepsilon(\mathcal{E}_0^\varepsilon, w) = w(K^{2/3}2^{1/3}).$$

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] de la Peña, V. and Giné, E. (1999). *Decoupling: From Dependence to Independence.* Springer-Verlag, New York etc.

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

[5] Freidlin, M. I and Wentzell, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. Springer-Verlag, New York etc.

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

[7] 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

[8] Grama, I. and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized models. *Probab. Theor. Relat. Fields* ***111*** *167-214*

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

[10] Karatzas, I and Shreve, S. E. (1997). *Brownian Motion and Stochastic Calculus* (2nd ed., Graduate Texts in Mathematics, Vol 113). Springer Verlag, New York etc.

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

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

[13] Korostelev, A. P. and Nussbaum, N. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli* **5** (6) 1099-1118

[14] Kushner, H. J. and Dupuis, P. G. (1992). *Numerical Methods for Stochastic Control Problems in Continuous Time.* Springer Verlag, New York etc.

[15] Kutoyants, Yu. (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

[16] Kutoyants, Yu. (1994). *Identification of Dynamical Systems With Small Noise* (Mathematics and Its Applications, Vol 300) Kluwer Academic Publishers, Dordrecht, The Netherlands.

[17] Kutoyants, Yu. (1997). Efficiency of the empirical distribution for ergodic diffusion. *Bernoulli* **3** 445- 456

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

[19] Leonov, S. L. (1997). On the solution of an optimal recovery problem and its application in nonparametric regression. Math. Meth. Stat. **6** 476-490

[20] Lepskii, O., Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512-2546
[21] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New-York etc.
[22] Milstein, G. N. (1998). On the mean-square approximation of a diffusion process in a bounded domain. *Stochastics Stochastics Rep.* ***64*** 211-233
[23] Milstein, G. and Nussbaum, M. (1998). Diffusion approximation for nonparametric autoregression. *Probab. Theor. Relat. Fields* ***112*** *535-543*
[24] Milstein, G. N. and Tretyakov, M. V. (1999). Simulation of a space-time bounded diffusion. *Ann. Appl. Probab.* ***9*** *732-779*
[25] Musiela, M. and Rutkowski, M. (1997). *Martingale Methods in Financial Modelling.* Springer Verlag, New York etc.
[26] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399-2430
[27] Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* (1980) 120-133
[28] Reiss, R. D. (1993). *A course on point processes.* Springer Verlag, New York etc.

(V. Genon-Catalot) Université de Marne-la-Vallée, Equipe d'Analyse et de Mathématiques Appliquées, cité Descartes 5, Bd Descartes, Champs-sur-Maine, F-77454 Marne-la-Vallée, Cedex 2, France

E-mail address: genon@univ-mlv.fr

(C. Laredo) Institut National de la Recherche Agronomique, Laboratoire de Biométrie, F-78350 Jouy-en-Josas, France

E-mail address: catherine.laredo@jouy.inria.fr

(M. Nussbaum) Department of Mathematics, Malott Hall, Cornell University, Ithaca NY 14852-4201, USA

E-mail address: nussbaum@math.cornell.edu