Institut für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Autoregression approximation of a nonparametric diffusion model

G. Milstein and M. Nussbaum

submitted: 29th December 1993

Institute of Applied Analysis
and Stochastics
Mohrenstraße 39
D – 10117 Berlin
Germany

Preprint No. 78
Berlin 1993

1991 Mathematics Subject Classification. Primary 62M10; Secondary 62G07, 62B15, 60J60, 60H10.
Key words and phrases. Nonparametric experiments, deficiency distance, likelihood ratio process, stochastic differential equation, autoregression, diffusion sampling, asymptotic sufficiency.

Herausgegeben vom
Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39
D — 10117 Berlin

Fax: + 49 30 2004975
e–mail (X.400): c=de;a=d400;p=iaas-berlin;s=preprint
e–mail (Internet): preprint@iaas-berlin.d400.de

Autoregression Approximation of a Nonparametric Diffusion Model

Grigori Milstein and Michael Nussbaum
Institute of Applied Analysis and Stochastics, Berlin

December 1993

Abstract

We consider a model of small diffusion type where the function which governs the drift term varies 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 an Euler difference scheme as a discrete version of the stochastic differential equation is asymptotically equivalent in the sense of Le Cam's deficiency distance, when the discretization step decreases with the noise intensity ϵ. We thus obtain a nonparametric version of diffusion limit results for autoregression. It follows that in the continuous diffusion model, discrete sampling on a uniform grid is asymptotically sufficient. The key technical step utilizes the notion of Hellinger process from semimartingale theory.

1 Introduction and Main Result

Consider the problem of estimating the function f from an observed diffusion process $y(t),\ t \in [0,1]$, which satisfies an Ito stochastic differential equation

$$dy(t) = f(y(t))dt + \epsilon\, dW(t),\ t \in [0,1],\ y_0 = 0 \tag{1}$$

where $dW(t)$ is Gaussian white noise and ϵ is a small parameter. Suppose that the function f belongs to some a priori set Σ, nonparametric in general. Kutoyants (1985) constructed estimates of the function f the squared L_2-risk of which converges with rate $\epsilon^{4m/(2m+1)}$ if the function f has m bounded derivatives. These are the standard

1991 *Mathematics Subject Classification.* Primary 62 M 10; Secondary 62 G 07, 62 B 15, 60 J 60, 60 H 10.
Key words and phrases. Nonparametric experiments, deficiency distance, likelihood ratio process, stochastic differential equation, autoregression, diffusion sampling, asymptotic sufficiency.

nonparametric rates of convergence which also hold in the problem of 'signal recovery in Gaussian white noise'

(2) $$dx(t) = f(t)dt + \epsilon\, dW(t),\ t \in [0,1]$$

Brown and Low (1992) found that the continuous model (2) is asymptotically equivalent to its discrete counterpart which is the nonparametric regression

(3) $$x_i = f(t_i) + \epsilon n^{-1/2}\xi_i,\ i = 1,\dots,n$$

with a uniform grid $t_i = (i-1)/n$ and standard normal variables ξ_i, provided that f varies in a nonparametric subset of $L_2(0,1)$ defined by a weak smoothness type condition and n tends to infinity not too slowly. The framework was asymptotic equivalence in the sense of Le Cam's deficiency distance Δ. *In this paper we address the analogous question with respect to discretizing the stochastic differential equation model (1).*
The discretization to consider is suggested by the theory of numerical solution of stochastic differential equations, see e. g. Kloeden, Platen (1993). In that part of probability theory one assumes the function f known and one tries to approximate the solution y from a discrete difference scheme where the continuous Gaussian white noise $dW(t)$ is substituted by a sequence of i. i. d. normal random variables. The simplest such difference scheme is the one of Euler type. To derive it, observe that the process y satisfies

(4) $$y(t_{i+1}) - y(t_i) = \int_{t_i}^{t_{i+1}} f(y(u))du + \epsilon(W(t_{i+1}) - W(t_i)).$$

Now for the approximation, the integral is substituted by $n^{-1}f(x_i)$ and $W(t_{i+1})-W(t_i)$ is written $n^{-1/2}\xi_i$ where $\xi_i,\ i = 1,\dots,n$ is a sequence of i. i. d. standard normal variables. Introducing an approximate solution process y_i defined on the grid only (y_i corresponding to grid point t_i), one gets a sequence of successive approximations

(5) $$y_{i+1} = y_i + n^{-1}f(y_i) + \epsilon\, n^{-1/2}\xi_i,\ i = 1,\dots,n,\ y_0 = 0.$$

This directly extends the classical Euler scheme for approximating solutions of deterministic differential equations, with random noise added in each step to mimick the random continuous solution y of (1). It is then shown that the process y_i on the discrete grid approximates the solution y of (1) in some probabilistic sense if one considers the ξ_i as actually coming from W via $\xi_i = n^{1/2}(W(t_{i+1}) - W(t_i))$.
The statistical task may be seen as the converse: given y satisfying (1), reconstruct f. A natural question which then arises is whether inference may be based on the grid values of the solution process $y(t_i)$ only. But these values satisfying (4) still depend on the whole trajectory of y via the integral over $[t_i, t_{i+1}]$. Going a step further, one might then ask whether estimating f in (1) is equivalent to estimating f from the 'truly' discretized process y_i in (5).
Our basic strategy for comparing two models, i. e. to find, for a sequence of experiments $\{\mathcal{E}_\epsilon^{(1)}\}$, an accompanying one $\{\mathcal{E}_\epsilon^{(2)}\}$ in the sense that the full deficiency distance tends to zero: $\Delta(\mathcal{E}_\epsilon^{(1)}, \mathcal{E}_\epsilon^{(2)}) \to 0$ as $\epsilon \to 0$, was developed by Le Cam (1985) for a parametric

model of independent observations and an accompanying sequence of Gaussian shift experiments. In that case the two experiments are said to be *asymptotically equivalent* as $\epsilon \to 0$. Later in Le Cam and Yang (1990) a general method was proposed for estimating the deficieny distance Δ: consider the two likelihood processes $\Lambda_\epsilon^{(i)}(f)$, $f \in \Sigma$, $i = 1,2$ as random variables under the respective dominating measure and put them on the same probability space; then show

$$\sup_{f\in\Sigma} E\left|\Lambda_\epsilon^{(1)}(f) - \Lambda_\epsilon^{(2)}(f)\right| \to 0,\ \epsilon \to 0. \tag{6}$$

That implies that the deficiency distance between the two experiments tends to zero. This was also the method adopted by Brown and Low (1992), with the Wiener space as natural common probability space for W and ξ_i. For a treatment of the nonparametric i. i. d. density model see Nussbaum (1992).

Accordingly, in our diffusion model, consider the density for (1) when the dominating measure is the distribution of ϵW: for $z = \epsilon W$

$$\Lambda_\epsilon^{(1)}(f) = \exp\left\{\frac{1}{\epsilon^2}\int_0^1 f(z(t))dz(t) - \frac{1}{2\epsilon^2}\int_0^1 f^2(z(t))dt\right\}. \tag{7}$$

For the discrete scheme (5) the analog is for $z_i = \epsilon\ (W(t_{i+1}) - W(t_i))$

$$\Lambda_\epsilon^{(2)}(f) = \exp\left\{\frac{1}{\epsilon^2}\sum_{i=1}^n f(z_i)(z_{i+1} - z_i) - \frac{1}{2\epsilon^2 n}\sum_{i=1}^n f^2(z_i)\right\}. \tag{8}$$

To state our main theorem, let us determine the parameter space Σ for our experiments. We assume only the standard condition for existence and uniqueness of a solution y of the SDE (1) (cp. Øksendal (1992), theorem 5.5) in a uniform version: let

$$\Sigma_M = \{f \text{ defined on } R,\ |f(x) - f(u)| \le M\ |x-u|,\ \ x,u \in R,\ |f(0)| \le M\}.$$

Observe that $f \in \Sigma_M$ implies the condition of linear growth which is commonly encountered in this connection: $|f(x)| \le M(1+|x|)$.

Theorem 1 *Suppose that for some $M > 0$ the parameter space Σ fulfills $\Sigma \subset \Sigma_M$, and that $n = n_\epsilon$ is chosen such that $\epsilon\ n_\epsilon \to \infty$. Then the experiments given by (1) and (5) are asymptotically equivalent as $\epsilon \to 0$.*

REMARK 1. The model (5) is of the *autoregressive* type, and corresponding diffusion limits have been studied extensively in parametric models. To see the connection, consider the case where the parameter space is

$$\Sigma = \{f,\ f(x) = \vartheta\ x,\ |\vartheta| \le M\} \subset \Sigma_M.$$

Thus we have the parametric model

$$dy(t) = \vartheta y(t)\ dt + \epsilon\ dW(t),\ t \in [0,1] \tag{9}$$

which has been investigated predominantly with an increasing interval of observation (see Kutoyants (1984), § 3.5). In our case of fixed interval and varying ϵ all experiments

(9) for different ϵ are equivalent (exactly). Indeed, multiplying the observations y by ϵ^{-1} yields an equivalent experiment, and the process $\tilde{y} = \epsilon^{-1}y$ satisfies (9) for $\epsilon = 1$. Thus our accompanying sequence of experiments is indeed constant, and theorem 1 for $n = \epsilon^{-2}$ establishes a *strong* ($\Delta-$) *convergence* of the autoregression experiments

$$y_{i+1} = y_i + n^{-1}\vartheta y_i + n^{-1}\xi_i, \ i = 0, \ldots, n, \ y_0 = 0 \tag{10}$$

to a diffusion limit (9). Define $\tilde{y}_i = ny_i$ and $\beta = 1 + n^{-1}\vartheta$; then (10) may be written

$$\tilde{y}_{i+1} = \beta \ \tilde{y}_i + \xi_i, \ i = 0, \ldots, n, \ y_0 = 0 \tag{11}$$

which is the familiar AR(1) model in the *nearly nonstationary* case where the parameter β is close to 1. Parametric inference in these models based on the diffusion limit has been studied by Chan and Wei (1987), Cox (1991); see also comments in Jeganathan (1988). The corresponding limit experiment argument is given in the forthcoming monograph of Shiryaev and Spokoiny (1993), based on the notion of λ-convergence of experiments. Thus theorem 1 appears as a nonparametric extension of the parametric diffusion limit results for nearly nonstationary AR(1) autoregression.

REMARK 2. It is now clear what the relation to *nonparametric autoregression* should be: define $\tilde{y}_i$ as above and a function $g(x) = x + f(n^{-1}x)$; then (5) may be written

$$\tilde{y}_{i+1} = g(\tilde{y}_i) + \xi_i, \ i = 1, \ldots, n, \ y_0 = 0. \tag{12}$$

Nonparametric inference for fixed, unknown g was studied by Doukhan and Ghindès (1983) under stationarity assumptions. They found that the theory of kernel type estimators parallels the signal plus white noise case (2), as regards rates of convergence. So obviously the nonparametric model (12) with g fixed and stationarity corresponds to parametric autoregression in the stable case where β is bounded away from 1 (and local asymptotic normality holds), while in the nearly critical case $g(x) = x + f(n^{-1}x)$, $f \in \Sigma$ the diffusion approximation of theorem 1 holds.

REMARK 3. Parametric asympotic results for autoregressive models are naturally available for nonnormal ξ_i, based on the limit experiment rationale. In particular the Gaussian diffusion limit in the nearly nonstationary case is known. This raises the question whether theorem 1 might be valid also for nongaussian ξ_i in the autoregression. The problem is open even for the regression (3), but the i. i. d. density case result (Nussbaum (1992)) does suggest a positive answer.

Theorem 1 is proved with a reasoning related to the one of Brown and Low (1992) in the 'signal plus noise' case. The two models (1) and (5), even when they are construed as being on the same probability space, are still of different type: one is discrete, the other continuous. To facilitate the treatment, it is convenient to compare the likelihoods of two continuous models. That is achieved by *interpolation* of the discrete process y_i, $i = 1, \ldots, n$. Recall the classical Euler scheme for solving the (deterministic) differential equation

$$\frac{d}{dt}y^0(t) = f(y^0(t)),\ t \in [0,1] \tag{13}$$

which is (5) without the random term, i. e.

$$n(y^0_{i+1} - y^0_i) = f(y^0_i),\ i = 1, \ldots, n,\ y^0_0 = 0. \tag{14}$$

The rationale is of course closely linked to linear interpolation: if $\bar{y}^0$ is a piecewise linear function on $[0,1]$ such that $\bar{y}^0(t_i) = y_i$ then its derivative on $[t_i, t_{i+1}]$ ís $f(\bar{y}^0(t_i))$. Ordinarily the linear interpolation $\bar{y}^0$ of the values y^0_i is indeed taken as the approximate solution of the differential equation.
For the stochastic Euler scheme (5) one might also consider a linear interpolation of the values y_i. But in order to account for the oscillating behaviour of the solution y of (1) inherited from the driving Wiener process, one might then *add a Brownian bridge* over and above the linear interpolation on each interval $[t_i, t_{i+1}]$. These Brownian bridges should be independent of the ξ_i; call this randomly interpolated process $\bar{y}$.
A convenient representation of $\bar{y}$ can be obtained as follows. Define a function $\bar{f}_n$ on $[0,1]$ which depends on a trajectory $z(t),\ t \in [0,1]$ as

$$\bar{f}_n(t,z) = \sum_{i=1}^{n} f(z(t_i))\ \chi_{(t_i,t_{i+1}]}(t)$$

Thus $\bar{f}_n$ is a piecewise constant function which interpolates $f(z(\cdot))$ in t_i.

Lemma 1 *The unique solution $\bar{y}$ of*

$$\bar{y}(t) = \int_0^t \bar{f}_n(u, \bar{y})\ du + \epsilon\ W(t) \tag{15}$$

may be represented as a randomly interpolated process as above, for an appropriate choice of the n Brownian bridges.

The proof is in section 3. It is clear that $\bar{y}$ contains as much information about f as $y_i,\ i = 1, \ldots, n$. Indeed $\bar{y}(t_i) = y_i$, but on the other hand linear interpolation and adding Brownian bridges which do not depend on f do not increase the informational content. Thus the experiments given by (5) and (15) are equivalent, for any parameter space Σ, and $\bar{y}(t_i),\ i = 1, \ldots, n$ is a sufficient statistic in (15). This can also be seen by looking at the likelihood for model (15) as follows.
Regard $\bar{y}$ formally as a diffusion type process defined by (15), see Liptser, Shiryaev, (1974), Chap. 4, §2, Definition 7. Indeed $\bar{f}_n(u, \bar{y})$ is for each u a non-anticipating functional, since it depends on $\bar{y}$ only via $\bar{y}(t_i),\ u \in (t_i, t_{i+1}]$. (But $\bar{y}$ is not a diffusion process in the strict sense, which would mean that $\bar{f}_n(u, \bar{y})$ depends $\bar{y}$ only via $\bar{y}(u)$). This process has a distribution which is absolutely continuous with respect to the distribution of $\epsilon\ W$ if almost surely $\int_0^1 \bar{f}_n^2(u, \bar{y})du < \infty$ (Liptser, Shiryaev, (1974), Chap. 7, §2, Theorem 7.6). But this is fulfilled since every value y_i from (5) is finite. Then the density is analogously to (7) for $z = \epsilon W$

$$\Lambda_\epsilon^{(3)}(f) = \exp\left\{ \frac{1}{\epsilon^2} \int_0^1 \bar{f}_n(t,z)\ dz(t) - \frac{1}{2\epsilon^2} \int_0^1 \bar{f}_n^2(t,z)\ dt \right\}. \tag{16}$$

Now observe

$$\int_0^1 \bar{f}_n(t,z)\, dz(t) = \sum_{i=1}^{n} f(z_i)(z_{i+1} - z_i),\; \int_0^1 \bar{f}_n^2(t,z)\, dt = \sum_{i=1}^{n} f^2(z_i) n^{-1}$$

so that we obtain $\Lambda_\epsilon^{(3)}(f) = \Lambda_\epsilon^{(2)}(f)$. This means that the density (16) depends on z via $z_i = z(t_{i+1}) - z(t_i)$ only, which again implies that the values $\bar{y}(t_i)$, $i = 1, \ldots, n$ are a sufficient statistic in (15) and the experiments given by (5) and (15) are equivalent, for any parameter space Σ. It remains to establish that the densities of the two diffusion type processes (1) and (15) fulfill

$$\sup_{f \in \Sigma} E \left| \Lambda_\epsilon^{(1)}(f) - \Lambda_\epsilon^{(3)}(f) \right| \to 0,\; \epsilon \to 0. \tag{17}$$

For the signal plus noise case (2) Brown and Low (1992) were able to use the explicit formula for the squared Hellinger distance $H^2(\cdot,\cdot)$ of two shifted Wiener measures: if Q_f, Q_{f^*} are two measures corresponding to (2) with trends f, f^*, say, then

$$H^2(Q_f, Q_{f^*}) = 2 - 2\exp\{-\frac{1}{8}\epsilon^{-2}\, \|f - f^*\|^2\} \tag{18}$$

(in analogy to the finite dimensional Gaussian case), and where f^* was also a step function approximant for f. In our more involved diffusion model, we use an inequality derived from the *Hellinger process* in semimartingale theory, cf. Jacod and Shiryaev (1987). Denote by P_f, P_f^* the probability measures on $C[0,1]$ given by the two processes (1) and (15), respectively; the notation P_f^* reflects the dependence of $\bar{f}_n(\cdot,\cdot)$ in (15) on f. Let h_f be the Hellinger process of order 1/2 between these two measures (see Jacod and Shiryaev (1987), §4b, p. 239): for a realization z

$$h_f(u) = \frac{1}{8\epsilon^2} \int_0^u (f(z(t)) - \bar{f}_n(t,z))^2 dt.$$

The inequality for the the total variation distance $\|\cdot\|_{TV}$ is

$$\left\| P_f - P_f^* \right\|_{TV} \left(= E \left| \Lambda_\epsilon^{(1)}(f) - \Lambda_\epsilon^{(3)}(f) \right| \right) \le 4\sqrt{E_f\; h_f(1)}, \tag{19}$$

where E_f denotes expectation wrt P_f, see Jacod and Shiryaev (1987), §4b, theorem 4.21, p. 279. A remarkable feature of inequality (19) is that it is nonsymmetric and holds both ways with expectation taken with respect to either P_f or P_f^*. The choice P_f somewhat facilitates evaluation of $E_f\; h_f$ which is our next task.

Lemma 2 *Under the conditions of theorem 1 we have*

$$E_f \int_0^1 (f(y(t)) - \bar{f}_n(t,y))^2 dt = o(\epsilon^2),\; \epsilon \to 0$$

uniformly over $f \in \Sigma$, where E_f denotes expectation with respect to the distribution P_f of the process y in (1).

This lemma establishes theorem 1; the proof is in section 3.

2 Sampling from a Diffusion

Let us now consider the situation where one has discrete data from the diffusion process y in (1). Various questions of inference based on sampled values $y(t_i), i = 1, \ldots, n$ have been treated in the literature: Prakasa Rao (1983), Dacunha-Castelle and Florens-Zmirou (1986), Florens-Zmirou (1989), Yoshida (1992). The question now arises whether these sampled values constitute an asymptotically equivalent experiment. Our answer will be based on the notion of an *asymptotically sufficient statistic.*
Consider experiments $\mathcal{E}_\epsilon^{(i)} = \{P_{f,\epsilon}^{(i)},\ f \in \Sigma\}$, $i = 1,2$, given on the same measurable space $(\Omega_\epsilon, \mathcal{F}_\epsilon)$, and suppose that T_ϵ is a sequence of measurable mappings from $(\Omega_\epsilon, \mathcal{F}_\epsilon)$ to some other measurable space $(\mathcal{T}_\epsilon, \mathcal{B}_\epsilon)$.

Definition 1 *The sequence $\{T_\epsilon\}$ is called asymptotically sufficient for the sequence of experiments $\mathcal{E}_\epsilon^{(1)}$ if*
i) T_ϵ is sufficient in the experiment $\mathcal{E}_\epsilon^{(2)}$,
ii) $\mathcal{E}_\epsilon^{(2)}$ approximates $\mathcal{E}_\epsilon^{(1)}$ in total variation: $\sup_{f\in\Sigma} \left\| P_{f,\epsilon}^{(1)} - P_{f,\epsilon}^{(2)} \right\|_{TV} \to 0$ *as* $\epsilon \to 0$.

It then immediately follows that in the experiment $\mathcal{E}_\epsilon^{(1)}$, data reduction by the statistic T_ϵ constitutes an asymptotically equivalent experiment. Indeed let $P_{f,\epsilon}^{(i)T}$ be the distribution of T_ϵ under $P_{f,\epsilon}^{(i)}$ and $\mathcal{E}_\epsilon^{(i)T} = \{P_{f,\epsilon}^{(i)T},\ f \in \Sigma\}$ be the corresponding experiment. Then

$$\left\| P_{f,\epsilon}^{(1)T} - P_{f,\epsilon}^{(2)T} \right\|_{TV} = \sup_{A\in\mathcal{B}_\epsilon} \left| P_{f,\epsilon}^{(1)T}(A) - P_{f,\epsilon}^{(2)T}(A) \right| = \sup_{A\in\mathcal{B}_\epsilon} \left| P_{f,\epsilon}^{(1)}(T_\epsilon^{-1}(A)) - P_{f,\epsilon}^{(2)}(T_\epsilon^{-1}(A)) \right|$$

$$\leq \sup_{B\in\mathcal{F}_\epsilon} \left| P_{f,\epsilon}^{(1)}(B) - P_{f,\epsilon}^{(2)}(B) \right| = \left\| P_{f,\epsilon}^{(1)} - P_{f,\epsilon}^{(2)} \right\|_{TV} \to 0 \text{ as } \epsilon \to 0, \text{ uniformly over } f \in \Sigma$$

so that $\mathcal{E}_\epsilon^{(i)T}$, $i = 1,2$ are asymptotically equivalent. But by sufficiency $\mathcal{E}_\epsilon^{(2)T}$ is exactly equivalent to $\mathcal{E}_\epsilon^{(2)}$.

Theorem 2 *Under the conditions of theorem 1, in the diffusion model (1) with $\epsilon \to 0$ the sampled values $y(t_1), \ldots, y(t_n)$ are an asymptotically sufficient statistic.*

Proof. Let $\mathcal{E}_\epsilon^{(1)}$ be the experiment given by observations y in (1) and $\mathcal{E}_\epsilon^{(2)}$ be given by $\bar{y}$ in (15). Define for a trajectory $z(t)$, $t \in [0,1]$ the statistic T_ϵ by $T_\epsilon(z) = (z(t_1), \ldots, z(t_n))$. Then T_ϵ is sufficient in $\mathcal{E}_\epsilon^{(2)}$, and the result follows from theorem 1. □

3 Proofs of lemmas

Proof of lemma 1. Recall that $\xi_i = n^{1/2}(W(t_{i+1}) - W(t_i))$; define

$$B_i(t) = W(t) - W(t_i) - n(t - t_i)\ (W(t_{i+1}) - W(t_i)) \text{ for } t \in [t_i, t_{i+1}]$$

and $B_i(t) = 0$ otherwise; thus $B_i(t)$ is a Brownian bridge on $[t_i, t_{i+1}]$ which is independent of ξ_i. Hence the $B_i(t)$, $i = 1, \ldots, n$ are mutually independent and independent of the ξ_i, $i = 1, \ldots, n$. Now we have $\bar{y}(0) = 0$ by definition of the interpolated process and also by (15). For $t \in [t_1, t_2] = [0, t_2]$ we have $\bar{f}_n(t, \bar{y}) = 0$, hence (15) implies

$$\bar{y}(t) = \epsilon\, W(t) = \epsilon\, B_1(t) + \epsilon\, n(t - t_i)\, W(t_i) = \epsilon\, B_1(t) + (t - t_i)\, \epsilon\, n^{1/2} \xi_1 \tag{20}$$

which is exactly the linear interpolation of 0 and y_1 on $[0, t_1]$ plus a Brownian bridge. Thus $\bar{y}(t_2) = y_2$, and for the induction step we may assume $\bar{y}(t_i) = y_i$. Then on $[t_i, t_{i+1}]$ we get similarly to (20)

$$\bar{y}(t) = y_i + \int_{t_i}^{t} \bar{f}_n(u, \bar{y})\, du + \epsilon\, (W(t) - W(t_i)) = y_i + (t - t_i)(f(y_i) + \epsilon\, n^{1/2} \xi_i) + \epsilon\, B_i(t).$$

Hence $\bar{y}(t_{i+1}) = y_i + n^{-1} f(y_i) + \epsilon\, n^{-1/2} \xi_i = y_{i+1}$ and the solution $\bar{y}(t)$ is again of the structure claimed.

Proof of lemma 2. We have

$$\int_0^1 (f(y(t)) - \bar{f}_n(t, y))^2 dt = \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} (f(y(t)) - f(y(t_i))^2\, dt$$

$$\leq \sum_{i=1}^{n} n^{-1} \sup_{t \in (t_i, t_{i+1}]} M\, |(y(t) - y(t_i)|^2$$

$$\leq 2n^{-1} \sum_{i=1}^{n} \sup_{t \in (t_i, t_{i+1}]} \left| \int_{t_i}^{t} f(y(u))\, du \right|^2 + 2n^{-1} \sum_{i=1}^{n} \sup_{t \in (t_i, t_{i+1}]} \epsilon^2 \left| W(t) - W(t_i) \right|^2 . \tag{21}$$

For the first term in (21), apply Cauchy-Schwartz and $|f(x)| \leq M(1 + |x|)$ to obtain

$$n^{-1} \sum_{i=1}^{n} \sup_{t \in (t_i, t_{i+1}]} \left| \int_{t_i}^{t} f(y(u))\, du \right|^2 \leq n^{-1} \sum_{i=1}^{n} n^{-1} \int_{t_i}^{t_{i+1}} f^2(y(u))\, du$$

$$\leq 2n^{-2} \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} M^2 (1 + y^2(u)) du. \tag{22}$$

To estimate the expectation of this term, we use the following relation: there is a constant C_M depending on M but not on ϵ such that for all $f \in \Sigma$

$$E_f\, y^2(t) \leq C_M,\ t \in [0, 1].$$

Indeed this follows from Øksendal (1992), exercise 5.6, which is connected with the existence and uniqueness theorem for SDE (note that in (1) we have an SDE with diffusion coefficient $\sigma(t, y) = \epsilon$). Taking an expectation in (22) we get

$$E_f n^{-2} \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} M^2 (1 + y^2(u)) du \leq n^{-2} M^2 (1 + C_M). \tag{23}$$

We have estimated the expectation of the first term in (21) as $O(n^{-2})$ uniformly. For the second term, note that this is an average of identically distributed random variables, which have the distribution of

$$\sup_{t\in[0,n^{-1})} \epsilon^2 |W(t)|^2 \simeq \epsilon^2 n^{-1} \sup_{t\in[0,1)} |W(t)|^2$$

where '$\simeq$' means equality in law. But $\sup_{t\in[0,1)} |W(t)|^2$ has finite expectation (Breiman (1968), ch. 13.7), so that the expectation of the second term in (21) is also $O(n^{-2})$ uniformly. Since $n^{-2} = o(\epsilon^{-2})$ by assumption, the proof is complete.

ACKNOWLEDGEMENT. The authors wish to thank Vladimir Spokoiny and Philippe Vieu for helpful discussions.

References

[1] Breiman, L. (1968). *Probability.* Addison-Wesley, New York.

[2] Brown, L. D. and Low, M. (1992). Asymptotic equivalence of nonparametric regression and white noise. Mss.

[3] Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15** 1050-1063.

[4] Cox, D. D. (1991). Gaussian likelihood estimation for nearly nonstationary AR(1) processes. *Ann. Statist.* **19** 1129-1142

[5] Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19** 263-284

[6] Doukhan, P. and Ghindès, M. (1983). Estimation de la probabilité de transition d'une chaîne de Markov Doëblin-récurrente. Etude du cas du processus autorégressif général d'ordre 1. *Stochastic Processes Appl.* **15** 271-292

[7] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **20** (4) 547-557

[8] Jacod, J., Shiryaev, A. (1987). *Limit theorems for Stochastic Processes.* Springer-Verlag, Berlin, Heidelberg, New York.

[9] Jeganathan, P. (1988). On the strong approximation of the distribution of estimators in linear stochastic models, I, II: Stationary and explosive AR models. *Ann. Statist.* **16** 1283-1314

[10] Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, Berlin, Heidelberg, New York.

[11] Kutoyants, Yu. A. (1984). *Parameter Estimation for Stochastic Processes.* Heldermann, Berlin

[12] 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

[13] Liptser, R. S., Shiryaev, A. (1977). *Statistics of Stochastic Processes.* Springer-Verlag, Berlin, Heidelberg, New York.

[14] Le Cam, L. (1985). Sur l'approximation de familles de mesures par des familles gaussiennes. *Ann. Inst. Henri Poincaré* **21** (3) 225-287

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

[16] Le Cam , L. and Yang, G. (1990). *Asymptotics in Statistics.* Springer-Verlag, Berlin, Heidelberg, New York.

[17] Nussbaum, M. (1992). Asymptotic equivalence of density estimation and white noise. Preprint No. 35, Institute of Applied Analysis and Stochastics (IAAS), Berlin.

[18] Øksendal, B. (1992). *Stochastic Differential Equations.* Springer-Verlag, Berlin, Heidelberg, New York.

[19] Prakasa Rao, B. L. S. (1983). Statistical inference from sampled data for stochastic processes. *Contemporary Mathematics*, Vol. 80, 249-284. Amer. Math. Soc., Providence.

[20] Shiryaev, A. N., Spokoiny, V. (1993). *Statistical Experiments and Decisions: Asymptotic Theory.* To appear.

[21] Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. J. Multivariate Analysis **41**, 220-242

INSTITUTE OF APPLIED ANALYSIS AND STOCHASTICS
MOHRENSTR. 39
D-10117 BERLIN
GERMANY
E-MAIL MILSTEIN@IAAS-BERLIN.D400.DE;
NUSSBAUM@IAAS-BERLIN.D400.DE

Veröffentlichungen des Instituts für Angewandte Analysis und Stochastik

Preprints 1992

1. D.A. Dawson, J. Gärtner: Multilevel large deviations.

2. H. Gajewski: On uniqueness of solutions to the drift–diffusion–model of semiconductor devices.

3. J. Fuhrmann: On the convergence of algebraically defined multigrid methods.

4. A. Bovier, J.-M. Ghez: Spectral properties of one–dimensional Schrödinger operators with potentials generated by substitutions.

5. D.A. Dawson, K. Fleischmann: A super–Brownian motion with a single point catalyst.

6. A. Bovier, V. Gayrard: The thermodynamics of the Curie–Weiss model with random couplings.

7. W. Dahmen, S. Prößdorf, R. Schneider: Wavelet approximation methods for pseudodifferential equations I: stability and convergence.

8. A. Rathsfeld: Piecewise polynomial collocation for the double layer potential equation over polyhedral boundaries. Part I: The wedge, Part II: The cube.

9. G. Schmidt: Boundary element discretization of Poincaré–Steklov operators.

10. K. Fleischmann, I. Kaj: Large deviation probability for some rescaled superprocesses.

11. P. Mathé: Random approximation of finite sums.

12. C.J. van Duijn, P. Knabner: Flow and reactive transport in porous media induced by well injection: similarity solution.

13. G.B. Di Masi, E. Platen, W.J. Runggaldier: Hedging of options under discrete observation on assets with stochastic volatility.

14. J. Schmeling, R. Siegmund–Schultze: The singularity spectrum of self-affine fractals with a Bernoulli measure.

15. A. Koshelev: About some coercive inequalities for elementary elliptic and parabolic operators.

16. P.E. Kloeden, E. Platen, H. Schurz: Higher order approximate Markov chain filters.

17. H.M. Dietz, Y. Kutoyants: A minimum–distance estimator for diffusion processes with ergodic properties.

18. I. Schmelzer: Quantization and measurability in gauge theory and gravity.

19. A. Bovier, V. Gayrard: Rigorous results on the thermodynamics of the dilute Hopfield model.

20. K. Gröger: Free energy estimates and asymptotic behaviour of reaction–diffusion processes.

21. E. Platen (ed.): Proceedings of the 1st workshop on stochastic numerics.

22. S. Prößdorf (ed.): International Symposium "Operator Equations and Numerical Analysis" September 28 – October 2, 1992 Gosen (nearby Berlin).

23. K. Fleischmann, A. Greven: Diffusive clustering in an infinite system of hierarchically interacting diffusions.

24. P. Knabner, I. Kögel–Knabner, K.U. Totsche: The modeling of reactive solute transport with sorption to mobile and immobile sorbents.

25. S. Seifarth: The discrete spectrum of the Dirac operators on certain symmetric spaces.

26. J. Schmeling: Hölder continuity of the holonomy maps for hyperbolic basic sets II.

27. P. Mathé: On optimal random nets.

28. W. Wagner: Stochastic systems of particles with weights and approximation of the Boltzmann equation. The Markov process in the spatially homogeneous case.

29. A. Glitzky, K. Gröger, R. Hünlich: Existence and uniqueness results for equations modelling transport of dopants in semiconductors.

30. J. Elschner: The h-p–version of spline approximation methods for Mellin convolution equations.

31. R. Schlundt: Iterative Verfahren für lineare Gleichungssysteme mit schwach besetzten Koeffizientenmatrizen.

32. G. Hebermehl: Zur direkten Lösung linearer Gleichungssysteme auf Shared und Distributed Memory Systemen.

33. G.N. Milstein, E. Platen, H. Schurz: Balanced implicit methods for stiff stochastic systems: An introduction and numerical experiments.

34. M.H. Neumann: Pointwise confidence intervals in nonparametric regression with heteroscedastic error structure.

35. M. Nussbaum: Asymptotic equivalence of density estimation and white noise.

Preprints 1993

36. B. Kleemann, A. Rathsfeld: Nyström's method and iterative solvers for the solution of the double layer potential equation over polyhedral boundaries.

37. W. Dahmen, S. Prössdorf, R. Schneider: Wavelet approximation methods for pseudodifferential equations II: matrix compression and fast solution.

38. N. Hofmann, E. Platen, M. Schweizer: Option pricing under incompleteness and stochastic volatility.

39. N. Hofmann: Stability of numerical schemes for stochastic differential equations with multiplicative noise.

40. E. Platen, R. Rebolledo: On bond price dynamics.

41. E. Platen: An approach to bond pricing.

42. E. Platen, R. Rebolledo: Pricing via anticipative stochastic calculus.

43. P.E. Kloeden, E. Platen: Numerical methods for stochastic differential equations.

44. L. Brehmer, A. Liemant, I. Müller: Ladungstransport und Oberflächenpotentialkinetik in ungeordneten dünnen Schichten.

45. A. Bovier, C. Külske: A rigorous renormalization group method for interfaces in random media.

46. G. Bruckner: On the regularization of the ill–posed logarithmic kernel integral equation of the first kind.

47. H. Schurz: Asymptotical mean stability of numerical solutions with multiplicative noise.

48. J.W. Barrett, P. Knabner: Finite element approximation of transport of reactive solutes in porous media. Part I: Error estimates for non–equilibrium adsorption processes.

49. M. Pulvirenti, W. Wagner, M.B. Zavelani Rossi: Convergence of particle schemes for the Boltzmann equation.

50. J. Schmeling: Most β shifts have bad ergodic properties.

51. J. Schmeling: Self normal numbers.

52. D.A. Dawson, K. Fleischmann: Super–Brownian motions in higher dimensions with absolutely continuous measure states.

53. A. Koshelev: Regularity of solutions for some problems of mathematical physics.

54. J. Elschner, I.G. Graham: An optimal order collocation method for first kind boundary integral equations on polygons.

55. R. Schlundt: Iterative Verfahren für lineare Gleichungssysteme auf Distributed Memory Systemen.

56. D.A. Dawson, K. Fleischmann, Y. Li, C. Müller: Singularity of super–Brownian local time at a point catalyst.

57. N. Hofmann, E. Platen: Stability of weak numerical schemes for stochastic differential equations.

58. H.G. Bothe: The Hausdorff dimension of certain attractors.

59. I.P. Ivanova, G.A. Kamenskij: On the smoothness of the solution to a boundary value problem for a differential–difference equation.

60. A. Bovier, V. Gayrard: Rigorous results on the Hopfield model of neural networks.

61. M.H. Neumann: Automatic bandwidth choice and confidence intervals in nonparametric regression.

62. C.J. van Duijn, P. Knabner: Travelling wave behaviour of crystal dissolution in porous media flow.

63. J. Förste: Zur mathematischen Modellierung eines Halbleiterinjektionslasers mit Hilfe der Maxwellschen Gleichungen bei gegebener Stromverteilung.

64. A. Juhl: On the functional equations of dynamical theta functions I.

65. J. Borchardt, I. Bremer: Zur Analyse großer strukturierter chemischer Reaktionssysteme mit Waveform–Iterationsverfahren.

66. G. Albinus, H.-Ch. Kaiser, J. Rehberg: On stationary Schrödinger–Poisson equations.

67. J. Schmeling, R. Winkler: Typical dimension of the graph of certain functions.

68. A.J. Homburg: On the computation of hyperbolic sets and their invariant manifolds.

69. J.W. Barrett, P. Knabner: Finite element approximation of transport of reactive solutes in porous media. Part 2: Error estimates for equilibrium adsorption processes.

70. H. Gajewski, W. Jäger, A. Koshelev: About loss of regularity and "blow up" of solutions for quasilinear parabolic systems.

71. F. Grund: Numerical solution of hierarchically structured systems of algebraic–differential equations.

72. H. Schurz: Mean square stability for discrete linear stochastic systems.

73. R. Tribe: A travelling wave solution to the Kolmogorov equation with noise.

74. R. Tribe: The long term behavior of a Stochastic PDE.

75. A. Glitzky, K. Gröger, R. Hünlich: Rothe's method for equations modelling transport of dopants in semiconductors.

76. W. Dahmen, B. Kleemann, S. Prößdorf, R. Schneider: A multiscale method for the double layer potential equation on a polyhedron.

77. H.G. Bothe: Attractors of non invertible maps.