Probab. Theory Relat. Fields **112**, 535–543 (1998)

Probability Theory and Related Fields

Diffusion approximation for nonparametric autoregression

Grigori Milstein, Michael Nussbaum

Weierstrass Institute, Mohrenstr. 39, D-10117 Berlin, Germany

Received: 12 April 1996 / Revised version: 29 October 1997

Abstract. A nonparametric statistical model of small diffusion type is compared with its discretization by a stochastic Euler difference scheme. It is shown that the discrete and continuous models are asymptotically equivalent in the sense of Le Cam's deficiency distance for statistical experiments, when the discretization step decreases with the noise intensity ϵ.

Mathematics Subject Classification (1991): Primary 62M10; Secondary 62G07, 62B15, 60J60, 60H10.

1 Introduction

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

$$dy(t) = f(y(t))dt + \epsilon\, dW(t), \quad t \in [0,1], \quad y(0) = 0 \qquad (1)$$

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 [12] constructed estimates of the function f the squared L_2-risk of which decreases with rate $\epsilon^{4m/(2m+1)}$ if the function f has m bounded derivatives. These are the standard

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'

$$dx(t) = f(t)\,dt + \epsilon\, dW(t), \quad t \in [0,1] \ . \tag{2}$$

Brown and Low [3] found that the continuous model (2) is asymptotically equivalent to its discrete counterpart, i.e. the nonparametric regression

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

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 moderate smoothness 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 Euler scheme for solving a stochastic differential equation. Introducing an approximate solution process y_i defined on a grid t_i, $i = 1,\dots,n$ only (y_i corresponding to grid point t_i), one gets a sequence of successive approximations

$$y_{i+1} = y_i + n^{-1}f(y_i) + \epsilon\, n^{-1/2}\xi_i, \quad i = 1,\dots,n, \quad y_1 = 0 \ . \tag{4}$$

It is then shown that the process y_i on the discrete grid approximates the solution y of (1) in some probabilistic sense as $n \to \infty$.

For statistical inference about f given y from (1), a natural question is whether inference may be based on the grid values of the solution process $y(t_i)$ only. But these values still depend on the whole path 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 discrete process y_i in (4).

Our strategy for comparing two models is based on the following basic concepts from asymptotic decision theory. Suppose we have two sequences of experiments $\mathbb{E}_{i,\epsilon} = (\Omega_{i,\epsilon}, \mathscr{B}_{i,\epsilon}, (P_{i,f,\epsilon}, f \in \Sigma))$, $i = 1,2$, $\epsilon \to 0$ having the same parameter space Σ. The *deficiency pseudodistance* $\Delta(\mathbb{E}_{1,\epsilon}, \mathbb{E}_{2,\epsilon})$ is always defined; $\mathbb{E}_{1,\epsilon}$ and $\mathbb{E}_{2,\epsilon}$ are *equivalent* or *of the same type* if $\Delta(\mathbb{E}_{1,\epsilon}, \mathbb{E}_{2,\epsilon}) = 0$ (Le Cam [14], chap. 2.3). Some further related concepts are the following.

Definition. (i) *Two sequences* $\mathbb{E}_{1,\epsilon}$, $\mathbb{E}_{2,\epsilon}$ *of experiments on a common measurable space* $(\Omega_{1,\epsilon}, \mathscr{B}_{1,\epsilon}) = (\Omega_{2,\epsilon}, \mathscr{B}_{2,\epsilon})$ *are asymptotically total variation equivalent if*

$$\sup_{f\in\Sigma}\left\|P_{1,\epsilon,f}-P_{2,\epsilon,f}\right\|_{TV}\to 0 \quad as\ \epsilon\to 0 \ .$$

where $\|\cdot\|_{TV}$ *is the total variation distance of measures.*

(ii) *For sequences* $\mathbb{E}_{1,\epsilon}$, $\mathbb{E}_{2,\epsilon}$ *as in* (i), *let* T_ϵ *be a sequence of statistics on* $(\Omega_{1,\epsilon},\mathscr{B}_{1,\epsilon})$ *with values in measurable spaces* $(X_\epsilon,\mathscr{F}_\epsilon)$. *The sequence* T_ϵ *is asymptotically sufficient in* $\mathbb{E}_{1,\epsilon}$ *if* T_ϵ *is sufficient in* $\mathbb{E}_{2,\epsilon}$ *and* $\mathbb{E}_{1,\epsilon},\mathbb{E}_{2,\epsilon}$ *are asymptotically total variation equivalent.*

(iii) *Two arbitrary sequences* $\mathbb{E}_{1,\epsilon},\mathbb{E}_{2,\epsilon}$ *are asymptotically equivalent in the sense of Le Cam if*

$$\Delta(\mathbb{E}_{1,\epsilon},\mathbb{E}_{2,\epsilon})\to 0 \quad as\ \epsilon\to 0 \ .$$

Asymptotic total variation equivalence implies asymptotic equivalence in the Le Cam sense. In the latter case, we also call $\mathbb{E}_{1,\epsilon},\mathbb{E}_{2,\epsilon}$ *accompanying sequences*; if all the $\mathbb{E}_{2,\epsilon}$, $\epsilon>0$ are of the same type we recover the notion of limit of experiments: the type of $\mathbb{E}_{1,\epsilon}$ has a (strong) limit. Henceforth the term asymptotic equivalence will be used in the sense of (iii).

The concept $\Delta(\mathbb{E}_{1,\epsilon},\mathbb{E}_{2,\epsilon})\to 0$ in statistical theory has found many applications for the case of limits of experiment types, in statistical models after localization (Le Cam [14]). For recent nonlocal results see Brown and Low [3], Nussbaum [17], Grama and Nussbaum [8].

2 Methodology and main result

Our basic method is coupling of likelihood processes; see Nussbaum [17], sect. 2 for details. Suppose each experiment $\mathbb{E}_i$, $i=1,2$ is dominated by a probability measure Π_i on $(\Omega_i,\mathscr{B}_i)$. Consider $\Lambda_i(f)=dP_{i,f}/d\Pi_i$ as a stochastic process indexed by f under Π_i. Now find versions Λ_i^* of Λ_i on a common probability space $(\Omega^*,\mathscr{B}^*,\Pi^*)$. Define probability measures $P_{i,f}^*$ by $dP_{i,f}^*=\Lambda_i^*(f)\,d\Pi^*$ and define experiments $\mathbb{E}_i^*=(\Omega^*,\mathscr{B}^*,(P_{i,f}^*,f\in\Sigma))$. If everything depends on ϵ, for asymptotic equivalence of $\mathbb{E}_{i,\epsilon}$ it then suffices that $\mathbb{E}_{i,\epsilon}^*$ are asymptotically total variation equivalent, i.e. that

$$\sup_{f\in\Sigma}E_{\Pi^*}\left|\Lambda_{1,\epsilon}^*(f)-\Lambda_{2,\epsilon}^*(f)\right|\to 0, \quad \epsilon\to 0 \ . \tag{5}$$

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

$$\Lambda_{1,\epsilon}(f)(z)=\exp\left\{\frac{1}{\epsilon^2}\int_0^1 f(z(t))\,dz(t)-\frac{1}{2\epsilon^2}\int_0^1 f^2(z(t))\,dt\right\} \ . \tag{6}$$

For the discrete scheme (4) the analog is for $z_i = \epsilon\ W(t_i)$

$$\Lambda_{2,\epsilon}(f)(z) = \exp\left\{ \frac{1}{\epsilon^2} \sum_{i=1}^{n} f(z_i)(z_{i+1} - z_i) - \frac{1}{2\epsilon^2 n} \sum_{i=1}^{n} f^2(z_i) \right\} \ . \qquad (7)$$

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 [18], 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: $|f(x)| \le M(1 + |x|)$.

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* (4) are asymptotically equivalent as $\epsilon \to 0$.

Remark 1. The model (4) is an autoregression scheme, 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 Ornstein-Uhlenbeck model

$$dy(t) = \vartheta y(t)\ dt + \epsilon\ dW(t), \quad t \in [0, 1] \qquad (8)$$

which has been investigated predominantly with fixed ϵ and an increasing interval of observation (see Kutoyants [11], §3.5). In our case of fixed interval and varying ϵ all experiments (8) 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 (8) for $\epsilon = 1$. Thus our accompanying sequence of experiments is constant with respect to type, and theorem 1 for $n = \epsilon^{-2}$ establishes a convergence of the autoregression experiments

$$y_{i+1} = y_i + n^{-1} \vartheta y_i + n^{-1} \xi_i, \quad i = 1, \ldots, n, \quad y_1 = 0 \qquad (9)$$

to a diffusion limit (8) with $\epsilon = 1$. Define $\tilde{y}_i = n y_i$ and $\beta = 1 + n^{-1} \vartheta$; then (9) may be written

$$\tilde{y}_{i+1} = \beta\ \tilde{y}_i + \xi_i, \quad i = 1, \ldots, n, \quad \tilde{y}_1 = 0 \qquad (10)$$

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 [4], Cox [5]; cp. also comments in Jeganathan [10], and Stockmarr and Jacobsen [21] for a multivariate AR(1) model. We have thus recovered the limit experiment argument for critical autoregression; the behaviour is known as "locally asymptotically Brownian functional" (Phillips [19]) or locally asymptotically quadratic (LAQ, Le Cam and Yang [15]). For a comprehensive discussion cf. Shiryaev and Spokoiny [20], chap V. Thus Theorem 1 appears as a natural nonparametric extension of the LAQ limit, the limit being substituted by a sequence of accompanying experiments.

Remark 2. It is now clear what the relation to *nonparametric autoregression* should be: define $\tilde{y}_i$ as above, $n = \epsilon^{-2}$ and a function $g(x) = x + f(n^{-1}x)$; then (4) may be written

$$\tilde{y}_{i+1} = g(\tilde{y}_i) + \xi_i, \quad i = 1, \ldots, n, \quad \tilde{y}_1 = 0 \ . \tag{11}$$

Nonparametric inference for fixed, unknown g was studied by Doukhan and Ghindés [6] 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 (cp. also Bosq [1], chap. 3.2). So obviously the nonparametric model (11) with g fixed and stationarity corresponds to parametric autoregression in the stable case where $|\beta| < 1$ (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 asymptotic results for autoregressive models are naturally available for nonnormal ξ_i, based on the limit experiment rationale. On the other hand, for nonnormal nonparametric regression models similar to (3), asymptotic equivalence to certain Gaussian models has been shown recently (Grama and Nussbaum [8]).

Proof of Theorem 1. The argument is related to the one of Brown and Low [3] in the 'signal plus noise' case. Define a function $\bar{f}_n$ on $[0, 1]$ which depends on a path $z(t),\ t \in [0, 1]$ 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. It is easy to see that there is a unique solution $\bar{y}(t)$, $t \in [0, 1]$ of

$$\bar{y}(t) = \int_0^t \bar{f}_n(u, \bar{y})\ du + \epsilon W(t)\ . \tag{12}$$

This solution $\bar{y}$ may be represented as a randomly interpolated process using the values $y_i,\ i = 1, \ldots, n$ from (4), construed as function values at t_i, where first a linear interpolation is carried out and then a Brownian bridge is added on each interval $[t_i, t_{i+1}]$. These Brownian bridges should be independent of the ξ_i in (4).

It is clear that $\bar{y}$ contains as much information about f as $y_i,\ i = 1, \ldots, n$, i.e. the respective experiments are equivalent. This can be seen formally by looking at the likelihood for model (12) as follows. Regard $\bar{y}$ as a diffusion type process defined by (12), see Liptser and Shiryaev [16], 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 $\bar{y}(t_i)$, $u \in (t_i, t_{i+1}]$. 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 and Shiryaev, [16], Chap. 7, §2, Theorem 7.6). But this is fulfilled since every value y_i from (4) is finite. Then the density is analogously to (6) for $z = \epsilon W$

$$\Lambda_{3,\epsilon}(f)(z) = \exp\left\{\frac{1}{\epsilon^2}\int_0^1 \bar{f}_n(t, z)\ dz(t) - \frac{1}{2\epsilon^2}\int_0^1 \bar{f}_n^2(t, z)\ dt\right\}\ . \tag{13}$$

Now observe (for $z_i = \epsilon W(t_i)$)

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

so that we obtain $\Lambda_{3,\epsilon}(f) = \Lambda_{2,\epsilon}(f)$. This means that the density (13) depends on z via $z_i = z(t_i)$ only, which implies that the values $\bar{y}(t_i),\ i = 1, \ldots, n$ are a sufficient statistic in (12). Since this experiment is dominated and the statistic has values in $\mathbb{R}^n$, it follows that the experiments given by (4) and (12) are equivalent, for parameter spaces $\Sigma \subset \Sigma_M$.

It remains to establish that the densities of the two diffusion type processes (1) and (12) fulfill

$$\sup_{f \in \Sigma} E\left|\Lambda_{1,\epsilon}(f) - \Lambda_{3,\epsilon}(f)\right| \to 0, \quad \epsilon \to 0\ . \tag{14}$$

We use an inequality involving the Hellinger process, cf. Jacod and Shiryaev [9]. Denote by $P_{1,f,\epsilon}$, $P_{3,f,\epsilon}$ the probability measures on $C[0, 1]$ given by the two processes (1) and (12), respectively; the notation $P_{3,f,\epsilon}$ reflects the dependence of $\bar{f}_n(\cdot, \cdot)$ in (12) on f. Let h_f be the Hellinger process of order $1/2$ between these two measures (see Jacod and Shiryaev [9], §4b, p. 239): for a realization y

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

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

$$\|P_{1,f,\epsilon} - P_{3,f,\epsilon}\|_{TV} \left(= E_{\epsilon,0}|\Lambda_{1,\epsilon}(f) - \Lambda_{3,\epsilon}(f)|\right) \le 4\sqrt{E_{\epsilon,f} \ h_f(1)} \ , \quad (15)$$

where $E_{\epsilon,f}$ denotes expectation wrt $P_{1,f,\epsilon}$, see Jacod and Shiryaev [9], 4b, theorem 4.21, p. 279. The following lemma now completes the proof of Theorem 1.

Lemma. *Suppose* Σ *is as in theorem* 1. *Then*

$$E_{f,\epsilon} \int_0^1 (f(y(t)) - \bar{f}_n(t,y))^2 \, dt = O(n^{-2} + \epsilon^2 n^{-1}), \quad \epsilon \to 0$$

uniformly over $f \in \Sigma$.

Proof. The Lipschitz condition on f and the linear growth condition $|f(x)| \le M(1 + |x|)$, both implied by $f \in \Sigma_M$, are used to infer

$$\int_0^1 (f(y(t)) - \bar{f}_n(t,y))^2 \, dt \le Mn^{-2} \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} M^2(1 + y^2(u)) \, du$$
$$+ 2Mn^{-1} \sum_{i=1}^{n} \sup_{t \in (t_i, t_{i+1}]} \epsilon^2 |W(t) - W(t_i)|^2 \ . \quad (16)$$

To estimate the first term in (16), we note the following: there is a constant C_M depending on M but not on ϵ such that for all $f \in \Sigma$

$$E_{f,\epsilon} \ y^2(t) \le C_M, \quad t \in [0,1]$$

(cp. Øksendal [18], exercise 5.6). The second term in (16) is the average of n i.i.d. random variables, distributed as

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

where '$\simeq$' means equality in law. But $\sup_{t \in [0,1)} |W(t)|^2$ has finite expectation (Breiman [2], ch. 13.7), so that the expectation of the second term in (16) is $O(\epsilon^2 n^{-1})$ uniformly. □

3 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, cf. Yoshida [22] and the references therein.

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 $\mathbb{E}_{1,\epsilon}$ be the experiment given by observations y in (1) and $\mathbb{E}_{3,\epsilon}$ be given by $\bar{y}$ in (12). Define for a path $z(t)$, $t \in [0,1]$ the statistic T_ϵ by $T_\epsilon(z) = (z(t_1), \ldots, z(t_n))$. Then T_ϵ is sufficient in $\mathbb{E}_{3,\epsilon}$, and the result follows from (14). □

Larédo [13] obtained asymptotic sufficiency of the sampled values in a parametric framework using a different method, based on the exponential family approximation related to the standard local asymptotic normality results. The asymptotic sufficiency then naturally is a local one, i.e. it refers to experiments restricted to neighborhoods shrinking at rate ϵ. In the parametric local approach other statistics like level crossings have also been treated, see Genon-Catalot and Larédo [7].

References

[1] Bosq, D.: Nonparametric Statistics for Stochastic Processes. Lecture Notes in Statistics 110, Springer Verlag, New York (1996)

[2] Breiman, L.: Probability. Addison-Wesley, New York (1968)

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

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

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

[6] Doukhan, P., and Ghindès, M.: 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 (1983)

[7] Genon-Catalot, V., and Larédo, C.: An asymptotic sufficiency property of observations related to the first hitting times of a diffusion. Statistics & Probab. Letters. **9**, 253–258 (1990)

[8] Grama, I., and Nussbaum, M.: Asymptotic equivalence of nonparametric generalized linear models. Probab. Theory Relat. Fields **111**, 167–214 (1998)

[9] Jacod, J., and Shiryaev, A.: Limit Theorems for Stochastic Processes. Springer-Verlag, Berlin, Heidelberg, New York (1987)

[10] Jeganathan, P.: 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 (1988)

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

[12] Kutoyants, Yu.A.: 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 (1985)

[13] Laredo, C.F.: A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. Ann. Statist. **18**, 1158–1171 (1990)

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

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

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

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

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

[19] Phillips, P.C.B.: Towards a unified asymptotic theory for autoregression. Biometrika **74**, 535–547 (1987)

[20] Shiryaev, A.N., and Spokoiny, V.: Statistical Experiments and Decisions: Asymptotic Theory. Unpublished manuscript

[21] Stockmarr, A., and Jacobsen, M.: Gaussian diffusions and autoregressive processes: weak convergence and statistical inference. Scand. J. Statist. **21**, 4 403–419 (1994)

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