ANNALES DE L'I. H. P., SECTION B

VALENTINE GENON-CATALOT

JEAN JACOD

On the estimation of the diffusion coefficient for multi-dimensional diffusion processes

Annales de l'I. H. P., section B, tome 29, nº 1 (1993), p. 119-151

<http://www.numdam.org/item?id=AIHPB_1993__29_1_119_0>

Ann. Inst. Henri Poincaré,

Vol. 29, n° 1, 1993, p. 119-151. *Probabilités et Statistiques*

On the estimation of the diffusion coefficient for multi-dimensional diffusion processes

by

Valentine GENON-CATALOT

U.F.R. de Mathématiques et U.R.A. n° 1321,
Université Paris-VII, Tour n° 45-55,
2, place Jussieu, 75251 Paris Cedex 05, France.

and

Jean JACOD

Laboratoire de Probabilités, U.R.A. n° 224,
Université Pierre-et-Marie-Curie, Tour n° 56,
4, place Jussieu, 75252 Paris Cedex 05, France

ABSTRACT. — If a diffusion process has a diffusion coefficient which depends on a parameter ϑ, one can construct consistent sequences of estimators of ϑ based on the observation of the process at only n times, as n goes to infinity. Here we construct such estimators, and study their asymptotic efficiency. Of special interest to us are the multi-dimensional case (for the process) and the consideration of rather general sampling schemes.

Key words : LAMN property, asymptotic estimation, estimation for diffusions.

RÉSUMÉ. — Soit un processus de diffusion dont le coefficient de diffusion dépend d'un paramètre ϑ. On construit ci-dessous des estimateurs de ϑ basés sur l'observation du processus à n instants, et on étudie le comportement asymptotique de ces estimateurs lorsque n tend vers l'infini. La difficulté du problème tient d'une part à ce que le processus de diffusion

est multi-dimensionnel (avec un coefficient de diffusion éventuellement dégénéré), d'autre part à ce qu'on considère des schémas de discrétisation quelconques.

---

1. INTRODUCTION

We are considering here an estimation problem for the (d-dimensional) solution X of a stochastic differential equation of the form

$$(*)\qquad dX_t = b(t, X)\,dt + a(\vartheta, t, X_t)\,dW_t, \qquad \mathscr{L}(X_0) = \nu,$$

where the diffusion coefficient a is a known function of the parameter ϑ in $\mathbb{R}^q$ and the drift b is a non-anticipative functional. Further, b and the initial distribution ν (a probability measure on $\mathbb{R}^d$) may depend or not on ϑ, and may be known or not (parametric or semi-parametric estimation). We observe the process at n distinct times $t(n, i)$, $i = 1, \ldots, n$ in the interval [0, 1] and we are looking for asymptotic properties of estimators of ϑ as n goes to infinity.

When X is 1-dimensional and $t(n, i) = i/n$ (regular sampling) and b is of the form $b(t, X) = b(\vartheta, X_t)$ and $a(\vartheta, t, X_t) = a(\vartheta, X_t)$ (homogeneous Markov case) and furthermore $a(\vartheta, x)$ does not vanish, Dohnal [7] has shown the LAMN (local asymptotic mixed normality) property for the likelihoods as $n \to \infty$, under some smoothness assumptions on the coefficients. This allows for versions of the convolution theorem and the minimax theorem of Hajek: *see* Jeganathan [17], [18] or Le Cam and Yang [20] for a recent account on the subject. In particular asymptotic lower bounds for the variance of the estimators can be drawn.

This result has several drawbacks: 1. it does not give feasible asymptotically efficient estimators because the likelihood at stage n is not explicitely known; 2. it can easily be extended to the non-homogeneous Markov case, but not to the non-Markovian situation where b depends on the whole past of the process X; 3. even in the homogeneous Markov case, it cannot be extended in general to multi-dimensional processes X, because there is no accurate enough expansion of the density of the semi-group of the process: in fact if the diffusion coefficient $c = aa^T$ is non-degenerate and if the coefficients derive from a potential (in a suitable sense: *see* [13]) such good expansions are available, due to an explicit expression of the densities in terms of Brownian bridge (*see* e. g. Dacunha-Castelle and Florens-Zmirou [5]); under the sole non-degeneracy assumption on c the classical expansions of Molchanov [22] and Azencott [2] are not sufficient, and in the degenerate case there is no expansion at all available.

Here we propose a family of explicit estimators, based on the minimization of suitable contrasts (*see* Dacunha-Castelle and Duflo [4] for a general account on this notion, and Dorogovcev [6], Florens-Zmirou [9], Genon-Catalot [11], [12], Kasonga [19], for examples of contrasts used in discretization problems for diffusion processes, mainly for estimating the drift coefficient). Our estimators $\hat{\vartheta}_n$ have a "LAMN" property, in the sense that $\sqrt{n}(\hat{\vartheta}_n-\vartheta)$ converges in law when ϑ is the true value of the parameter to a variable S which, conditionally on the path of X, is centered Gaussian.

When $c=aa^{\mathrm{T}}$ is everywhere invertible, we construct estimators $\hat{\vartheta}_n$ which are optimal in this family, in the sense that if $\hat{\vartheta}'_n$ is another sequence converging after normalization to S′, then S′ − S is independent of S, conditionally on the path of X (hence S′ is more spread out than S): this is a sort of convolution theorem.

Let us stress the fact that we do not consider only "regular samplings" $t(n, i)=i/n$. Indeed, the only properties which we require is that the $t(n, i)$'s are distinct for different i's, and that the "empirical sampling measures" $\mu_n=\frac{1}{n}\sum_{1\leq i\leq n}\varepsilon_{t(n,i)}$ weakly converge to a measure μ on [0, 1] (μ is Lebesgue measure for regular samplings), and this property is indeed necessary to obtain limit theorems. Then, when $c=aa^{\mathrm{T}}$ is invertible and if we consider the optimal estimators $\hat{\vartheta}_n$ mentionned above, the covariance matrix of the limit law of $\sqrt{n}(\hat{\vartheta}_n-\vartheta)$, conditionally on the path of X, is $2\mathrm{B}(\vartheta)^{-1}$ where

$$\left\{\begin{array}{l} \mathrm{B}(\vartheta)=\displaystyle\int \mathrm{B}(\vartheta)_s\,\mu(ds), \qquad \text{and} \qquad \mathrm{B}(\vartheta)_s \\ \text{is the random } q\times q \text{ matrix whose } (i, j)\text{th} \\ \text{entries is the trace of } \dfrac{\partial c}{\partial\vartheta_i}c^{-1}\dfrac{\partial c}{\partial\vartheta_j}c^{-1}, \text{ evaluated at } (\vartheta, s, \mathrm{X}_s). \end{array}\right.$$

Hence instead of taking equally spaced sampling times, it may be wiser to concentrate all observations around some given time (μ is then a Dirac measure). In fact, we intend to describe in a further paper [13] an "optimal" sampling scheme, based on "random" sampling schemes in the spirit of [15] (it is then an adaptive sampling procedure). Let us also mention that using irregular sampling schemes allows to deal with missing data.

All results are stated in Section 2, the proofs being given in sections 3-6, and some examples are displayed in Section 7. Let us add a few words about these proofs: unlike many proofs of LAMN property or more generally of convergence of triangular arrays of variables to a mixed normal variable (*see* e. g. Hall and Heyde [14], Feigin [8], etc.) which concern the asymptotic behaviour as the observation times goes to infinity, we cannot use classical martingale limit theorems based on a nesting condition of the filtrations involved. We use an ad-hoc proof (simpler than Dohnal's one) based on a "functional" martingale-type limit theorem. A sizeable portion of the difficulties is due to the fact that we consider general sampling schemes.

2. STATEMENT OF RESULTS

2*a*. *Notation and assumptions.* – We first describe the general setting. The parameter space is a compact subset Θ of $\mathbb{R}^q$. The basic d-dimensional process $X=(X^i)_{1\leqq i\leqq d}$ is a solution to the following equation

$$(1) \qquad dX_t = b(t, X)\,dt + a(\vartheta, t, X_t)\,dW_t, \qquad \mathscr{L}(X_0)=\nu,$$

where:

– ν is a probability measure on $\mathbb{R}^d$, possibly unknown or possibly depending on the parameter ϑ,

– b is a non-anticipative $\mathbb{R}^d$-valued map, possibly unknown or possibly depending on the parameter ϑ,

– a is a map: $\Theta\times[0, 1]\times\mathbb{R}^d\to\mathbb{R}^d\otimes\mathbb{R}^m$,

– W is an m-dimensional Brownian motion.

An important role is played by the "diffusion coefficient" $c=aa^T$ (where a^T denotes the transpose of a): this is a function from $\Theta\times[0, 1]\times\mathbb{R}^d$ into the set $\mathscr{S}_d^+$ of symmetric nonnegative $d\times d$ matrices.

Below we introduce two hypotheses on the model: H1 will be in force throughout the paper, while H2 will only be assumed occasionally.

*Hypothesis H*1. (i) $t\to b(t, X)$ is continuous.

(ii) The partial derivatives $\nabla_x^2 a$, $\nabla_x\nabla_\vartheta a$, $\nabla_\vartheta^2 a$, $\nabla_x \partial a/\partial t$ exist and are continuous on $\Theta\times[0, 1]\times\mathbb{R}^d$.

(iii) Equation (1) admits a *non-exploding* strong solution on [0, 1] ("strong" means that the solution is adapted to the filtration generated by the Wiener process W and the initial value X_0). ■

Since a is locally Lipschitz, if b were identically 0 the strong solution (hence the weak solution as well) would be unique; then by Girsanov's Theorem, under H1 Equation (1) has a unique weak solution $P_{\nu,b}^{\vartheta}$ (*see* Stroock and Varadhan [23] or Liptser and Shiryayev [21]): that is, $P_{\nu,b}^{\vartheta}$ is the law of the solution on the canonical space $\Omega=C([0, 1], \mathbb{R}^d)$ equipped with the canonical filtration $(\mathscr{F}_t)_{t\in[0, 1]}$ and the canonical process $X=(X_t)_{t\in[0, 1]}$. From the statistical point of view, it is more natural to work with these weak solutions.

*Hypothesis H*2. For all ϑ, outside a $P_{\nu,b}^{\vartheta}$-null set, the matrices $c(\zeta, t, X_t)$ are invertible for all $t\in[0, 1]$, $\zeta\in\Theta$. ■

In most applications the initial distribution ν is the Dirac measure at some known point x_0 of $\mathbb{R}^d$, so that X_0 is known. If ν is unknown, we gain no insight on ϑ by observing X_0. However, in order to unify the presentation, we suppose that we observe the process at time $t(n, 0)=0$ and at n other times $t(n, 1)<\ldots<t(n, n)$ in $(0, 1]$ (asymptotically as $n\to\infty$, it makes no difference to have n or $n+1$ observations). In order to obtain limit results as $n\to\infty$, we obviously need some kind of nice behaviour for the sampling times. This is expressed by the

Hypothesis H3. The probability measures $\mu_n=\frac{1}{n}\sum_{1\leq i\leq n}\varepsilon_{t(n,i)}$ on $[0, 1]$ weakly converge to a limiting measure μ. ■

Since we wish to have (at least) consistent estimators for ϑ, we need an *identifiability* assumption on the model. Under H3, this can be expressed as

Hypothesis H4. For all $\vartheta\in\Theta$, for $P^{\vartheta}_{\nu,b}$-almost all ω, for all $\zeta\neq\vartheta$ the two functions $t\to c(\vartheta, t, X_t(\omega))$ and $t\to c(\zeta, t, X_t(\omega))$ are not μ-a. s. equal. ■

Denoting by $\mathscr{F}^n$ the σ-field generated by the observations $X_{t(n,i)}$ at stage n, Hypothesis H4 amounts to saying that, under H3, for all $\zeta\neq\vartheta$ the two sequences $\{P^{\zeta}_{\nu,b}$ on $(\Omega, \mathscr{F}^n)\}_{n\geq 1}$ and $\{P^{\vartheta}_{\nu,b}$ on $(\Omega, \mathscr{F}^n)\}_{n\geq 1}$ are entirely separated: this is exactly the condition for having consistent estimators.

Remarks on these hypotheses. — 1. The reason for stating H2 as above is that we want to accomodate linear coefficients: for example when $d=m=1$, we want to consider $a(\vartheta, t, x)=A(\vartheta, t)x$ with A never vanishing; then c can take the value 0, however if $\nu((-\infty, 0])=0$ the process takes its values in $(0, \infty)$ and H2 is satisfied (*see* § 7 *b* below).

2. The sequence (μ_n) is always tight. If H3 fails, this sequence has several limit points, and as we will see our sequences of estimators would have accordingly several limiting points (for convergence in law).

3. Assumption H4 is difficult to check in general, because it depends on the path of the process. However, it is satisfied in the following cases:

– μ is the Dirac measure $\mu=\varepsilon_s$ (the observation times are concentrated around s) and $c(\vartheta, s, x)\neq c(\zeta, s, x)$ for all x and $\zeta\neq\vartheta$,

– μ has a topological support equal to $[0, 1]$ and for all x and $\vartheta\neq\zeta$, there is a $t\in[0, 1]$ with $c(\vartheta, t, x)\neq c(\zeta, t, x)$,

– μ has a topological support including 0, and ν is the Dirac measure $\nu=\varepsilon_{x_0}$ and $c(\vartheta, 0, x_0)\neq c(\zeta, 0, x_0)$ for all $\vartheta\neq\zeta$. ■

2 *b*. *The contrasts*. – In order to ease the notation, we write $t_i = t(n, i)$ if there is no ambiguity. Set also

$$(2) \qquad \begin{cases} \delta(n, i) = \delta_i = t_i - t_{i-1} \\ X_i^n = \dfrac{1}{\sqrt{\delta_i}} (X_{t_i} - X_{t_{i-1}}) \end{cases} \quad \text{for} \quad 1 \leqq i \leqq n.$$

Now we want to estimate ϑ from the observation of the σ-field $\mathscr{F}^n$. However we do not know explicitely the relative densities $dP^{\vartheta}_{\nu, b}/dP^{\vartheta'}_{\nu', b'}$ in restriction to $\mathscr{F}^n$ (even in the Markov case, these densities are functions of the transition densities, which are unknown). Hence we are led to construct the estimators by minimizing suitable contrasts.

In order to motivate our choice of contrasts, let us first assume that $b = 0$ and $\nu = \varepsilon_{x_0}$ and $c(\vartheta, t, x) = c(\vartheta, t)$ does not depend on x and is invertible. Then X is a continuous Gaussian process with independent increments, and the log-likelihood of the family $(X_{t(n, i)})_{1 \leqq i \leqq n}$ w. r. t. Lebesgue measure on $(\mathbb{R}^d)^n$, is given by

$$-\frac{d}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} [\log \det c(\vartheta)_i^n + (X_{t_i} - X_{t_{i-1}})^T (c(\vartheta)_i^n)^{-1} (X_{t_i} - X_{t_{i-1}})],$$

where $c(\vartheta)_i^n = \int_{t_{i-1}}^{t_i} c(\vartheta, s)\, ds$. The MLE (maximum likelihood estimator) maximizes the above expression, and due to the continuity of $t \to c(\vartheta, t)$ the asymptotic optimality of the MLE is shared by the estimator maximizing the same expression, but with $c(\vartheta)_i^n$ replaced by $\delta_i c(\vartheta, t_{i-1})$. That is, the estimator minimizing the following contrast

$$\frac{1}{n} \sum_{i=1}^{n} [\log \det c(\vartheta, t_{i-1}) + (X_i^n)^T c(\vartheta, t_{i-1})^{-1} X_i^n].$$

has the same asymptotic properties than the MLE.

When c depends on x, but remains invertible (*i. e.* under H2), it is thus natural to use the following contrast to minimize:

$$(3) \qquad \frac{1}{n} \sum_{i=1}^{n} [\log \det c(\vartheta, t_{i-1}, X_{t_{i-1}}) + (X_i^n)^T c(\vartheta, t_{i-1}, X_{t_{i-1}})^{-1} X_i^n].$$

In general H2 does not hold. To mimick the previous approach we consider the class of all contrasts of the form

$$(4) \qquad U^n(\vartheta) = \frac{1}{n} \sum_{i=1}^{n} f(c(\vartheta, t(n, i-1), X_{t(n, i-1)}), X_i^n).$$

Here f should be a "nice" function of the pair (G, x), where $x \in \mathbb{R}^d$ and G belongs to some subset $\mathscr{S}$ of $\mathscr{S}_d^+$, such that the set

$\bigcup_{t,\zeta}\{\omega : c(\zeta, t, X_t(\omega)) \notin \mathcal{S}\}$ is $P^{\vartheta}_{\nu, b}$-null for all ϑ. The biggest possible class of contrasts is obtained for $\mathcal{S}$ as small as possible (for instance under H2, it is naturel to take $\mathcal{S}=\mathcal{S}_d^{++}$, the set of all invertible matrices in $\mathcal{S}_d^+$, so that (3) is of the form (4) with $f(G, x)=\log\det(G)+x^T G^{-1} x$). The regularity conditions on the pair $(\mathcal{S}, f)$ are expressed by the following:

Condition C1. (i) $\mathcal{S}=\bigcup \mathcal{S}_l$ where $(\mathcal{S}_l)$ is an increasing sequence of compact subsets of $\mathcal{S}_d^+$ with $\lim_{l\to\infty} P^{\vartheta}_{\nu, b}[c(\zeta, t, X_t)\in\mathcal{S}_l \text{ for all } \zeta, t]=1$ for all ϑ.

(ii) $f=f(G, x)$ is a function: $\mathcal{S}\times\mathbb{R}^d\to\mathbb{R}$ such that for i and $j=0, 1, 2$ the partial derivatives $\nabla^i_G\nabla^j_x f$ are continuous and satisfy $|\nabla^i_G\nabla^j_x f(G, x)|\leqq\gamma(G)(1+|x|^{\gamma(G)})$ for some continuous function γ on $\mathcal{S}$ (here ∇ denotes the gradient and $|\,.\,|$ is the Euclidian norm). ■

One possible choice is $\mathcal{S}_l=\{G\in\mathcal{S}_d^+ : |G|\leqq l\}$, so that $\mathcal{S}=\mathcal{S}_d^+$; under H2, we also have C1-(i) with $\mathcal{S}_l=\{G\in\mathcal{S}_d^+ : |G|\leqq l \text{ and } 1/l\leqq\det(G)\leqq l\}$, so that $\mathcal{S}=\mathcal{S}_d^{++}$.

H1, H3 and C1 will insure the convergence of $U^n(\vartheta)$ to a limit. To obtain consistency and asymptotic mixed normality for the related minimum contrast estimators, additional conditions are needed:

Condition C2. The function f in C1 is even in its second argument. ■

Condition C3. With ρ_G denoting the centered Gaussian law on $\mathbb{R}^d$ with covariance matrix $G\in\mathcal{S}_d^+$ for every $G\in\mathcal{S}$ the function $G'\to\int f(G', x)\,\rho_G(dx)$ on $\mathcal{S}$ has a unique minimum at $G'=G$. ■

Remarks. – 1. For each value of ϑ, $U^n(\vartheta)$ is observable. We do not want to include the function b in the contrast because it is unknown. Even when b is known, we cannot include it in the contrast since it is a function of the whole past of X and thus is not observable [except in the Markov case $b(t, X)=b(t, X_t)$].

2. It is possible to exhibit even more general contrasts of the form $U^n(\vartheta)=\frac{1}{n}\sum_{i\leqq i\leqq n} f(\vartheta, t_{i-1}, X_{t_{i-1}}, X_i^n)$ with similar results. But it is difficult to think of "natural" functions f other than depending on (ϑ, t, X_t) through $c(\vartheta, t, X_t)$. ■

2 *c. Convergence of contrasts.* – Introduce the following notation:

$$(5)\qquad \begin{cases} \rho_{\vartheta, t}=\rho_{c(\vartheta, t, X_t)} & \text{(recall the notation } \rho_G \text{ in Condition C3)} \\ F(\vartheta, t, x)=f(c(\vartheta, t, X_t), x), \end{cases}$$

$$(6)\qquad U_t(\vartheta, \zeta)=\int F(\zeta, t, x)\,\rho_{\vartheta, t}(dx).$$

THEOREM 1. — *Under* H1, H3, C1, *and for each* $\vartheta \in \Theta$, $U^n(\zeta)$ *converges uniformly in* $\zeta \in \Theta$ *in* $P^{\vartheta}_{v,b}$*-measure to the random variable*

$$U(\vartheta, \zeta) = \int U_t(\vartheta, \zeta)\, \mu(dt).$$

Let $\hat{\vartheta}_n$ be a minimum contrast estimator associated with U^n, *i. e.* $\hat{\vartheta}_n$ is any solution of the equation $U^n(\hat{\vartheta}_n) = \min_{\vartheta \in \Theta} U^n(\vartheta)$ (such a solution always exists, because $\vartheta \to U^n(\vartheta)$ is continuous and Θ is compact). In order to deduce the convergence $\hat{\vartheta}_n \to \vartheta$ in $P^{\vartheta}_{v,b}$-measure from the above, it is enough that $\zeta = \vartheta$ be the unique minimum of $\zeta \to U(\vartheta, \zeta)$ $P^{\vartheta}_{v,b}$-a. s. This is clearly true under H4 and C3 (indeed H4 is necessary for that, in connection with the sequence of sampling times satisfying H3; on the other hand C3 is a bit too strong, in the sense that it works for all sequences of sampling times). Then we have:

COROLLARY 2. — *Under* H1, H3, H4, C1, C3, *then* $\hat{\vartheta}_n \to \vartheta$ *in* $P^{\vartheta}_{v,b}$*-measure.*

We will see at the end of Section 4 (Remark 10) that we can substantially weaken the regularity assumption H1 for the two above results.

2 *d. Asymptotic mixed normality.* — We need still more notation:

$$\left\{ \begin{aligned} B(\vartheta)_t &= \int_{\mathbb{R}^d} \nabla^2_{\vartheta} F(\vartheta, t, x)\, \rho_{\vartheta, t}(dx), \\ D(\vartheta)_t &= \int_{\mathbb{R}^d} \nabla_{\vartheta} F(\vartheta, t, x)\, \nabla_{\vartheta} F(\vartheta, t, x)^T \rho_{\vartheta, t}(dx), \end{aligned} \right. \tag{7}$$

$$\left\{ \begin{aligned} B(\vartheta) &= \int_{[0,1]} B(\vartheta)_t\, \mu(dt), \\ D(\vartheta) &= \int_{[0,1]} D(\vartheta)_t\, \mu(dt). \end{aligned} \right. \tag{8}$$

Observe that $B(\vartheta)_t$ and $D(\vartheta)_t$ are processes taking values in the set of symmetric $q \times q$ matrices, and $D(\vartheta)_t$ is non-negative.

THEOREM 3. — *Assume* H1, H3, H4, C1, C2, C3. *Let* ϑ *be an interior point of* Θ, *such that* $B(\vartheta)$ *is* $P^{\vartheta}_{v,b}$*-a. s. invertible. Then the random vectors* $S_n = \sqrt{n}(\hat{\vartheta}_n - \vartheta)$ *converge in law under* $P^{\vartheta}_{v,b}$ *to a "mixed normal variable"* S *defined on an extension of the space* $(\Omega, \mathscr{F}_1, P^{\vartheta}_{v,b})$ *and which, conditionally on* $\mathscr{F}_1$, *is centered Gaussian with covariance matrix*

$$\Gamma(\vartheta) = B(\vartheta)^{-1} D(\vartheta) B(\vartheta)^{-1}.$$

Remark. — If C2 fails, we do not know in general whether S_n converges. It does, however, when $t(n, i) = i/n$; but, conditionally on $\mathscr{F}_1$, the limit S is a non-centered Gaussian vector, with a mean that depends on the coefficients a and b. ■

Remark. – Even if the drift coefficient b is a function of ϑ, we cannot improve on this theorem. ■

Remark. – If the initial measure ν is a known function of ϑ, say ν_ϑ, it may be possible to construct other estimators $\hat{\vartheta}'_n$ (based on the observation of X_0 only, for example) which converge to ϑ faster than above (*i. e.* with smaller limiting variance, or even with a faster rate than $1/\sqrt{n}$): this depends on the structure of the statistical model $(\mathbb{R}^d, (\nu_\vartheta)_{\vartheta \in \Theta})$; of course, if this model is regular (as it is the case when $\nu_\vartheta = \varepsilon_{x_0}$), it is impossible to improve Theorem 3 in this way. ■

In order to compare two different constrasts, we prove also a bivariate version of this result, which reads as follows. Let $U'^n(\zeta)$ be the contrast associated with another pair $(\mathscr{S}', f')$ and add a 'prime' to all related quantities: $B'(\vartheta)_t$, $\hat{\vartheta}'_n$, S'_n, etc. Now let $\bar{B}(\vartheta)_t$ and $\bar{D}(\vartheta)_t$ be the $2q \times 2q$ symmetric matrix-valued processes defined by

$$(9)\quad \left\{\begin{aligned} &\bar{B}(\vartheta)_t = \begin{bmatrix} B(\vartheta)_t & 0 \\ 0 & B'(\vartheta)_t \end{bmatrix}, \qquad \bar{D}(\vartheta)_t = \begin{bmatrix} D(\vartheta)_t & D''(\vartheta)_t \\ D''(\vartheta)_t^T & D'(\vartheta)_t \end{bmatrix}. \\ &\text{where} \\ &\qquad D''(\vartheta)_t = \int \nabla_\vartheta F(\vartheta, t, x) \nabla_\vartheta F'(\vartheta, t, x)^T \rho_{\vartheta, t}(dx), \end{aligned}\right.$$

$$(10)\quad \left\{\begin{aligned} \bar{B}(\vartheta) &= \int_{[0,1]} \bar{B}(\vartheta)_t \,\mu(dt), \\ \bar{D}(\vartheta) &= \int_{[0,1]} \bar{D}(\vartheta)_t \,\mu(dt). \end{aligned}\right.$$

PROPOSITION 4. – *Assume that all hypotheses of Theorem* 3 *are met, for both contrasts. Then the pair* (S_n, S'_n) *converges in law under* $P^\vartheta_{\nu, b}$ *to a variable* $\bar{S} = (S, S')$ *defined on an extension of the space* $(\Omega, \mathscr{F}_1, P^\vartheta_{\nu, b})$, *which conditionally on* $\mathscr{F}_1$ *is centered Gaussian with covariance matrix*

$$(11)\qquad \bar{\Gamma}(\vartheta) = \bar{B}(\vartheta)^{-1} \bar{D}(\vartheta) \bar{B}(\vartheta)^{-1}.$$

2*e*. *Comparison of contrasts.* – Here we assume H2, so that we may take $\mathscr{S} = \mathscr{S}_d^{++}$ (the set of invertible matrices in $\mathscr{S}_d^+$) in C1. The contrast (3) is of the form (4), with

$$(12)\qquad f(G, x) = \log\det(G) + x^T G^{-1} x.$$

Theorem 5 below shows that this contrast is best in the class of contrasts of type (4), the asymptotic conditional variance of minimum contrast estimators being minimal for f given by (12). In fact we even have a form of Hajek's *convolution theorem* [10], although we have not proved the LAMN property for the likelihoods (it is of course a strong indication that the LAMN property is true: in the 1-dimensional Markov case, the LAMN property has been proved by Dohnal [7], and the estimators associated with (12) achieve the asymptotic variance bound deduced from the LAMN property).

THEOREM 5. – *Assume* H1, H2, H3, H4.

(*a*) *The pair* [$\mathscr{S}=\mathscr{S}_d^{++}$, *f given by* (12)] *satisfies* C1, C2, C3, *and*

$$(13)\qquad D(\vartheta)_t = 2\,B(\vartheta)_t,\qquad B(\vartheta)_t^{ij} = \mathrm{tr}\left[\left(\frac{\partial c}{\partial\vartheta_i}c^{-1}\frac{\partial c}{\partial\vartheta_j}c^{-1}\right)(\vartheta, t.\ X_t)\right].$$

and if $B(\vartheta)$ *is* $P^{\vartheta}_{\nu,b}$-*a.s. invertible the asymptotic conditional covariance matrix of the associated estimators is then*

$$(14)\qquad \Gamma(\vartheta) = 2\,B(\vartheta)^{-1}.$$

(*b*) *Let* $(\mathscr{S}', f')$ *be another pair satisfying* C1, C2, C3 (*add a 'prime' to all quantities related to* f'), *and assume that both* $B(\vartheta)$ *and* $B'(\vartheta)$ *are* $P^{\vartheta}_{\nu,b}$-*a.s. invertible, where* ϑ *is an interior point of* Θ, *and call* (S, S′) *the limit of the pair* (S_n, S'_n): see *Proposition* 4. *Then conditionally on* $\mathscr{F}_1$, *the variables* S *and* S′-S *are independent.*

Remark. – It is worth noticing that if $I(\vartheta, t, x)$ is the Fisher information matrix at ϑ for the model $(\mathbb{R}^d, (\rho_{c(\vartheta,t,x)})_{\vartheta\in\Theta})$, then $B(\vartheta)_t = 2\,I(\vartheta, t, X_t)$: *see* Section 6. Then (14) becomes $\Gamma(\vartheta) = \left[\int_{[0,1]} I(\vartheta, t, X_t)\,\mu(dt)\right]^{-1}$. ■

3. SOME ESTIMATES

In the remainder of the paper, we fix ϑ, b and ν and write $P = P^{\vartheta}_{\nu,b}$.

Up to enlarging the space Ω, we can and will assume that there is a standard Wiener process $W = (W^i)_{1\leqq i\leqq m}$ on Ω with respect to which the canonical process X is a solution to (1) and is adapted to the filtration $(\mathscr{G}_t)$ generated by W (*i.e.* $\mathscr{F}_t \subseteqq \mathscr{G}_t$), because the solution is strong.

We complement our notation [recall that we have set $t_i = t(n, i)$ when there is no ambiguity on the value of n]:

$$(15)\qquad \begin{cases} \mathscr{G}_i^n = \mathscr{G}_{t(n,i)} = \sigma(W_s : s \leqq t_i), \\ Y_i^n = \dfrac{1}{\sqrt{\delta_i}}\,a(\vartheta, t_{i-1}, X_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}), \end{cases}$$

so that

(16) conditionally on $\mathscr{G}_{i-1}^n$, Y_i^n is centered Gaussian with covariance matrix $c(\vartheta, t_{i-1}, X_{t_{i-1}})$.

3 *a*. *The basic estimates.* — Let us replace for a moment H1 by a stronger assumption:

Hypothesis H1'. H1 holds, and there is a constant K such that $|b|$, $|a|$, $|\nabla_x a|$, $|\nabla_\vartheta a|$, $|\nabla_x^2 a|$, $|\nabla_\vartheta^2 a|$, $|\nabla_x \nabla_\vartheta a|$, $|\partial a/\partial t|$, $|\nabla_x \partial a/\partial t| \leqq \mathrm{K}$. ■

Denote by $\varepsilon(u) = \sup_{t \in [s, s+u]} |b(t, \mathrm{X}) - b(s, \mathrm{X})|$ the (random) modulus of continuity of the function $t \to b(t, \mathrm{X})$. Doob's inequality applied to the $(\mathscr{G}_i^n)$-martingale $\mathrm{M}_i = \mathrm{E}[\varepsilon(u)^2 | \mathscr{G}_i^n]$ yields $\mathrm{E}(\sup_{1 \leqq i \leqq n} \mathrm{M}_i) \leqq 2\, \mathrm{E}(\varepsilon(u)^4)^{1/2}$, which goes to 0 as $u \to 0$ because $\varepsilon(u) \to 0$ and $\varepsilon(u) \leqq 2\mathrm{K}$. Therefore

$$\tau(u) := u + \sup_{0 \leqq i \leqq n} \mathrm{E}[\varepsilon(u)^2 | \mathscr{G}_i^n] \xrightarrow{\mathrm{P}} 0 \quad \textit{as} \quad u \to 0. \tag{17}$$

Below, we condider a C^2 function g on $\mathbb{R}^d$, satisfying

$$|g(x)| + |\nabla g(x)| + |\nabla^2 g(x)| \leqq \gamma (1 + |x|^\gamma) \tag{18}$$

for some constant γ. Our aim is to prove the following proposition.

PROPOSITION 6. — *Assume* H1' *and* (18). *There is a constant* C *depending on* K *and* γ *only, with* [recall (2) for X_i^n]

$$\mathrm{E}[(g(\mathrm{X}_i^n) - g(\mathrm{Y}_i^n))^2 | \mathscr{G}_{i-1}^n] \leqq \mathrm{C}\, \delta(n, i) \tag{19}$$

and, if g is an even function,

$$|\mathrm{E}[g(\mathrm{X}_i^n) - g(\mathrm{Y}_i^n) | \mathscr{G}_{i-1}^n]| \leqq \mathrm{C} \sqrt{\delta(n, i)\, \tau(\delta(n, i))}. \tag{20}$$

The proof is broken into several steps, for which we use the following simplifying notation. We fix n, i and set $s = t(n, i-1)$ and

$$b_u = b(s+u, \mathrm{X}), \qquad a_u = a(\vartheta, s+u, \mathrm{X}_{s+u}),$$
$$\mathrm{W}'_t = \mathrm{W}_{s+t} - \mathrm{W}_s, \qquad \mathrm{X}'_t = \mathrm{X}_{s+t} - \mathrm{X}_s, \qquad \mathrm{Y}_t = a_0 \mathrm{W}'_t, \qquad \mathrm{Z}_t = \mathrm{X}'_t - \mathrm{Y}_t,$$
$$\mathrm{A}_t = \int_0^t b_u\, du, \qquad \mathrm{N}_t = \int_0^t (a_u - a_0)\, d\mathrm{W}'_u, \qquad \mathrm{M}_t = \int_0^t a_u\, d\mathrm{W}'_u.$$

For any process V we also write $\mathrm{V}_t^* = \sup_{u \leqq t} |\mathrm{V}_u|$.

LEMMA 7. — *For each* $p \geqq 1$ *there is a constant* K_p *depending on* K *and* p *only, such that*

$$\mathrm{E}(\mathrm{X}_t'^{*p} | \mathscr{G}_s) \leqq \mathrm{K}_p t^{p/2}, \tag{21}$$

$$\mathrm{E}(\mathrm{Z}_t^{*p} | \mathscr{G}_s) \leqq \mathrm{K}_p t^p. \tag{22}$$

Proof. — (*a*) Below, all constants depending on K and p are denoted by K_p. From the Burkhölder-Davis-Gundy inequality we deduce

$$\mathrm{E}(\mathrm{M}_t^{*p} | \mathscr{G}_s) \leqq \mathrm{K}_p \mathrm{E}\left[\left(\int_0^t |a_u|^2\, du\right)^{p/2} \Big| \mathscr{G}_s\right] \leqq \mathrm{K}_p t^{p/2}$$

(use H1′), and obviously $|A_t| \leqq K\,t$. Since $X' = M + A$, (21) follows.

(*b*) From H1′ again, we obtain $|a_u - a_0| \leqq K(u + X'^*_u)$. Then as above,

$$E(N^{*p}_t | \mathcal{G}_s) \leqq K_p\, E\left[\left\{\int_0^t (u^2 + X'^{*2}_u)\,du\right\}^{p/2} | \mathcal{G}_s\right]$$
$$\leqq K_p\left[t^{3p/2} + t^{p/2-1}\int_0^t E(X'^{*p}_u | \mathcal{G}_s)\,du\right]$$

if $p \geqq 2$ (use Hölder's inequality for the last estimate), so (22) follows from (21) for $p \geqq 2$. Finally (22) for $p \in [1, 2)$ follows from the Schwarz inequality applied to (22) with $2\,p$. ■

Set $\bar{X}'_t = X'_t/\sqrt{t}$ and $\bar{Y}_t = Y_t/\sqrt{t}$ and $\bar{Z}_t = Z_t/\sqrt{t} = \bar{X}'_t - \bar{Y}_t$. Below, all constants depending only on K and γ are denoted by C. By Taylor's formula and (18) we get

$$(23) \qquad |g(\bar{X}'_t) - g(\bar{Y}_t) - \nabla g(\bar{Y}_t)^T \bar{Z}_t| \leqq C(1 + |\bar{Y}_t|^C + |\bar{Z}_t|^C)\,|\bar{Z}_t|^2.$$

First, this yields (with a different C):

$$|g(\bar{X}'_t) - g(\bar{Y}_t)|^2 \leqq C(1 + |\bar{Y}_t|^C + |\bar{Z}_t|^C)\,|\bar{Z}_t|^2,$$

while (21) and (22) yield $E(|\bar{Z}_t|^p | \mathcal{G}_s) \leqq K_p\, t^{p/2}$ and $E(|\bar{Y}_t|^p | \mathcal{G}_s) \leqq K_p$ for all $p \geqq 1$. Thus we readily deduce (19) (recall that $t \leqq 1$).

Second, if we can prove that

$$|E[\nabla g(\bar{Y}_t)^T \bar{Z}_t | \mathcal{G}_s] \leqq C\{t + [t\,E(\varepsilon(t)^2 | \mathcal{G}_s)]^{1/2}\},$$

we will deduce (20) from (21), (22), (23). But if g is even, then ∇g is odd. Hence (20) follows from the

LEMMA 8. – *If h is an odd function with $|h(x)| \leqq \gamma(1 + |x|^\gamma)$, then (with $\bar{Z}^i$ being the ith coordinate of $\bar{Z}$)*:

$$|E(h(\bar{Y}_t)\,\bar{Z}^i_t | \mathcal{G}_s)| \leqq C\{t + [t\,E(\varepsilon(t)^2 | \mathcal{G}_s)]^{1/2}\}.$$

Proof. – (*a*) One of the key points of the proof is that $\bar{Y}_t$ has, conditionally on $\mathcal{G}_s$, the same distribution than $a_0 W'_1$, which in particular does not depend on t. Hence with I = the identity $d \times d$ matrix, and recalling the notation ρ_G of Condition C3,

$$E[h(\bar{Y}_t)^2 | \mathcal{G}_s] = \int h(a_0 x)^2\, \rho_I(dx) \leqq C$$

for some constant C, in virtue of $|h(x)| \leqq \gamma(1 + |x|^\gamma)$.

(*b*) We write $A^i_t = B_t + B'_t$ with $B_t = b^i_0\, t$ and $B'_t = \int_0^t (b^i_u - b^i_0)\,du$. First, since h is odd,

$$E[h(\bar{Y}_t)\,B_t | \mathcal{G}_s] = t b^i_0 \int h(a_0 x)\,\rho_I(dx) = 0.$$

On the other hand, $|B'_t| \leqq t\,\varepsilon(t)$, so by Schwarz inequality and (*a*):

$$|E[h(\bar{Y}_t) B'_t \mid \mathscr{G}_s]| \leqq \{E[h(\bar{Y}_t)^2 \mid \mathscr{G}_s] E(B'^2_t \mid \mathscr{G}_s)\}^{1/2} \leqq C\,t\,E[\varepsilon(t)^2 \mid \mathscr{G}_s]^{1/2}.$$

Putting these together yields

$$|E[h(\bar{Y}_t) A^i_{tl}]| \leqq C\,t\,E[\varepsilon(t)^2 \mid \mathscr{G}_s]^{1/2}.$$

(*c*) Recall that i is fixed. Set

$$\alpha^{jk}_u = (\partial a^{ij}/\partial x_k)(\vartheta, s+u, X_{s+u}), \qquad \beta^{jkh}_u = (\partial^2 a^{ij}/\partial x_k\,\partial x_h)(\vartheta, s+u, X_{s+u}),$$
$$\delta^j_u = (\partial a^{ij}/\partial t)(\vartheta, s+u, X_{s+u}), \qquad c_u = c(\vartheta, s+u, X_{s+u}).$$

By H1′ we have

(24) $$|\alpha_u|, |\beta_u|, |\delta_u|, |b_u|, |c_u| \leqq K, \qquad |\alpha_u - \alpha_0| \leqq K(u + X'^*_u).$$

Ito's formula yields

$$a^{ij}_u - a^{ij}_0 = D^j_u + F^j_u + G^j_u,$$

where

$$D^j_u = \int_0^u \left[\delta^j_v + \sum_k \alpha^{jk}_v b^k_v + \sum_{k,h} \frac{1}{2} \beta^{jkh}_v c^{kh}_v\right] dv$$

$$F^j_u = \int_0^u \sum_k (\alpha^{jk}_v - \alpha^{jk}_0)\, dW'^k_v, \qquad G^j_u = \sum_k \alpha^{jk}_0 W'^k_u.$$

Using (24) and (21), we obtain (recall that C changes from line to line):

$$|D^j_u| \leqq C u, \qquad E[(F^j_u)^2 \mid \mathscr{G}_s] \leqq C u^2.$$

Now, we set $N^i = N' + N''$ with

$$N'_t = \int_0^t \sum_j G^j_u\, dW'^j_u, \qquad N''_t = \int_0^t \sum_j (D^j_u + F^j_u)\, dW'^j_u.$$

From what precedes,

$$E(N''^2_t \mid \mathscr{G}_s) = E\left(\int_0^t \sum_j (D^j_u + F^j_u)^2\, du \mid \mathscr{G}_s\right) \leqq C\,t^3,$$

and thus by (*a*) and Schwarz inequality,

$$|E[h(\bar{Y}_t) N''_t \mid \mathscr{G}_s]| \leqq C\,t^{3/2}.$$

(*d*) We have

$$E[h(\bar{Y}_t) N'_t \mid \mathscr{G}_s] = E\left[h(a_0 W'_t/\sqrt{t}) \sum_{j,k} \alpha^{jk}_0 \int_0^t W'^k_u\, dW'^j_u \mid \mathscr{G}_s\right].$$

But, conditionally on $\mathscr{G}_s$, the distributions of the processes W′ and $-W'$ are the same; since h is odd, we deduce

$$E[h(\bar{Y}_t) N'_t \mid \mathscr{G}_s] = 0.$$

(*e*) It remains to observe that $\bar{Z}_t^i = (A_t^i + N_t' + N_t'')/\sqrt{t}$: then (*b*), (*c*) and (*d*) prove the claim. ■

3 *b. Applications.* – Here we show how to apply the previous results when H1′ does not hold. We suppose that the pair $(\mathscr{S}, f)$ satisfies C1, and set

$$\text{(25)}\quad R_l = \inf(t: |X_t| \geqq l \text{ or } |b(t, X)| \geqq l \text{ or } c(\zeta, t, X_t) \notin \mathscr{S}_l \text{ for some } \zeta \in \Theta)$$

(with $\inf(\varnothing) = \infty$): then R_l is a stopping time taking its values in $[0, 1] \cup \{\infty\}$ and C1-(i) implies

$$\text{(26)}\qquad P(\Omega_l) \to 1 \quad \text{as } l \to \infty, \qquad \text{where} \quad \Omega_l = \{R_l = \infty\}.$$

With φ_l being a C^∞ function: $\mathbb{R}^d \to \mathbb{R}$ such that

$$1_{\{|x| \leqq l\}} \leqq \varphi_l(x) \leqq 1_{\{|x| \leqq l+1\}},$$

we set

$$\text{(27)}\qquad a_l(\zeta, t, x) = a(\zeta, t, x)\varphi_l(x), \qquad b_l(t, X) = b(t \wedge R_l, X_{.\wedge R_l}).$$

We can consider the following equation, on the space $(\Omega, (\mathscr{G}_t), P)$ and with respect to the Wiener process W (recall that here ϑ is fixed):

$$dX(l)_t = b_l(t, X(l))\,dt + a_l(\vartheta, t, X(l)_t)\,dW_t, \qquad X(l)_0 = X_0.$$

Clearly, it has a unique strong solution, which satisfies $X(l)_t = X_t$ if $t \leqq R_l$. In accordance with (2) and (15), we set

$$X_i^n(l) = \frac{1}{\sqrt{\delta_i}}(X(l)_{t_i} - X(l)_{t_{i-1}}),$$

$$Y_i^n(l) = \frac{1}{\sqrt{\delta_i}} a_l(\vartheta, t_{i-1}, X(l)_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}),$$

so that

$$\text{(28)}\qquad X_i^n(l) = X_i^n \quad \text{if } t_i \leqq R_l, \qquad Y_i^n(l) = Y_i^n \quad \text{if } t_{i-1} \leqq R_l.$$

Further, (a_l, b_l) satisfies H1′. Then (21) yields constants $K_{l,p}$ such that

$$\text{(29)}\qquad E[|X_i^n(l)|^p \,|\, \mathscr{G}_{i-1}^n] \leqq K_{l,p}.$$

Next, let $(\mathscr{S}, f)$ be a pair satisfying C1, and F associated with f by (5). Let H be any of the following functions:

H is the kth power for $k = 1, 2, 3, 4$ of any component of F, $\nabla_\vartheta F$ or $\nabla_\vartheta^2 F$.

If $\gamma(.)$ is the function appearing in C1, $\gamma_l := \sup(\gamma(G): G \in \mathscr{S}_l)$ is finite and $|H(\zeta, t, x)| \leqq \gamma_l'(1 + |x|^{\gamma_l'})$ on the set $\{R_l \geqq t\}$ for all H as above, with $\gamma_l' = 8(1 + \gamma_l)^4$. It is obvious that Proposition 6 can be applied for the process $X(l)$, not only to "deterministic" functions g, but also to each $x \to H(\zeta, t_{i-1}, x) 1_{\{R_l > t_{i-1}\}}$ which, as a function of ω, is $\mathscr{G}_{i-1}^n$-measurable and, as a function of x, satisfies (18) for $\gamma = \gamma_l'$. Then if C_l below denotes a constant depending only on l and on the functions a, b, f, (19) and (28) yield

$$\text{(30)} \quad \mathrm{E}[[\mathrm{H}(\zeta, t_{i-1}, \mathrm{X}_i^n(l)) - \mathrm{H}(\zeta, t_{i-1}, \mathrm{Y}_i^n)]^2 \, 1_{\{t_{i-1} < \mathrm{R}_l\}} \mid \mathcal{G}_{i-1}^n] \leqq \mathrm{C}_l \, \delta_i.$$

Applying Schwarz inequality gives

$$\text{(31)} \quad |\mathrm{E}[\mathrm{H}(\zeta, t_{i-1}, \mathrm{X}_i^n(l)) - \mathrm{H}(\zeta, t_{i-1}, \mathrm{Y}_i^n) \mid \mathcal{G}_{i-1}^n] \, 1_{\{t_{i-1} < \mathrm{R}_l\}}| \leqq \sqrt{\mathrm{C}_l \, \delta_i}.$$

Since $\sup_{t \leqq \mathrm{R}_l, \, \zeta \in \Theta, \, \omega \in \Omega} |c(\zeta, t, \mathrm{X}_t)| < \infty$, we get

$$\text{(32)} \qquad \mathrm{E}[|\mathrm{H}(\zeta, t_{i-1}, \mathrm{Y}_i^n)| \, 1_{\{t_{i-1} < \mathrm{R}_l\}} \mid \mathcal{G}_{i-1}^n] \leqq \mathrm{C}_l.$$

Finally, apply (20) and (28) to obtain, under C2 (which implies that H is even in x) and if τ_l denotes the function τ of (17) for the coefficients (a_l, b_l):

$$\text{(33)} \quad |\mathrm{E}[\mathrm{H}(\zeta, t_{i-1}, \mathrm{X}_i^n(l)) - \mathrm{H}(\zeta, t_{i-1}, \mathrm{Y}_i^n) \mid \mathcal{G}_{i-1}^n] \, 1_{\{t_{i-1} < \mathrm{R}_l\}}| \leqq \mathrm{C}_l \sqrt{\delta_i \, \tau_l(\delta_i)}.$$

4. CONVERGENCE OF CONTRASTS

Recall once more that ϑ, b, ν are fixed and $\mathrm{P} = \mathrm{P}_{\nu, b}^{\vartheta}$. Let us first prove a more or less well-known result about triangular arrays of random variables.

LEMMA 9. – *Let χ_i^n, U be random variables, with χ_i^n being $\mathcal{G}_i^n$-measurable. The following two conditions imply $\sum_{i=1}^n \chi_i^n \xrightarrow{\mathrm{P}} \mathrm{U}$:*

$$\text{(33)} \qquad \sum_{i=1}^n \mathrm{E}(\chi_i^n \mid \mathcal{G}_{i-1}^n) \xrightarrow{\mathrm{P}} \mathrm{U},$$

$$\text{(34)} \qquad \sum_{i=1}^n \mathrm{E}[(\chi_i^n)^2 \mid \mathcal{G}_{i-1}^n] \xrightarrow{\mathrm{P}} 0.$$

Proof. – Set

$$\xi_i^n = \chi_i^n - \mathrm{E}(\chi_i^n \mid \mathcal{G}_{i-1}^n), \qquad \mathrm{B}_n = \sum_{1 \leqq i \leqq n} \xi_i^n,$$

$$\mathrm{C}_n = \sum_{1 \leqq i \leqq n} \mathrm{E}[(\xi_i^n)^2 \mid \mathcal{G}_{i-1}^n], \qquad \text{and} \qquad \mathrm{D}_n = \sum_{1 \leqq i \leqq n} \mathrm{E}[(\chi_i^n)^2 \mid \mathcal{G}_{i-1}^n].$$

By (33) it is enough to prove $\mathrm{B}_n \xrightarrow{\mathrm{P}} 0$. But $\mathrm{E}(\xi_i^n \mid \mathcal{G}_{i-1}^n) = 0$ by construction, so $(\sum_{1 \leqq i \leqq .} \xi_i^n)^2 - \sum_{1 \leqq i \leqq .} \mathrm{E}[(\xi_i^n)^2 \mid \mathcal{G}_{i-1}^n]$ is a local martingale w.r.t. $(\mathcal{G}_i^n)_{0 \leqq i \leqq n}$. Hence we can apply Lenglart inequality (*see* e.g. [16], I-3.30), which gives

$P(B_n^2>a)\leq \frac{b}{a}+P(C_n>b)$ for all $a, b>0$. Since $C_n\leq D_n$, (34) yields $P(C_n>b)\to 0$ for all $b>0$, hence the result. ■

Proof of Theorem 1. – (*a*) For each $l\in\mathbb{N}$ we set $V_l^n(\zeta)=\sum_{1\leq i\leq n}\chi_i^n(l)$, with $\chi_i^n(l)=\frac{1}{n}[F(\zeta, t_{i-1}, X_i^n(l))-U_{t_{i-1}}(\vartheta, \zeta)]\, 1_{\{t_{i-1}<R_l\}}$. Observe that by (6) and (16) we have $U_{t_{i-1}}(\vartheta, \zeta)=E[F(\zeta, t_{i-1}, Y_i^n)\,|\,\mathcal{G}_{i-1}^n]$, hence (31) gives

$$|E(\chi_i^n(l)\,|\,\mathcal{G}_{i-1}^n)|\leq \frac{1}{n}\sqrt{C_l\,\delta_i}$$

similarly (30) and $\delta_i\leq 1$ and (32) applied to $H=F^2$ yield for some constant C_l':

$$E[\chi_i^n(l)^2\,|\,\mathcal{G}_{i-1}^n]\leq C_l' n^{-2}.$$

Since $\sum\sqrt{\delta_i}\leq\sqrt{n}$ we deduce from Lemma 9 that $V_l^n(\zeta)\xrightarrow{P} 0$.

(*b*) Further, if γ_l is the constant showing in § 3*b* and if we apply (32) to $H=\nabla_\zeta F$ and (6), we have $|\nabla_\zeta \chi_i^n(\zeta)|\leq \frac{1}{n}[\gamma_l(1+|X_i^n(l)|^{\gamma_l})+C_l]$. Then (29) yields $\sup_n E[\sup_\zeta |\nabla_\zeta V_l^n(\zeta)|]<\infty$: hence the convergence in (a) is uniform in ζ, *i.e.* $\sup_\zeta |V_l^n(\zeta)|\xrightarrow{P} 0$.

(*c*) Set $V^n(\zeta)=U^n(\zeta)-\bar{U}^n(\zeta)$, where $\bar{U}^n(\zeta)=\frac{1}{n}\sum_{1\leq i\leq n}U_{t_{i-1}}(\vartheta, \zeta)$. In view of (4) and (28), $V^n(\zeta)=V_l^n(\zeta)$ on the set Ω_l, so (*b*) and (26) imply $\sup_\zeta |V^n(\zeta)|\xrightarrow{P} 0$. Hence it remains to prove that $\bar{U}^n(\zeta)$ converges to $U(\vartheta, \zeta)$ (*see* Theorem 1) uniformly in ζ, for each ω. To see this, it is enough to notice first that $\sup_{\zeta, n}|\nabla\bar{U}^n(\zeta)|<\infty$ and $|\nabla_\zeta U(\vartheta, \zeta)|<\infty$, and second that since $\bar{U}^n(\zeta)=\int U_t(\vartheta, \zeta)\,\mu_n(ds)-\frac{1}{n}(U(\vartheta, \zeta)_{t_n}-U(\vartheta, \zeta)_0)$ and $t\to U_t(\vartheta, \zeta)$ is continuous, then $(\bar{U}^n(\zeta)\to U(\vartheta, \zeta)$ for every ζ. ■

Remark 10. – It will be useful for a forthcoming paper on random sampling to observe that we have not used here the full force of H1: if $\Omega_l'=\{\sup_t |X_t|\leq l\}$ (hence $P(\Omega_l')\to 1$), it would be enough to have $E[(X_i^n)^p 1_{\Omega_l'}]\leq K_{l,p}$ and also (30) and (32) on each Ω_l' for $H=F$ and $H=\nabla F$, with X_i^n instead of $X_i^n(l)$, and without the indicator $1_{\{t_{i-1}\leq R_l\}}$. Moreover, in (30) and (32) one could replace $\mathscr{G}^n_{i-1}$ by $\mathscr{F}^n_{i-1}=\mathscr{F}_{t(n,\,i-1)}$, hence we do not need $\mathscr{F}_t \subseteq \mathscr{G}_t$.

For example, suppose that we take for $f(G, x)$ a polynomial in x, with coefficients being C^1-functions of G. Then Theorem 1 and its Corollary 2 are still valid if H1 is replaced by the much weaker requirements that

(*a*) $\partial_\vartheta c$ exists and is locally bounded on $\Theta \times [0, 1] \times \mathbb{R}^d$ (hence c as well),

(*b*) there is a unique non-exploding weak solution,

(*c*) and $\left| E\left[\prod_{j=1}^{d} (X^{n,j}_i)^{k_j} - \prod_{j=1}^{d} (Y^{n,j}_i)^{k_j} \mid \mathscr{F}^n_{i-1}\right]\right| \leq C_l(k_1, \ldots, k_d)\sqrt{\delta_i}$ on Ω'_l for some constant $(C_l(k_1, \ldots, k_d)$ depending only on the integers $l, k_1, \ldots, k_d$. ■

5. ASYMPTOTIC MIXED NORMALITY

5*a*. Here we assume H1, H3, H4, C1, C2 and C3, and as before $P=P^\vartheta_{v,\,b}$.

The proof of Theorem 3 goes along a traditional route. We assume that ϑ is an interior point of Θ, and that $\hat{\vartheta}_n$ minimizes $U^n(.)$. Let A_n be the set where $\hat{\vartheta}_n$ belongs to the interior of Θ, so that Corollary 2 yields

$$P(A_n) \to 1. \tag{35}$$

Recall also that $S_n=\sqrt{n}(\hat{\vartheta}_n-\vartheta)$. On A_n we have $\nabla U^n(\hat{\vartheta}_n)=0$, hence by Taylor's formula

$$B_n S_n = -L_n \quad \text{on } A_n,$$

where

$$B_n=\int_0^1 \nabla^2 U^n(\vartheta+u(\hat{\vartheta}_n-\vartheta))\,du$$

$$L_n=\sqrt{n}\,\nabla U^n(\vartheta).$$

LEMMA 11. – *If* $B(\vartheta)$ *is a.s. invertible, the following two properties imply the claim of Theorem* 3*:*

$$B_n \xrightarrow{P} B(\vartheta), \tag{36}$$

(37) $(B_n, L_n) \xrightarrow{\mathscr{L}} (B(\vartheta), L)$ *where* L *is defined on an extension of the space* Ω *and is* $\mathscr{N}(0, D(\vartheta))$, *conditionally on* $\mathscr{G}_1$.

Proof. – $S=-B(\vartheta)^{-1}L$ has the specifications given in Theorem 3, because $\mathscr{F}_1 \subseteq \mathscr{G}_1$ and $\Gamma(\vartheta)$ is $\mathscr{F}_1$-measurable. So we only have to prove that $S_n \xrightarrow{\mathscr{L}} S$.

Let $A'_n = (\omega \in A_n : B_n(\omega)$ is invertible$\}$, and $B'_n = B_n$ on A'_n, and $B'_n = I_q$ (the $q \times q$ identity matrix) on A'^c_n. Then (35) and (36) yield $P(A'_n) \to 1$ so (37) implies $(B'_n, L_n) \xrightarrow{\mathscr{L}} (B(\vartheta), L)$. This in turn implies $B'^{-1}_n L_n \xrightarrow{\mathscr{L}} S$, and since $S_n = B'^{-1}_n L_n$ on A'_n the result follows from $P(A'_n) \to 1$. ■

LEMMA 12. — *Property* (36) *holds.*

Proof. — (*a*) First we remark that $\nabla^2 U^n(\vartheta) \xrightarrow{P} B(\vartheta)$: this is proved like in parts (*a*) and (*c*) of the proof of Theorem 1, upon substituting $F(\vartheta, .)$ with each component of $\nabla^2_\vartheta F(\vartheta, .)$.

(*b*) Since $\hat{\vartheta}_n \xrightarrow{P} \vartheta$ it remains to show that $N_n(\varepsilon_n) \xrightarrow{P} 0$ as $n \to \infty$ for any sequence $\varepsilon_n \to 0$, where $N_n(\varepsilon) = \sup_{\zeta : |\zeta| \leqq \varepsilon} |\nabla^2 U^n(\vartheta + \zeta) - \nabla^2 U^n(\vartheta)|$.

Let γ_l be as in § 3*b*. Set $\gamma'_l = \gamma_l + 1$ and $\bar{f}(G, x) = f(G, x)/(1 + |x|^{\gamma'_l})$. The functions $\bar{f}$, $\nabla_G \bar{f}$ and $\nabla^2_G \bar{f}$ are smaller than $3\gamma'_l/(1 + |x|)$, hence by C 1-(ii) the families of functions $G \to \bar{f}(G, x)$, $G \to \nabla_G \bar{f}(G, x)$ and $G \to \nabla^2_G \bar{f}(G, x)$ (indexed by $x \in \mathbb{R}^d$) are uniformly equi-continuous on $\mathscr{S}_l$. By H1 the functions c, $\nabla_\vartheta c$, $\nabla^2_\vartheta c$ are uniformly continuous on $\Theta \times [0, 1] \times \{x : |x| \leqq l\}$. Then

$$\eta_l(\varepsilon) = \sup_{|\zeta| \leqq \varepsilon,\, t \in [0, 1],\, x \in \mathbb{R}^d,\, \omega \in \Omega_l} |\nabla^2_\vartheta F(\vartheta + \zeta, t, x, \omega) - \nabla^2_\vartheta F(\vartheta, t, x, \omega)| / (1 + |x|^{\gamma'_l})$$

has $\eta_l(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Now $N_n(\varepsilon) \leqq \eta_l(\varepsilon) Z^l_n$ on Ω_l, where $Z^l_n = \frac{1}{n} \sum_{1 \leqq i \leqq n} (1 + |X^n_i(l)|^{\gamma'_l})$. By (29) each sequence $(Z^l_n)_{n \in \mathbb{N}}$ is bounded in $L^1(P)$, hence if $\varepsilon_n \to 0$ we have $N_n(\varepsilon_n) 1_{\Omega_l} \leqq \eta_l(\varepsilon_n) Z^l_n \xrightarrow{P} 0$ as $n \to \infty$. The result follows from (26). ■

As a first step to obtain (37), we set

$$L'_n = \sum_{i=1}^{n} \xi^n_i, \qquad \text{where} \quad \xi^n_i = n^{-1/2} \nabla_\vartheta F(\vartheta, t_{i-1}, Y^n_i). \tag{38}$$

LEMMA 13. — *We have* $L'_n - L_n \xrightarrow{P} 0$.

Proof. — Fix $j \leqq q$. For each $l \in \mathbb{N}$ we set $V^n_l = \sum_{1 \leqq i \leqq n} \chi^n_i(l)$, with

$$\chi^n_i(l) = n^{-1/2} \left[\frac{\partial F}{\partial \vartheta_j}(\vartheta, t_{i-1}, X^n_i(l)) - \frac{\partial F}{\partial \vartheta_j}(\vartheta, t_{i-1}, Y^n_i) \right] 1_{\{t_{i-1} < R_l\}}.$$

By (28) we have $V_l^n = L_n^j - L_n'^j$ on Ω_l. Hence by (26) it is enough to prove that $V_l^n \xrightarrow{P} 0$ for each l. Since C2 holds, we can apply (30) and (33) to $H = \partial F / \partial \vartheta_j$ to obtain:

$$|E(\chi_i^n(l) \,|\, \mathscr{G}_{i-1}^n)| \leqq n^{-1/2} C_l \sqrt{\delta_i \tau_l(\delta_i)},$$
$$E[\chi_i^n(l)^2 \,|\, \mathscr{G}_{i-1}^n] \leqq C_l \frac{1}{n} \delta_i.$$

Since $\sum \delta_i \leqq 1$, we can apply Lemma 9 to obtain $V_l^n \xrightarrow{P} 0$, provided

$$\alpha_n := n^{-1/2} \sum_{i=1}^{n} \sqrt{\delta_i \tau_l(\delta_i)} \xrightarrow{P} 0 \quad \text{as } n \to \infty. \tag{39}$$

We have $\sum_{1 \leqq i \leqq n} (\delta_i^n)^{1/2} \leqq \sqrt{n}$ and $\operatorname{card}\{i : \delta_i^n > n^{-1/4}\} \leqq n^{1/4}$. Then the left-hand side of (39) is smaller than $n^{-1/4} \tau_l(1)^{1/2} + \tau_l(n^{-1/4})^{1/2}$, so that (39) follows from (17) and the Lemma is proved. ■

5*b*. At this point, and in view of Lemmas 11, 12, 13, it remains to prove:

(40) $(B_n, L_n') \xrightarrow{\mathscr{L}} (B(\vartheta), L)$ where L is defined on an extension of the space Ω and is $\mathcal{N}(0, D(\vartheta))$, conditionally on $\mathscr{G}_1$.

To get an idea about how to get (40), let us first note that by C3, the first partial derivatives of $G' \to \int f(G', x) \rho_G(dx)$ vanish at $G' = G$. Using notation (5), this writes as

$$\int \nabla_\vartheta F(\vartheta, t, x) \rho_{\vartheta, t}(dx) = 0.$$

Then, recalling (7), we deduce from (16) and (32) applied to $H = |\nabla_\vartheta F|^4$ that

$$\left\{\begin{array}{l} E(\xi_i^n \,|\, \mathscr{G}_{i-1}^n) = 0 \\ E(\xi_i^{n,j} \xi_i^{n,k} \,|\, \mathscr{G}_{i-1}^n) = \dfrac{1}{n} D(\vartheta)_{t_{i-1}}^{jk} \\ E(|\xi_i^n|^4 \,|\, \mathscr{G}_{i-1}^n) \leqq n^{-2} C_l \quad \text{on } \Omega_l, \end{array}\right. \tag{41}$$

and thus [recall H3 and (26)]:

$$(42)\qquad \begin{cases} \sum_{i=1}^{n} \mathrm{E}(\xi_i^n \mid \mathscr{G}_{i-1}^n) \xrightarrow{\mathrm{P}} 0 \\ \sum_{i=1}^{n} \mathrm{E}(\xi_i^{n,j}\xi_i^{n,k} \mid \mathscr{G}_{i-1}^n) \xrightarrow{\mathrm{P}} \mathrm{D}(\vartheta) \\ \sum_{i=1}^{n} \mathrm{E}(|\xi_i^n|^2 1_{\{|\xi_i^n|>\varepsilon\}} \mid \mathscr{G}_{i-1}^n) \xrightarrow{\mathrm{P}} 0. \end{cases}$$

Then we are under the conditions for the convergence of L'_n to an $\mathcal{N}(0, \mathrm{D}(\vartheta))$ vector, except that $\mathrm{D}(\vartheta)$ is random. Further, we do not have the "nesting" condition of Hall and Heyde [14] on the filtrations $(\mathscr{G}_i^n)$ which are necessary to accomodate random limits in (42). So we need a different sort of proof, which goes as follows:

Suppose that $t_i^n = i/n$. Consider the process $\mathrm{V}_t^n = \sum_{1 \leq i \leq [nt]} \xi_i^n$, where $[.]$ denotes the integer part. Then summing up to $[nt]$ in (42) we get convergence for all t: this does not *a priori* imply the convergence of the processes V^n, but it does *imply their tightness*. Taking a convergent subsequence, *we can identify the limit through some martingale characterization.* The same idea works if t_i^n is not i/n, provided the distribution function of the limiting measure μ in H3 is continuous and strictly increasing.

In general, the measure μ in H3 is arbitrary. So we need first to add (fictitious) observation times, and then change the time-scale, so as to obtain a suitable modification of μ. This is a bit complicated, so for the reader interested only in regular sampling we first give the proof in this simple case: then § 5 *c* can be skipped.

5 *c*. *The regular sampling case.* – In this subsection we assume that $t(n, i) = i/n$. For consistency with the proof in the general case, we use here slightly complicated notation, to be presently introduced [recall (38) for ξ_i^n]:

$$\Phi_n(t) = \frac{[nt]}{n}, \qquad \mathscr{H}_t^n = \mathscr{G}_{\Phi_n(t)}, \qquad \alpha_i^n = \mathrm{W}_{i/n} - \mathrm{W}_{(i-1)/n},$$
$$\mathrm{W}_t'^n = \sum_{i=1}^{[nt]} \alpha_i^n = \mathrm{W}_{\Phi_n(t)}, \qquad \mathrm{V}_t^n = \sum_{i=1}^{[nt]} \xi_i^n.$$

The processes W'^n and V^n are clearly martingales relatively to the piecewise constant filtration $(\mathscr{H}_t^n)$: use (41) for V^n. We introduce their brackets:

$$(43)\qquad \begin{cases} \mathrm{G}^{n,jk} = \langle \mathrm{W}'^{n,j}, \mathrm{W}^{n,k} \rangle, \qquad \mathrm{H}^{n,jk} = \langle \mathrm{V}^{n,j}, \mathrm{V}^{n,k} \rangle, \\ \mathrm{K}^{n,jk} = \langle \mathrm{W}'^{n,j}, \mathrm{V}^{n,k} \rangle, \end{cases}$$

and write

$$G^n=(G^{n,jk})_{1\leqq j,k\leqq m},\qquad H^n=(H^{n,jk})_{1\leqq j,k\leqq q},$$
$$K^n=(K^{n,jk})_{1\leqq j\leqq m,\ 1\leqq k\leqq q}.$$

Observe that (with I_m = the identity $m\times m$ matrix):

$$G_t^n=I_m\,\Phi_n(t),\qquad H_t^n=\sum_{i=1}^{[nt]}D(\vartheta)_{(i-1)/n},\qquad K_t^n=0$$

[the first equality is trivial, the second one follows from (41), the third one comes from $E(\alpha_i^{n,j}\xi_i^{n,k}\mid\mathscr{G}_{i-1}^n)=0$, which in turn comes out from the fact that $\alpha_i^{n,j}\xi_i^{n,k}$ is an odd function of α_i^n, since f is an even function of x.]

Therefore

(44) W'^n, G^n, H^n, K^n converge uniformly in time on [0, 1] (for all ω) to W', G, H, K, where $W'=W$, $G_t=I_m\,t$, $H_t=\int_0^t D(\vartheta)_s\,ds$, $K=0$.

We consider the terms $\tau_n=(B_n, W'^n, G^n, H^n, K^n, V^n)$, taking their values in $\Omega'=\mathbb{R}^{q\times q}\times\mathbb{D}([0,1],\mathbb{R}^m\times\mathbb{R}^{m\times m}\times\mathbb{R}^{q\times q}\times\mathbb{R}^{m\times q}\times\mathbb{R}^q)$; this space is endowed with the product topology, with $\mathbb{D}(.)$ equipped with the Skorohod topology. Below, convergence in law refers to this topology. We also denote by $(\bar{B}, \bar{W}', \bar{G}, \bar{H}, \bar{K}, \bar{V})$ the canonical variable on this space Ω'.

Now we are ready to give the *essential steps* for proving Theorem 3:

(*a*) Because of (44) we can apply Aldous' tightness criterion to the V^n's: the sequence (V^n) is tight (*cf.* Aldous [1], or Theorem VI-4.13 of [16]). Furthermore, the last property in (42) is a Lindeberg condition which, translated in terms of V^n, asserts that the supremum of the jumps of V^n goes to 0 as $n\to\infty$: then the sequence (V^n) is even C-tight.

Combining this with (44) and Lemma 11, we obtain that the sequence (τ_n) is tight. Then, *up to taking a subsequence*,

(45) The laws of τ_n weakly converge to a probability measure P' on Ω', such that:

(i) $(\bar{B}, \bar{W}', \bar{G}, \bar{H}, \bar{K})$ has under P' the same law as $(B(\vartheta), W', G, H, K)$ (recall that $K=0$);

(ii) $\bar{W}'$, $\bar{G}$, $\bar{H}$, $\bar{K}$, $\bar{V}$ are P'-a.s. continuous in time.

(*b*) Denote by $(\mathscr{F}_t')$ the filtration generated by $(\bar{W}', \bar{G}, \bar{H}, \bar{K}, \bar{V})$ on Ω'.

LEMMA 14. – *On the space* $(\Omega', (\mathscr{F}_t'), P')$, $\bar{W}'$ *and* $\bar{V}$ *are local martingales with brackets*

$$\langle \bar{W}'^j, \bar{W}'^k\rangle=\bar{G}^{jk},\qquad \langle \bar{V}^j, \bar{V}^k\rangle=\bar{H}^{jk},\qquad \langle \bar{W}'^j, \bar{V}^k\rangle=\bar{K}^{jk}(=0\ a.s.).$$

Proof. – We have to prove that all processes $\bar{W}'^j$, $\bar{W}'^j\bar{W}'^k-\bar{G}^{jk}$, etc. are local martingales. We prove this for $\bar{M}=\bar{V}^j\bar{V}^k-\bar{H}^{jk}$ (for the others it is the same, or simpler).

Let $M^n=V^{n,j}V^{n,k}-H^{n,jk}$ and also $A_t^n=\sum_{1\leq i\leq n\Phi_n(t)} E(|\xi_i^n|^4\,|\,\mathscr{G}_{i-1}^n)$, and

$$S(n,y)=\inf(t:|M_t^n|+|V_t^n|>y),$$
$$S(y)=\inf(t:|\bar{M}_t|+|\bar{V}_t|>y),$$
$$T(n,y)=\inf(t:A_t^n>y).$$

Observe that $R(n,y)=S(n,y)\wedge\left[T(n,y)-\frac{1}{n}\right]^+$ is a stopping time and that $|M_t^n|\leq y$ if $t<S(n,y)$ and

$$|M_{S(n,y)}^n|\leq y+2y^2+2|\xi_i^n|^2+E(|\xi_i^n|^2\,|\,\mathscr{G}_{i-1}^n)$$

if $S(n,y)=i/n$. Thus $E[|M_{t\wedge R(n,y)}^n|^2]\leq 4y^2+16y^4+20y$, from which we deduce the uniform integrability of each sequence $(M_{t\wedge R(n,y)}^n)_{n\geq 1}$.

Set $\bar{\tau}=(\bar{B},\bar{W}',\bar{G},\bar{H},\bar{K},\bar{V})$. Combining (45) and the last inequality in (41) yields the convergence in law of (τ_n, M^n, A^n) to $(\bar{\tau},\bar{M},0)$. Then (*see* e.g. Propositions VI-2.11 and VI-2.12 of [16]) for all y in a dense subset of $\mathbb{R}_+$ we have first that $\omega'\to\bar{M}_{.\wedge S(y)}(\omega')$ is P′-a.s. continuous for the Skorohod topology and second that $(\tau_n, M^n_{.\wedge R(n,y)}, R(n,y))$ converges in law as $n\to\infty$ to $(\bar{\tau},\bar{M}_{.\wedge S(y)},S(y))$. From the uniform integrability above and from the fact that M^n is a local martingale on $(\Omega,(\mathscr{H}_t^n),P)$, we then deduce (*see* Proposition IX-1.12 of [16]) that for y in our above dense set, $\bar{M}_{.\wedge S(y)}$ is a martingale for the filtration generated by $(\bar{\tau},\bar{M}_{.\wedge S(y)})$, *i.e.* for $(\mathscr{F}_t')$. Since $S(y)\to\infty$ as $y\to\infty$, it follows that $\bar{M}$ is a local martingale. ■

(*c*) To fit once more our notation with the general case to be studied later, we introduce the measure η on $\mathbb{R}_+$ having $\eta([0,t])=t\wedge 1$ (so here it is the Lebesgue measure restricted to $[0,1]$).

We can always enlarge the space Ω' and the filtration $(\mathscr{F}_t')$ so that there is a continuous Gaussian martingale $\bar{Z}=(\bar{Z}^j)_{1\leq j\leq q}$ with $\bar{Z}_0=0$ and $\langle\bar{Z}^j,\bar{Z}^k\rangle_t=\delta_{jk}\eta([0,t])$ (so here it is a Wiener process), which is independent of $(\bar{B},\bar{W}',\bar{G},\bar{H},\bar{K},\bar{V})$. We denote by $(\mathscr{G}_t')$ the filtration on Ω' generated by $\bar{W}'$. Due to (44) and (45) and to the $(\mathscr{G}_t)$-predictability of $(D(\vartheta)_t)_{t\geq 0}$, we have $\bar{H}_t=\int_0^t h_s\,ds$ up to P′-null sets for some $\mathscr{S}_q^+$-valued process h which is predictable w.r.t. the filtration $(\mathscr{G}_t')$ on Ω' [*Remark:* this is the only place where we use the fact that the solution to (1) is strong].

Now we diagonalize h, writing $h_t=\pi_t^{\mathrm{T}}\lambda_t\pi_t$ where π_t is an orthogonal $q\times q$ matrix-valued process and λ_t is a diagonal $q\times q$ matrix-valued process, both predictable w.r.t. $(\mathscr{G}'_t)$. Set

$$\bar{Z}'^i_t=\int_0^t 1_{\{\lambda_s^{ii}>0\}}(\lambda_s^{ii})^{-1/2}\sum_{j=1}^q \pi_s^{ij}\,d\bar{V}_s^j+\int_0^t 1_{\{\lambda_s^{ii}=0\}}\,d\bar{Z}_s^i.$$

By standard calculations, we check that

$$\langle \bar{Z}'^i, \bar{Z}'^j\rangle_t=\delta_{ij}\,t,\qquad \langle \bar{Z}'^i, \bar{W}'^j\rangle_t=0,$$

so that on $(\Omega', (\mathscr{F}'_t), P')$:

(46) $\bar{Z}'=(\bar{Z}'^j)_{1\leq j\leq q}$ is a Gaussian continuous martingale, independent from $\bar{W}'$.

We also easily get

$$\bar{V}_t^i=\int_0^t\sum_{j=1}^q \pi_s^{ji}(\lambda_s^{jj})^{1/2}\,d\bar{Z}'^j_s.$$

Since π_s and λ_s are $(\mathscr{G}'_s)$-adapted, if follows from (46) and $\pi_t^{\mathrm{T}}\lambda_t\pi_t=h_t$ that

(47) Conditionally on $\mathscr{G}'_1$, $\bar{V}$ is a Gaussian continuous martingale with (deterministic) bracket $\bar{H}$.

(*d*) Now we are almost finished: the same argument as above shows us that $\bar{B}$ is $\mathscr{G}'_1$-measurable, and of course $\bar{G}$, $\bar{K}$ are the same. Then, using (47) and (45)-(i), we see that the measure P′ is uniquely determined. That is, the laws of τ_n converge to this unique P′. In particular, since $H_1=D(\vartheta)$ and $V_1^n=L'_n$, we deduce (40) from (45) and (47): that is, Theorem 3 is proved in the regular sampling case.

5 *d. Proof of Theorem* 3 *in the general case.* — Our first task is to add observation points outside the support of the limiting measure μ in H3 and change time. For this we change the time interval from [0, 1] to $\mathbb{R}_+$, and all measures below (μ_n, μ, etc.) are considered as defined on $\mathbb{R}_+$.

Let D be the support of μ and $D^\varepsilon=\{x\in\mathbb{R}_+:\ d(x, D)\geqq\varepsilon\}$. Since $\mu(D^\varepsilon)=0$ we have $\mu_n(D^\varepsilon)\to 0$ for all $\varepsilon>0$, hence there is a sequence (n_p) increasing to $+\infty$, such that $\mu_n(D^{1/p})\leqq 1/p$ for $n\geqq n_p$. Set $p_n=\sup(p: n_p\leqq n)$ and $\alpha(n)=n^{-1/2}+1/p_n$, and denote by (u_k, v_k) the intervals contiguous to D, in $\mathbb{R}_+$. Set

$$\left\{\begin{aligned} &C_n=\bigcup_{k\geq 1}\{s=u_k+i/n \text{ for } i\in\mathbb{N} \text{ and } u_k-\alpha(n)\leqq s\leqq v_k-\alpha(n)\}\\ &J_n=\{i: 1\leqq i\leqq n,\ C_n\cap[t(n, i-1), t(n, i)]=\varnothing\},\qquad J'_n=\{1, \ldots, n\}\setminus J_n,\\ &C'_n=\{t(n, i): 1\leqq i\leqq n\},\qquad C''_n=\{t(n, i): i\in J_n\}.\end{aligned}\right.$$

If $i \in J'_n$, then either $\delta(n, i) \geqq \alpha(n)$ [the number of these being smaller than $\sqrt{n}$ because $\sum_i \delta(n, i) \leqq 1$], or $t(n, i) \in D^{1/p_n}$. Thus

$$\mathrm{card}\,(J'_n) \leqq \sqrt{n} + n\,\mu_n(D^{1/p_n}) \leqq \sqrt{n} + n/p_n$$

and so

$$\frac{1}{n}\,\mathrm{card}\,(J'_n) \to 0, \qquad \frac{1}{n}\,\mathrm{card}\,(J_n) \to 1. \tag{48}$$

Next, set $t'(n, 0) = 0$ and call $t'(n, 1) < \ldots < t'(n, i) < \ldots$ the points of $C_n \cup C''_n$. Set

$$\begin{gathered} \delta'(n, i) = t'(n, i) - t'(n, i-1), \qquad t''(n, i) = t'(n, i) + \frac{i}{n}, \\ \Sigma(n, t) = \{ i : t''(n, i) \leqq t \}, \qquad \sigma(n, t) = \mathrm{card}\,(\Sigma(n, t)), \\ K_n = \{ i : \text{there exists } j(n, i) \in J_n \text{ with } t'(n, i) = t(n, j(n, i)) \}, \\ \Phi_n(t) = \sum_{i \in \Sigma(n, t)} \delta'(n, i) = t'(n, \sigma(n, t)). \end{gathered} \tag{49}$$

The next two lemmas gather all necessary results on the changes of time.

LEMMA 15. — *Φ_n converges to a continuous non-decreasing function Φ.*

Proof. — (*a*) Consider the following measures on $\mathbb{R}_+$:

$$\left\{ \begin{aligned} &\mu'_n = \frac{1}{n} \sum_{i \geqq 1} \varepsilon_{t'(n, i)} \qquad \mu''_n = \frac{1}{n} \sum_{i \geqq 1} \varepsilon_{t''(n, i)} \\ &\lambda_n = \frac{1}{n} \sum_{s \in C_n} \varepsilon_s \qquad \nu_n = \frac{1}{n} \sum_{s \in C''_n} \varepsilon_s. \end{aligned} \right. \tag{50}$$

In view of (48) and H3, we have $\nu_n \to \mu$ weakly, while $\lambda_n \to 1_{D^c}.\lambda$ weakly (λ denotes the Lebesgue measure on $\mathbb{R}_+$) by definition of C_n. Then $\mu'_n = \nu_n + \lambda_n \to \mu' := \mu + 1_{D^c}.\lambda$.

Denote by F'_n, F''_n, F' the distribution functions of μ'_n, μ''_n, μ' respectively, and by $F'^{-1}_n, \ldots$ their right-continuous inverses.

(*b*) From $\mu'_n \to \mu'$ we have $F'_n(t) \to F'(t)$ for all continuity points t of F'. Since F' is strictly increasing, we deduce $F'^{-1}_n(t) \to F'^{-1}(t)$ for all t. Now, $F'^{-1}_n(t) = t'(n, [nt]+1)$ and

$$F''^{-1}_n(t) = t''(n, [nt]+1) = t'(n, [nt]+1) + ([nt]+1)/n;$$

thus $F''^{-1}_n(t) \to F''^{-1}(t) := F'^{-1}(t) + t$ for all t. The function F''^{-1} is strictly increasing and continuous, so its right-continuous inverse F'' has the same properties, and $F''_n(t) \to F''(t)$ for all t (then $\mu''_n \to \mu''$ weakly, if μ'' is the measure having F'' for distribution function).

(c) It remains to observe that $F''_n(t)=\sigma(n, t)/n$, so that

$$\Phi_n(t)=F'^{-1}_n\left[\left(F''_n(t)-\frac{1}{n}\right)^+\right].$$

Since the convergence $F'^{-1}_n \to F'^{-1}$ is locally uniform (because F'^{-1} is continuous), it follows that $\Phi_n(t)\to\Phi(t)$, where $\Phi(t)=F'^{-1}(F''(t))$ is clearly continuous and non-decreasing. ■

Next, consider the measures

$$\eta_n=\frac{1}{2}\sum_{i\in K_n}\varepsilon_{t''(n, i)}.$$

We have $\eta_n(\mathbb{R}_+)\leqq 1$ and the support of η_n is included in [0,3]: then the sequence (η_n) is tight for the narrow topology. We could prove that η_n does converge, but the following will be enough:

LEMMA 16. – *Consider a subsequenc $\eta_{n'}$ converging to η. Then η has no atom, and if f is a continuous bounded function on $\mathbb{R}_+$ we have*

$$\frac{1}{n'}\sum_{i\in K_{n'}\cap\Sigma(n', t)} f(t'(n, i-1))\to\int_0^t f(s)\,\mu(ds), \tag{51}$$

$$\int_0^\infty f[\Phi(s)]\,\eta(ds)=\int_0^\infty f(s)\,\mu(ds). \tag{52}$$

Proof. – (a) The measure η_n is smaller than the measure μ''_n of (50); since $\mu''_n\to\mu''$ where μ'' has no atom, we deduce that η has no atom.

In order to prove(51) we can of course replace $t'(n, i-1)$ by $t'(n, i)$ in the left-hand side, because $\sup_i \delta'(n, 1)\to 0$ as $n\to\infty$. Then this left-hand side becomes

$$\int_0^t f[\Phi_{n'}(s)]\,\eta_{n'}(ds).$$

Since $\Phi_n\to\Phi$ locally uniformly and $\eta_{n'}\to\eta$ weakly, (51) is obvious. Further the above integral is the same for all $t\geqq 3$, in which case it also equals $\int f(s)\,\nu_n(ds)$ [*see* again (50)]. Since $\nu_n\to\mu$, we deduce (52) from (51). ■

Now we can start proving Theorem 3. We will simply indicate the modifications which should be done with respect to § 5 *b*.

Without loss of generality we can assume that the Wiener process W is defined on $\mathbb{R}_+$. Set

$$\alpha_i^n=W_{t_i'^n}-W_{t_{i-1}'^n},\qquad \beta_i^n=\begin{cases}\xi_i^n & \text{if } i\in K_n\\ 0 & \text{if not.}\end{cases}$$

$$W_t'^n = \sum_{i \in \Sigma(n, t)} \alpha_i^n = W_{\Phi_n(t)}, \qquad V_t^n = \sum_{i \in \Sigma(n, t)} \beta_i^n.$$

Consider also the filtration $\mathscr{H}_t^n = \mathscr{G}_{\Phi_n(t)}$. The processes W'^n and V^n are still martingales relatively to the filtration $(\mathscr{H}_t^n)$: for V^n, use again (41), and also the following fact [*cf.* (49)]: if $i \in K_n$ and $t = t''(n, i)$, then $\Phi_n(t) = t'(n, i)$ and $\mathscr{H}_{t-}^n = \mathscr{G}_{j(n, i)-1}^n$ and $\Delta V_t^n = \xi_{j(n, i)}^n$. The brackets defined by (43) are now

$$G_t^n = I_m \Phi_n(t), \qquad H_t^n = \frac{1}{n} \sum_{i \in K_n \cap \Sigma(n, t)} D(\vartheta)_{t'(n, i-1)}, \qquad K_t^n = 0;$$

In view of Lemma 16, and up to taking a subsequence, we can assume that the sequence η_n converges weakly to a limit η.

All the above processes being constant after time 3, we can proceed, *replacing everywhere the time* 1 *by* 3. We have (44) with $G_t = I_m \Phi(t)$, $H_t = \int_0^t D(\vartheta)_{\Phi(s)} \eta(ds)$ (*see* Lemma 16), $K = 0$. Aldous' tightness criterion still applies here, so up to taking a further subsequence we still have (45). Then Lemma 16 holds. Part (*c*) of § 5*b* applies [except that $\bar{Z}$ is no longer a Wiener process, and that $\langle \bar{Z}'^i, \bar{Z}'^j \rangle = \delta_{ij} \eta([0, t])$].

As in Part (*d*), we arrive at $(B_n, V_3^n) \xrightarrow{\mathscr{L}} (B(\vartheta), L)$, but $V_3^n = L_n$ is no longer true. However, $L_n' - V_3^n = \sum_{i \in J_n'} \xi_i^n$, and (41) and (48) readily allow to apply Lemma 9 with $U = 0$, so that $L_n' - V_3^n \xrightarrow{P} 0$ and (40) follows.

5*e*. *Proof of Proposition* 4. — Up to considering $\mathscr{S}'' = \mathscr{S} \cap \mathscr{S}'$ and $\mathscr{S}_l'' = \mathscr{S}_l \cap \mathscr{S}_l'$, which meet C1 (i), we can assume that both f, f' are defined on the same product $\mathscr{S} \times \mathbb{R}^d$. We add a "prime" to all quantities relative to f'.

Then, it is enough to reproduce the proof of Theorem 3, with the following modifications: in (35) replace A_n by the set $\bar{A}_n$ where both $\hat{\vartheta}_n$ and $\hat{\vartheta}_n'$ are in the interior of Θ. Then $\bar{B}_n \bar{S}_n = -\bar{L}_n$ on $\bar{A}_n$, with the following vectors and matrices:

$$\bar{B}_n = \begin{bmatrix} B_n & 0 \\ 0 & B_n' \end{bmatrix}, \qquad \bar{S}_n = \begin{bmatrix} S_n \\ S_n' \end{bmatrix}, \qquad \bar{L}_n = \begin{bmatrix} L_n \\ L_n' \end{bmatrix}.$$

Lemme 12 applied to B_n and B_n' separately yields $\bar{B}_n \xrightarrow{P} \bar{B}(\vartheta)$, and the rest of the proof is exactly similar, with q replaced by $2q$ and $D(\vartheta)_t$ by $\bar{D}(\vartheta)_t$ and $\nabla_\vartheta F$ by the vector

$$\overline{\nabla_\vartheta F} = \begin{bmatrix} \nabla_\vartheta F \\ \nabla_\vartheta F' \end{bmatrix}.$$

6. COMPARISON BETWEEN CONTRASTS

Our aim here is to prove Theorem 5. Below, we assume H2. We take $\mathcal{S}=\mathcal{S}_d^{++}$, which is the union of all $\mathcal{S}_l=\{G\in\mathcal{S}_d^+ : 1/l\leqq \det(G)\leqq l\}$, so that by H1 and H2 we obviously have C1 (i). It is also evident that the function f of (12) satisfies C1 (ii), relative to these $\mathcal{S}_l$, and C2.

Let G be a C2 function on Θ, with values in $\mathcal{S}=\mathcal{S}_d^{++}$, and write $H(\vartheta)=G(\vartheta)^{-1}$. For each ϑ write $\rho_\vartheta=\rho_{G(\vartheta)}$. Let f be given by (12), and set

$$(53)\qquad \begin{cases} F(\vartheta, x) := f(G(\vartheta), x)=\log\det G(\vartheta)+x^T H(\vartheta)x, \\ v(\vartheta,\zeta)=\int F(\zeta, x)\,\rho_\vartheta(dx). \end{cases}$$

If λ denotes Lebesgue measure on $\mathbb{R}^d$, then $Z(\vartheta)=d\rho_\vartheta/d\lambda$ has the following form:

$$(54)\qquad Z(\vartheta, x)=(2\pi)^{-d/2}e^{-F(\vartheta, x)/2}.$$

LEMMA 17. – *The function f of* (12) *satisfies* C3.

Proof. – Let $U=Z(\zeta)/Z(\vartheta)$. Then (53) and (54) yield

$$\int \log U(x)\,\rho_\vartheta(dx)=\frac{1}{2}[v(\vartheta,\vartheta)-v(\vartheta,\zeta)].$$

Now, $\int U(x)\,\rho_\vartheta(dx)=1$; hence by Jensen's inequality $v(\vartheta,\vartheta)\leqq v(\vartheta,\zeta)$, with equality iff $U=1$ ρ_ϑ-a. s., that is iff $\rho_\vartheta=\rho_\zeta$, that is iff $G(\zeta)=G(\vartheta)$. Applying this to $G(\vartheta)=G$ and $g(\zeta)=G'$ in C3 gives the result. ■

Remark. – With a slightly different formulation, this lemma is well known since $\int \log U(x)\,\rho_\vartheta(dx)$ is indeed the Kullback information of ρ_ζ w.r.t. ρ_ϑ. Accordingly, the next lemma reduces in fact to the explicit computation of the Fisher information matrix $I(\vartheta)$ of the model $(\rho_\zeta)_{\zeta\in\Theta}$ at ϑ, which is $I(\vartheta)=\beta/2=\delta/4$ with the notation below. ■

Lemma 17 implies in particular $\nabla_\zeta v(\vartheta,\zeta)=0$ for $\zeta=\vartheta$, that is

$$(55)\qquad \int \nabla_\vartheta F(\vartheta, x)\,Z(\vartheta, x)\,dx=0.$$

Differentiating in ϑ and using (54), we obtain

$$(56)\qquad \int\left[\nabla_\vartheta^2 F(\vartheta, x)-\frac{1}{2}\nabla_\vartheta F(\vartheta, x)\nabla_\vartheta F(\vartheta, x)^T\right]\rho_\vartheta(dx)=0.$$

Now we fix ϑ and we set

$$(57)\quad \beta=\int \nabla_\vartheta^2 F(\vartheta, x)\,\rho_\vartheta(dx), \qquad \delta=\int \nabla_\vartheta F(\vartheta, x)\,\nabla_\vartheta F(\vartheta, x)^T\,\rho_\vartheta(dx).$$

LEMMA 18. – *We have* $\delta=2\beta$ *and*

$$\beta^{uv}=\operatorname{tr}\left[\frac{\partial G}{\partial\vartheta_u}(\vartheta)\,H(\vartheta)\,\frac{\partial G}{\partial\vartheta_v}(\vartheta)\,H(\vartheta)\right].$$

Proof. – That $\delta=2\beta$ follows from (56) and (57). We can also write (56) as $\int [Z(\vartheta, x)\,\nabla_\vartheta^2 F(\vartheta, x)+\nabla_\vartheta Z(\vartheta, x)\,\nabla_\vartheta F(\vartheta, x)^T]\,dx=0$, hence

$$\beta=-\int \nabla_\vartheta Z(\vartheta, x)\,\nabla_\vartheta F(\vartheta, x)^T\,dx.$$

Now $\int \nabla_\vartheta Z(\vartheta, x)\,g(x)\,dx$ is the gradient of $\vartheta \to \rho_\vartheta(g)$. In particular, $\int \nabla_\vartheta Z(\vartheta, x)\,dx=0$ and $\int \nabla_\vartheta Z(\vartheta, x)\,x^T K x\,dx=\operatorname{tr}[K\,\nabla G(\vartheta)]$ for every $d\times d$ matrix K. Observing that $\nabla_\vartheta F(\vartheta, x)=\nabla_\vartheta(\log\det G(\vartheta))+x^T\,\nabla_\vartheta H(\vartheta)\,x$, we get

$$\beta^{uv}=-\operatorname{tr}\left[\frac{\partial G}{\partial\vartheta_u}(\vartheta)\,\frac{\partial H}{\partial\vartheta_v}(\vartheta)\right].$$

But $G(\zeta)\,H(\zeta)=I_d$, hence $\partial H/\partial\vartheta_v=-H\,\partial G/\partial\vartheta_v\,H$ and the result follows. ■

Next, we consider a second function f' satisfying C1, C2 and C3. As above, we write $F'(\vartheta, x)=f'(G(\vartheta), x)$, and we set

$$(58)\quad \begin{cases} \beta'=\int \nabla_\vartheta^2 F'(\vartheta, x)\,\rho_\vartheta(dx), \\ \delta'=\int \nabla_\vartheta F'(\vartheta, x)\,\nabla_\vartheta F'(\vartheta, x)^T\,\rho_\vartheta(dx) \\ \delta''=\int \nabla_\vartheta F(\vartheta, x)\,\nabla_\vartheta F'(\vartheta, x)^T\,\rho_\vartheta(dx). \end{cases}$$

Since f' meets C3, we see that (55) is satisfied with F′ instead of F. Differentiating in ϑ gives $\delta''=2\beta'$.

Proof of Theorem 5. – For the claim (*a*) it remains to prove (13). To this effect, we apply Lemma 18 to each family $G(\vartheta)=c(\vartheta, t, X_t)$, for which $\beta=B(\vartheta)_t$ and $\delta=D(\vartheta)_t$.

(*b*) Use again $G(\vartheta)=c(\vartheta, t, X_t)$, and add the subscript "t" to all above quantities: β_t, δ_t, β'_t, δ'_t, δ''_t. Then we write $\underline{\beta}$, $\underline{\delta}$, etc. for the integrals of these functions w. r. t. the measure μ. Comparing (57) and (58) with (9) (10), we get

$$\bar{B}(\vartheta)=\begin{bmatrix}\underline{\beta} & 0\\ 0 & \underline{\beta}'\end{bmatrix},\qquad \bar{D}(\vartheta)=\begin{bmatrix}\underline{\delta} & \underline{\delta}''\\ \underline{\delta}''^{T} & \underline{\delta}'\end{bmatrix},$$

which yields [see (13) and use $\underline{\delta}''=2\underline{\beta}'$ and $\underline{\delta}=2\underline{\beta}$]:

$$\bar{\Gamma}(\vartheta)=\begin{bmatrix}\underline{\beta}^{-1}\underline{\delta}\,\underline{\beta}^{-1} & \underline{\beta}^{-1}\underline{\delta}''\underline{\beta}'^{-1}\\ \underline{\beta}'^{-1}\underline{\delta}''^{T}\underline{\beta}^{-1} & \underline{\beta}'^{-1}\underline{\delta}'\underline{\beta}'^{-1}\end{bmatrix}=\begin{bmatrix}2\underline{\beta}^{-1} & 2\underline{\beta}^{-1}\\ 2\underline{\beta}^{-1} & \underline{\beta}'^{-1}\underline{\delta}'\underline{\beta}'^{-1}\end{bmatrix}.$$

This is the conditional covariance of the pair (S, S'), and a simple computation shows that the conditional covariance of the pair (S, S' − S) is the following block-diagonal matrix, which yields the claim (*b*) and Theorem 5:

$$\begin{bmatrix}2\underline{\beta}^{-1} & 0\\ 0 & \underline{\beta}'^{-1}\underline{\delta}'\underline{\beta}'^{-1}-2\underline{\beta}^{-1}\end{bmatrix}. \quad \blacksquare$$

7. EXAMPLES

7 *a*. *The one-dimensional case.* — Here we quickly study the case where $d=m=1$; for simplicity (but it is not essential), we also assume that $q=1$, and we write $\dot{c}, \ldots$ for the partial derivatives w. r. t. ϑ.

We have $\mathscr{S}_1^{+}=[0, \infty)$ and $\mathscr{S}_1^{++}=(0, \infty)$. If H2 holds, we can use the contrast based on the function $f(G, x)=\log G+x^2/G$ of (12), and (13) writes as

$$(59)\qquad D(\vartheta)_t=2\,B(\vartheta)_t,\qquad B(\vartheta)_t=(\dot{c}/c)^2(\vartheta, t, X_t).$$

In general (*i. e.* when H2 does not hold), we cannot use this f, but we can always use the following function:

$$(60)\qquad f'(G, x)=(x^2-G)^2.$$

Clearly, f' satisfies C1, C2 and C3, with $\mathscr{S}=\mathscr{S}_1^{+}$. Moreover, elementary computations yield [by (7) and (8)]:

$$(61)\qquad D'(\vartheta)_t=8\,(c\dot{c})^2(\vartheta, t, X_t),\qquad B'(\vartheta)_t=2\,\dot{c}^2(\vartheta, t, X_t).$$

If μ in H3 is a Dirac measure and if H2 holds, we deduce from (59) and (61) that the asymptotic conditional variances $\Gamma(\vartheta)$ and $\Gamma'(\vartheta)$ of the estimators $\hat{\vartheta}_n$ and $\hat{\vartheta}'_n$ based on f and f' respectively are equal. In view of Theorem 5 *b*, the following readily follows:

PROPOSITION 19. — *When* $d=m=q=1$ *and* H2 *holds, and if the limiting measure* μ *in* H3 *is a Dirac measure, then* (*with the above notation*)

$S_n = \sqrt{n}(\hat{\vartheta}_n - \vartheta)$ *and* $S'_n = \sqrt{n}(\hat{\vartheta}'_n - \vartheta)$ *have the same limiting distribution, and even* $S_n - S'_n$ *tends to* 0 *in* $P^{\vartheta}_{\nu, b}$*-measure.*

The heuristic conclusion that can be drawn from the above is that, even when μ is not a Dirac measure, the estimators based on f' cannot be drastically bad. Hence if H2 fails, we have reasonnable estimators.

7 *b. One-dimensional linear equation.* – Suppose here that $d=m=q=1$ and that a has the form $a(\vartheta, t, x) = A(\vartheta, t)x$, and that $\nu = \varepsilon_{x_0}$ for $x_0 \neq 0$, that is we consider the equation:

$$dX_t = b(t, X)\,dt + A(\vartheta, t) X_t\,dW_t \qquad X_0 = x_0 \neq 0.$$

Then $c(\zeta, t, x) = C(\zeta, t)x^2$ were $C = A^2$. Suppose also that $A \neq 0$ identically. First $P^{\vartheta}_{\nu, b}$ and $P^{\vartheta}_{\nu, 0}$ (for $b \equiv 0$) are equivalent, second under $P^{\vartheta}_{\nu, 0}$ the solution X_t a.s. never hits 0. Then this property is true also under $P^{\vartheta}_{\nu, b}$, and H2 holds. Therefore we can use the optimal contrast associated with f in (12): $f(G, x) = \log G + x^2/G$, and (59) yields

$$B(\vartheta)_t = (\dot{C}/C)(\vartheta, t)^2, \qquad \Gamma(\vartheta) = 2\Big/\int B(\vartheta)_t\,\mu(dt),$$

which are non-random. Then the estimators based on this contrast are asymptotically normal (instead of mixed normal).

We can observe here that, since we should have the asymptotic variance $\Gamma(\vartheta)$ as small as possible, the best choice of the observation times $t(n, i)$ are those leading to a limiting μ measure which maximizes $\int B(\vartheta)_t\,\mu(dt)$. (Note however that we do not know the true value of ϑ.)

In particular if $A(\vartheta, t) = A(\vartheta)$ does not depend on t (if further $b=0$, we have a so-called *multiplicative model*), $B(\vartheta)_t$ does not depend on t either, and *all sampling schemes are thus equivalent.*

7 *c. A truly non invertible situation.* – Here $d=m=q=1$ and we consider the equation

$$dX_t = b\,dt + a(\vartheta, X_t)\,dW_t, \qquad \mathscr{L}(X_0) = \nu,$$

where $b>0$, and $a(\vartheta, x) \geq 0$, and $a(\vartheta, x) = 0$ iff $x \geq \alpha$, and $x_0 < \alpha$. Then c is not invertible, but the non-invertibility set does not depend on ϑ.

We use the contrasts based on the function f' of (60). If $\dot{c}(\vartheta, 0, x_0) \neq 0$, and provided 0 belongs to the support of μ. $B'(\vartheta)$ is a.s. invertible and the asymptotic conditional variance of the estimators is

$$\Gamma(\vartheta) = 2\,\frac{\int (\dot{c}c)^2(\vartheta, t, X_t)\,\mu(dt)}{\left(\int \dot{c}^2(\vartheta, t, X_t)\,\mu(dt)\right)^2}.$$

The integrals above can be restricted to the subset of [0, 1] where $X_t < \alpha$.

7 *d*. Here is our first 2-dimensional example, where $d=2$, $m=1$, $q=1$. We consider the equation (with ν a Dirac measure):

$$\begin{cases} dX_t^1 = b^1(t, X)\,dt + \alpha(\vartheta, X_t)\,dW_t, & X_0^1 = x_0^1 \\ dX_t^2 = b^2(t, X)\,dt, & X_0^2 = x_0^2 \end{cases}$$

where $\alpha > 0$ identically. Setting $\beta = \alpha^2$ we have $c = \begin{bmatrix} \beta & 0 \\ 0 & 0 \end{bmatrix}$ and H2 cannot be satisfied. We take for $\mathscr{S}$ the set of matrices $G = \begin{bmatrix} g & 0 \\ 0 & 0 \end{bmatrix}$ with $g > 0$. Set

$$f(G, x) = \log g + (x^1)^2/g \quad \text{if} \quad G = \begin{bmatrix} g & 0 \\ 0 & 0 \end{bmatrix}.$$

Then the estimators $\hat{\vartheta}_n$ minimizing the contrasts based on f have the following asymptotic conditional variance

$$\Gamma(\vartheta) = 2/B(\vartheta), \qquad B(\vartheta) = \int B(\vartheta)_t\,\mu(dt), \qquad B(\vartheta)_t = (\dot{\beta}/\beta)(\vartheta, X_t).$$

In fact, these estimators are even optimal in the sense of Theorem 5 *b*, which can be seen as follows: the asymptotic properties of all our contrasts do not depend on the drift b, so that we can take $b^1 = b^2 = 0$; then $X_t^2 = x_0^2$ for all t, so we are actually looking at a 1-dimensional problem and we can apply Theorem 5.

7 *e*. *An example coming from mathematical finance.* – This example is borrowed to Courtadon [3]. We have $m = d = 2$ and we consider the linear equation:

$$\begin{cases} dX_t^1 = \gamma(X_t^2 - X_t^1)\,dt + \beta X_t^1\,dW_t^1, \\ dX_t^2 = \alpha X_t^2\,dt + \rho\sigma X_t^2\,dW_t^1 + \sigma X_t^2\,dW_t^2, \end{cases}$$

where all parameters α, β, γ, σ, ρ are unknown, but where we are really interested in estimating β, σ, ρ only, and we know that the triple $\vartheta = (\beta, \sigma, \rho)$ takes its values in $\Theta = (0, \infty)^2 \times \mathbb{R}$: this is not a compact subset of $\mathbb{R}^3$, but it is easily seen that the previous results apply nevertheless here.

The same discussion as in § 7-*b* shows that H2 holds, so we can take the function f of (12). The associated contrast at stage n is

$$U^n(\vartheta) = \log(\beta^2\sigma^2) + \frac{1+\rho^2}{\beta^2}\Sigma_{11}^n - \frac{2\rho}{\beta\sigma}\Sigma_{12}^n + \frac{1}{\sigma^2}\Sigma_{22}^n + \frac{1}{n}\sum_{i=1}^{n} \log(X_{t_i}^1 - X_{t_{i-1}}^1)^2 (X_{t_i}^2 - X_{t_{i-1}}^2)^2,$$

where

$$\Sigma_{jk}^{n}=\frac{1}{n}\sum_{i=1}^{n}\frac{(X_{t_i}^{j}-X_{t_{i-1}}^{j})(X_{t_i}^{k}-X_{t_{i-1}}^{k})}{\delta_i X_{t_{i-1}}^{j} X_{t_{i-1}}^{k}}.$$

Then the minimum contrast estimator is $\hat{\vartheta}_n=(\hat{\beta}_n, \hat{\rho}_n, \hat{\sigma}_n)$, where

$$\hat{\beta}_n=(\Sigma_{11}^{n})^{1/2}, \qquad \hat{\rho}_n=\Sigma_{12}^{n}(\Sigma_{11}^{n}\,\Sigma_{22}^{n}-(\Sigma_{12}^{n})^2)^{1/2},$$
$$\hat{\sigma}_n=[(\Sigma_{11}^{n}\,\Sigma_{22}^{n}-(\Sigma_{12}^{n})^2)/\Sigma_{11}^{n}]^{1/2}.$$

Finally, the asymptotic conditional covariance matrix for these estimators at point $\vartheta=(\beta, \rho, \sigma)$ is given by

$$\Gamma(\vartheta)=\begin{bmatrix} \beta^2/2 & \rho\beta/2 & 0 \\ \rho\beta/2 & 1+\rho^2 & -\rho\sigma/2 \\ 0 & -\rho\sigma/2 & \sigma^2/2 \end{bmatrix}.$$

In particular, it is deterministic: that is, as for all linear models (*see* § 7*b* above), the normalized estimators converge to a normal (and not mixed normal) distribution. Further, $\Gamma(\vartheta)$ does not depend on μ, that is all sampling schemes are equivalent: this comes from the fact that the model is homogeneous in time.

REFERENCES

[1] D. J. ALDOUS, Stopping times and tightness, *Ann. Probab.*, Vol. **6**, 1978, pp. 335-340.

[2] R. AZENCOTT, Densité des diffusions en temps petit : développements asymptotiques (partie I), *Sém. Proba. XVIII*, 1984, pp. 402-498; *Lect. Notes in Math.*, n° 1059, Springer Verlag; Berlin.

[3] G. COURTADON, Une synthèse des modèles d'évaluation d'options sur obligations, *Finance*, **6**, Vol. **2**, 1985.

[4] D. DACUNHA-CASTELLE and M. DUFLO, *Probabilités et statistiques II*, Masson; Paris, 1983.

[5] D. DACUNHA-CASTELLE and D. FLORENS-ZMIROU, Estimation of the coefficient of a diffusion from discrete observations, *Stochastics*, Vol. **19**, 1986, pp. 263-284.

[6] A. Ja. DOROGOVCEV, The consistency of an estimate of a parameter of a stochastic differential equation, *Theory Probab. and Math. Statist.* Vol. **10**, 1976, pp. 73-82.

[7] G. DOHNAL, On estimating the diffusion coefficient, *J. Appl. Prob.*, Vol. **24**, 1987, pp. 105-114.

[8] P. FEIGIN, Asymptotic theory of conditional inference for stochastic processes, *Stochastic Proc. Appl.*, Vol. **22**, 1985, pp. 89-102.

[9] D. FLORENS-ZIMROU, Approximate discrete-time schemes for statistics of diffusion processes, *Statistics*, Vol. **20**, 1989, pp. 547-557.

[10] J. HAJEK, A characterization of limiting distributions of regular estimates, *Z. für Warsch. Theo. Geb.*, Vol. **14**, 1970, pp. 324-330.

[11] V. GENON-CATALOT, *Thèse*, Université Paris-Sud, Orsay, 1987.

[12] V. GENON-CATALOT, Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, Vol. **21**, 1990, pp. 99-116.

[13] V. Genon-Catalot and J. Jacod, On the diffusion of the diffusion coefficient for diffusion processes: random sampling, *Preprint* #100, Lab. Probabilités, Univ. Paris-VI, 1992.

[14] P. Hall and C. Heyde, *Martingale limit theory and its applications*, Academic Press; New York, 1980.

[15] J. Jacod, Random sampling in estimation problems for continuous Gaussian processes with independent increments, *Stoch. Processes and Appl.*, 1993 (to appear).

[16] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*, Springer Verlag; Berlin, 1987.

[17] P. Jeganathan, On the asymptotic theory of estimation when the limit of the loglikelihood is mixed normal, *Sankya*, Series A, Vol. **44**, 1982, pp. 173-212.

[18] P. Jeganathan, Some asymptotic properties of risk functions when the limit of the experiment is mixed normal, *Sankya*, Series A, Vol. **45**, 1983, pp. 66-87.

[19] R. A. Kasonga, The consistency of a non-linear least-squares estimate from diffusion processes, *Stoch. Processes and Appl.*, Vol. **30**, 1988, pp. 263-275.

[20] L. Le Cam and G. L. Yang, *Asymptotics in Statistics*, Springer Verlag, Berlin, 1990.

[21] R. S. Liptser and A. N. Shiryayev, *Statistics of random processes, I. General theory*, Springer Verlag, Berlin, 1977.

[22] S. Molchanov, Diffusion processes and Riemannian geometry, *Russ. Math. Survey*, Vol. **30**, 1975, pp. 1-63.

[23] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Springer Verlag, Berlin, 1979.

(*Manuscript received November* 13, 1991;
revised April 1*st*, 1992.)