On Some Estimation Problems Arising in Environmental and Pollution Models

By

François DELEBECQUE* and Norihiko ADACHI**

(Received June 28, 1979)

Abstract

The problem of estimating some unknown parameters in a dynamical system described by stochastic differential equations is discussed. A method of estimation based on non-linear filtering and maximum of likelihood theories is presented. A numerical procedure for the estimation is developed, and the condition for convergence of the method is obtained.

1. Introduction

Let us consider a river water quality model[1]. The model is represented by a system of ordinary differential equations. It has five state variables. Two of them are concentrations of bacteria, and the others are those of substrates (ammonium nitrogen, oxidized nitrogen, organic material and dissolved oxygen). The model also contains twenty parameters. Two of them are physical and the others are biological constants. Measurement of concentrations of substrates is possible, but it is almost impossible to evaluate the biomass of a certain class of bacteria in a specific area. Moreover, there is little information available concerning biological parameters. Therefore, we have to identify biological parameters from the observation of the substrates' concentrations. This is a typical difficult problem often encountered in the modelling of pollution systems. The characteristics of such models are, in gneeral, the presence of non-linearities and rather few state variables. Also, the stochastic aspect seems important here because many phenomena appear which are difficult to represent in a deterministic framework. In this paper a dynamical system is described by stochastic differential equations, and we discuss the problem of estimating some unknown parameters in the system. We present here a method of estimation based on non-linear filtering and maxi-

* Institut de Recherche D'Informatique et D'Automatique. Domaine de Volceau-Rocquencourt 78150 Le Chesnay, France

** Department of Applied Mathmatics and Physics

mum of likelihood theories. The limitation of the method is imposed by the size of the system (no more than 2 or 3 state variables). On the other hand, we can take into account non-linearities. Moreover, a general stochastic model allows us the use of a general statistical framework to show the asymptotic properties of the estimates.

2. Computation of Maximum of Likelihood of Bayesian Estimates

2.1 The non-degenerate case

We assume that the state of the system is given by a diffusion process:

$$dx_t = b(x_t, \theta) + \sigma(x_t) dw_t \tag{1}$$

where θ is an unknown parameter.
We denote the matrix $\sigma(x)\sigma'(x)$ by $a(x)$ and we assume the usual hypothesis on the functions a and b:
They are supposed to be continuous functions of $x\cdot$, and $\lambda(x)$, the smallest eigenvalue of a satisfies: $\lambda(x) \geq c > 0$.
Then (1) admits a unique "weak" solution. That is, there exsists a probability measure P_θ and a pair of stochastic processes (x_t), (w_t), with continuous trajectories whereby under P_θ, (1) holds a.s., and w_t is a Brownian motion. We shall also introduce:

$$y_t \text{ is a } P_\theta\text{-Brownian motion independent of } (x_t)\,. \tag{2}$$

(This is always possible by replacing P_θ by $P_\theta \times W$, where W is the Wiener measure). Therefore, under P_θ, the σ-algebras $F_t^x = \sigma(x_s, s \leqq t)$ and $F_t^y = \sigma(y_s, s \leqq t)$ are independent. We denote $F_t = F_t^x \vee F_t^y$.
We now describe the process of observation:

$$dy_t = h(x_t)dt + d\bar{w}_t\,. \tag{3}$$

h is a continuous function of x, and $\bar{w}_t$ is the observation noise, a Brownian motion.

By the Cameron-Martin Girsanov formula[2], we know that if $B' = B + AC$, where B (resp. A) is the drift term (resp. the diffusion term) of a diffusion process P, then the diffusion process Q with the drift term B' and the same diffusion term A is absolutely continuous with respect to P, both measures being restricted to F_t, and

$$\frac{dQ}{dP}\Big|_{F_t} = \exp\Big\{\int_0^t \langle C(x_s), dx_s - Bds\rangle - \frac{1}{2}\int_0^t \langle AC, C\rangle(x_s)ds \tag{4}$$

(Here, F_t denotes the σ-algebra $\sigma(x_s, s\leq t)$).
The following proposition is then easily derived with:

$$X = (x, y),\ B' = \begin{pmatrix} b \\ h \end{pmatrix},\ B = \begin{pmatrix} b \\ 0 \end{pmatrix},\ C = \begin{pmatrix} 0 \\ h \end{pmatrix},\ A = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}.$$

Proposition 1: The probability law Q of the process (x_t, y_t) given by (1) and (3) is absolutely continuous with respect to P with both measures restricted to F_t and $L_t = \frac{dQ}{dP}|_{F_t}$ is given by

$$L_t = \exp\int_0^t h(x_s)dy_s - \frac{1}{2}\int_0^t h^2(x_s)ds. \tag{5}$$

L_t is a positive martingale for (F_t, P) because for $A\in F_s$, $t\geq s$ we have

$$\int_A L_s dP = Q(A) = \int_A L_t dP.$$

Moreover, Ito's formula gives:

$$L_t = 1 + \int_0^t h(x_s)L_s dy_s. \tag{5$'$}$$

The problem is now to "compute" the probability law of the observed process (y_t), that is $Q\,|\,F_t^y$ for all $t\geq 0$.

Proposition 2: The probability law of the process (y_t) given by (3) is absolutely continuous with respect to $P|F_t^y$ and

$$\frac{dQ}{dP}|_{F_t^y} = \underset{P}{E}(L_t|F_t^y) = \tilde{L}_t.$$

(Indeed, if $A\in F_t^y$, $Q(A) = \int_A L_t dP = \int_A \underset{P}{E}(L_t|F_t^y)dP$)
Now we give a recursive equation for $\tilde{L}_t$. For that we introduce a smooth R-valued function (bounded):

$$u_t(f) = \underset{P}{E}(L_t f(x_t)\,|\,F_t^y).$$

Then it is shown in reference (3) that one has:

$$u_t(f) = u_0(f) + \int_0^t u_s(Af)ds + \int_0^t u_s(hf)dy_s. \tag{7}$$

A is the generator of the Markov process (x_t)

$$Af(x) = \sum_i b_i \frac{\partial f}{\partial x_i}(x) + \frac{1}{2}\sum_{ij} a_{ij}\frac{\partial^2}{\partial x_i \partial x_j}. \tag{8}$$

We remark that $\tilde{L}_t = u_t(1)$, and for the density of the "measure" u_t (also denoted by $u_t(x)$) we have the following:

$$\begin{cases} u_t(x) = A^* u_t(x) + u_t(x) h(x) dy_t \\ u_0(x) = \mu_0(x) \end{cases} \tag{9}$$

μ_0 is the law of the random variable x_0 (given).

Let π_t denote the non-linear filter:

$$\pi_t(f) = \underset{Q}{E}(f(x_t) \mid F_t^y) . \tag{10}$$

Then we have the Bucy's lemma:

Proposition 3:

$$\pi_t(f) = \frac{u_t(f)}{u_t(1)} \tag{11}$$

Proof: If $A \in F_s^y$, we have:

$$\begin{aligned} \int_A \pi_t(f) u_t(1) dP &= \int_A \pi_t(f) L_t dP & &\text{(definition of } u_t(1)) \\ &= \int_A \pi_t(f) dQ & &\text{(construction of } Q) \\ &= \int_A f(x_t) dQ & &\text{(definition of } \pi_t) \\ &= \int_A f(x_t) L_t dP . \end{aligned}$$

Thanks to (11) we can give another expression for $\tilde{L}_t$. From (7) with $f=1$, we obtain (A1=0):

Comparing with (5'), we derive:

$$\tilde{L}_t = \exp\left\{\int_0^t \pi_s(h)\, dy_s - \frac{1}{2}\int_0^t (\pi_s(h))^2 ds\right\} . \tag{12}$$

Thus we have shown:

Proposition 4: The probability law of the process (y_t) is a weak solution of:

$$dy_s = \pi_s(h) ds + d\bar{w}_s \tag{13}$$

where $\bar{w}_s$ is a Q-Brownian motion.

So the statistical structure of the problem of estimation of θ by the observation of (y_t) can be wrewritten:

$$\begin{aligned} \Omega &= C(0, \infty; R^m) \\ y_t(\omega) &= \omega_t \\ F_t &= \sigma(y_s, s \leq t) . \end{aligned}$$

Q_θ is a weak solution of

$$dy_s = \hat{h}_s(\theta)ds + d\bar{w}_s \tag{13'}$$

with $h_s(\theta)$ being the function $\pi_s(h)$, where π_s is the non-linear filter of (1) and (3). In particular, we have a structure dominated with the Wiener measure as a reference measure, and we also have an exponential family. The factorization theorem shows immediately that $\hat{h}_s(\theta)$ is a sufficient statistic of θ.

Maximum of Likelihood Estimators

The maximum of likelihood estimator of θ is given by:

$$\max_\theta \int u_t^\theta(x)dx$$

$$du_t(x) = -\sum_{ij} \frac{\partial}{\partial x_i} (b_i^\theta u)(x) + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial y_j} (a_{ij}u)(x) + u_t(x)h(x)dy_t . \tag{14}$$

This problem can be seen as the problem of controlling the partial differential equation (14) with the "control" θ and the "cost function" $\int u_t(x)dx$.

Remark 1: This problem is equivalent to

$$\max_\theta \int \left\{ \int_0^t \pi_s(h)dy_s - \frac{1}{2}\int_0^t (\pi_0(h))^2 ds \right\} , \tag{15}$$

with the partial differential equation of the non-linear filtering problem satisfied by $\pi_t(x)$. This last problem seems, from a numerical point of view, much more difficult: the cost function (15) and partial differential equations for π_t are non-linear.

Remark 2: We can consider other kinds of state equations instead of (1). Indeed, if (x_t) is a Markov process with generator A, we obtain for (7) and (14) the same kind of equation with the generator A. This remark will be useful later. Problem (14) can be solved by a gradient-like method:

$$\theta(n+1) = \theta(n) + \rho \frac{\partial J}{\partial \theta} (\hat{\theta}_n) ,$$

for instance (for $n=1$), in the case of a parameter which appears linearly $b^\theta(x) = \theta \cdot b(x)$ (b a known function):

$$dv_t(x) = -\frac{\partial}{\partial x} (b \cdot v)(x) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (av)(x) + v_t(x)h(x)dy_t$$

$$\frac{\partial J}{\partial \theta} = \int v_t(x)dx .$$

Bayesian estimate:

In the Bayesian approach, we are given an "a priori" probability law of $q(d\theta)$ of the parameter θ. The probability law of (y_t) is then $\int Q_\theta q(d\theta)$, and θ is estimated at time t by its "a posteriori" mean that is $\underset{Q}{E}(\theta \mid F_t^y) = \tilde{\theta}_t$.

Then by the Bayes's formula, we have

$$\tilde{\theta}_t = \frac{\int \theta u_t(x) dx \, q(d\theta)}{\int u_t(x) dx \, q(d\theta)}, \tag{15}$$

$u_t(x)$ given once more by (14).

In this case, we have to solve a family of partial differntial equations indexed by θ, and then apply formula (15).

Remark 3: It is possible to consider an observation such as

$$dy_t = h(x_t)dt + \beta(y_t)d\bar{w}_t \tag{16}$$

instead of (3). However, for our purpose (estimation) this is not really an improvement. Indeed, the function β can be estimated "along the observed trajectory" by the quadratic variation:

$$\lim_{t_i \in t(0,s)} \sum_i (y_{t_{i+1}} - y_{t_i})^2 \to \int_0^s \beta^2(y_s)ds\,.$$

2.2 The degenerate case

So far we have considered the case of a state equation of type (1), and we have supposed that the observation is given by a diffusion Markov process (3). In particular, the observation process is in the state equation. Now we release this assumption: this corresponds to the case of a 2-dimensional Markov process (of a diffusion type), one component of which is observed without noise. To be more specific, we consider instead of (1) and (3).

$$dx_t = b^\theta(x_t, y_t)dt + \sigma(x_t, y_t)dw_t \tag{17}$$

$$dy_t = h(x_t, y_t)dt + d\bar{w}_t\,. \tag{18}$$

(Remark 3 is also valid here. We can modify the reference measure in consequence.)

We now derive the analog of equation (14).

To simplify the notations we assume $n = m = 1$.

As in part 1, we consider a *reference probability measure* P such that, under P, (x_t) is a solution of (17) and (y_t) is a standard Brownian motion independent of (w_t).

Then we define Q by:

$$\frac{dQ}{dP}\Big|_{F_t} = L_t = \exp\left\{\int_0^t h(x_s, y_s)\,dy_s - \frac{1}{2}\int_0^t h^2(x_s, y_s)\,ds\right\}.$$

In this way, the Cameron-Martin Girsanov theorem (Prop. 1) claims that, under Q, the pair (x_t, y_t) is a weak solution of (17) and (18).

It is shown in (ref. 6) theorems 7–12 that (y_t) permits the following representation:

$$dy_t = \pi_t(h)ds + d\bar{w}_t \tag{19}$$

where $(\bar{w}_t)$ is a standard Brownian motion (the innovation process) adapted to F_t^y, and $\pi_t(h) = \underset{Q}{E}(h(x_t, y_t)\,|\,F_t^y)$.

Let $f(x, y)$ be a bounded R-valued function.

We introduce:

$$u_t(f) = \underset{Q}{E}(f(x_t, y_t)L_t\,|\,F_t^y)\,. \tag{20}$$

In particular, $u_t(1)$ gives the density with respect to the Wiener measure of the observation process (y_t).

Thanks to (19) we also have:

$$u_t(1) = \exp\left\{\int_0^t \pi_s(h)dy_s - \frac{1}{2}\int_0^t \pi_s^2(h)ds\right\} \tag{21}$$

or equivalently (Ito's formula):

$$u_t(1) = 1 + \int_0^t u_s(1)\pi_s(h)dy_s\,. \tag{22}$$

The generator of the Markov process (17), (18) is:

$$Af(x, y) = b\frac{\partial f}{\partial x} + h\frac{\partial f}{\partial y} + \frac{1}{2}\,a\frac{\partial^2 f}{\partial x^2} + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(x, y)\,. \tag{23}$$

We also introduce $Nf(x, y) = \frac{\partial f}{\partial y}(x, y)$.

Now we can derive a recursive equation for u_t.

Proposition 5: The "measure" u_t satisfies the following diffusion-type equation:

$$u_t(f) = u_0(f) + \int_0^t u_s(Af)ds + \int_0^t u_s(hf + Nf)\,dy_s\,. \tag{24}$$

Proof. From Theorem 8.3 in ref. [8], we have for the non-linear filter $\pi_t(f)$ the equation:

$$\pi_t(f) = \pi_0(f) + \int_0^t \pi_s(Af)ds + \int_0^t [\pi_s(hf + Nf) - \pi_s(h)\pi_s(f)]\,(dy_s - \pi_s(h)ds)\,,$$

where $Nf(x,y)=\frac{\partial f}{\partial y}(x,y)$.

From (22) we have $u_t(1)=u_0(1)+\int_0^t u_s(1)\pi_s(h)dy_s$.

We can apply Ito's formula to $u_t(f)=u_t(1)\pi_t(f)$:

$$\begin{aligned}u_t(f) &= u_0(f)+\int_0^t \pi_s(f)u_s(1)\pi_s(h)dy_s+\int_0^t \pi_s(1)\pi_s(Af)(dy_s-\pi_s(h)ds)\\ &\quad+\int_0^t u_s(1)[\pi_s(hf)-\pi_s(h)\pi_s(f)+\pi_s(Nf)](dy_s-\pi_s(h)ds)\\ &\quad+\int_0^t \langle u_s(1),\ \pi_s(f)\rangle ds\,.\end{aligned}$$

Now, $\int_0^t \langle u_s(1), \pi_s(f)\rangle ds=\int_0^t u_s(1)\pi_s(h)[\pi_s(hf)-\pi_s(f)\pi_s(h)+\pi_s(Nf)]ds$ and so we obtain

$$\begin{aligned}u_t(f) &= u_0(f)+\int_0^t u_s(Af)ds+\int_0^t \pi_s(f)u_s(1)\pi_s(h)\,dy_s\\ &\quad+\int_0^t u_s(1)[\pi_s(hf)-\pi_s(h)\pi_s(f)+\pi_s(Nf)]dy_s\,,\end{aligned}$$

that is:

$$u_t(f) = u_0(f)+\int_0^t u_s(Af)ds+\int_0^t [u_s(Nf)+u_s(hf)]dy_s\,.$$

We are now in a position to settle the maximum of likelihood and the Bayesian estimate.

By transposition, (24) gives the analog of equation (14):

$$du_t(x,y) = A^*u_t(x,y)dt+N^*u_t(x,y)dy_t+u_t(x,y)h(x,y)dy_s \tag{25}$$

with

$$\begin{cases}A^*f= -\frac{\partial}{\partial x}[b^\theta f]-\frac{\partial}{\partial y}[hf]-\frac{\partial^2}{\partial x^2}[af]+\frac{1}{2}\frac{\partial^2 f}{\partial y^2}\\ N^*f=Nf=\frac{\partial f}{\partial y}.\end{cases}$$

The maximum of likelihood estimate $\hat{\theta}_t$ is then given by:

$$\max_\theta \int u_t(x,y)dx\,dy \tag{26}$$

under (25), and the Bayesian estimate $\tilde{\theta}_t$ is

$$\tilde{\theta}_t = \frac{\int \theta u_t(x,y)dx\,dy\,q(d\theta)}{\int u_t(x,y)dx\,dy\,q(d\theta)}\,. \tag{27}$$

We note that in this degenerate case we have to deal with a two dimensional partial differential equation. We also observe that eq. (25) is linear and simpler than the non-linear fitlering equation of $\pi_t(f)$.

3. A Numerical Procedure for Computation (The Non Degenerate Case)

In this section, we give an approximation of equation (14).

First, we recall the definition of a "pure-jump" Markov process with the discrete state space $E=\{x_1, x_2, \cdots\}$. Such a process is characterized by a family of positive numbers $\{q(x), x\in E\}$, together with a probability transition matrix $p(x,y)$ $(\sum_y p(x,y)=1, p(x,x)=0)$. When in state x, the process x_t spends some (random) time $T(x)$ at x, $T(x)$ is distributed as a geometric random variable with the mean $1/q(x)$, and then jumps to another (random) state y with the probability $p(x,y)$.

For such a process the generator is given by:

$$Af(x) = q(x)\{\sum_{y\neq x} p(x,y)f(y)-f(x)\} . \tag{28}$$

We consider a mesh of size k, and we discretize the operator in (8). (This is done with $n=1$ to simplify the notations. See Ref. [5] for a general case.)

$$\begin{aligned} A_kF(x) &= b^+(x)\frac{f(x+k)-f(x)}{k} - b^-(x)\frac{f(x)-f(x-k)}{k} \\ &\quad + \frac{1}{2}a(x)\frac{f(x+k)+f(x-k)-2f(x)}{k^2}, \end{aligned}$$

or equivalently:

$$A_kf(x) = q_k(x)\{p_k(x, x+k)f(x+k)+p_k(x, x-k)-f(x)\}$$

with

$$\begin{aligned} q_k(x) &= \frac{a(x)}{k^2}+\frac{|b(x)|}{k} \\ p_k(x, x+k) &= \left(\frac{b^+(x)}{k}+\frac{1}{2}\frac{a(x)}{k^2}\right)/q_k(x) \\ p_k(x, x-k) &= \left(\frac{b^-(x)}{k}+\frac{1}{2}\frac{a(x)}{k^2}\right)/q_k(x) . \end{aligned} \tag{29}$$

We see that $p_k\geq 0 \sum_{y\neq x} p_k(x,y)=1$, so we can interpret A_k as a pure-jump Markov process generator. Moreover, it can be shown that the family of measure $\{P_k\}$ generated by the process is tight. (See Ref. [3] for definition.) Now we can use as an approximation for (14) the equation:

$$du_t(x) = A_k^* u_k(x)dt + u_t(x)h(x)dy_t\,. \tag{14}_k$$

It turns out that this equation gives precisely the "density" of the random "measure":

$$u_t(f) = \underset{P}{E}(f(x_t)L_t \mid F_t^y)$$

where (x_t) is now a pure-jump Markov process with the discrete state space given by the mesh of size k. Equation $(14)_k$ is an ordinary diffusion process driven by (y_t), and can be used as an approximation of (14). Note that we in no way change the observation process (y_t). This very useful method of computation cannot be extended to a degenerate case without modifying the observation process.

Remark 4: It is worthwhile to settle the discrete time version of the previous equations introduced for continuous time stochastic processes. Here, P is a probability measure which makes (X_n) a Markov chain with the transition probability p, and (ΔY_n) is a sequence of the R^m-valued independent and identically distributed random variables with a common distribution $p_0(y)dy$ (where dy denotes the Lebesgue measure), and also independent of (X_n). Then

$$L_n = \prod_{m=1}^{n} \frac{p_0(\Delta Y_m - h(x_m))}{p_0(\Delta Y_m)} \tag{30}$$

is a (P, F_n) positive martingale and we can define Q by

$$\frac{dQ}{dP}\Big|_{F_n} = L_n\,.$$

With this choice of L_n, Q makes (Y_n) a process admitting the representation

$$\Delta Y_n = h(x_n) + W_n \tag{31}$$

where W_n are i. i.d. with distribution p_0. In a case where p_0 is gaussian $N(0, \sqrt{\Delta t})$ and $\Delta Y_n = h(x_n)\Delta t + W_n$
we get:

$$L_n = \exp\left\{\sum_{m=1}^{n} h(X_m)\Delta Y_m - \frac{1}{2}\sum_{n=1}^{n} h^2(X_m)\Delta t\right\},$$

a formula like (5).
Now we also introduce:

$$u_n(f) = \underset{P}{E}(L_n f(X_{n+1}) \mid F_n^y) \tag{32}$$

and the "density"

$$u_n(x) = u_n(1_{\{x\}})$$

The situation here is different from the continous time case, mainly because it is not true that we have

$$\Delta Y_n = \hat{h}_n + \bar{W}_n$$

independence where $\bar{W}_n$ admits the *same* distribution as W_n under Q.
Because of the P independence of (X_n) and (ΔY_n), the Markov property of X_n gives immediately the following recursive equation for u_n which replaces (14):

$$\begin{aligned} u_{n+1}(x) &= \sum_{x'} u_n(x') \frac{p_0(\Delta Y_n - h(x'))}{p_0(\Delta Y_n)} p(x', x) \\ u_0(x) &= \text{the initial law of } (X_n). \text{ (given)} \end{aligned} \tag{33}$$

and, as in the continuous-time case, if the unknown parameter is in $p_\theta(x,x')$, the maximum of likelihood estimation is given by:

$$\max_\theta \sum_x u_n(x) .$$

4. Properties of Estimates

We have seen in the preceding section that the statistical structure of the problem is given by: $dy_t = \pi_s(h)dt + d\bar{w}_t$ where $\bar{w}_t$ is a standard Brownian motion. The density of the probability law Q of the process (y_t) (on the space of continuous function) with respect to the Wiener measure is

$$\tilde{L}_t(\theta) = \exp\left\{\int_0^t \pi_s(h)\, dy_s - \frac{1}{2}\int_0^t \pi_s(h)\, ds\right\} \tag{34}$$

where $\pi_s(h)$ depends on θ in a complicated way. It is the non-linear filter of the function h computed for a Markov process with a generator depending on θ. Now we recall some facts about the maximum of likelihood estimators in the simplest case of random variables with a discrete parameter time. Let X be the sequence of real valued random variables, assumed to be distributed according to one particular density (with respect to the Lebesgue measure) in the family $\{f(x,\theta), \theta \in \theta\}$, depending continuously on θ. We assume that θ is a compact set and that the process (X_n) is ergodic. (for example, the X_n are i.i.d.) The method of the maximum of likelihood consists in maximizing

$$L_n(\theta) = \frac{1}{n} \sum_{m=1}^n \log f(X_m, \theta) .$$

The basic argument to show the convergence of the method is the ergodic theorem[9], applied to the family of random process (with a p-dimensional "time" θ)

$$\{\log f(X_n, \theta),\ n = 1, 2, \cdots\}\ .$$

By this theorem we have:

$$\sup_{\theta}\ |L_n(\theta) - \int \log f(x, \theta) p(x) dx| \to 0 \qquad \text{a.s.}$$

where $p(x)$ is the true distribution of X_1.
Let $\theta(p)$ satisfy $\sup_{\theta} \int \log f(x, \theta) p(x) dx = \int \log f(x, \theta(p)) p(dx) \equiv K(\theta(p), \theta)$.
Hence, we have $K(\hat{\theta}_n, p) \to K(\theta(p), p)$
and then by a continuity argument $\hat{\theta}_n \to \theta(p)$ a.s.
If the observed distribution p belongs to the selected family $\{f(x, \theta)\}$ i.e. $p(x) = f(x, \theta_0)$, then by the Information's inequality $\theta(p) = \theta_0$
($\int f(x, \theta_0) \log f(x, \theta) dx$ reaches its maximum for $\theta = \theta_0$).
This shosw that in the case of an ergodic process the maximum likelihood method is consistent. The preceding argument also shows that for $\theta \neq \theta'$, the two probability distributions of the *process* (X_n), say P_θ and $P_\theta{}'$, are disjoint. (However, they are mutually absolutely continuous when restricted to $\sigma(X_1, \cdots, X_n)$ for finite n.)
$L_n(\theta)$ converges P_θ a.s. to $\int f(x, \theta) \log f(x, \theta) dx$
and $P_{\theta'}$, a.s. to $\int f(x, \theta') \log f(x, \theta) dx$. These two quantities will be different unless $\theta = \theta'$.
Moreover, $\dfrac{L_n(\theta)}{L_n(\theta')}$ converges P_θ a.s. to

$$\frac{\int f(x, \theta) \log f(x, \theta) dx}{\int f(x, \theta) \log f(x, \theta) dx} > 1 \qquad \forall \theta' \neq \theta\ .$$

This shows that if θ is assumed to take only a finite number of values $\{\theta, \theta_1, \theta_2, \cdots, \theta_k\}$), then for a (random) N large enough, $\hat{\theta}_N = \theta$ when P_θ is obtained.

Through these nice propoerties do not extend to the case we are interested in, the following proposition shows that, "in general" for $\theta \neq \theta'$, the two probability laws asosciated with the observed process (y_t) are disjoint and therefore, "in general", the maximum of the likelihood estimate converges.

Let $\hat{h}_s = \pi_s^\theta(h)$, $\hat{h}'_s = \pi_s^{\theta'}(h)$

and Q (resp. Q') a weak solution of

$$dy_s = \hat{h}_s\, ds + d\bar{w}_s \qquad (\text{resp. } dy_s = \hat{h}'_s ds + d\bar{w}_s)\,. \tag{35}$$

With these notations, (34) becomes

$$\tilde{L}_t = \exp\left\{\int_0^t \hat{h}_s\, dy_s - \frac{1}{2}\int \hat{h}_s^2 ds\right\}.$$

Proposition 6: One has:

$$L_t \begin{cases} \longrightarrow 0 \quad Q \text{ a.s.} \\ \longrightarrow \infty \quad Q' \text{ a.s.} \end{cases} \qquad \text{on set } A = \left\{\int_0^\infty (\hat{h}'_s - \hat{h}_s)^2 ds = \infty\right\}$$

Proof: By the Cameron-Martin formula we have

$$\frac{dQ(h)}{dQ}\Big|_{F_t} = \exp\left\{\int_0^t \langle h_s - \hat{h}'_s,\, dx_s - \hat{h}_s ds\rangle - \frac{1}{2}\int_0^t (h_s - \hat{h}_s)^2 ds\right\}$$

for all h_s, F_t^y-adapted. With $h_s = \frac{1}{2}(h_s' + \hat{h}_s)$ we have that

$$Z_t = \exp\left\{\int_0^t \left\langle \frac{h_s - \hat{h}_s}{2},\, dx_s - \hat{h}_s ds\right\rangle - \frac{1}{2}\int_0^t \left(\frac{h_s - \hat{h}_s}{2}\right)^2 ds\right\}$$

is a positive martingale for Q. Hence, it admits a Q a.s. limit Z. We then note that:

$$\tilde{L}_t = (Z_t)^2 \exp - \frac{1}{4}\int_0^t (\hat{h}'_s - \hat{h}_s)^2 ds$$

which converges Q a.s· to 0 on set A. The other convergence is proved by the same way. The previous proposition shows that on set A the measures induced by Q and Q' are disjoint. This is known to be a necessary condition for the convergence of the M.L.E. estimators in a general framework. This result intuitively means that if the non-linear filters $\pi_t(h)$ of the function h computed with θ and θ' verify almost surely the previous condition, then we can distinguish the true parameter between θ and θ'.

Now we compute the conditional Fisher's information given by a trajectory. First, we remark that

$$\frac{d}{d\theta}\log \tilde{L}_t(\theta) = \int_0^t \pi'_s(h)\,(dy_s - \pi_s(h)\, ds) \tag{37}$$

where $\pi'_s(h) = \frac{d}{d\theta}(\pi_s(h))$.

Because of the representation (35), the right-hand side of (37) can be written:

$$\int_0^t \pi'_s(h) d\bar{w}_s$$

with $\bar{w}_s$ is a (F_t^y, Q) Brownian motion.

It follows that $M_t = \frac{d}{d\theta} \log \tilde{L}_t(\theta)$, is, for F_t^y and $Q(\theta)$, a martingale. It is well known that the increasing process associated with M_t, that is the unique continuously increasing process I_t, such that $M_t^2 - I_t$ is a martingale given by $I_t = \int_0^t [\pi_s'(h)]^2 ds$. It is the conditional Fisher information and the relation $I_\infty = \infty$ can be seen as the "derivative" of the condition

$$\int_0^\infty (\hat{h}_s' - \hat{h}_s)^2 ds = \infty \ .$$

We now have $\hat{h}_s(\theta + d\theta) - \hat{h}_s(\theta) \sim d\theta \ \pi_s'(h)$

The condition $[I_\infty = \infty]$ implies that for all $d\theta$ with $\theta' = \theta + d\theta$ we obtain $\int_0^\infty (\hat{h}_s' - \hat{h}_s)^2 ds = \infty$. That is, the two measures $Q(\theta + d\theta)$ and $Q(\theta)$ are disjoint. Then, thanks to Prop. 6,

$$\frac{L_t(\theta)}{L_t(\theta')} \rightarrow \infty$$

(resp. 0) $Q(\theta)$ a.s. (resp. $Q(\theta')$ a.s.). This property implise that the maximum of likelifiood estimate $\hat{\theta}_t$ converges in a case where θ is assumed to take a finite number of values. The proof can be achieved for θ in a compact set. (See e.g. Ref. [2]) The previous argument, which of course requires some additional assumptions to be rigorous, emphasizes however the main difference between the i.i.d. case, where the information takes the form $L_n(\theta) = nL_1(\theta)$, and so always diverges to infinity as the sample size increases, and the case considered here where such a condition is not always fulfilled.

Remark 5: In Ref. [4], the problem of the estimation of the drift function b has been studied in the much simpler case of complete observation $(x = y)$. In that paper, the case of a parameter appearing in a multiplicative way is first studied. This case would correspond here to the assumption $\pi_s^\theta(h) = \theta \hat{h}_s$. In this case, a direct computation of $\hat{\theta}_t$ is possible, and we obtain

$$\hat{\theta}_t = \frac{\int_0^t \hat{h}_s dy_s}{\int_0^t \hat{h}_s^2 ds} .$$

Hence, (35) gives: $\hat{\theta}_t = \theta + \frac{\int_0^t \hat{h}_s d\bar{w}_s}{\int_0^t \hat{h}_s^2 ds}$

and by the convergence theorem for L^2-Martingales, (see Ref. [7] prop. 7–2–4), we have:

$$\theta_t \to \theta \quad Q_\theta \text{ a.s. on the set } \int_0^\infty \hat{h}_s^2 ds = \infty .$$

Remark 6: If X_t is a stationary process which admits a unitque invariant measure, then π_t (as a Markov process with values in the set of all probability measures ν over the state space of X_t) it also admits a unique invariant measure ϕ, and one has

$$\frac{1}{t}\int_0^t \psi(\pi_s(f))ds \to \int \psi(\nu(f)\phi(d\nu) .$$

In this case we have

$$\frac{1}{t}\int_0^t \hat{h}_s^2 ds \to \infty$$

because

$$\frac{1}{t}\int_0^t \hat{h}_s^2 ds \to \int [\nu(h)]^2 \phi(d\nu) > 0 .$$

In that latter case, by (35) we have:

$$l_t(\theta) \equiv \frac{1}{t}\int_0^t \log \tilde{L}_t(\theta) = \frac{1}{t}\int_0^t h_s(\theta)\,[\hat{h}_s(\theta_0)\,ds + d\bar{w}_s] - \frac{1}{2t}\int_0^t \hat{h}_s^2(\theta)ds$$

and $\lim\, l_t(\theta) = \int \nu(h(\theta_0))\,[\nu(h(\theta_0)) - \frac{1}{2}\,\nu(h(\theta))]\phi_{\theta_0}(d\nu) \equiv K(\theta_0, \theta)$.

It is easily seen that $K(\theta_0, \theta)$ reaches its maximum value for $\theta = \theta_0$. Then we can use the same proofs as in the i.i.d. case to show that $\hat{\theta}_t$, the maximum of likelihood estimate, converges to θ_0, $Q(\theta_0)$ a.s.

In this case, the equation $\lim \frac{1}{t}\int_0^t I_s(\theta_0)\,ds = \int [\nu(h')]^2 \phi_{\theta_0}(d\nu) = \bar{I}(\theta_0)$ shows that $I_t(\theta_0) \to \infty$ $Q(\theta_0)$ a.s.. Also, by the same techniques as in the i.i.d. case, $\sqrt{t}\,(\hat{\theta}_t - \hat{\theta}_0)$ has the asymptotically normal distribution $N(0, [I(\theta_0)]^{-1})$.

References

1) M. Shima, F. Delebecque and N. Adachi: A Dynamic Model of Quality of River Water with considerations of Nitrification Process, Proceedings of IFIP Working Conference on Modeling and Simulation of Land, Air and Water Resources System, North-Holland Pub. Comp. (1978)
2) R. R. Bahadur: Limit theorem in Math. Statistical SIAM Publications (1971)
3) P. Brémaud: Filtrage récursif pour une observation mixte par la méthode de Probabilité de réference-Thèse Paris-6 (1976)
4) F. Delebecque and J. P. Quadrat: Sur I'Estimation des characteristique locales d'un processus de diffusion anec sauts-Rapport IRIA Laboria 311–1978.

5) H. Kunita: Asymptotic behavior of non-linear filtering error of Markov process—J. of Mult. Analysis (1971)
6) R. S. Liptser-Shiryayev: Statistics of Random Process I-II Spring. Verl. (1977)
7) J. Neveu: Martingales a temps discret—Masson Paris (1974)
8) P. Priouret: Processus de diffusion. In Lect. Notes in Math. 390. Spring. Verl. (1974)
9) K. Parthasarathy: Probability Measures on Metric spaces—Acad, Press. New York (1967)