# LIKELIHOOD BASED STATISTICS FOR PARTIALLY OBSERVED DIFFUSION PROCESSES*

F. Campillo, F. Le Gland
INRIA Sophia–Antipolis
route des Lucioles
F–06565 VALBONNE Cédex

## Abstract

The purpose of this paper is to study some statistical problems : parameter estimation, binary detection, change detection (disorder problem), etc. for partially observed diffusion processes, using the likelihood approach.

It is shown that the stochastic PDE related to the state estimation problem, provides also a way to compute the likelihood function/ratio.

A recent result on consistency of the MLE, in the small noise asymptotics, is also presented.

## 1 Introduction

Consider the following partially observed stochastic differential system, defined on some probability space

$$
\begin{aligned}
dX_t &= b(X_t)\,dt + \sigma(X_t)\,dW_t \\
dY_t &= h(X_t)\,dt + dV_t
\end{aligned}
\tag{1}
$$

where the non observed process $\{X_t\,,\,t \geq 0\}$ and the observation $\{Y_t\,,\,t \geq 0\}$ takes values in $\mathbf{R}^m$ and $\mathbf{R}^d$ respectively. $\{W_t\,,\,t \geq 0\}$ and $\{V_t\,,\,t \geq 0\}$ are independent Wiener processes of appropriate dimension, with covariance matrix $I$, and the random variable $X_0$ is independent of the Wiener processes, with probability distribution $p_0(x)\,dx$. The available information at time $t$ is contained in the $\sigma$–algebra $\mathcal{Y}_t = \sigma(Y_s\,,\,0 \leq s \leq t)$.

The first problem one is faced with, is state estimation : to estimate recursively the state $X_t$ given observations $\mathcal{Y}_t$ up to time $t$. The solution to this first problem is given by the Zakai equation, a stochastic partial differential equation which computes recursively the conditional density of $X_t$ given $\mathcal{Y}_t$. This assumes that the partially observed dynamical system (1) is completely identified, which usually is not the case.

*Research partially supported by USACCE under Contract DAJA45–90–C–0008, and by CNRS–GR Automatique.

Therefore, a second problem is to assume that the model (1) is parametrized by some unknown parameter $\theta$ in $\Theta \subset \mathbf{R}^p$, and to estimate $\theta$ on the basis of observations $\mathcal{Y}_t$. Several statistical problems are introduced in Section 2 for the parametrized model (1). Off–line statistical procedures based on likelihood are presented in Section 3. It is shown in Section 4 that the Zakai equation provides also a way to compute these likelihood statistics.

Another issue is to prove that the statistical algorithms based on the likelihood approach, actually provide good estimates, in some asymptotic sense. A recent result in this direction has been obtained for the consistency of the MLE, in the small noise asymptotics, see James–LeGland [4].

## 2 Statistical problems

Let $\Theta \subset \mathbf{R}^p$ denote the parameter space. Assume that observations $\{Y_t\,,\,0 \leq t \leq T\}$ are available from the following model

$$
\begin{aligned}
dX_t &= b_\theta(X_t)\,dt + \sigma(X_t)\,dW_t^\theta \\
dY_t &= h_\theta(X_t)\,dt + dV_t^\theta
\end{aligned}
$$

The statistical problems to be considered in this paper are

(a) parameter estimation : estimate $\theta \in \Theta$.

(b) binary detection : decide between the two simple hypotheses

$$
\begin{aligned}
\mathrm{H}_0 &: \quad \theta = \theta_0\ , \\
\mathrm{H}_1 &: \quad \theta = \theta_1\ .
\end{aligned}
$$

Another related problem is the sequential binary detection problem.

(c) change detection (disorder problem) : decide between the two multiple hypotheses

$$\mathrm{H}_0 \;:\quad \theta = \theta_0 \ ,$$

$$\mathrm{H}_1 \;:\quad \text{there exists } 0 \leq r \leq T, \text{ such that } \begin{cases} \theta = \theta_0 & \text{on} \quad 0 \leq t < r \ , \\ \theta = \theta_1 & \text{on} \quad r \leq t \leq T \ . \end{cases}$$

In case $\mathrm{H}_1$ has been decided, another problem of interest is to estimate the change time $r$.

A variant of this problem, is when only $\theta_0$ is known : the alternate hypothesis $\mathrm{H}_1$ is multiple with respect to both $r$ and $\theta_1$. In case $\mathrm{H}_1$ has been decided, both $r$ and $\theta_1$ are to be estimated.

(d) Bayesian change detection (jump Markov parameter) : recursively estimate $\theta_t$ given $\mathcal{Y}_t$, assuming that $\{\theta_t \, , \, t \geq 0\}$ is a finite state Markov process, independent of the Wiener processes, with jump intensity matrix $Q$. This problem is closely related to state estimation, see Loparo–Roth–Eckert [7].

For each of the problems listed above, the first step is to provide an expression, in terms of conditional expectation, for the likelihood function (LF), the likelihood ratio (LR), or the generalized likelihood ratio (GLR), depending on the problem.

# 3 Likelihood based off–line statistics

**Statistical model** On the canonical space $\Omega = C([0,T] \, ; \mathbf{R}^{m+d})$ are given

- a pair of stochastic processes $\{X_t \, , \, 0 \leq t \leq T\}$ and $\{Y_t \, , \, 0 \leq t \leq T\}$ taking values in $\mathbf{R}^m$ and $\mathbf{R}^d$ respectively,
- a family $\mathcal{M} = \{P_\theta \, , \, \theta \in \Theta\}$ of probability measures,

such that under $P_\theta$

$$\begin{aligned} dX_t &= b_\theta(X_t) \, dt + \sigma(X_t) \, dW_t^\theta \\ dY_t &= h_\theta(X_t) \, dt + dV_t^\theta \end{aligned} \tag{2}$$

where $\{W_t^\theta \, , \, 0 \leq t \leq T\}$ and $\{V_t^\theta \, , \, 0 \leq t \leq T\}$ are independent Wiener processes of appropriate dimension, with covariance matrix $I$, and the random variable $X_0$ is independent of the Wiener processes, with probability distribution $p_0^\theta(x) \, dx$. Throughout the paper, the coefficients are assumed to be continuous and bounded functions on $\mathbf{R}^m$.

The main assumption is that all the available information is contained in $\mathcal{Y}_T = \sigma(Y_t \, , \, 0 \leq t \leq T)$.

Introduce

$$Z_t^s[\theta] \triangleq \exp \left\{ \int_s^t h_\theta^*(X_\tau) \, dY_\tau - \tfrac{1}{2} \int_s^t |h_\theta(X_\tau)|^2 \, d\tau \right\}$$

and $Z_t[\theta] \triangleq Z_t^0[\theta]$ . Provided the probability measures on $\mathbf{R}^m$ with densities $\{p_0^\theta \, , \, \theta \in \Theta\}$ are mutually absolutely continuous, the statistical model defined above is dominated by some probability measure $P^\dagger$. Indeed, it is proved in [2] that

**Proposition 3.1** *The probability measures in $\mathcal{M}$ are mutually absolutely continuous. In addition*

$$\left. \frac{dP_\theta}{dP^\dagger} \right|_{\mathcal{Y}_T} = \mathbf{E}_\theta^\dagger (Z_T[\theta] \mid \mathcal{Y}_T) \ ,$$

*where $P_\theta^\dagger$ is the* reference probability *measure.*

## Parameter estimation

The likelihood function for the estimation of $\theta$, based on observations in $\mathcal{Y}_T$, is given by

$$L[\theta] \triangleq \left. \frac{dP_\theta}{dP^\dagger} \right|_{\mathcal{Y}_T} = \mathbf{E}_\theta^\dagger (Z_T[\theta] \mid \mathcal{Y}_T) \ , \tag{3}$$

and the maximum likelihood estimate (MLE) is defined as

$$\widehat{\theta} \in \operatorname*{argmax}_{\theta \in \Theta} L[\theta] \ .$$

To find $\widehat{\theta}$, one can use an iterative optimization algorithm for the maximization of the likelihood function $L[\theta]$. An alternative approach is to use the EM algorithm, as proposed by Dembo–Zeitouni [3]. This algorithm is based on the following immediate consequence of the Jensen inequality, where $\ell[\theta]$ denotes the log–likelihood function

$$\begin{aligned} \ell[\theta] - \ell[\theta'] &= \log \mathbf{E}_{\theta'}^\dagger \Big( \frac{Z_T[\theta]}{Z_T[\theta']} \mid \mathcal{Y}_T \Big) \\ &\geq \mathbf{E}_{\theta'}^\dagger \Big( \log \frac{Z_T[\theta]}{Z_T[\theta']} \mid \mathcal{Y}_T \Big) \triangleq Q[\theta, \theta'] \ . \end{aligned}$$

The idea of the EM algorithm is to replace the direct maximization of $L[\theta]$, by the iterative maximization of the auxiliary function, i.e.

$$\widehat{\theta}^{n+1} \in \operatorname*{argmax}_{\theta \in \Theta} Q[\theta, \widehat{\theta}^n] \ .$$

Under mild hypotheses, the sequence $\{\widehat{\theta}^n \, , \, n \geq 0\}$ converges to a stationary point of the original likelihood function $L[\theta]$. See Campillo–LeGland [2] for a comparison between the two approaches.

## Binary detection

The likelihood ratio for deciding between hypotheses $\mathrm{H}_0$ and $\mathrm{H}_1$, based on observations in $\mathcal{Y}_T$, is given by

$$R \triangleq \left. \frac{dP_1}{dP_0} \right|_{\mathcal{Y}_T} = \frac{L[\theta_1]}{L[\theta_0]} \ , \tag{4}$$

where $P_i = P_{\theta_i}$ for $i = 0, 1$. The likelihood ratio test is defined by the following reject region for the null hypothesis $\mathrm{H}_0$

$$R \geq c \ ,$$

where $c > 1$ is the threshold.

## Sequential binary detection

In this problem, the horizon is not fixed. Let $R_t$ denote the likelihood ratio for deciding between hypotheses $\mathrm{H}_0$ and $\mathrm{H}_1$, based on observations in $\mathcal{Y}_t$. An *admissible decision policy* for the sequential binary detection problem, is defined by a stopping time $\tau$ and a $\mathcal{Y}_\tau$–measurable $\{0,1\}$–valued decision random variable $\delta$ : if $\delta = 0$ (resp. $\delta = 1$) the null hypothesis $\mathrm{H}_0$ is accepted (resp. rejected). In other words, $\delta$ defines a reject region for the null hypothesis $\mathrm{H}_0$. A *threshold decision policy* is defined by a stopping time of the form

$$\tau \stackrel{\triangle}{=} \inf\{t \geq 0 \ : \ R_t \notin (A, B)\}$$

and a reject region for the null hypothesis $\mathrm{H}_0$ of the form

$$\delta = \begin{cases} 1 \ , & \text{if } R_\tau \geq B \ , \\ 0 \ , & \text{if } R_\tau \leq A \end{cases}$$

where $0 < A < 1 < B < \infty$ are the constant thresholds. This problem has been studied by Baras–LaVigna [1], following Liptser–Shiryayev [6].

## Change detection (disorder)

The statistical model for this problem can be described through the introduction of time dependent coefficients : for $0 \leq r \leq T$ , let $P_r$ denote the probability measure on the canonical space $\Omega$, under which

$$\begin{aligned} dX_t &= b_r(t, X_t)\, dt + \sigma(X_t)\, dW_t^r \\ dY_t &= h_r(t, X_t)\, dt + dV_t^r \end{aligned} \qquad (5)$$

where $\{W_t^r \ , \ 0 \leq t \leq T\}$ and $\{V_t^r \ , \ 0 \leq t \leq T\}$ are independent Wiener processes of appropriate dimension, with covariance matrix $I$, and

$$b_r(t,x) \stackrel{\triangle}{=} b_0(x) + [b_1(x) - b_0(x)]\ 1_{\{r \leq t \leq T\}} \ ,$$

$$h_r(t,x) \stackrel{\triangle}{=} h_0(x) + [h_1(x) - h_0(x)]\ 1_{\{r \leq t \leq T\}} \ ,$$

where $b_i(x) = b_{\theta_i}(x)$ and $h_i(x) = h_{\theta_i}(x)$ for $i = 0, 1$

Introducing for $0 \leq r \leq T$

$$Z_t^s[r] \stackrel{\triangle}{=} \exp\left\{ \int_s^t h_r^*(\tau, X_\tau)\, dY_\tau - \tfrac{1}{2} \int_s^t |h_r(\tau, X_\tau)|^2\, d\tau \right\}$$

and $Z_t[r] \stackrel{\triangle}{=} Z_t^0[r]$ , it holds that the probability measures $\{P_r^\dagger \ , \ 0 \leq r \leq T\}$ are mutually absolutely continuous. Moreover

$$\left.\frac{dP_r}{dP^\dagger}\right|_{\mathcal{Y}_T} = \mathbf{E}_r^\dagger(Z_T[r] \mid \mathcal{Y}_T) \ .$$

Note that, with this definition, the probability associated with the null hypothesis $\mathrm{H}_0$ is $P_T$.

The generalized likelihood ratio for deciding between hypotheses $\mathrm{H}_0$ and $\mathrm{H}_1$, based on observations in $\mathcal{Y}_T$, is given by

$$R \stackrel{\triangle}{=} \max_{0 \leq r \leq T} \left.\frac{dP_r}{dP_T}\right|_{\mathcal{Y}_T} = \max_{0 \leq r \leq T} \frac{L[r]}{L[T]} \ , \qquad (6)$$

where

$$L[r] \stackrel{\triangle}{=} \left.\frac{dP_r}{dP^\dagger}\right|_{\mathcal{Y}_T} = \mathbf{E}_r^\dagger(Z_T[r] \mid \mathcal{Y}_T) \ ,$$

is the likelihood function for the estimation of the change time $r$, based on observations in $\mathcal{Y}_T$. The generalized likelihood ratio test is defined by the following reject region for the null hypothesis $\mathrm{H}_0$

$$R \geq c \ ,$$

where $c > 1$ is the threshold.
Moreover, in case $\mathrm{H}_1$ has been decided, the maximum likelihood estimate of the change time $r$, based on observations in $\mathcal{Y}_T$, is given by

$$\widehat{r} \in \operatorname*{argmax}_{0 \leq r \leq T} L[r] \ .$$

Introducing the $\sigma$–algebras $\mathcal{F}_t = \sigma(X_s \ , \ 0 \leq s \leq t)$ and $\mathcal{Y}_t^s = \sigma(Y_\tau - Y_s \ , \ s \leq \tau \leq t)$, the following decomposition holds for the likelihood function $L[r]$

$$\begin{aligned} L[r] &= \mathbf{E}_r^\dagger(Z_T[r] \mid \mathcal{Y}_T) = \mathbf{E}_r^\dagger(Z_r[0] \cdot Z_T^r[1] \mid \mathcal{Y}_T) \\ &= \mathbf{E}_r^\dagger\left(\mathbf{E}_r^\dagger(Z_r[0] \cdot Z_T^r[1] \mid \mathcal{F}_r \vee \mathcal{Y}_r \vee \mathcal{Y}_T^r) \mid \mathcal{Y}_T\right) \\ &= \mathbf{E}_r^\dagger\left(Z_r[0] \cdot \mathbf{E}_r^\dagger(Z_T^r[1] \mid \mathcal{F}_r \vee \mathcal{Y}_T^r) \mid \mathcal{Y}_T\right) \\ &= \mathbf{E}_r^\dagger\left(Z_r[0] \cdot \mathbf{E}_1^\dagger(Z_T^r[1] \mid \mathcal{F}_r \vee \mathcal{Y}_T^r) \mid \mathcal{Y}_T\right) \ , \end{aligned}$$

where $Z_t^s[i] = Z_t^s[\theta_i]$ for $i = 0, 1$. Defining

$$v_t^1(x) \stackrel{\triangle}{=} \mathbf{E}_1^\dagger(Z_T^t[1] \mid \mathcal{Y}_T^t \vee \{X_t = x\}) \ ,$$

it holds

$$\begin{aligned} L[r] &= \mathbf{E}_r^\dagger(Z_r[0] \cdot v_r^1(X_r) \mid \mathcal{Y}_T) \\ &= \mathbf{E}_0^\dagger(Z_r[0] \cdot v_r^1(X_r) \mid \mathcal{Y}_T) \ . \end{aligned} \qquad (7)$$

The purpose of the next section is to provide some computational procedure, that would allow to numerically compute the likelihood based statistics introduced so far.

# 4 Computational likelihood statistics

In this section, the link between the likelihood based statistical problems introduced above, and the state estimation problem, will be investigated. At this point, it is necessary to introduce some notations and definitions related to nonlinear filtering and smoothing.

For the sake of simplicity, any reference to the parameter $\theta$ will be dropped for the time being.

□ *Filtering:* Let $p_t$ denote the unnormalized conditional density of the random variable $X_t$ given $\mathcal{Y}_t$, defined by

$$(p_t, \phi) \triangleq \mathbf{E}^\dagger(\phi(X_t) Z_t \mid \mathcal{Y}_t) \tag{8}$$

for any test–function $\phi$. The unnormalized conditional density $\{p_t \, , \, t \geq 0\}$ satisfies the Zakai equation [8]

$$dp_t = L^* p_t \, dt + h^* p_t \, dY_t \ , \tag{9}$$

where $L$ is the following partial differential operator, associated with the stochastic differential system (1)

$$L \triangleq \tfrac{1}{2} \sum_{i,j=1}^m a^{i,j}(\cdot) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^m b^i(\cdot) \frac{\partial}{\partial x_i} \ .$$

□ *Smoothing (fixed–interval):* Let $T > 0$ denote the fixed end–time, and $q_t$ denote the unnormalized conditional density of the random variable $X_t$ given $\mathcal{Y}_T$, defined by

$$(q_t, \phi) \triangleq \mathbf{E}^\dagger(\phi(X_t) Z_T \mid \mathcal{Y}_T) \ .$$

Introducing the backward Zakai equation

$$dv_t + L v_t \, dt + h^* v_t \, dY_t = 0 \ , \qquad v_T \equiv 1 \ , \tag{10}$$

it is proved in Pardoux [8, 9] that $(p_t, v_t)$ is independent of $t$ and $q_t = p_t \cdot v_t$ . In addition

$$v_t(x) = \mathbf{E}^\dagger(Z_T^t \mid \mathcal{Y}_T^t \vee \{X_t = x\}) \ .$$

Existence, uniqueness and regularity results for stochastic PDE can be found in Pardoux [8].

Let now $L_\theta$ and $L_r(t)$ denote the partial differential operators

$$L_\theta \triangleq \tfrac{1}{2} \sum_{i,j=1}^m a^{i,j}(\cdot) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^m b_\theta^i(\cdot) \frac{\partial}{\partial x_i} \ ,$$

$$L_r(t) \triangleq \tfrac{1}{2} \sum_{i,j=1}^m a^{i,j}(\cdot) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^m b_r^i(t, \cdot) \frac{\partial}{\partial x_i} \ ,$$

associated with the stochastic differential equation (2) and (5) respectively.

## Parameter estimation

The following expression holds for the likelihood function (3)

$$L[\theta] = (p_T^\theta, 1) \ ,$$

where the unnormalized conditional density $\{p_t^\theta \, , \, t \geq 0\}$ solves the Zakai equation

$$dp_t^\theta = L_\theta^* p_t^\theta \, dt + h_\theta^* p_t^\theta \, dY_t \ .$$

## Binary detection

A similar expression holds for the likelihood ratio (4)

$$R = \frac{(p_T^1, 1)}{(p_T^0, 1)} \ .$$

Here the unnormalized conditional density $\{p_t^i \, , \, t \geq 0\}$ solves the Zakai equation

$$dp_t^i = L_i^* p_t^i \, dt + h_i^* p_t^i \, dY_t \ , \tag{11}$$

where $L_i = L_{\theta_i}$ and $h_i = h_{\theta_i}$ for $i = 0, 1$ .

## Change detection (disorder)

Let $\{p_t^r \, , \, t \geq 0\}$ and $\{v_t^r \, , \, t \geq 0\}$ denote the solution of

$$dp_t^r = L_r^*(t) p_t^r \, dt + h_r^*(t) p_t^r \, dY_t \ ,$$

and

$$dv_t^r + L_r(t) v_t^r \, dt + h_r^*(t) v_t^r \, dY_t = 0 \ , \qquad v_T^r \equiv 1 \ ,$$

respectively, where $h_r(t)$ is shorthand for $h_r(t, \cdot)$. The generalized likelihhod ratio (6) is given by

$$R = \max_{0 \leq r \leq T} \frac{(p_T^r, 1)}{(p_T^T, 1)} \ . \tag{12}$$

However, a much more efficient expression can be obtained. Indeed, for all $0 \leq t \leq T$

$$L[r] = (p_T^r, 1) = (p_t^r, v_t^r) \ ,$$

and in particular for $t = r$

$$L[r] = (p_r^r, v_r^r) = (p_r^0, v_r^1) \ , \tag{13}$$

where

$$dp_t^0 = L_0^* p_t^0 \, dt + h_0^* p_t^0 \, dY_t \ , \tag{14}$$

and

$$dv_t^1 + L_1 v_t^1 \, dt + h_1^* v_t^1 \, dY_t = 0 \ , \qquad v_T^1 \equiv 1 \ . \tag{15}$$

Therefore, it is enough to solve two stochastic PDE, the forward equation (14) with parameter $\theta_0$, and the backward equation (15) with parameter $\theta_1$. This gives the following expression for the generalized likelihood ratio

$$R = \max_{0 \leq r \leq T} \frac{(p_r^0, v_r^1)}{(p_T^0, 1)} \ .$$

which is much more efficient than the original expression (12), which would require to solve an infinite number of stochastic PDE (see Figure 1).

**Remark 4.1** The expression (13) for the likelihood ratio could also be obtained from the previous expression (7) obtained by decomposition.

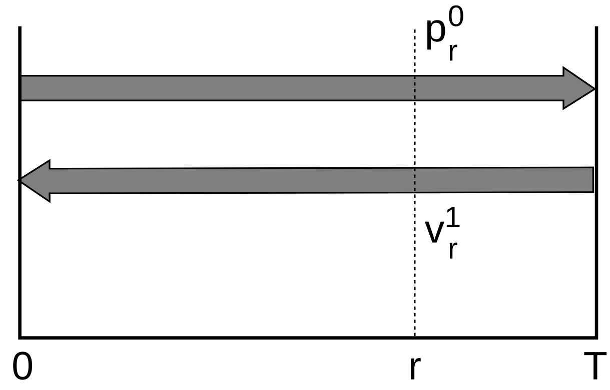


Figure 1: Stochastic PDE for the disorder problem.

It can also be proved that the likelihood function $r \mapsto L[r]$ is smooth, provided the change can only occur in the drift coefficient, i.e. $h_0 = h_1$. Actually, using the two–sided stochastic calculus developped in Pardoux [10]

$$\begin{aligned} d(p_t^0, v_t^1) &= (L_0^* p_t^0, v_t^1)\, dt + (h_0^* p_0^t, v_t^1)\, dY_t \\ &\quad - (p_t^0, L_1 v_t^1)\, dt - (p_t^0, h_0^* v_t^1)\, dY_t \\ &= (p_t^0, [L_0 - L_1]\, v_t^1)\, dt\ . \end{aligned}$$

## Bayesian change detection

The unnormalized conditional distribution of the compound state $(X_t, \theta_t)$ given observations in $\mathcal{Y}_t$, is defined by

$$(p_t^i, \phi) = \mathbf{E}^\dagger(\phi(X_t)\ 1_{\{\theta_t\, =\, i\}}\ Z_t[\theta] \mid \mathcal{Y}_t)\ ,$$

for $i = 1, 2, \ldots, N$ , where in this section the process $\{Z_t[\theta]\, , 0 \le t \le T\}$ is defined by

$$Z_t^s[\theta] \triangleq \exp\left\{ \int_s^t h_{\theta_\tau}^*(X_\tau)\, dY_\tau - \tfrac{1}{2} \int_s^t |h_{\theta_\tau}(X_\tau)|^2\, d\tau \right\}$$

and $Z_t[\theta] \triangleq Z_t^0[\theta]$ . In addition $\{p_t^i\, , 0 \le t \le T\}$ satisfies the following system of coupled Zakai equations

$$dp_t^i = L_i^* p_t^i\, dt + h_i^* p_t^i\, dY_t + \sum_{j=1}^N q_{j,i}\, p_t^j\, dt\ , \qquad (16)$$

where $Q = \{q_{i,j}\}$ is the jump intensity matrix for the Markov process $\{\theta_t\, , 0 \le t \le T\}$. Note that in system (16) the coupling occurs only through zero–order state–independent coefficients.

The unnormalized marginal conditional distribution

$$(p_t^i, 1) = c_t \cdot P(\theta_t = i \mid \mathcal{Y}_t)\ ,$$

can be used to compute the maximum a posteriori (MAP) estimate

$$\widehat{\theta}_t^{MAP} \in \underset{1 \le i \le N}{\operatorname{argmax}}\ (p_t^i, 1)\ .$$

Assuming that the jump intensity matrix is of the form $\varepsilon \cdot Q$ where $\varepsilon > 0$ is a small parameter – which means that the frequency of the jumps is small – it is possible to obtain an asymptotic expansion of the unnormalized conditional distribution in powers of $\varepsilon$.

# 5 Asymptotic statistics

Some off–line statistical procedures based on likelihood, have been presented. Whether these statistical procedures actually provide good results, has to be investigated in some asymptotic sense. Two kind of asymptotics are generally considered in the statistics of random processes, see Kutoyants [5]

- the small noise asymptotics, where the noise covariances are of order $\sqrt{\varepsilon}$, and $\varepsilon$ is sent to zero,
- the long–time asymptotics, where the observation horizon $T$ is sent to infinity.

This section is devoted to presenting a recent result on the consistency of the MLE in the small noise asymptotics, see James–LeGland [4].

**Statistical model** On the canonical space $\Omega = C([0,T]\, ; \mathbf{R}^{m+d})$ are given

- a pair of stochastic processes $\{X_t\, , 0 \le t \le T\}$ and $\{Y_t\, , 0 \le t \le T\}$ taking values in $\mathbf{R}^m$ and $\mathbf{R}^d$ respectively,
- for each $\varepsilon > 0$, a family $\mathcal{M}^\varepsilon = \{P_{\theta,\varepsilon}\, , \theta \in \Theta\}$ of probability measures,

such that under $P_{\theta,\varepsilon}$

$$\begin{aligned} dX_t &= b_\theta(X_t)\, dt + \sqrt{\varepsilon}\, dW_t^{\theta,\varepsilon} \\ dY_t &= h_\theta(X_t)\, dt + \sqrt{\varepsilon}\, dV_t^{\theta,\varepsilon} \end{aligned} \qquad (17)$$

where $\{W_t^{\theta,\varepsilon}\, , 0 \le t \le T\}$ and $\{V_t^{\theta,\varepsilon}\, , 0 \le t \le T\}$ are independent Wiener processes of appropriate dimension, with covariance matrix $I$, and the random variable $X_0$ is independent of the Wiener processes, with probability distribution $p_0^{\theta,\varepsilon}(x)\, dx$. It is assumed that the initial density is of the form

$$p_0^{\theta,\varepsilon}(x) = C_{\theta,\varepsilon} \cdot \exp\{-\frac{1}{\varepsilon} S_0^\theta(x)\}\ ,$$

where the function $S_0^\theta$ has an unique minimizer $\overline{x}_0^\theta$.

**Limiting deterministic system** For any $\theta \in \Theta$, consider the following deterministic differential system

$$(\Sigma^\theta) \qquad \left\{ \begin{aligned} \dot{x}_t^\theta &= b_\theta(x_t^\theta)\ , & x_0^\theta &= \overline{x}_0^\theta \\ \dot{y}_t^\theta &= h_\theta(x_t^\theta)\ , & y_0^\theta &= 0 \end{aligned} \right.$$

which is obtained from (17) by sending $\varepsilon$ to zero. This defines a family $\mathcal{M}^0 = \{(\Sigma^\varepsilon) , \theta \in \Theta\}$ of deterministic differential systems.

Actually, the following convergence in probability of the experiments holds

$$P_{\theta,\varepsilon}( \sup_{0\leq t\leq T} |Y_t - y_t^\theta| > \delta) \xrightarrow{\varepsilon\downarrow 0} 0 .$$

**Deterministic parameter estimation** Assume that a trajectory $\{y_t^\alpha , t \geq 0\}$ is observed, which is actually the output of some deterministic differential system $(\Sigma^\alpha)$ in the model $\mathcal{M}^0$, for some unknown $\alpha$. The problem is to estimate the parameter $\theta \in \Theta$, based on observations $\{y_t^\alpha , t \geq 0\}$.

The model $\mathcal{M}^0$ is said *identifiable* on $[0,T]$, if for all $\theta' \neq \theta$, there exists $t \in [0,T]$ such that $y_t^{\theta'} \neq y_t^\theta$, i.e. different values of the parameter give different output trajectories. In other words, the mapping $\theta \mapsto \{y_t^\theta , 0 \leq t \leq T\}$ is injective. The deterministic parameter estimation problem consists of inverting this mapping. This can be expressed in terms of the following variational problem.

Introduce the cost functional

$$J_\alpha^\theta(\xi,t) \;=\; S_0^\theta(\xi_0) + \tfrac{1}{2} \int_0^t |\dot{\xi}_s - b_\theta(\xi_s)|^2 \, ds$$

$$+ \tfrac{1}{2} \int_0^t |\dot{y}_s^\alpha - h_\theta(\xi_s)|^2 \, ds - \tfrac{1}{2} \int_0^t |\dot{y}_s^\alpha|^2 \, ds \;,$$

for absolutely continuous $\xi \in C([0,T] ; \mathbf{R}^m)$, and the following least–squares functional

$$\ell_\alpha[\theta] = \inf\{J_\alpha^\theta(\xi,T) \; : \; \xi \in C([0,T] ; \mathbf{R}^m)\} .$$

The deterministic parameter estimate (DPE) is defined by

$$M_\alpha = \operatorname*{argmin}_{\theta\in\Theta} \ell_\alpha[\theta] .$$

The following consistency result can be proved, which relies on PDE techniques for large deviations (*vanishing viscosity* theorem) and on the convergence in probability of the experiments.

**Theorem 5.1** *For all $\alpha \in \Theta$, if the deterministic model $\mathcal{M}^0$ is identifiable, then any MLE sequence $\{\widehat{\theta}^\varepsilon , \varepsilon > 0\}$ converges in $P_{\alpha,\varepsilon}$–probability to the "true" parameter : for all $\delta > 0$*

$$P_{\alpha,\varepsilon}(|\widehat{\theta}^\varepsilon - \alpha| > \delta) \xrightarrow{\varepsilon\downarrow 0} 0 .$$

Another issue is the rate of convergence of the MLE sequence to the true value of the parameter. The solution to this question relies on proving a local asymptotic normality (LAN) result. This is currently under investigation.

Other problems should also be considered, including : large time asymptotics, recursive (on–line) estimation, and adaptive filtering.

# References


[1] J.S. BARAS and A. LA VIGNA, Real–time sequential hypotheses testing for diffusion signals, in: *26th IEEE CDC*, Los Angeles–1987, 1153–1157, IEEE (1987).

[2] F. CAMPILLO and F. LE GLAND, MLE for partially observed diffusions : direct maximization vs. the EM algorithm, *Stochastic Processes and Applications* **33** (2) 245–274 (1989).

[3] A. DEMBO and O. ZEITOUNI, Parameter estimation of partially observed continuous–time stochastic processes via the EM algorithm, *Stochastic Processes and Applications* **23** (1) 91–113 (1986).

[4] M. JAMES and F. LE GLAND, Consistent parameter estimation for partially observed diffusions with small noise, INRIA Research Report #1223 (May 1990).

[5] Yu.A. KUTOYANTS, *Parameter estimation for stochastic processes*, Heldermann Verlag (1984).

[6] R.S. LIPTSER and A.N. SHIRYAYEV, *Statistics of Random Processes*, Springer–Verlag (1977).

[7] K.A. LOPARO, Z. ROTH and S.J. ECKERT, Nonlinear filtering for systems with random structure, *IEEE Transactions* **AC–31** (11) 1164–1168 (1986).

[8] E. PARDOUX, Stochastic PDEs and filtering of diffusion processes, *Stochastics* **3** (2) 127–167 (1979).

[9] E. PARDOUX, Equations du lissage non–linéaire, in: *Filtering and Control of Random Processes*, Paris–1983 (eds. H.Korezlioglu, G.Mazziotto and J.Szpirglas) 206–218, Springer–Verlag (LNCIS–61) (1984).

[10] E. PARDOUX, Two–sided stochastic calculus for SPDEs, in: *Stochastic PDEs and Applications (Trento–1985)* (eds. G.DaPrato and L.Tubaro) 200–207, Springer–Verlag (LNM–1236) (1987).

[11] A.S. WILLSKY and H.L. JONES, A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems, *IEEE Transactions* **AC–21** (1) 108–112 (1976).

[12] A.I. YASHIN, On a problem of sequential hypothesis testing, *Theory of Probability and Applications* **28** (1) 157–165 (1983).