# On the maximum likelihood estimation in the case of dependent random variables

Wojciech Pieczynski

*Pub. Inst. Stat. Univ.*
*XXXIII, fasc. 2, 1988, 79 à 87*

# ON THE MAXIMUM LIKELIHOOD ESTIMATION IN THE CASE OF DEPENDENT RANDOM VARIABLES

Wojciech PIECZYNSKI
Département Mathématiques et Systèmes de Communication
Ecole Nationale Supérieure des Télécommunications de Bretagne
BP 832 - 29285 BREST CEDEX

**Abstract**

Our approach, based on a general theorem about asymptotic separation by J. Geffroy, requires neither any regularity assumption of the likelihood function nor any version of the law of large numbers for martingales. Furthermore, it is shown that the rate of convergence of the estimator is exponential.

## I - Introduction

Several authors have shown interest in the problem of the behaviour of the m.l.e. in the case of dependent observation S. Silvey (1961) gave some conditions under which the m.l.e. is asymptotically normal and weakly consistent, but they do not seem to be readily applicable. Bars-Shalom (1972) and Bhat (1974) pointed out sufficient conditions for existency and asymptotical efficiency. A form of the law of large numbers for martingales is the main tool of their approach. Asymptotic normality is studied with a form of the limit central theorem. Crowder (1975) uses a two term Taylor expansion of the likelihood function.

The approach of this paper is quite different : it is based on Geffroy's decantation theorem. This theorem, recently improved by Moche (7), specifies, in the case of dependent random variables, the construction of a consistent estimator whose exponential rate of convergence is ensured by some mild conditions expressed in terms of conditional probabilities.

## II - Theoretical investigations.

We consider a statistical model $(X,B,P_\theta)_{\theta\in\Theta}$, with $(X,B) = \bigotimes_{n=1}^{\infty} (X_n,B_n)$.

For each $n \geq 1$ let $\mu_n$ be a positive $\sigma$ finite measure on $(X_n, B_n)$, $P_\theta^{(n)}$ the projection of $P_\theta$ on $(X^{(n)}, B^{(n)}) = \bigotimes_{1\leq i\leq n} (X_i, B_i)$. $P_\theta^{(n)}$ is assumed to be

defined by the conditional densities $f_{\theta,x(n-1)}$ relatively to $\mu_n$. We will denote $P_{\theta,x(n-1)} = f_{\theta,x(n-1)}\,\mu_n$ the corresponding conditional probabilities. The conditionning is symbolic for n = 1 (i.e $f_{\theta,x(o)} = f_{\theta,1}$). $(\Theta, d)$ is a metric compact space. For each $\delta > O$ let us state:

$$f_{\theta, x^{(n-1)}, \delta} = \sup_{d(\theta,\theta')<\delta} f_{\theta', x^{(n-1)}} \tag{1}$$

$$a_{\theta, x^{(n-1)}, \delta} = \int_{X_n} f_{\theta, x^{(n-1)}, \delta}\, d\mu_n \tag{2}$$

$$a_{\theta, n, \delta} = \sup_{x^{(n-1)} \in X^{(n-1)}} a_{\theta, x^{(n-1)}, \delta} \tag{3}$$

We shall denote $(H_1)$ the following property :

$(H_1)$ $\forall\ \theta \in \Theta \quad \forall \varepsilon > 0 \quad \exists\ \delta > 0$ such that $\forall\ n \geq 1 \quad a_{\theta,n,\delta} \leq 1+\varepsilon$

Our second hypothesis will be :

For any $\theta_1 \neq \theta_2$ there exists $\alpha > O$ and $(w_n)$ a sequence of elements in $[O,1]$ such that for every $n \geq 1$, $x^{(n-1)} \in X^{(n-1)}$ one can define $B_{x^{(n-1)}} \in B_n$ for which

$$P_{\theta_1, x^{(n-1)}}[B_{x^{(n-1)}}] \geq w_n + \alpha$$

and

$(H_2)$

$$P_{\theta_2, x^{(n-1)}}[B_{x^{(n-1)}}] \leq w_n$$

Supposing the existence of the m.l.e., denoted $(\hat{\theta}_n)$, we can state the following result :

**Theorem :**

$(H_1)$ and $(H_2)$ imply the existence, for each $\theta \in \Theta$ and $\varepsilon > 0$, of $k > 0$ and $c > 0$ such that

$$P_\theta^{(n)}[d(\hat{\theta}_n, \theta) \geq \varepsilon] \leq ce^{-kn} \qquad (4)$$

for each $n \in N^*$

Proof.

Let $f^{(n)}_\theta$ be a density of $P^{(n)}_\theta$ relatively to $\mu^{(n)} = \mu_1 \otimes \ldots \otimes \mu_n$. For each $\delta > 0$ we shall denote

$$f_{\theta}^{(n),\delta} = \sup_{d(\theta',\theta)<\delta} f_{\theta'}^{(n)} \quad \text{and, for } \theta_1 \neq \theta_2 :$$

$$E^{(n),\delta}_{\theta_1,\theta_2} = \left\{ f^{(n)}_{\theta_1} > f^{(n),\delta}_{\theta_2} \right\} \qquad (5)$$

Let's show that $(H_1)$ and $(H_2)$ imply the following property :

For every $\theta_1 \neq \theta_2$ there exists $\delta > 0$, $c_1 > 0$, $k_1 > 0$ such that :

$$(H_3) \qquad P^{(n)}_{\theta_1}\left[E^{(n),\delta}_{\theta_1,\theta_2}\right] \geq 1 - c_1 e^{-k_1 n}$$

for each $n \in N^*$

(which is a sufficient condition of (4)).

For each $n \geq 1$, $x^{(n-1)}$, $\theta \in \Theta$, $\delta > 0$ let's put

$$f^*_{\theta, x^{(n-1)}, \delta} = \frac{1}{a} f_{\theta, x^{(n-1)}, \delta} \; f_{\theta, x^{(n-1)}, \delta} \qquad (6)$$

we obtain probability's densities - let us denote $P^*_{\theta,x^{(n-1)},\delta}$ the corresponding probabilities. Using $(H_1)$ we can easily prove that for every $B \in B_n$

$$\left| P^{*}_{\theta, x^{(n-1)}, \delta}[B] - P_{\theta, x^{(n-1)}}[B] \right| \xrightarrow[\delta \to 0]{} 0 \qquad (7)$$

and the above convergence is uniform relatively to n and $x^{(n-1)}$.

Let $\theta_1 \neq \theta_2$. $(H_2)$ and (7) allov us to affirm :

There exist $\delta > 0$, $\alpha > 0$ and $(w_n)$ a sequence of elements in $[0,1]$ such that for every $n \geq 1$, $x^{(n-1)} \in X^{(n-1)}$ one can define $B \in B_n$ for which

$$P_{\theta_1, x^{(n-1)}}[B] \geq w_n + \alpha \quad \text{and} \quad P^{*}_{\theta_2, x^{(n-1)}, \delta}[B] \leq w_n + \frac{\alpha}{2}$$

This property implies the existence of a sequence $(D_n)$ $(D_n \in B^{(n)}$ for every n such that :

$$P^{(n)}_{\theta_1}[D_n] \geq 1 - e^{-\alpha_1 n} \qquad (9)$$

$$P^{(n)*}_{\theta_2, \delta}[D_n] \leq e^{-\alpha_1 n} \qquad (10)$$

where $P^{(n)*}_{\theta_2, \delta}$ is the probability defined on $B^{(n)}$ by the family $P^{*}_{\theta_2, x^{(n-1)}, \delta}$ and $\alpha_1 = 1/8\ \alpha^2$. For the demonstration see [5] or [7]. (The sequence $(D_n)$ can be constructed from the sets $B_{X^{(n-1)}}$).

Let $\theta_1 \neq \theta_2$ . We can choose, in accordance with $(H_1)$, $\delta > 0$ such that for every $n \in N^*$:

$$a^{n, \delta}_{\theta_2} \leq e^{\frac{\alpha_1}{2}} \qquad (11)$$

Furthermore, we can suppose that this $\delta > 0$ also verifies (10). Let us consider $P_{\theta}{}^{(n),\delta}$ - measures defined on $(X^{(n)}, B^{(n)})$ by the densities $f_{\theta}{}^{(n),\delta}$ (defined at the beginning of the proof) and $Q_{\theta}{}^{(n),\delta}$ - measures defined on $(X^{(n)}, B^{(n)})$ by the conditional densities $f_{\theta, x^{(n-1)}, \delta}$. Knowing that the density of $P_{\theta}{}^{(n),\delta}$ (respectively $Q_{\theta}{}^{(n),\delta}$) is :

$$\sup_{d(\theta', \theta) < \delta} \prod_{i=1}^{n} f_{\theta', x^{(i-1)}}(x_i)$$

respectively : $$\prod_{i=1}^{n} \sup_{d(\theta', \theta) < \delta} f_{\theta', x^{(i-1)}}(x_i)$$

We can state :

$$P_{\theta_2}^{(n), \delta} \leq Q_{\theta_2}^{(n), \delta} \tag{12}$$

Now, (3), (6) and (11) imply :

$$f_{\theta_2, x^{(n-1)}, \delta} \leq a_{\theta_2, n, \delta} f^*_{\theta_2, x^{(n-1)}, \delta} \leq e^{\frac{\alpha_1}{2}} f^*_{\theta_2, x^{(n-1)}, \delta} \tag{13}$$

This last inequality involves :

$$Q_{\theta_2}^{(n), \delta} \leq e^{\frac{\alpha_1}{2} n} P^{(n)*}_{\theta_2, \delta} \tag{14}$$

Finally, (14), (12) and (10) allow us to write :

$$P_{\theta_2}^{(n), \delta}[D_n] \leq e^{\frac{\alpha_1}{2} n} e^{-\alpha_1 n} = e^{-\frac{\alpha_1}{2} n} \tag{15}$$

Now, the very definition of $E^{(n), \delta}_{\theta_1, \theta_2}$ (see (5)) implies :

$$P_{\theta_1}^{(n)}\left[E^{(n), \delta}_{\theta_1, \theta_2}\right] - P_{\theta_2}^{(n), \delta}\left[E^{(n), \delta}_{\theta_1, \theta_2}\right] \geq P_{\theta_1}^{(n)}[D_n] - P_{\theta_2}^{(n), \delta}[D_n] \tag{16}$$

so, (9), (15) and (16) imply

$$P_{\theta_1}^{(n)}\left[E_{\theta_1,\theta_2}^{(n),\delta}\right] \geq 1-2e^{-\frac{\alpha_1}{2}n}$$

so $(H_3)$ is verified. Let us show that $(H_3)$ is a sufficient condition of (4). Let $\theta \in \Theta$, $\varepsilon > 0$. Writing $D=\{\theta' \in \Theta / d(\theta,\theta') \geq \varepsilon\}$ we shall denote, for every $\theta' \in D$, $\delta_{\theta,\theta'}$ a $\delta > 0$ associated with $(\theta, \theta')$ by $(H_3)$. The set of open balls $[B(\theta', \delta_{\theta,\theta'})]_{\theta' \in D}$ is a cover of D which is compact - let $[B(\theta_i, \delta_i)]_{1\leq i\leq N}$ (with $\delta_i = \delta_{\theta,\theta_i}$) be a finite subcover. Denoting $c_i$, $k_i$ the numbers associated with $(\theta,\theta_i)$ by $(H_3)$ we can write :

$$\forall\, 1 \leq i \leq N \qquad P_{\theta}^{(n)}\left[E_{\theta,\theta_i}^{(n),\delta_i}\right] \geq 1-c_i e^{-k_i n} \tag{17}$$

Stating $c = c_1 + \ldots + c_N$ and $k = \inf_{1\leq i\leq N} k_i$, (17) implies :

$$P_{\theta}^{(n)}\left[\bigcap_{1\leq i\leq N} E_{\theta,\theta_i}^{(n),\delta_i}\right] \geq 1-ce^{-kn} \tag{18}$$

Now, $[B(\theta_i, \delta_i)]_{1\leq i\leq N}$ being a cover of D, the very definition of sets $E_{\theta_1,\theta_2}^{(n),\delta}$ implies :

$$\bigcap_{1\leq i\leq N} E_{\theta,\theta_i}^{(n),\delta_i} \subset \{d(\hat{\theta}_n, \theta) < \varepsilon\} \tag{19}$$

(18) and (19) yield :

$$P_{\theta}^{(n)}[d(\hat{\theta}_n, \theta) < \varepsilon] \geq 1-ce^{-kn}$$

which ends the proof.

Remark

Hypothesis $(H_1)$ denotes a kind of continuity of the conditional probabilities with regard to the parameter. If the set of parameters $\Theta$ is finite, this hypothesis becomes useless and the demonstration is highly simplified : hypothesis $(H_2)$ alone implies the following property $(H_4)$ (to compare with $(H_3)$) :

For every $\theta_1 \neq \theta_2$ there exist $c_1 > 0$, $k_1 > 0$ such that

$$(H_4) \qquad P_{\theta_1}^{(n)}\left[f_{\theta_1}^{(n)} \geq f_{\theta_2}^{(n)}\right] \geq 1 - c_1 e^{-k_1 n}$$

for each $n \in N^*$

which implies (4).

## III - Application to gaussian processes.

Let $(X_n)_{n \geq 1}$ be a real gaussian process with $\theta = E[X_n]$, $\theta$ being in a compact set $\Theta$. For each $x^{(n-1)} = (x_1, \ldots, x_{n-1})$ the conditional probabilities are normal with mean

$$m_{\theta, x^{(n-1)}} = d_n \theta + \sum_{k=1}^{n-1} c_k x_k$$

and variance $\sigma^2_n$ independent from $\theta$ and $x^{(n-1)}$. $d_n$ is defined by :

$$d_n = 1 - \sum_{k=1}^{n-1} c_k$$

We can state :

**Proposition .**

If there exist $\alpha_1$, $\alpha_2$, $\beta_1$, $\beta_2$ strictly positive numbers such that :

$$\alpha_1 \leq |d_n| \leq \beta_1, \quad \alpha_2 \leq |\sigma_n| \leq \beta_2$$

then $(H_1)$ and $(H_2)$ are verified.

Demonstration :

(1), (2), (3) imply :

$$a_{\theta, n, \delta} = \frac{|d_n|\delta}{\sqrt{2\pi\sigma_n^2}}$$

so :

$$a_{\theta, n, \delta} \leq \frac{\beta_1}{\alpha_2\sqrt{2\pi}}\,\delta$$

and $(H_1)$ is obviously verified. Let F be the distribution function of N(0,1). Putting, for $\theta_1 \neq \theta_2$, $\theta_1 < \theta_2$ :

$$B_{x^{(n-1)}} = ]-\infty, d_n \frac{\theta_1 + \theta_2}{2} + \sum_{k=1}^{n-1} c_k x_k]$$

We have :

$$P_{\theta_1, x^{(n-1)}}\left[B_{x^{(n-1)}}\right] = F\left[-\frac{d_n(\theta_2 - \theta_1)}{2\sigma_n}\right] \geq F\left[\frac{\alpha_1(\theta_2 - \theta_1)}{2\beta_2}\right]$$

$$P_{\theta_2, x^{(n-1)}}\left[B_{x^{(n-1)}}\right] = F\left[-\frac{d_n(\theta_2 - \theta_1)}{2\sigma_n}\right] \leq F\left[-\frac{\alpha_1(\theta_2 - \theta_1)}{2\beta_2}\right]$$

So $(H_1)$ holds with $W_n = F[-t]$, $\alpha = F[t] - F[-t]$, $t = \frac{\alpha_1}{2\beta_2}(\theta_2 - \theta_1)$ which ends the proof.

We can apply the above property to an AR process : let $(U_n)$ be a gaussian white noise, $X_n = a_1 X_{n-1} + \ldots + a_k X_{n-k} + U_n$ and $Y_n = X_n + \theta$, $\theta$ being to be estimated. $d_n = 1 - (a_1 + \ldots + a_k)$ is independent from n : if $d_n \neq 0$ and $0 < \alpha_2 < \mathrm{Var}(U_n) < \beta_2$, $(H_1)$ and $(H_2)$ hold.

## REFERENCES

[1] BAR-SHALOM, Y. (1972). On the asymptotic properties of the maximum likelihood estimate obtained from dependent observations. J.R. Statist, Soc. B, 33, 72-77.

[2] BHAT, B.R. (1974). On the method of maximum likelihood for dependent observations. J.R. Statist. Soc. B, 36, 48-53.

[3] BIRGE, L. (1983). Approximation in metric spaces and the theory of estimation. Z. Wahrsch. Verw. Gebiete 65, n°2, 181-237.

[4] CROWDER, M. (1975). Maximum likelihood estimation for dependent observations. J.R.Statist. Soc. B, 38, 45-53.

[5] GEFFROY, J. (1977). Sufficient convergence conditions for some tests in the case of not necessarily independent or equidistributed data. Recent developments in Statistics. J.R. Barra and al. editors. North-Holland Publishing Company.

[6] HILLION, A. (1980). Propriétés des distances en variations et d'Hellinger entre lois de probabilités sur des espaces produits. Thesis, University Paris VI.

[7] MOCHE, R. (1986). La décantation en estimation statistique. To appear in Pub.Inst.Stat.Univ.Paris.

[8] PIECZYNSKI, W; (1987). Sur la méthode du maximum de vraisemblance dans le cas des observations dépendantes. C.R.A.S. Paris, T. 304, Série I, n°13.

[9] SILVEY, S.D (1961). A note on Maximum-Likelihood in the case of dependent random variables. J.R. Statist. Soc. B, 23, 444-452.

[10] WALD, A. (1948). Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20, 595-601.

Reçu en Juillet 1988