Using local tests to estimate convergence rates for identification

Albert Benveniste
IRISA, Campus de Beaulieu,
35042 Rennes cedex, France,
benveniste@irisa.fr

Bernard Delyon
IRMAR, Campus de Beaulieu,
35042 Rennes cedex, France;
delyon@maths.univ-rennes1.fr

Abstract

Convergence rates and related Central Limit theorems have been the subject of numerous papers. In [1] a systematic link was first established between 1/ system identification, and 2/ model validation or testing for small changes. Similarities and relations are discussed in [2], part I, chap 5, and this so-called *local approach* has proved very successful in practical applications. In this paper, we first revisit and clarify this relationship, and propose in addition new simple proofs for stationary systems. Except for specific algorithms (e.g., regression models) estimating the convergence rate of an identification procedure is difficult. Even if there are general Central Limit Theorems available, building the corresponding estimators in practice is not easy. We propose a practical alternative, based on the relation between identification and local testing, and we propose a bootstrap-like estimator for the convergence rate.

Keywords : identification, local asymptotic normality, testing, model validation.

1 Introduction

In this paper we examine relationships between the consistency of M-estimators, the local asymptotic normality of statistics and the asymptotic normality of estimators. And we use this relationship for proposing an alternative approach to estimate the convergence rate of an identification procedure for general M-estimators.

1.1 M-estimators

M-estimators have been introduced in the 60s by Huber (cf [6]). We consider an observation process $(Y_1, Y_2, ...)$ and try to estimate a parameter θ of its distribution as a solution $\widehat{\theta}_n$ of $\sum_{i=1}^n H(\theta, Y_i) = 0$:

$$\widehat{\theta}_n \quad = \quad \arg_\theta \quad : \quad \sum_{i=1}^n H(\theta, Y_i) = 0 \ . \tag{1}$$

Of course this is based on our knowledge that the true parameter $\theta_\star$ satisfies

$$E[H(\theta_\star, Y_i)] = 0. \tag{2}$$

The obtained estimator is called an M-estimator [7]. This estimator corresponds sometimes to a *minimum contrast* or *quasi-likelihood* estimate

$$\widehat{\theta}_n = \arg\min_\theta \sum_{i=1}^n G(\theta, Y_i) \tag{3}$$

where G is the quasi-likelihood [5]. Obviously, if G is smooth enough, the solution of (3) is a particular solution of (1) by taking[1]: $H = \nabla G$. However, in some cases equations like (1) are solved and H is not the gradient of any function G, cf section 5. We prove in the first part of this paper the consistency of the procedure (1) under reasonably weak assumptions on H, and in the context of dependent sequences.

1.2 Local asymptotic normality

The concept of *local asymptotic normality* was introduced by Lucien LeCam in the 50s for comparing parameterized distributions P_θ over infinite sequences $(Y_1, Y_2, ...)$. The idea is that generically, under suitable conditions, although the P_θs for different θs are singular with respect to each other, the marginal density Λ_n of $P_{\theta+\widetilde{\theta}/\sqrt{n}}(Y_1, ..Y_n)$ with respect to $P_\theta(Y_1, ..Y_n)$ exists; furthermore it is a random variable $\Lambda_n(Y_1, ..Y_n)$ which converges in distribution, under P_θ, when n tends to infinity to some limit (in this context $\widetilde{\theta}$ is fixed and $\sqrt{n}$ happens to be the correct normalization for obtaining such limit theorems). This limit may be seen as a kind of differential which characterizes the way in which P_θ varies in the direction $\widetilde{\theta}$; LeCam's theory shows how the identification of this limit leads to various limit-theorems concerning general statistics based on samples drawn from P_θ, when θ varies in such a $O(n^{-1/2})$-rescaled neighborhood of a given value. We will not go that far since we will only study the distribution under $P_{\theta+\widetilde{\theta}/\sqrt{n}}$ of the following statistics, which is directly related to M-estimator (1):

$$U_n \quad = \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n H(\theta, Y_i) \tag{4}$$

We show that, for n large,

$$\text{under } P_\theta \quad : \quad U_n \sim \mathcal{N}(\, 0 \, , \Sigma(\theta)\,) \tag{5}$$

---

[1] For $f(x, y)$ a function, we shall denote by $\nabla_x f$ the partial derivative of f w.r.t. x, when it exists. Also, the derivative of f will sometimes be denoted by ∇f.

$$\text{under } P_{\theta+\widetilde{\theta}/\sqrt{n}} \quad : \quad U_n \sim \mathcal{N}\left(-E_\theta[\nabla_\theta H(\theta,Y)]\widetilde{\theta}\,,\, \Sigma(\theta)\right) \tag{6}$$

where the important point is that covariance matrix $\Sigma(\theta)$ is the same in both cases. This remark is the basis for the testing methods built upon general M-estimators, as introduced in [1, 2]. The advantage of this direct approach is to lead to similar conclusions for the behavior of statistics in a much simpler framework, allowing to deal with dependent sequences (LeCam's theory for dependent processes is quite complicated, and difficult to use in practice since it is generally very difficult to have a formula for the score, and even for the information matrix; however interesting results appear in [8]).

1.3 The relationship between identification and testing

Here follows a sketch of the main steps of our argument on the relationship between identification and testing in the asymptotic local framework. First, we need that the law of large numbers $1/n\sum_1^n H(\theta,Y_i) \to E[H(\theta,Y)]$ holds uniformly in θ, this is guaranteed under mild conditions, as stated in Section 2. This allows us to establish, in the same section, the consistency of the M-estimators for stationary dependent sequences. Section 3 focuses on testing under the local approach: the asymptotic behavior of U_n is given under $P_{\theta+\widetilde{\theta}/\sqrt{n}}$ distribution, using mild mixing conditions. The key step in enlightening the link between local tests and identification is given in section 4, it consists of the following elementary argument: first, make θ explicit in

$$U_n(\theta) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H(\theta,Y_i)\;.$$

Assuming that the (Y_i) are drawn from distribution $P_{\theta_\star}$ (i.e., $\theta_\star$ is the "true" underlying system parameter), then, for $\theta-\theta_\star$ small, one has

$$\begin{aligned} & U_n(\theta) \qquad\qquad (7)\\ \approx\;\; & U_n(\theta_\star)+\frac{1}{n}\left(\sum_1^n\;\nabla_\theta H(\theta_\star,Y_i)\right)\;(\theta-\theta_\star)\sqrt{n}\end{aligned}$$

Since $\widehat{\theta}_n$ is characterized by the equation $U_n(\widehat{\theta}_n)=0$, we get, using (7) and the law of large numbers,

$$\sqrt{n}(\widehat{\theta}_n-\theta_\star) \;\;\approx\;\; -E\left[\nabla_\theta H(\theta_\star,Y)\right]^{-1}\; U_n(\theta_\star) \tag{8}$$

and we deduce immediately the Central Limit Theorem for $\widehat{\theta}_n$ from the asymptotic behavior of the statistics U_n. This argument is detailed in section 4. The use of these theorems is illustrated in section 5 on the Instrumental Variable method for estimating the AR part of an ARMA process with time-varying MA parameters, a problem first investigated in [10] in a more difficult setting. Deeper use on so-called "subspace methods" for eigenstructure identification in vibration monitoring will be reported elsewhere.

1.4 A practical consequence: estimating convergence rates for identification

Besides its interest for its own, this relationship between $\widehat{\theta}_n$ and $U_n(\theta_\star)$ has a practical interest. The Central Limit Theorem for M-estimators states (see theorem 4):

$$\sqrt{n}\;(\widehat{\theta}_n-\theta_\star) \quad \xrightarrow{d} \quad \mathcal{N}(0,h'(\theta_\star)^{-T}\Sigma(\theta_\star)h'(\theta_\star)^{-1}) \tag{9}$$

where

$$\begin{aligned} h'(\theta) &= E[\nabla_\theta H(\theta,Y)]\,,\ \text{and} && (10)\\ \Sigma &= R(0)+\sum_{i>0} R(i)+R(i)^T\,,\ \text{where} &&\\ R(i) &= \mathrm{Cov}(H(\theta_\star,Y_1),H(\theta_\star,Y_{i+1}))\;, && (11)\end{aligned}$$

and h' denotes the derivative of h. While estimating or calculating the sensitivity matrix $E\left[\nabla_\theta H(\theta_\star,Y)\right]$ is most of the time easy, estimating or calculating matrix Σ often leads to illconditioned estimates. In contrast formula (8) suggests that it is enough to estimate the covariance matrix of $U_n(\theta_\star)$, which can be safely achieved from drawing samples of this statistics. In fact,

> a good estimate for Σ is obtained by taking an empirical estimate for the covariance matrix of $U_n(\theta')$, for θ' close enough to the true system $\theta_\star$. In particular, one can bootstrap this estimate by taking for θ' the actual estimate θ_n for $\theta_\star$, where n is large.

1.5 An example

The following example is analysed in details in section 5. Consider the following scalar ARMA(m,r) process:

$$\begin{aligned} y_n &= \underbrace{\sum_{i=1}^{m} a_i y_{n-i}}_{\theta^T\overline{y}_{n-1}}+\underbrace{\sum_{j=0}^{r-1} b_j\xi_{n-j}}_{e_n}\\ \theta &= (a_1,\ldots a_m)^T\;,\;\overline{y}_{n-1} \;=\; (y_{n-1},\ldots y_{n-m})^T\end{aligned}$$

where ξ_n is an i.i.d sequence.

For some nominal model θ_0, consider the following statistics, for some $p\geq m$:

$$\begin{aligned} U_n(\theta_0) &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n} H(\theta_0,Y_k) && (12)\\ H(\theta_0,Y_k) &= (y_k-\theta_0^T\overline{y}_{k-1})\check{y}_{k-q} &&\\ \check{y}_{k-q} &= (y_{k-q},\ldots y_{k-q-p+1})^T &&\\ Y_k &= (y_k,\ldots y_{k-q-p+1})^T\;, &&\end{aligned}$$

note the close relationship with the Instrumental Variable method, see [2], part I, p 183. Formulæ (5,6) describe the behaviour of $U_n(\theta_0)$, for n large.

On the other hand, the instrumental method for estimating θ consists in

$$\text{solving for } \theta \text{ the equation} \quad U_n(\theta) = 0\,,$$

with $p = m$. The classical formula for the convergence rate of this estimator is given by formulæ (9,10,11). An estimate of Σ based on formula (11) is unfortunately illconditioned, and typically not positive definite. Instead, based on the remark at the end of subsection 1.4, it is much preferred to draw an empirical estimate of the covariance of the statistics $U_n(\widehat{\theta})$, where $\widehat{\theta}$ is some large sample estimate of parameter θ, since guaranteing a good conditioning for such an estimate is straightforward.

2 Consistency of the quasi-likelihood procedure

We state here separately the Uniform Law of Large Numbers since this result is of major importance in the theory of M-estimators.

We focus on proposing simple and "optimal" assumptions, namely: ergodicity (A2) and regularity (A3). These assumptions may be compared to others proposed in the litterature, such as [4] where the ergodicity is replaced with a mixing condition and (A3) is replaced with the A-smoothness (Theorem 1 p.515).

In this paper we consider the case where Θ is bounded, if it is not the case, an additional lemma has to be used for proving first the boundedness of the sequence of estimates; this latter lemma, unlike the following one, involves quite technical assumptions (cf [7] p.131), and that is why we feel simpler to stick to the compact case, for which we give a rather direct proof.

We consider the following assumptions:

(A1) Θ is a compact subset of the Euclidean space $\mathbb{R}^n$ and $H(\theta, y)$ is a vector-valued measurable function defined on $\Theta \times \mathbb{R}^m$.

(A2) Process $(Y_i)_{1\leq i}$ is a strictly stationary ergodic sequence of random variables and

$$E[\ \sup_{\theta\in\Theta} |H(\theta, Y_1)|\ ] \ < \ \infty.$$

(A3) For each θ the function $\theta \longrightarrow H(\theta, Y)$ is continuous at θ for almost all value of Y (i.e. this full probability set may depend on θ) ^{2}.

Lemma 1 (Uniform Law of Large Numbers) *Set $h(\theta) = E[H(\theta, Y)]$. Then, under assumptions (A1,A2,A3),*

$$\lim_n \sup_{\theta\in\Theta} \left| h(\theta) - \frac{1}{n}\sum_{i=1}^{n} H(\theta, Y_i) \right| = 0 \tag{13}$$

holds.

Proof: See [3]. ∎

Theorem 1 *Let assumptions of (A1,A2,A3) hold. Then any sequence $(\widehat{\theta}_n)$ of random vectors, $\sigma(Y_1, ...Y_n)$-measurable, such that*

$$\lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} H(\widehat{\theta}_n, Y_i) = 0$$

satisfies

$$\lim_n h(\widehat{\theta}_n) = 0.$$

Note. Obviously, adding an identifiability assumption like (C2) below implies the consistency of $\widehat{\theta}_n$.

Proof: Theorem 1 is an immediate consequence of lemma 1. ∎

3 Local asymptotic normality and testing

In this section, we consider a family of probability measures P_θ on the space Ω of the sequences of m-dimensional vectors; these measures are indexed by $\theta \in \Theta$. We are interested in the asymptotic distribution of the statistics

$$U_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H(\theta_0, Y_{n,i})$$

when θ_0 is fixed (θ_0 represents the nominal value of the parameter) and $(Y_{n,i})_{1\leq i}$ is drawn from the distribution P_{θ_n} where

$$\theta_n = \theta_0 + \frac{\widetilde{\theta}}{\sqrt{n}}\ ,$$

and $\widetilde{\theta}$ is fixed. The normalization $1/\sqrt{n}$ is necessary in order to get a non-trivial limit distribution. The obtained asymptotic variance will allow us to build change tests for the parameter θ_0.

Let us state first some prerequisite result on central limit-theorem for arrays of variables.

Definition 1 (uniformly mixing array) *We shall say that a process $(Y_{n,i})_{1\leq i}$, $n \geq 1$, is a uniformly mixing array if for some $C, \varepsilon, \delta > 0$, and all $k, l, n \geq 1$*

$$\|Y_{n,k}\|_{2+\varepsilon} < C \tag{14}$$

$$\|E[Y_{n,k+l}|\mathcal{F}_{n,l}]\|_2 \leq C\ k^{-\delta-1/2} \tag{15}$$

---

[2] This assumption allows us to handle a $H(\theta, Y)$ involving quantizers, see [2], p. 33.

where $\mathcal{F}_{n,i} = \sigma(Y_{n,1}, \dots Y_{n,i})$*. We shall say that* $(Y_i)_{1\le i}$*, is a mixing sequence if* $(Z_{n,i}) = (Y_i)$ *is a uniformly mixing array.*

Theorem 2 *Let* $(Y_{n,i})_{1\le i\le n}$ *be a strictly stationary uniformly mixing array and*

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_{n,j}.$$

If the limit

$$V = \lim_n E[S_n S_n^T] \tag{16}$$

exists, then

$$S_n \xrightarrow{d} \mathcal{N}(0, V). \tag{17}$$

Proof: This theorem is established in [9] in its one dimensional version with a more general setting (the correspondence in notations is: $X_{n,i} = Y_{n,i}/\sqrt{n}$, $k_n(t) = [nt]$, $\sigma_{n,i} = 1/n$, $\psi_k = k^{-\delta-1/2}$). If we apply this one dimensional version to the inner product $\langle S_n, u\rangle$ for any vector u, we obtain the convergence in distribution of $\langle S_n, u\rangle$ to $\mathcal{N}(0, u^T V u)$, which implies, using characteristic functions, the convergence of S_n to $\mathcal{N}(0, V)$. ∎

Let us now introduce the assumptions we shall need for the local asymptotic normality of U_n.

(B1). Θ is an arbitrary subset of the Euclidean space $\mathbb{R}^n$ and $H(\theta, y)$ is a vector-valued measurable function defined on $\Theta \times \mathbb{R}^m$. Given any $\theta \in \Theta$, process (Y_i) drawn from P_θ is stationary; for any sequence θ_n converging to θ, and $(Y_{n,i})_{1\le i}$, $n \ge 1$ processes drawn from the distributions P_{θ_n}, the array of variables $(Z_{n,i})_{1\le i}$, $n \ge 1$ defined by:

$$Z_{n,i} = H(\theta_n, Y_{n,i}) - E_{\theta_n}[H(\theta_n, Y_{n,i})],$$

is uniformly mixing with $\delta > 1/2$ (cf equation (15)) .

(B2) The function

$$(\theta, \theta_\star) \longrightarrow h_{\theta_\star}(\theta) = E_{\theta_\star}[H(\theta, Y)]$$

is continuous on $\Theta \times \Theta$, its gradient $\nabla_\theta h$ w.r.t. θ is a continuous function of $(\theta, \theta_\star)$, and

$$h_\theta(\theta) = 0, \quad \forall \theta \in \Theta.$$

We shall denote by

$$h'(\theta) \quad = \quad \nabla_\theta h_{\theta_\star}(\theta) \quad \text{for } \theta_\star = \theta$$

the value of $\nabla_\theta h$ on the diagonal of $\Theta \times \Theta$.

(B3) The function $H(\theta, y)$ satisfies

$$\sup_{\theta_\star \in \Theta} E_{\theta_\star}[|H(\theta, Y)|^2] < \infty, \quad \forall \theta \in \Theta.$$

And for each θ the function $\theta \longrightarrow H(\theta, Y)$ is continuous at θ for almost all value of Y (i.e. this full probability set may depend on θ).

The condition $\delta > 1/2$ in assumption (B1) will allow us to check directly that condition (16) is satisfied for our processes. Since next theorem involves only values of θ close to some nominal value θ_0 these assumptions will actually need to be satisfied only for Θ being some neighborhood of θ_0.

Theorem 3 (Asymptotic normality of test) *Assume that* (B1,B2,B3) *are satisfied. Consider*

$$\theta_n = \theta_0 + \frac{\widetilde{\theta}}{\sqrt{n}}$$

for $\widetilde{\theta}$ *fixed (we assume that* $\theta_n \in \Theta$, $n > 0$*), and*

$$U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H(\theta_0, Y_{n,i})$$

where $(Y_{n,i})$ *is drawn from the distribution* P_{θ_n}*, and set*

$$R_\theta(i) = \mathrm{Cov}_\theta(H(\theta_0, Y_1), H(\theta_0, Y_{i+1}))\ ,$$

where Cov_θ *denotes the covariance under distribution* P_θ*. Then*

$$U_n \xrightarrow{d} \mathcal{N}(-h'(\theta_0)\widetilde{\theta}, \Sigma(\theta_0)),$$

where

$$\Sigma(\theta) = R_\theta(0) + \sum_{i>0} \left(R_\theta(i) + R_\theta(i)^T\right)\ .$$

Proof: See delyon97.

The reader is referred to [3] for further discussions on Assumptions B1–B3 and related examples.

4 Asymptotic normality of M-estimators

In this section we prove a asymptotic normality result on the estimates of $\theta_\star$ from observations (Y_i) generated under a fixed probability measure P. We are thus back in the semi-parametric setting. Consider the following set of assumptions:

(C1) Θ is an arbitrary subset of the Euclidean space $\mathbb{R}^m$ and $H(\theta, y)$ is a vector-valued measurable function defined on $\Theta \times \mathbb{R}^m$.

For any $\theta \in \Theta$, the sequence of variables

$$Z_i = H(\theta, Y_i) - E[H(\theta, Y_i)],$$

where $(Y_i)_{1\leq i}$ are drawn from the distribution P, form a stationary mixing sequence (in the sense of definition 1).

(C2) Set $h(\theta) = E[H(\theta, Y)]$; the unique solution $\theta_\star$ of $h(\theta) = 0$ is an interior point of Θ. Moreover, for any sequence θ_n, $\lim_n h(\theta_n) = 0$ implies $\lim_n \theta_n = \theta_\star$.

(C3) For any compact subset $\mathcal{K}$ of Θ, the function $H(\theta, y)$ satisfy

$$E[|H(\theta, Y)|^2] < \infty, \quad \forall\theta \in \Theta \ ,$$

and $h(\theta) = E[H(\theta, Y)]$ is continuously differentiable. Moreover, for each θ, the function $H(\theta, y)$ is, for almost all value of y, continuously differentiable w.r.t. θ and its derivative $\nabla_\theta H(\theta, y)$ w.r.t. θ satisfies

$$E[\sup_\theta |\nabla_\theta H(\theta, Y)|] < \infty \ , \tag{18}$$

$$\text{and} \quad h'(\theta) = E[\nabla_\theta H(\theta, Y)] \ , \tag{19}$$

where h' denotes the derivative of h.

Assumption (C2) is clearly an identifiability condition. Note that we have taken the dimension of θ to be the same as that of $H(\theta, Y)$ for reasons that will become clear in the theorem to follow.

Theorem 4 (asymptotic normality of M-estimators) *We assume that (C1,C2,C3) are satisfied. Consider a sequence $(Y_i)_{1\leq i}$ drawn under P; then any sequence $(\widehat{\theta}_n)$ which is a solution of*

$$\sum_{i=1}^n H(\widehat{\theta}_n, Y_i) = 0$$

which remains a.s. in a compact subset of Θ, converges almost surely to $\theta_\star$. Furthermore, h is continuously differentiable and if $h'(\theta_\star)$ is invertible, the following limits in distribution hold :

$$\begin{aligned} 0 \ = \ & U_n(\theta_\star) \ + \\ & \left(\frac{1}{n}\sum_{i=1}^n \nabla_\theta H(\theta_\star, Y_i) + \delta_n(\widehat{\theta}_n)\right)(\widehat{\theta}_n - \theta_\star)\sqrt{n}, \end{aligned} \tag{20}$$

where

$$|\delta_n(\theta)| \leq \zeta_n(\varepsilon) \quad , \quad \lim_{\varepsilon\to 0}\sup_n \zeta_n(\varepsilon) = 0, \ \textit{a.s.},$$

and

$$\sqrt{n}\ (\widehat{\theta}_n - \theta_\star) \quad \xrightarrow{d} \quad \mathcal{N}(0, h'(\theta_\star)^{-T}\Sigma h'(\theta_\star)^{-1})$$

where

$$\begin{aligned} \Sigma &= R(0) + \sum_{i>0} R(i) + R(i)^T \\ R(i) &= \operatorname{Cov}(H(\theta_\star, Y_1), H(\theta_\star, Y_{i+1})) \ . \end{aligned}$$

Proof : See delyon97.

5 Application–example

Consider the following scalar process

$$y_n = \sum_{i=1}^m a_i y_{n-i} + e_n \tag{21}$$

where the noise e_n is a q-dependent sequence[3] of zero-mean variables such that

$$E[e_n] = 0 \tag{22}$$

$$E[e_n^4] < \infty \tag{23}$$

and the polynomial $A(z) = 1 - \sum a_i z^i$ has its roots strictly outside the unit circle. For example, equation (21) may represent an ARMA(m,r) process with time-varying MA part :

$$y_n \ = \ \sum_{i=1}^m a_i y_{n-i} + \sum_{j=0}^{r-1} b_j(n)\xi_{n-j}$$

where ξ_n is an i.i.d sequence and $b(n) = (b_0(n), ...b_{r-1}(n))$ is some q-dependent random process with values in $\mathbb{R}^r$. Rewrite y_n as

$$\begin{aligned} y_n &= \theta^T \overline{y}_{n-1} + e_n \\ \theta &= (a_1, ...a_m)^T \\ \overline{y}_{n-1} &= (y_{n-1}, ...y_{n-m})^T, \end{aligned}$$

and make θ explicit whenever needed in

$$A_\theta(z) \ = \ 1 - \sum a_i z^i.$$

We shall assume that these variables are observed under their stationary distribution, i.e.,

$$\overline{y}_n = \sum_{k=0}^\infty \mathbf{A}_\theta^k (e_{n-k}, 0, \ldots, 0)^T \ , \tag{24}$$

where $\mathbf{A}_\theta$ is the $m \times m$ companion matrix built with θ on its first row.

5.1 Parameter testing

We would like to decide between the hypotheses :

$$\begin{aligned} H_0 &: \ \theta = \theta_0 \\ H_1 &: \ \theta = \theta_0 + \widetilde{\theta}/\sqrt{n}. \end{aligned}$$

For this purpose, we consider, for some $p \geq m$, the statistics

$$U_n \ = \ \frac{1}{\sqrt{n}} \sum_{k=1}^n H(\theta_0, Y_k) \tag{25}$$

$$\begin{aligned} H(\theta_0, Y_k) &= (y_k - \theta_0^T \overline{y}_{k-1})\check{y}_{k-q} \\ \check{y}_{k-q} &= (y_{k-q}, ...y_{k-q-p+1})^T \\ Y_k &= (y_k, ...y_{k-q-p+1})^T \ , \end{aligned}$$

---

[3] This means that for any $l \geq q$, e_{n+l} is independent of $(e_n, e_{n-1}, \ldots)$.

Note the close relationship with the Instrumental Variable method, see [2], part I, p 183. Note that under H_1, the sequence Y_k is drawn under the $O(n^{-1/2})$-rescaled distribution P_{θ_n}, $\theta_n = \theta_0 + \widetilde{\theta}/\sqrt{n}$.

In order to use theorem 3, we need an estimate of $h'(\theta_0)$ and $\Sigma(\theta_0)$. The following result proves the consistence of a simple estimate; we emphasize the dependency w.r.t. n by writing $Y_{n,k}$ instead of Y_k :

Lemma 2 *Under H_0 as well as under H_1, the estimators*

$$\begin{aligned} \widehat{K}_n &= \frac{1}{n}\sum_{k=1}^{n} \nabla_\theta H(\theta_0, Y_{n,k}) = -\frac{1}{n}\sum_{k=1}^{n} \check{y}_{k-q}\overline{y}_{k-1}^T, \\ \widehat{\Sigma}_n &= \frac{1}{n}\sum_{k=1}^{n}\sum_{j=-m}^{m} H(\theta_0, Y_{n,k})H(\theta_0, Y_{n,k+j})^T \end{aligned}$$

satisfy

$$\widehat{K}_n \xrightarrow{P} h'(\theta_0), \quad \widehat{\Sigma}_n \xrightarrow{P} \Sigma(\theta_0).$$

(We set $H(\theta_0, Y_{n,i}) = 0$ for $i < 1$ or $i > n$).

Proof: See [3]. The proof uses the uniform law of large numbers. ∎

Theorem 5 *Under the above assumptions (equations (21,22,23) and the stability of polynomial A_{θ_0}), the statistics U_n (equation (25)) satisfy*

$$\begin{aligned} \widehat{\Sigma}_n^{-1/2} U_n &\xrightarrow{d} \mathcal{N}(0, I) && \textit{under} \quad H_0 \\ \widehat{\Sigma}_n^{-1/2}(U_n - \widehat{K}_n\widetilde{\theta}) &\xrightarrow{d} \mathcal{N}(0, I) && \textit{under} \quad H_1. \end{aligned}$$

Proof: See [3]. The proof uses theorem 3 and lemma 2. ∎

5.2 Parameter estimation

The instrumental method for estimating θ consists in solving for θ the equation $U_n(\theta) = 0$, with $p = m$, where

$$U_n(\theta) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} H(\theta, Y_k).$$

Since $p = m$, $\check{y}_n$ coincides with $\overline{y}_n$ and the estimate $\widehat{\theta}_n$ satisfies:

$$\begin{aligned} \sum_{k=2m}^{n} H(\widehat{\theta}_n, Y_k) &= 0 \\ H(\theta, Y_k) &= (y_k - \theta^T\overline{y}_{k-1})\overline{y}_{k-q} \\ Y_k &= (y_k, \ldots y_{k-q-m+1})^T. \end{aligned}$$

We have the following convergence theorem

Theorem 6 *If the matrix*

$$h'(\theta_\star) = \begin{pmatrix} r_{q-1} & \cdots & \cdots & r_{q+m-2} \\ \vdots & \ddots & \ddots & \vdots \\ r_{q-m+1} & \ddots & \ddots & \vdots \\ r_{q-m} & r_{q-m+1} & \cdots & r_{q-1} \end{pmatrix},$$

where $r_i = \mathrm{Cov}(y_k, y_{k-i})$, is invertible, then

$$\sqrt{n}\ (\widehat{\theta}_n - \theta_\star) \xrightarrow{d} \mathcal{N}(0, h'(\theta_\star)^{-T}\Sigma h'(\theta_\star)^{-1})$$

where

$$\begin{aligned} \Sigma &= R(0) + \sum_{i>0} R(i) + R(i)^T \\ R(i) &= \mathrm{Cov}(H(\theta_0, Y_1), H(\theta_0, Y_{i+1})). \end{aligned}$$

Proof: We shall use theorem 4; note that (C1) and (C3) are satisfied (cf section 5.1). We have only to check (C2). Since $h(\theta)$ is here a linear function of θ, (C2) follows from the invertibility of $h'(\theta_\star)$. ∎

References

[1] A. Benveniste, M. Basseville, G. Moustakides, The asymptotic local approach to change detection and model validation, *IEEE Trans. on Automatic Control,* AC-32, 583-592, 1987.

[2] A. Benveniste, M. Métivier, P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer–Verlag, 1990.

[3] B. Delyon, A. Juditsky, A. Benveniste. On the relationship between identification and local tests. Irisa Res. Rep. 1104, May 1997. ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz

[4] A.R.Gallant, *Nonlinear Statistical Model,* Wiley, 1987.

[5] C. C. Heyde, Quasi-likelihood and Optimality for Estimating Functions: Some Current Unifying Themes, International Statistical Institute, Sept. 1989, Paris.

[6] P. J. Huber, The behavior of maximum likelihood estimates under nonstandard conditions, *Proc. Fifth Berk. Symp. Math. Statist. and Prob.* , 221-233, Univ. California Press, Berkeley, 1967.

[7] P. J. Huber, *Robust statistics*, Wiley, 1981.

[8] A. F. Kushnir, A. I. Pinskii Asymptotically Optimal Tests for Testing Hypotheses for Independent Samples of Observations, *Theory of Probability and its Applications, vol XVI, No 2, 1971.*

[9] D. L. McLeish, On the Invariance Principle for Non-stationary Mixingales, *Ann. Prob.* 1977, vol 5, No 4, 616-621.

[10] G. Moustakides, A. Benveniste, Detecting changes in the AR part of a nonstationary ARMA process, *Stochastics,* 16, 137-155, 1986.