A filtering-based algorithm for change detection in dynamic models with unknown parameter after change

Nadine Hilgert[1], Ghislain Verdier[2], and Jean-Pierre Vila[1]

[1] Analyse des Systèmes et Biométrie, UMR 729, INRA/SUPAGRO,
2 place Viala,
34060, Montpellier, France
nadine.hilgert@supagro.inra.fr
jean-pierre.vila@supagro.inra.fr

[2] Ecole Nationale Supérieure des Mines de Saint-Etienne,
Centre de Microélectronique de Provence,
880 avenue de Mimet,
13541, Gardanne, France
verdier@emse.fr

Abstract. A statistical method for change detection in autoregressive models is proposed. The model change is supposed to be due to a parameter change. As it is often the case in practice, the value characterizing the change is unknown. The proposed approach, called filtering detection rule, is derived from the well-known CUSUM rule and is based on a new class of particle filters to estimate the unknown conditional likelihood in the change mode. The asymptotic optimality of the procedure is obtained in the sense of Lorden (1971): the rule asymptotically minimizes the worst mean delay for detection under a constraint on the mean time between two false alarms. Implementation is easy and computation time is rather short, making the filtering detection rule an interesting alternative to the GLR rule, to which it is compared in simulations.
Keywords. Filtering-detection rule, GLR rule, Model change detection, Particle filtering

1 Introduction

In recent years, model change detection has become a crucial issue for various industrial fields: quality control in agro-food industries, navigation system in aeronautics, fault detection in biotechnological processes... Ensuring the safety of the installation or the quality of the production process requires thus developing methods that allow early detection of abnormal situations.

Statistical methods are among the most widely used approaches for change detection. They were introduced by Shewart (1931) and later, Page (1954) proposed the now well-known CUSUM rule (see Basseville and Nikiforov (1993) for an overview). The CUSUM rule can detect a change in a dynamic model describing the time evolution of a stochastic process, by using a conditional-likelihood ratio between the two functioning modes (nominal and fault modes) as test statistics. This model change is supposed to be due to a parameter change from a nominal value θ_0 (the process is "in control", hypothesis H_0) to a value θ_1 (the process is "out of control", hypothesis H_1). This rule can be applied as soon as the model and the two values θ_0 and θ_1 are known. But in practice, the value θ_1 characterizing the fault mode is often unknown but lies in a known area Θ_1. The CUSUM rule can then no longer be used. To address this problem, Lorden (1973) proposed the GLR detection rule which relies on the maximization of the conditional-likelihood ratio over Θ_1 as well as over all possible change times. The GLR rule is efficient in practice and optimality results were obtained, in particular by Lai (1998). However, the non-recursive writing and the maximization process of the likelihood ratio can cause too heavy computing time in many applications.

The aim of this presentation is to propose an alternative to the GLR rule for the detection of a parameter change in any autoregressive model when the parameter value θ_1 characterizing the out-of-control mode is unknown. The new detection rule is an adaptation of the Filtering Detection Rule for general nonlinear state-space models proposed by Verdier et al. (2008). A kernel convolution filter (Rossi and Vila (2006)) is applied to a simple equivalent state-space model in order to estimate the value θ_1. Optimality results are showed.

The paper is organized as follows. In section 2, the problem of model change detection is presented and the GLR detection rule is reminded. In section 3, the filtering detection rule is detailed and finally, in section 4, simulation trials are performed in order to evaluate the performance of the new rule compared to the GLR.

2 The change detection problem and the GLR rule

The objective of any model change detection rule is to detect a sudden change in the distribution of process variables $(X_i)_{i\geq 1}$, which could be represented as a parameter change in the dynamic model of the process, and therefore, in the conditional density of the observations.

The models considered in this presentation are of the following form:

$$X_n = f(X_{n-1}, \theta) + \epsilon_n, \tag{1}$$

where $X_n \in \mathbb{R}^d$ is the observation vector, ϵ_n a white noise and f a general function from $\mathbb{R}^{d+1}$ to $\mathbb{R}^d$. The probability distribution function of the white noise, $\mathcal{L}_\epsilon$, and the model function f are supposed to be known. θ is the parameter characterizing the change, moving at an unknown but not random change time t_0, from a known value θ_0 to an unknown value θ_1 lying in a compact set Θ_1. We note $P^{(t_0)}$ the probability measure of $(X_i)_{i\geq 1}$ corresponding to a change at time t_0 and P_{θ_0} denotes the case $t_0 = \infty$ (no change). The well-known CUSUM rule cannot be applied in this case. The GLR detection rule was proposed as an alternative by Lorden (1973). The GLR test statistic is constructed from the conditional-likelihood ratio between the two functioning modes and is defined as:

$$\forall\, n \geq 1, \quad g_n = \max_{1\leq j\leq n} \sup_{\theta\in\Theta_1} \sum_{i=j}^{n} \log \frac{p_\theta(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})}\ , \tag{2}$$

with $p_\theta(X_i|X_{1:i-1})$ the conditional density of X_i given the past $X_1, ..., X_{i-1}$. Due to the likelihood ratio, g_n tends to be negative in average under H_0 and positive and increasing under H_1. The stopping time is then:

$$t_a = \inf\{n : g_n \geq h\}\ ,$$

with h a given threshold satisfying a constraint, for example, on the false alarm rate. Lai (1998) obtained optimality results for this detection rule: it minimizes the worst mean delay for detection (see section 3) over all rules with a given time between two false alarms.

The implementation of this rule requires the maximization of the likelihood ratio over the set Θ_1. This maximization can lead to too large computing time even if a window-limited GLR rule is used (see Lai (1998)) to reduce the complexity of the procedure.

3 The new filtering-detection rule

3.1 Description of the approach

The aim of this section is to present a new detection rule based on the ratio of two quantities: the observation conditional likelihood under hypothesis H_0, $p_{\theta_0}(X_n|X_{1:n-1})$, which is known from model (1) since θ_0 is known, and a quantity which tends, under H_1, to the unknown conditional likelihood $p_{\theta_1}(X_n|X_{1:n-1})$. To address this problem, we used tools from the filtering theory. In recent years, particle filtering methods have been used to estimate the series of conditional likelihoods by sequential Monte Carlo simulations in state-space models (Li and Kadirkamanathan (2004), Azimi-Sadjadi and Krishnaprasad (2002)). A state-space modeling is generally used when the variables X_n of a dynamic system are not observed directly but through observation variables Y_n that are linked to X_n by way of an observation equation of the form $Y_n = h(X_n) + \nu_n$. For this type of model, and when θ_1 is supposed to be known, Verdier et al (2008) proposed the Filtering Detection Rule (FDR, based on a kernel convolution filter) and proved the asymptotic optimality of this procedure. In the present study, an adaptation of the FDR rule is applied to model (1). Here the state variables X_n are directly measured, thus yielding the simple observation equation $Y_n = X_n$. Furthermore, the parameter θ (unknown under H_1) is considered as a constant state variable in order to be estimated through a particle filter. Under H_1, the model takes then the following form:

$$\begin{cases} \begin{bmatrix} X_n \\ \theta_n \end{bmatrix} = \begin{bmatrix} f(X_{n-1}, \theta_{n-1}) + \epsilon_n \\ \theta_{n-1} \end{bmatrix} \\ Y_n = X_n \end{cases} \tag{3}$$

The conditional likelihood for model (3) is decomposed as follows:

$$p_{H_1}(X_n|X_{1:n-1}) = \int p_{H_1}(X_n|\theta_{n-1}, X_{1:n-1}).p_{H_1}(\theta_{n-1}|X_{1:n-1})\, d\theta_{n-1},$$

and is estimated by Monte Carlo simulation as:

$$\hat{l}_n^{N,m} = \frac{1}{m}\sum_{j=1}^{m} p_{H_1}(X_n|\theta_{n-1}^N(j), X_{1:n-1}), \tag{4}$$

with a sample $(\theta_{n-1}^N(j))_{j=1,\ldots,m}$ generated from the density $\bar{p}_{H_1}^N(\theta_{n-1}|X_{1:n-1})$: the normalized truncation over Θ_1 of a convergent filter estimate $\hat{p}_{H_1}^N(\theta_{n-1}|X_{1:n-1})$ of the density $p_{H_1}(\theta_{n-1}|X_{1:n-1})$. This estimate $\hat{p}_{H_1}^N(\theta_{n-1}|X_{1:n-1})$ is obtained with a new particle filtering method: the kernel convolution filtering, introduced by Rossi and Vila (2006) who proved convergence results under conditions less restrictive than that of most particle filters met in literature. In formula (4), N stands for the number of particles used in the estimation.

Thereby, the stopping time of the proposed detection rule is defined as:

$$\hat{t} = \inf\left\{ n : \max_{1\le j\le n} \sum_{i=j}^{n} \log \frac{\hat{l}_i^{N,m}}{p_{\theta_0}(X_i|X_{1:i-1})} \ge h \right\},$$

where h is a given threshold.

3.2 The filtering-detection algorithm

Initialization : $n = 1$

- Generation of N simulated $\bar{\theta}_0(i), i = 1, ..., N$ according to a given a priori distribution over Θ_1.
- Generation of N simulated noise realizations $\epsilon_1(i),\ i = 1, ..., N$ according to the distributions $\mathcal{L}_\epsilon$.
- Updating of the N particles through model (3):

$$\tilde{\theta}_1(i) = \bar{\theta}_0(i) \quad \text{and} \quad \tilde{Y}_1(i) = \tilde{X}_1(i) = f(X_0, \bar{\theta}_0(i)) + \epsilon_1(i)$$

- The N particle couples $(\tilde{\theta}_1(i), \tilde{X}_1(i))$ together with the first observation X_1, are used to compute the kernel estimate $\hat{p}_{H_1}^N(\theta_1|X_1)$ of the conditional density $p_{H_1}(\theta_1|X_1)$ (similarly to (5), see below).
- Determination of the normalized truncation $\bar{p}_{H_1}^N(\theta_1|X_1)$ of $\hat{p}_{H_1}^N(\theta_1|X_1)$ over Θ_1.

Time n : $n > 1$

<u>*Step 1* :</u>

- For $i = 1, \ldots, N$ let us simulate:
$\bar{\theta}_{n-1}(i) \sim \bar{p}_{H_1}^N(\theta_{n-1}|X_{1:n-1}),\ \epsilon_n(i) \sim \mathcal{L}_\epsilon$
- Then

$$\tilde{\theta}_n(i) = \bar{\theta}_{n-1}(i) \quad \text{and} \quad \tilde{Y}_n(i) = \tilde{X}_n(i) = f(X_{n-1}, \bar{\theta}_{n-1}(i)) + \epsilon_n(i)$$

<u>*Step 2* :</u>

- For $j = 1, \ldots, m$ let us simulate:
$\theta_{n-1}^N(j) \sim \bar{p}_{H_1}^N(\theta_{n-1}|X_{1:n-1})$

- Then, from the observation X_n, $\hat{l}_n^{N,m}$ is computed according to (4).

<u>*Step 3* :</u>

- From $p_{\theta_0}(X_n|X_{1:n-1})$ and $\hat{l}_n^{N,m}$, we compute

$$\hat{g}_n = \max_{1\leq j\leq n} \sum_{i=j}^{n} \hat{Z}_i \quad \text{with} \quad \hat{Z}_i = \log \frac{\hat{l}_i^{N,m}}{p_{\theta_0}(X_n|X_{1:n-1})} \; .$$

An alarm is set up if the statistics $\hat{g}_n$ exceeds the chosen threshold h. Otherwise one goes to Step 4.

<u>*Step 4* :</u>

- From the N couples $(\tilde{\theta}_n(i), \tilde{X}_n(i)), i = 1, ..., N$ and the observation X_n, one gets, following Rossi and Vila (2006), a kernel estimate of the conditional state variable density $p_{H_1}(\theta_n|X_{1:n})$:

$$\hat{p}_{H_1}^N(\theta_n|X_{1:n}) = \frac{\sum_{i=1}^N K_{\delta_x}^x(\tilde{X}_n(i) - X_n).K_{\delta_\theta}^\theta(\tilde{\theta}_n(i) - \theta_n)}{\sum_{i=1}^N K_{\delta_x}^x(\tilde{X}_n(i) - X_n)} \; , \tag{5}$$

where K^x and K^θ are two convolution kernels with bandwidth parameters δ_x and δ_θ respectively and with the usual notation: $K_\delta(x) = K(x/\delta)/\delta^{dim(x)}$.
- Determination of the normalized truncation $\bar{p}_{H_1}^N(\theta_n|X_{1:n})$ of $\hat{p}_{H_1}^N(\theta_n|X_{1:n})$ over Θ_1.
- $n = n + 1$ and go back to step 1

Remark 1. The test statistic can be equivalently written in a recursive manner : $\hat{g}_n = (\hat{g}_{n-1} + \hat{Z}_n)^+$ with the notation $m^+ = m$ if $m > 0$ and 0 otherwise. This is a useful form in practice.

3.3 Properties of the rule

The aim of this section is to study the optimality properties of the Filtering Detection Rule applied to detect a parameter change in model (1). The optimality of a rule is usually established by showing that it minimizes a given criterion about the detection delay, among all the rules satisfying a constraint on the false alarm rate. One of the first optimality results for dependent variables was obtained by Lai (1998), who proved that the CUSUM rule asymptotically (as $\gamma \to \infty$, see below) minimizes the worst mean delay for detection (introduced by Lorden (1971)):

$$\bar{E}_{\theta_1}(T) = \sup_{t_0\geq 1} \sup \text{ess}\, E^{(t_0)}[(T - t_0 + 1)^+|X_{1:t_0-1}] \tag{6}$$

among all detection rules T with an average time between false alarm such that, for a given $\gamma > 0$

$$E_{\theta_0}(T) \geq \gamma \quad \text{(ARL constraint).} \tag{7}$$

(The essential supremum in (6) is taken with respect to $X_1, ..., X_{t_0-1}$). Lai established a lower bound for the worst mean detection delay. Let us note:

$$Z_i(\theta) = \log \frac{p_\theta(X_i|X_{1:i-1})}{p_{\theta_0}(X_i|X_{1:i-1})}.$$

Theorem 1 (Lai (1998)). *If there exists a constant I such that*

$$\lim_{n\to\infty} \sup_{t_0\geq 1} \textit{ess sup}\; P^{(t_0)} \left\{ \max_{t\leq n} \sum_{i=t_0}^{t_0+t} Z_i(\theta_1) \geq I(1+\delta)n|X_{1:t_0-1} \right\} = 0, \tag{8}$$

then, as $\gamma \to \infty$, it holds

$$\inf\{\bar{E}_{\theta_1}(T) : E_{\theta_0}(T) \geq \gamma\} \geq (I^{-1} + o(1)) \log \gamma. \tag{9}$$

Under some general assumptions, the Filtering Detection Rule reaches the lower bound (9) and therefore is asymptotically optimal. Let us show how.

Suppose the following assumptions hold:

(A1) There exist two constants c and C such that:

$$\forall\, x_i,\ \forall \theta \in \Theta_1, \quad c \leq p_\theta(x_i|X_{1:i-1}) \leq C.$$

(A2) The log-likelihood ratio satisfies: $\exists\ I > 0$ such that,

$$\frac{1}{n} \sum_{i=t_0}^{t_0+n-1} Z_i(\theta_1) \to I \quad \text{in probability under } P^{(t_0)} \text{ when } n \to \infty,$$

and

$$\forall \delta > 0, \exists \Theta_\delta \subset \Theta_1\ , \exists n(\delta) \geq 1 \text{ such that} : \theta_1 \in \Theta_\delta \text{ and}$$
$$\sup_{n \geq n(\delta)} \sup_{k \geq t_0 \geq 1} \operatorname{ess\,sup} P^{(t_0)} \left\{ \inf_{\theta \in \Theta_\delta} \sum_{i=k}^{k+n} Z_i(\theta) \leq (I-\delta)n | X_{1:k-1} \right\} \leq \delta.$$

Remark 2. The first assumption is for example satisfied by model (1) with a bounded white noise, which is quite realistic in practice.

The second assumption was used by Lai (1998) to obtain the optimality of the Mixture Likelihood Ratio rule.

The optimality of the Filtering Detection Rule is stated in the following theorem:

Theorem 2. *Suppose that (8) and assumptions (A1) and (A2) are satisfied. Under the assumptions required for the convergence of the kernel convolution filter (Rossi et Vila (2006)), it holds:*
i-

$$E_{\theta_0}(\hat{t}) \geq e^h.$$

ii- As h, N and m tend to infinity,

$$\bar{E}_{\theta_1}(\hat{t}) \leq (I^{-1} + o(1))h,$$

Therefore, with $h = \log \gamma$, $\hat{t}$ satisfies the Lorden's constraint (by i-) and reaches the lower bound (9) (by ii-).

Proof : The proof of this theorem (see Verdier (2007)) is an adaptation of the proof proposed by Lai for the optimality of the Mixture Likelihood Ratio rule.

4 Simulation trials

We considered the following autoregressive linear model:

$$X_n = 0.5X_{n-1} + \theta + \epsilon_n, \tag{10}$$

with ϵ_n a truncated Gaussian white noise with variance $\sigma^2 = 0.1$. The two functioning modes are defined by:

$$H_0 : \theta = \theta_0 = 0.71 \quad \text{and} \quad H_1 : \theta = \theta_1 \in \Theta_1 = [0.2; 0.65]$$

We compared the performances of the GLR and the Filtering Detection Rule on 2000 trajectories simulated according to model (10) with a change at $t_0 = 50$. For each trajectory the value θ_1 was sampled from a uniform law on Θ_1. The filtering-detection rule was applied with $N = 500$, $m = 500$ and using a Gaussian kernel with a bandwidth parameter of the form : $\delta^x_N = 0.2N^{-1/5}$. The thresholds of the two rules were determined by simulation such that the mean time before the first false alarm equals 100, and we obtained: $h_{GLR} = 3.2$ and $h_{FDR} = 1.3$. We consider as a non-detection a trajectory without alarm until time step $t = 150$.

The mean detection delay of the Filtering Detection Rule was 7.03, and we recorded one non-detection (the θ_1 value corresponding to the non-detected change is $\theta_1 = 0.6037$). For the GLR rule, there were four non-detections (corresponding to $\theta_1 = 0.6236$, 0.6085, 0.6173 and 0.6418) and the mean time before detection was 7.41.

This simulation with a simple model showed that these two rules have similar behavior with an advantage for the Filtering Detection Rule concerning the computation time. Other simulations, in which the FDR was compared to Mixture Likelihood Ratio approaches (Lai(1998)), have also shown the effectiveness of our method. Indeed, the performances of MLR rules heavily depend on the choice of an a priori law for the parameter θ_1. This is not the case with our rule for which the "distribution" of θ_1 is readjusted at each time step under H_1.

5 Conclusions

We proposed in this paper an alternative to classical rules such as GLR and MLR rules for detecting a parameter change in auto-regressive models when the value characterizing the out-of-control mode, θ_1, is unknown. Our approach is an adaptation of the Filtering Detection Rule (Verdier et al. (2008)) and is based on the application of a convolution kernel filter to an equivalent simple state-space model. We proved the asymptotic optimality of the rule in the sense of Lorden, under classical assumptions in change detection theory. The first simulations gave interesting results in terms of computation time and efficiency in detecting changes, in comparison with more classical approaches applied in the case of unknown parameter after a change, such as GLR and MLR rules.

References

Azimi-Sadjadi, B. and Krishnaprasad, P. (2002). Change detection for nonlinear systems; a particle filtering approach. *Proceedings of the American Control Conference*, 4074–4079.

Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Changes. Theory and Applications*, Prentice-Hall.

Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory*, **44**, 2917–2929.

Li, P. and Kadirkamanathan, V. (2004). Fault detection and isolation in nonlinear stochastic systems - a combined adaptive Monte Carlo filtering and likelihood approach. *Int. Journal of Control*, **77**, 1101–1114.

Lorden, G. (1971). Procedures for reacting to a change in distribution *Annals of Mathematical Statistics*, **42**, 1897–1908.

Lorden, G. (1973). Open-ended tests for Koopman-Darmois families. *Annals of Statistics*, **1**, 633–643.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100–115.

Rossi, V. and Vila J-P. (2006). Nonlinear filtering in discrete time: a particle convolution approach. *Pub. Inst. Statist. Univ. Paris*, **L**, 71–102.

Shewhart, W. A. (1931). *Economic Control of Quality Manufactured Product*, Mac-millan, New-York.

Verdier, G. (2007) Détection statistique de rupture de modèle dans les systèmes dynamiques - Application à la supervision de procédés de dépollution biologique. *PhD thesis, Université Montpellier 2, France.*

Verdier, G., Hilgert, N. and Vila J-P. (2008). Optimality of CUSUM rule approximations in change-point detection problems: application to nonlinear state–space systems. *IEEE Transactions on Information Theory*, **54**, 5102–5112.