

Fault Detection in Hidden Markov Models : A Local Asymptotic Approach*

François LeGland and Laurent Mevel
IRISA / INRIA
Campus de Beaulieu
35042 RENNES Cédex, France
e-mail : {legland,lmevel}@irisa.fr

Abstract

In this paper, the problem of detecting a change in the transition probability matrix of a hidden Markov chain is addressed, using the local asymptotic approach. The score function, evaluated at the nominal value, is used as the residual, and is expressed as an additive functional of the extended Markov chain consisting of the hidden state, the observation, the prediction filter and its gradient w.r.t. the parameter. The problem of residual evaluation is solved using available limit theorems on the extended Markov chain, which allow to replace the original detection problem by the simpler problem of detecting a change in the mean of a Gaussian r.v.

Keywords : fault detection, local test, residual generation, residual evaluation, hidden Markov model, prediction filter, likelihood function, score function.

1 Introduction

The purpose of this paper is to design statistical tests to decide whether $\theta = \alpha$, corresponding to a nominal behaviour of the system, or $\theta \neq \alpha$, on the basis of observations (Y_0, \dots, Y_n) , in the following parametric model.

- Under \mathbb{P}^θ , the hidden state sequence $\{X_n, n \geq 0\}$ is a Markov chain with values in the finite set $S = \{1, \dots, N\}$, with *primitive* transition probability matrix $Q_\theta = (q_\theta^{i,j})$, i.e. for any $i, j \in S$

$$q_\theta^{i,j} = \mathbb{P}^\theta[X_{n+1} = j \mid X_n = i] ,$$

and initial probability distribution $p_0 = (p_0^i)$ independent of θ , i.e. for any $i \in S$

$$p_0^i = \mathbb{P}^\theta[X_0 = i] .$$

*This work was partially supported by the Commission of the European Communities, under the TMR project *System Identification*, project number FMRX-CT98-0206, and under the IHP research training network *Statistical Methods for Dynamical Stochastic Models*, network number HPRN-CT2000-00100.

- The d -dimensional observations $\{Y_n, n \geq 0\}$ are mutually independent given the sequence of states of the Markov chain, i.e.

$$\begin{aligned} & \mathbb{P}^\theta[Y_n \in dy_n, \dots, Y_0 \in dy_0 \mid X_n = i_n, \dots, X_0 = i_0] \\ &= \prod_{k=0}^n \mathbb{P}^\theta[Y_k \in dy_k \mid X_k = i_k] . \end{aligned}$$

For any $n \geq 0$, and for any $i \in S$, the conditional probability distribution of the observation Y_n given that $(X_n = i)$, is absolutely continuous w.r.t. a positive and σ -finite measure λ on \mathbb{R}^d , i.e.

$$\mathbb{P}^\theta[Y_n \in dy \mid X_n = i] = b^i(y) \lambda(dy) ,$$

with a λ -a.e. positive density independent of θ . For any $y \in \mathbb{R}^d$, let

$$b(y) = (b^i(y)) \quad \text{and} \quad B(y) = \text{diag}(b^i(y)) .$$

Throughout the paper, the *true* value of the parameter is denoted by α , and the following assumptions hold.

Assumption A : For the *true* value α , the transition probability matrix $Q_\alpha = (q_\alpha^{i,j})$ is positive.

Assumption B : The mapping $\theta \mapsto Q_\theta$ is two-times differentiable, with bounded first and second derivatives, and Lipschitz continuous second derivative.

2 Prediction filters, log-likelihood function, and their derivatives

For any $n \geq 1$, and any value θ of the parameter, let $p_n^\theta = (p_n^i)$ denote the *prediction filter*, i.e. the conditional probability distribution under \mathbb{P}^θ of the state X_n given observations (Y_0, \dots, Y_{n-1}) : for any $i \in S$

$$p_n^i = \mathbb{P}^\theta[X_n = i \mid Y_0, \dots, Y_{n-1}] .$$

The random sequence $\{p_n^\theta, n \geq 0\}$ takes values in the set $\mathcal{P} = \mathcal{P}(S)$ of probability distributions over the finite set S , and satisfies the forward Baum equation

$$p_{n+1}^\theta = \frac{Q_\theta^* B(Y_n) p_n^\theta}{b^*(Y_n) p_n^\theta} \triangleq Q_\theta^* f[Y_n, p_n^\theta], \quad (1)$$

for all $n \geq 0$, where for any $y \in \mathbb{R}^d$, and any $p \in \mathcal{P}$

$$f[y, p] = \frac{B(y) p}{b^*(y) p}.$$

Here and throughout the paper, the notation $*$ denotes the transpose of a matrix. Differentiating (1) w.r.t. the i -th component of the p -dimensional parameter θ yields

$$\partial_i p_{n+1}^\theta = Q_\theta^* F[Y_n, p_n^\theta] \partial_i p_n^\theta + \partial_i Q_\theta^* f[Y_n, p_n^\theta], \quad (2)$$

where for any $y \in \mathbb{R}^d$, and any $p \in \mathcal{P}$

$$F[y, p] = \left[I - \frac{B(y) p e^*}{b^*(y) p} \right] \frac{B(y)}{b^*(y) p},$$

is the Jacobian matrix at point $p \in \mathcal{P}$ of the mapping $p \mapsto f[y, p]$. The sequence $\partial p_n^\theta = (\partial_i p_n^\theta)$ belongs to Σ^p , where

$$\Sigma = \{w \in \mathbb{R}^N : e^* w = 0\},$$

is the linear tangent space to \mathcal{P} , and $e = (1, \dots, 1)^*$. Differentiating further (2) w.r.t. the j -th component of the p -dimensional parameter θ yields

$$\begin{aligned} \partial_{i,j}^2 p_{n+1}^\theta &= Q_\theta^* F_\theta[Y_n, p_n^\theta] \partial_{i,j}^2 p_n^\theta \\ &\quad - Q_\theta^* F_\theta[Y_n, p_n^\theta] \left[\partial_j p_n^\theta \frac{b^*(Y_n) \partial_i p_n^\theta}{b^*(Y_n) p_n^\theta} \right. \\ &\quad \left. + \partial_i p_n^\theta \frac{b^*(Y_n) \partial_j p_n^\theta}{b^*(Y_n) p_n^\theta} \right] \\ &\quad + \partial_i Q_\theta^* F_\theta[Y_n, p_n^\theta] \partial_j p_n^\theta \\ &\quad + \partial_j Q_\theta^* F_\theta[Y_n, p_n^\theta] \partial_i p_n^\theta \\ &\quad + \partial_{i,j}^2 Q_\theta^* f_\theta[Y_n, p_n^\theta]. \end{aligned} \quad (3)$$

Following Arapostathis and Marcus [1], define $Z_n^\theta = (X_n, Y_n, p_n^\theta, \partial p_n^\theta)$ for any $n \geq 0$, and any value θ of the parameter. Using techniques of Benveniste, Métivier and Priouret [4], the following result has been proved in LeGland and Mevel [7, 8].

Geometric ergodicity : for any values θ and θ' of the parameter, the extended Markov chain $\{Z_n^\theta, n \geq 0\}$ with values in $S \times \mathbb{R}^d \times \mathcal{P} \times \Sigma^p$, is geometrically ergodic under $\mathbb{P}^{\theta'}$. For simplicity, the marginals of the invariant measure $\mu_{\theta'}^\theta$, are also denoted by $\mu_{\theta'}^\theta$.

The log-likelihood function (suitably normalized) for the estimation of the parameter θ based on observations (Y_0, \dots, Y_n) can be expressed as an additive functional of the extended Markov chain $\{Z_n^\theta, n \geq 0\}$ as follows

$$\ell_n(\theta) = \frac{1}{n} \sum_{k=0}^n \log[b^*(Y_k) p_k^\theta].$$

Similarly, the score function, i.e. the derivative of the log-likelihood function w.r.t. the parameter θ , can also be expressed as an additive functional of the extended Markov chain $\{Z_n^\theta, n \geq 0\}$ as follows

$$\partial_i \ell_n(\theta) = \frac{1}{n} \sum_{k=0}^n \frac{b^*(Y_k) \partial_i p_k^\theta}{b^*(Y_k) p_k^\theta}.$$

Finally, the Hessian of the log-likelihood function is defined by

$$\begin{aligned} \partial_{i,j}^2 \ell_n(\theta) &= \frac{1}{n} \sum_{k=0}^n \left[\frac{b^*(Y_k) \partial_{i,j}^2 p_k^\theta}{b^*(Y_k) p_k^\theta} \right. \\ &\quad \left. - \frac{b^*(Y_k) \partial_i p_k^\theta}{b^*(Y_k) p_k^\theta} \frac{b^*(Y_k) \partial_j p_k^\theta}{b^*(Y_k) p_k^\theta} \right]. \end{aligned}$$

The following results have been proved in Mevel [9], see also LeGland and Mevel [6].

- (i) Kullback–Leibler information : for any values θ and θ' of the parameter

$$\ell_n(\theta) - \ell_n(\theta') \longrightarrow -K_{\theta'}(\theta), \quad \mathbb{P}^{\theta'}\text{-a.s.}$$

where the following expression holds

$$\begin{aligned} K_{\theta'}(\theta) &= \int \log[b^*(y) p] \mu_{\theta'}^\theta(dy, dp) \\ &\quad - \int \log[b^*(y) p] \mu_{\theta'}^{\theta'}(dy, dp) \geq 0. \end{aligned}$$

- (ii) Fisher information matrix :

$$\partial^2 \ell_n(\alpha) \longrightarrow -I(\alpha), \quad \mathbb{P}^\alpha\text{-a.s.}$$

where the matrix $I(\alpha) = (I^{i,j}(\alpha))$ is defined by

$$\begin{aligned} I^{i,j}(\alpha) &= \int \frac{b^*(y) w^i}{b^*(y) p} \frac{b^*(y) w^j}{b^*(y) p} \\ &\quad \mu_\alpha^\alpha(dy, dp, dw^i, dw^j). \end{aligned}$$

- (iii) Asymptotic normality of the score :

$$\sqrt{n} \partial \ell_n(\alpha) \Longrightarrow \chi, \quad \text{under } \mathbb{P}^\alpha,$$

where χ is a p -dimensional Gaussian r.v. with zero mean and covariance matrix $I(\alpha)$.

- (iv) Local Lipschitz continuity of the Hessian :

$$\mathbb{E}^\alpha \left[\sup_{\theta : |\theta - \alpha| < r} \|\partial^2 \ell_n(\theta) - \partial^2 \ell_n(\alpha)\| \right] \leq C r.$$

3 Nonlocal test

To decide whether $\theta = \alpha$ or $\theta \neq \alpha$, the first (nonlocal) approach is to design a test to decide, on the basis of observations (Y_0, \dots, Y_n) , between the two hypotheses

$$H_0 : \theta = \alpha,$$

$$H_1 : \theta \in \Theta_{\text{fault}}, \text{ where } \alpha \notin \Theta_{\text{fault}}.$$

This can be achieved by the following generalized likelihood ratio test, see van Trees [10, Section 2.5]

$$\sup_{\theta \in \Theta_{\text{fault}}} \ell_n(\theta) - \ell_n(\alpha) \geq c,$$

i.e. the following condition

$$\sup_{\theta \in \Theta_{\text{fault}}} \ell_n(\theta) - \ell_n(\alpha) > c,$$

is used to reject the null hypothesis, and c is a threshold to be selected. The asymptotic behaviour of the test statistics under the null and alternative hypotheses can be obtained, hence the problem of threshold selection can be solved, but this method requires to compute the log-likelihood function for all possible values of the parameter in Θ_{fault} . Indeed, a uniform version of the law of large numbers (ii) yields

$$\sup_{\theta \in \Theta_{\text{fault}}} \ell_n(\theta) - \ell_n(\alpha) \longrightarrow - \inf_{\theta \in \Theta_{\text{fault}}} K_\alpha(\theta),$$

in \mathbb{P}^α -probability, and

$$\sup_{\theta \in \Theta_{\text{fault}}} \ell_n(\theta) - \ell_n(\alpha) \longrightarrow - \inf_{\theta \in \Theta_{\text{fault}}} K_{\theta'}(\theta) + K_{\theta'}(\alpha),$$

in $\mathbb{P}^{\theta'}$ -probability, hence the probability of false alarm

$$F = \mathbb{P}^\alpha \left[\sup_{\theta \in \Theta_{\text{fault}}} \ell_n(\theta) - \ell_n(\alpha) > c \right] \longrightarrow 0,$$

provided the threshold c satisfies

$$\inf_{\theta \in \Theta_{\text{fault}}} K_\alpha(\theta) > -c,$$

and the probability of nodetection

$$N = \sup_{\theta' \in \Theta_{\text{fault}}} \mathbb{P}^{\theta'} \left[\sup_{\theta \in \Theta_{\text{fault}}} \ell_n(\theta) - \ell_n(\alpha) < c \right] \longrightarrow 0,$$

provided the threshold c satisfies

$$\sup_{\theta' \in \Theta_{\text{fault}}} \left[\inf_{\theta \in \Theta_{\text{fault}}} K_{\theta'}(\theta) - K_{\theta'}(\alpha) \right] < -c.$$

As a result, the probability of false alarm and the probability of nodetection can go simultaneously to zero, provided the following *detectability* condition

$$\sup_{\theta' \in \Theta_{\text{fault}}} \left[\inf_{\theta \in \Theta_{\text{fault}}} K_{\theta'}(\theta) - K_{\theta'}(\alpha) \right] < \inf_{\theta \in \Theta_{\text{fault}}} K_\alpha(\theta),$$

holds.

4 Local test

Following Benveniste, Basseville and Moustakides [3], see also Zhang, Basseville and Benveniste [11] and Basseville [2], the second (local) approach is to design a test to decide, on the basis of observations (Y_0, \dots, Y_n) , between the two hypotheses

$$H_0 : \theta = \alpha,$$

$$H_1 : \theta = \alpha + \Delta/\sqrt{n}, \text{ for some } \Delta \neq 0.$$

This can be achieved by the following procedure.

- The first step, called *residual generation*, is to propose a statistics, called the *residual*, which depends only the observations and on the nominal value α , and which ideally should be close to zero under the null hypothesis, and significantly different from zero under the alternative hypothesis.
- The second step, called *residual evaluation*, is to study the asymptotic behaviour of the residual under the null hypothesis and under a contiguous alternative hypothesis, and to design a simple test based on the residual.

Introducing the score function evaluated at the nominal value α ,

$$\zeta_n = \sqrt{n} \partial \ell_n(\alpha),$$

as the residual, the central limit theorem (iii) yields immediately that

$$\zeta_n \Longrightarrow \chi, \quad \text{under } \mathbb{P}^\alpha,$$

where χ is a p -dimensional Gaussian r.v. with zero mean and covariance matrix $I(\alpha)$. To study the asymptotic behaviour of the residual ζ_n under the contiguous probability measure $\mathbb{P}^{\alpha+\Delta/\sqrt{n}}$, we introduce the log-likelihood ratio

$$\begin{aligned} \lambda_n(\alpha, \Delta) &= \log \frac{d\mathbb{P}_n^{\alpha+\Delta/\sqrt{n}}}{d\mathbb{P}_n^\alpha} \\ &= n [\ell_n(\alpha + \Delta/\sqrt{n}) - \ell_n(\alpha)], \end{aligned}$$

where for any value θ of the parameter, \mathbb{P}_n^θ denotes the marginal on the set of observations (Y_0, \dots, Y_n) of the probability measure \mathbb{P}^θ . The key result is the following local asymptotic normality (LAN) property, which has been obtained by Bickel and Ritov [5] in the stationary case, and using different techniques.

Theorem 4.1 *The family $\{\mathbb{P}_n^{\alpha+\Delta/\sqrt{n}}, n \geq 0\}$ of probability measures is locally asymptotically normal, and moreover*

$$(\zeta_n, \lambda_n(\alpha, \Delta)) \Longrightarrow (\chi, \Delta^* \chi - \frac{1}{2} \Delta^* I(\alpha) \Delta),$$

under \mathbb{P}^α , where χ is a p -dimensional Gaussian r.v. with zero mean and covariance matrix $I(\alpha)$.

Proof: Using a Taylor expansion yields

$$\begin{aligned} \lambda_n(\alpha, \Delta) &= \sqrt{n} \Delta^* \partial \ell_n(\alpha) + \frac{1}{2} \Delta^* \partial^2 \ell_n(\alpha) \Delta \\ &+ \Delta^* \int_0^1 [\partial^2 \ell_n(\alpha + u \Delta / \sqrt{n}) \\ &- \partial^2 \ell_n(\alpha)] (1 - u) du \Delta . \end{aligned}$$

Combining the law of large numbers (ii), the central limit theorem (iii), and the Lipschitz property (iv) yields

$$(\zeta_n, \lambda_n(\alpha, \Delta)) \implies (\chi, \Delta^* \chi - \frac{1}{2} \Delta^* I(\alpha) \Delta) ,$$

under \mathbb{P}^α , where χ is a p -dimensional Gaussian r.v. with zero mean and covariance matrix $I(\alpha)$. \square

It is now straightforward to obtain the asymptotic behaviour of the residual ζ_n under the contiguous alternative $\mathbb{P}^{\alpha+\Delta/\sqrt{n}}$.

Proposition 4.2

$$\zeta_n \implies I(\alpha) \Delta + \chi , \quad \text{under } \mathbb{P}^{\alpha+\Delta/\sqrt{n}} ,$$

where χ is a p -dimensional Gaussian r.v. with zero mean and covariance matrix $I(\alpha)$.

Proof: Let $F_{\mu,R}$ denote the Gaussian probability distribution on \mathbb{R}^p , with mean μ and covariance matrix R , and notice that $F_{R,\mu,R}$ is absolutely continuous w.r.t. $F_{0,R}$, with density

$$\frac{dF_{R,\mu,R}}{dF_{0,R}}(x) = \exp\{\mu^* x - \frac{1}{2} \mu^* R \mu\} .$$

Therefore, the LAN property proved in Theorem 4.1 yields

$$\begin{aligned} &\mathbb{E}^{\alpha+\Delta/\sqrt{n}}[\phi(\zeta_n)] \\ &= \mathbb{E}^\alpha[\phi(\zeta_n) \exp\{\lambda_n(\alpha, \Delta)\}] \\ &\longrightarrow \int \phi(x) \exp\{\Delta^* x - \frac{1}{2} \Delta^* I(\alpha) \Delta\} F_{0,I(\alpha)}(dx) \\ &= \int \phi(x) F_{I(\alpha) \Delta, I(\alpha)}(dx) , \end{aligned}$$

for any test function ϕ defined on \mathbb{R}^p . \square

The original problem of designing a test to detect a change in the transition probability matrix of a HMM is now replaced by the simpler problem of designing a test to detect a change in the mean of a Gaussian r.v.

$$H_0 : \zeta_n \sim \mathcal{N}(0, I(\alpha)) ,$$

$$H_1 : \zeta_n \sim \mathcal{N}(I(\alpha) \Delta, I(\alpha)) , \text{ for some } \Delta \neq 0 .$$

This can be achieved by the following generalized likelihood ratio test

$$\sup_{\Delta \neq 0} [\Delta^* \zeta_n - \frac{1}{2} \Delta^* I(\alpha) \Delta] \geq c ,$$

i.e. the following condition

$$\sup_{\Delta \neq 0} [\Delta^* \zeta_n - \frac{1}{2} \Delta^* I(\alpha) \Delta] > c ,$$

is used to reject the null hypothesis, and c is a threshold to be selected. If the Fisher information matrix $I(\alpha)$ is invertible, this reduces to

$$\frac{1}{2} \zeta_n^* I^{-1}(\alpha) \zeta_n > c .$$

An additional advantage of the local test is that both the residual and the Fisher information matrix are evaluated, or estimated, for the nominal value α only. In particular, the exact value $I(\alpha)$ of the Fisher information matrix can be replaced by the approximation $I_n(\alpha) = (I_n^{i,j}(\alpha))$ defined by

$$I_n^{i,j}(\alpha) = \frac{1}{n} \sum_{k=0}^n \frac{b^*(Y_k) \partial_i p_k^\alpha}{b^*(Y_k) p_k^\alpha} \frac{b^*(Y_k) \partial_j p_k^\alpha}{b^*(Y_k) p_k^\alpha} ,$$

for observations (Y_0, \dots, Y_n) collected under the nominal model. Indeed, it is straightforward to show that

$$I_n(\alpha) \longrightarrow I(\alpha) , \quad \mathbb{P}^\alpha\text{-a.s.}$$

5 Conclusion

Two statistical tests have been presented for detecting a change in the transition probability matrix of a hidden Markov chain.

- The *nonlocal* test is very difficult to use in practice, because there is no analytical expression for the supremum of the log-likelihood function over the set Θ_{fault} of possible values taken by the parameter in the alternative hypothesis.
- In opposition, the *local* test is much simpler to use, since the computations are done for a single value of the parameter, i.e. the nominal value α . In addition, many observations are usually available under the nominal model, which makes it possible to estimate the Fisher information matrix $I(\alpha)$.

References

- [1] A. Arapostathis and S. I. Marcus. Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. *Mathematics of Control, Signals, and Systems*, 3(1):1–29, 1990.
- [2] M. Basseville. On-board component fault detection and isolation using the statistical local approach. *Automatica*, 34(11):1391–1416, Nov. 1998.
- [3] A. Benveniste, M. Basseville, and G. V. Moustakides. The asymptotic local approach to change detection and model validation. *IEEE Transactions on Automatic Control*, AC-32(7):583–592, July 1987.
- [4] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations*, volume 22 of *Applications of Mathematics*. Springer-Verlag, New York, 1990.
- [5] P. J. Bickel and Y. Ritov. Inference in hidden Markov models. I : Local asymptotic normality in the stationary case. *Bernoulli*, 2(3):199–228, 1996.
- [6] F. Le Gland and L. Mevel. Asymptotic behaviour of the MLE in hidden Markov models. In *Proceedings of the 4th European Control Conference, Bruxelles 1997*, July 1997. Paper FRA-F6.
- [7] F. Le Gland and L. Mevel. Exponential forgetting and geometric ergodicity in HMM's. In *Proceedings of the 36th Conference on Decision and Control, San Diego 1997*, pages 537–542. IEEE-CSS, Dec. 1997.
- [8] F. Le Gland and L. Mevel. Exponential forgetting and geometric ergodicity in hidden Markov models. *Mathematics of Control, Signals, and Systems*, 13(1):63–93, 2000.
- [9] L. Mevel. *Statistique Asymptotique pour les Modèles de Markov Cachés*. Thèse de Doctorat, Université de Rennes 1, Rennes, Nov. 1997. <ftp://ftp.irisa.fr/techreports/theses/1997/mevel.ps.gz>.
- [10] H. L. van Trees. *Detection, Estimation and Modulation Theory. Part I : Detection, Estimation and Linear Modulation Theory*. John Wiley & Sons, New York, 1968.
- [11] Q. Zhang, M. Basseville, and A. Benveniste. Early warning of slight changes in systems. *Automatica*, 30(1 (Special issue on Statistical Methods in Signal Processing and Control)):95–113, Jan. 1994.