Change Detection for Hidden Markov Models

László Gerencsér and Cecilia Prosdocimi

Abstract— Hidden Markov Models (HMM-s) are widely used in a number of application areas. In this paper we consider the problem of detecting changes in the statistical pattern of a hidden Markov process. We adapt the so-called Hinkley-detector that was first introduced for independent observations. Assuming that the dynamics before and after the change is known, we are lead to the problem of analyzing the Hinkley-detector with an L-mixing input. It is shown that, under suitable technical conditions, the output process is also L-mixing. The result yields a rigorous upper bound for the false-alarm frequency. The limitations and potentials of the result will also be discussed.

I. INTRODUCTION

Hidden Markov Models (HMM-s) are widely used to model stochastic systems in many different application areas such as speech recognition, genetics, communication networks, volatility modelling. The statistical analysis of HMM-s is a basic problem in which significant progress has been achieved in recent years. In this paper we focus on the problem of change detection for HMM-s. Detection of changes in the statistical pattern of a hidden Markov process is of interest in a number of applications. In the case of speech processing we may wish to identify the moments of switching from one speaker to another one. In the case of distributed, aggregated sensing we may wish to identify the events of sensor failure.

A basic method for detecting temporal changes in an independent sequence is the Cumulative Sum (CUSUM) algorithm or Hinkley-detector, which will be presented in Section II. It was first used for independent observations, but its range of applicability can be extended. We propose an adaptation of the Hinkley-detector to HHM-s. The key problem here is to generate an appropriate residual process. The adaptation of the Hinkley detector to the case when the dynamics of the HMM before and after the change is unknown was considered in [16], using a number of heuristic arguments.

In the present paper we consider the mathematically cleaner case when the dynamics before and after the change is *known*. The residuals will then be defined via

$$\xi_n = -\log P(Y_n \mid Y_{n-1}, \ldots, Y_0; \theta_i), \quad i = 1, 2 \quad (1)$$

where Y_k denotes the observation sequence. It has been shown in [17] and [18] that, under reasonable technical

L. Gerencsér is with MTA SZTAKI (Computer and Automation Institute Hungarian Academy of Sciences, Budapest, Hungary `gerencser@sztaki.hu`
C. Prosdocimi is with DIMAD, University of Florence, Italy `cecilia.prosdocimi@gmail.com`

conditions for the transition probability matrix and the read-out probability, ξ_n is an L-mixing process. (For the definition of L-mixing see the Appendix). Thus we are lead to the study of the Hinkley-detector with an L-mixing input sequence. The main result of the paper is that, under suitable technical conditions, the output of the Hinkley-detector is also L-mixing. A direct corollary of the result is a rigorous conceptual upper bound for the almost sure false alarm frequency.

Stability results for the Hinkley-statistics for the i.i.d case have been established among others in [9] and [24]. The result of this paper a nice addition to these earlier results.

II. CHANGE DETECTION

Detection of changes of statistical patterns is a fundamental problem in many applications, for a survey see [5] and [7]. From the technical point of view change detection is no easier than identification of system dynamics. On the other hand it is well-known that detecting changes in the dynamics is possible even with grossly under-specified models. Thus change detection is often robust against modelling errors.

A basic method for detecting temporal changes is the Cumulative Sum (CUSUM) test or *Hinkley*-detector, introduced by Page [25] and analyzed later, among others, by Hinkley [20] and Lorden [23]. The Hinkley-detector is defined via a sequence of random variables R_n, often called residuals in the engineering literature, such as likelihood ratios, such that

$$\mathbb{E}(R_i) < 0 \quad \text{for } i \leq \tau^* - 1, \quad \text{and} \quad \mathbb{E}(R_i) > 0 \quad \text{for } i \geq \tau^*,$$

with τ^* denoting the change point. To give an example, in the case of i.i.d. samples with densities $f(x, \theta_0)$ and $f(x, \theta_1)$ before and after the change point, we would set

$$R_i = -\log f(x_i, \theta_0) + \log f(x_i, \theta_1)$$

where x_i is the i-th sample. It is easy to see, by using the Kullback-Leibler inequality, that R_i does satisfy the inequalities above. Now letting, with $S_0 = 0$,

$$S_n := \sum_{i=1}^{n} R_i,$$

the Hinkley-detector is defined for $n \geq 0$ as

$$g_n := S_n - \min_{0 \leq k \leq n} S_k = \max_{0 \leq k \leq n} (S_n - S_k). \quad (2)$$

An alarm is generated if g_n exceeds a pre-fixed threshold $\delta > 0$. Thus the moment of alarm is defined by

$$\hat{\tau} = \inf\{n \mid S_n - \min_{0 \leq k \leq n} S_k > \delta\}. \quad (3)$$

The Hinkley-detector was first used for independent observations, but its range of applicability has been extended for

dependent sequences, such as stationary ergodic processes in [4]. The applicability of the Hinkley detector to ARMA systems, with unknown dynamics, has been demonstrated in [3]. Much later, it was adapted to Hidden Markov Models, with unknown dynamics, in [16].

Assuming that actually there is no change at all, i.e. $\tau^* = +\infty$, the Hinkley detector can still be used to monitor the process, and we may occasionally get an alarm. The quality of the Hinkley-detector is partially characterized by the almost sure frequency of these *false alarms*, as a function of the threshold δ. Formally, we have to consider

$$\limsup_N \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{g_n > \delta\}}. \tag{4}$$

A potentially useful result for estimating the above quantity will be presented.

III. HIDDEN MARKOV MODELS: DEFINITION AND A CHANGE DETECTION ALGORITHM

We consider HHM-s with a finite state space χ, say $| \chi |= N$ and a general observation or read-out space $\mathcal{Y}$, which is assumed to be a Polish space (i.e. a complete and separable metric space).

Definition 1: The pair (X_n, Y_n) is a Hidden Markov process if (X_n) is a homogenous Markov chain with state space χ and the observation sequence (Y_n) is conditionally independent and identically distributed given the σ-field generated by the process (X_n).

Let Q^* be the transition probability matrix of the unobserved Markov chain (X_n), i.e. let

$$Q^*_{ij} = P(X_{n+1} = j \mid X_n = i).$$

We write $Q^* > 0$ if all the entries of Q^* are positive. If $\mathcal{Y}$ is finite, say $| \mathcal{Y} |= M$, we use the following notation

$$P(Y_n = y \mid X_n = x) = b^x(y). \tag{5}$$

For continuous read-outs the read-out densities are defined as

$$P(Y_n \in dy \mid X_n = x) = b^x(y)\lambda(dy), \tag{6}$$

for a suitable non-negative σ-finite measure λ. A key quantity in the statistical analysis of HMM-s is the predictive filter defined as

$$p^j_{n+1} = P(X_{n+1} = j \mid Y_N, \dots, Y_0). \tag{7}$$

One of the first contributions in the statistical analysis of hidden Markov processsesis [6], in which the maximum-likelihood (ML) estimation of the parameters of a finite state space and finite read-out HMM is studied. Strong consistency of the maximum-likelihood estimator for finite state space and binary read-outs has been established in [1]. The extension of these results to continuous read-outs requires new insights.

The first step in proving consistency of the maximum-likelihood method would be to show the validity of the strong law of large numbers for the log-likelihood function. This can be achieved showing the validity of the strong law of large numbers for a function of an extended Markov chain (X_n, Y_n, p_n). This has been investigated in the literature basically with three different methods: using the subadditive ergodic theorem in [22], using geometric ergodicity arguments in [21], and using L-mixing processes in [17], [18]. For more recent contributions see [8] and [10].

Let θ^* be the true parameter driving the dynamics of a HHM (X_n, Y_n) and let

$$\theta^* = \begin{cases} \theta_1 & \text{for } n \leq \tau^* - 1 \\ \theta_2 & \text{for } n \geq \tau^*, \end{cases} \tag{8}$$

for an unknown τ^*, but for given $\theta_1 \neq \theta_2$. Our goal is to estimate τ^*. To state our change-detection algorithm, first note that the negative of the log-likelihood function can be interpreted as a code-length, modulo a constant. Thus we first set for any feasible θ

$$C_n(Y_n; \theta) := -\log p(Y_n \mid Y_{n-1}, \dots, Y_0; \theta), \tag{9}$$

and then define the *residual-process*

$$R_n := C_n(Y_n; \theta_1) - C_n(Y_n; \theta_2). \tag{10}$$

We certainly get, by the Kullback-Leibler inequality,

$$\mathbb{E}_{\theta^*}(R_n) < 0 \quad \text{for} \quad n \leq \tau^* - 1,$$

and also, in the case of $\tau^* = 0$,

$$\mathbb{E}_{\theta^*}(R_n) > 0 \quad \text{for} \quad n \geq \tau^*.$$

The Hinkley-detector defined in terms of this residual process (R_n) yields the desired change detection algorithm.

To study the probabilistic properties of the resulting Hinkley detector, we first note, that under suitable conditions, the process $C_n(Y_n; \theta)$ is L-mixing (see Theorem 5.2 in [17]). More precisely: define

$$\delta(y) = \frac{\max_x b^x(y)}{\min_x b^x(y)}. \tag{11}$$

Theorem 1: Consider a hidden Markov process (X_n, Y_n). Assume that the transition probability matrices Q^* and Q, corresponding to θ^* and θ, respectively, are primitive, and that for all $x \in \mathcal{X}$ we have $b^x(y) > 0$ for λ almost all $y \in \mathcal{Y}$. Furthermore assume that for all $s \geq 1$ and for all $i, j \in \mathcal{X}$

$$\int | \log b^j(y)|^s \ b^{*i}(y)\lambda(dy) < \infty, \tag{12}$$

and also that for all $s \geq 1$ and for all $i \in \mathcal{X}$

$$\int |\delta(y)|^s b^{*i}(y)\lambda(dy) < \infty. \tag{13}$$

Then the process $C_n(Y_n; \theta)$ is L-mixing.

We conclude that under the assumption of no change, i.e. $\tau^* = \infty$, the residual (R_n) defined in equation (10) is L-mixing. The conditions (12) and (13) are certainly satisfied for a finite read-out space $\mathcal{Y}$, assuming that $b^x(y) > 0$ for all x, y. If, in addition, Q^* and Q are positive, then (R_n) is a *bounded sequence*. In the next section we prove a remarkable

input-output property of the Hinkley-detector, when the input is L-mixing.

IV. STABILITY PROPERTIES OF THE *Hinkley detector*

The Hinkley detector (g_n) in equation (2) can be equivalently defined via a *non-linear dynamical system*, with $a_+ = \max\{0, a\}$, as follows:

$$g_n = (g_{n-1} + R_n)_+ \quad \text{with} \quad g_0 = 0. \qquad (14)$$

From a system-theoretic point of view this system is not stable in any sense. E.g., for a constant, positive input (g_n) becomes unbounded, and the effect of initial perturbations may not vanish. On the other hand, for an i.i.d. input sequence (R_n), with $\mathrm{E}(R_n) < 0$ some stability of the output process (g_n) can be expected. The resulting non-linear stochastic system is a standard object in queuing theory and in the theory of risk processes (see [26]). In this case the process (g_n) is clearly a homogenous Markov chain. A number of stability properties of (g_n) have been established in [2], [24], and [9].

In particular it can be shown that for random i.i.d. inputs with negative expectation, and finite exponential moments of some positive order the (g_n) statistics is L-mixing (see [19]). Theorem 2 is a generalization of this result for L-mixing input, when no change occurs. The result allows to give an upper bound for the false alarm frequency.

To state our result we need two technical assumptions on the input sequence (R_n). The first one is fairly mild, requiring that (R_n) is an L-mixing process with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$ such that

$$\sum_{\tau=0}^{+\infty} \tau \gamma_q(\tau, R) < +\infty \qquad \text{for all} \quad 1 \le q < +\infty. \qquad (15)$$

The second assumption is much more restrictive, saying that

$$M_\infty(R) < +\infty \qquad \text{and} \qquad (16)$$

This condition will be discussed in the remark following Lemma 1. We define a critical exponent in terms of $M_\infty(R)$ and $\Gamma_\infty(R)$ as follows:

$$\beta^* := \varepsilon/(4M_\infty(R)\Gamma_\infty(R)). \qquad (17)$$

Then, for any $\beta' \le \beta^*$ define

$$\mu = \mu(\beta') := \exp\left(4M_\infty(R)\Gamma_\infty(R)(\beta')^2 - \beta'\varepsilon\right). \qquad (18)$$

Note that for the critical value β^* we have $\mu(\beta^*) = 1$, and for $\beta' < \beta^*$ we have $\mu(\beta') < 1$.

Theorem 2: Let (R_n) be an L-mixing process w.r.t. $(\mathcal{F}_n, \mathcal{F}_n^+)$ such that (15) and (16) are satisfied, and

$$\mathbb{E}(R_n) \le -\varepsilon < 0 \qquad \text{for all} \quad n \ge 0. \qquad (19)$$

Let (g_n) be defined as in (2). Then (g_n) is L-mixing w.r.t. $(\mathcal{F}_n, \mathcal{F}_n^+)$. In addition, for any β, β' such that $0 < \beta < \beta' < \beta^*$, we have with $\mu = \mu(\beta')$

$$\mathbb{E}(\exp \beta g_n) \le 1 + \left(\frac{\beta}{\beta' - \beta}\right)\frac{\mu}{1 - \mu} =: K_{\beta,\beta'}. \qquad (20)$$

For the proof we use the following equivalent formulation:

$$g_n = \max_{1 \le i \le n}(R_i + \cdots + R_n)_+. \qquad (21)$$

Then, define the auxiliary process

$$g_{n,n-\tau}(R) := \max_{1 \le i \le n-\tau}(R_i + \cdots + R_n)_+. \qquad (22)$$

Lemma 1: Let (R_n) and β, β' and μ be as in Theorem 2. Then

$$\mathbb{E}(\exp \beta\, g_{n,n-\tau}(R)) \le 1 + \left(\frac{\beta}{\beta' - \beta}\right)\frac{\mu^{\tau+1}}{1 - \mu}. \qquad (23)$$

For the proof of the lemma we need an exponential inequality for partial sums of L-mixing processes. We do have such an inequality, see [13], for the case $M_\infty(R) < +\infty$ and $\Gamma_\infty(R) < +\infty$. Unfortunately, it is not clear if this inequality can be extended to unbounded processes. It seems that the boundedness of (R_n) is a common assumption for exponential inequalities for partial sums of mixing processes, see Section 1.4.2. in [11].

To show that (g_n) is L-mixing we a key step in the proof is to approximate g_n by

$$g_{n,n-\tau}^{++} := \max_{n-\lceil \frac{\tau}{2}\rceil+1 \le i \le n}(R_{i,n-\tau}^+ + \cdots + R_{n,n-\tau}^+)_+, \quad (24)$$

where

$$R_{i,n-\tau}^+ := \mathbb{E}(R_i | \mathcal{F}_{n-\tau}^+).$$

Note that $g_{n,n-\tau}^{++}$ is $\mathcal{F}_{n-\tau}^+$ measurable, as required. Then we estimate the residual $g_n - g_{n,n-\tau}^{++}$ using an intermediate approximation of g_n, defined as

$$\bar{g}_{n,n-\tau} := \max_{n-\lceil \frac{\tau}{2}\rceil+1 \le i \le n}(R_i + \cdots + R_n)_+, \qquad (25)$$

and the lemma above.

A. False alarm frequency

As a corollary to Theorem 2 we can get an upper bound for the a.s. false alarm rate, when $\tau^* = +\infty$, defined as

$$\limsup_N \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_{\{g_n > \delta\}}. \qquad (26)$$

Theorem 3: Let (R_n) and β^* be as in Theorem 2, and let (g_n) be defined as in (2). Then for any $\delta > 0$, and any $0 < \beta < \beta' < \beta^*$ we have

$$\limsup_N \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_{\{g_n > \delta\}} \le K_{\beta,\beta'} \exp(-\beta\delta) \qquad (27)$$

where $K_{\beta,\beta'}$ is defined in Theorem 2.

Proof: The proof is obtained by applying the strong law of large numbers for centered L-mixing processes to the sequence $(f(g_n))$, with f being a Lipschitz-continuous upper bound of the characteristic function $I_{\{g>\delta\}}$, and finding an upper bound for the false alarm probability $\mathbb{P}\{g_n > \delta\}$ using our main theorem and an exponential Markov-inequality. ∎

V. Conclusions and Comments

We have presented a Hinkley-detector for HHM-s, with know dynamics before and after the change. This has lead us to the study of the Hinkley-detector with L-mixing input. We have shown that, under certain technical conditions, the output of the Hinkley detector is L-mixing.

It should be admitted though, that this result is not directly applicable to the change detction of HMM-s. Namely, the Hinkley-scores defined for HMM-s are typically not bounded. A notable exception is the case of finite state and read-out space with all transition and read-out probabilities being positive. But even in this case the condition $\Gamma_\infty(R) < +\infty$ can not be guaranteed. The technical difficulty in extending this result is the apparent lack of an appropriate exponential inequality for the partial sums of unbounded L-mixing processes, given as Theorem 5.1 in [13]. Boundedness of (R_n) is also a common assumption for exponential inequalities for partial sums of other kinds of mixing processes, see Section 1.4.2. in [11]. A notable exception is the case when $|R_n| \leq |R_n^*|$, where R_n^* is the stationary response of a finite dimensional linear stochastic system driven by Gaussian white noise.

Appendix

We summarize a few definitions given in [12]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition 2: We say that a stochastic process (X_n) is M-*bounded* if for all $1 \leq q < +\infty$

$$M_q(X) := \sup_{n \geq 0} \| X_n \|_q < +\infty.$$

We can also define $M_q(X)$ for $q = +\infty$ as

$$M_\infty(X) := \sup_{n \geq 1} \operatorname{ess\,sup} |X_n|.$$

Let $(\mathcal{F}_n)_{n \geq 1}$ be an increasing family of σ-fields and let $(\mathcal{F}_n^+)_{n \geq 1}$ be a decreasing family of σ-fields, $\mathcal{F}_n \subseteq \mathcal{F}$ and $\mathcal{F}_n^+ \subseteq \mathcal{F}$ for any n. Assume that $\mathcal{F}_n$ and $\mathcal{F}_n^+$ are independent for all n. Let $\tau > 0$ be an integer, and let for $1 \leq q < +\infty$

$$\gamma_q(\tau, X) = \gamma_q(\tau) := \sup_{n \geq \tau} \| X_n - \mathbb{E}(X_n | \mathcal{F}_{n-\tau}^+) \|_q,$$

$$\Gamma_q(X) := \sum_{\tau=0}^{+\infty} \gamma_q(\tau).$$

We can also define

$$\gamma_\infty(\tau, X) := \sup_{n \geq \tau} \operatorname{ess\,sup} |X_n - \mathbb{E}(X_n | \mathcal{F}_{n-\tau}^+)|,$$

$$\Gamma_\infty(X) := \sum_{\tau=0}^{+\infty} \gamma_\infty(\tau, X).$$

Definition 3: A process (X_n) is L-*mixing* with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$ if X_n is $\mathcal{F}_n$-measurable for all $n \geq 1$, (X_n) is M-bounded, and $\Gamma_q(X) < +\infty$ for all $1 \leq q < +\infty$.

A prime example of L-mixing process is the output process of a stable linear stochastic system driven by a M-bounded i.i.d. sequence. A useful observation: let (X_n) be an L-mixing process with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$. Let f be a real Lipschitz-continuous bounded function. Then $(f(X_n))_{n \geq 1}$ is L-mixing with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$. Centered L-mixing processes satisfy the strong law of large numbers, see Corollary 1.3 in [12].

References

[1] Arapostathis, A. and Marcus, I. (1990) *Analysis of an Identification Algorithm Arising in the Adaptive Estimation of Markov Chains* in *Math. Control Signals and Systems* Vol.3, pp. 1-29

[2] Asmussen, S. (2003) *Applied Probability and Queues*, Springer-Verlag, New York

[3] Baikovicius, J. and Gerencsér, L. (1992) *Change-point detection as Model selection*, in *Informatica*, Vol.3, No. 1, pp. 3-20

[4] Bansal, R. and Papantoni-Kazakos, P. (1986) *An Algorithm for Detecting a Change in a Stochastic Process* in *IEEE Transactions on information Theory*, Vol. 32, No. 2, 227-235

[5] Basseville, M. and Nikiforov, I.V. (1993) *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall

[6] Baum, L.E. and Petrie, T. (1966) *Statistical inference for probabilistic functions of finte state Markov chains* in *The Annals of Mathematical Statistics*, Vol. 37, pp. 1554-1563

[7] Brodsky, B.E. and Darkhovsky, B.S., (1993) *Nonparametric Methods in Change-Point Problems*, Kluwer Academic Publishers,

[8] Chigansky, P. (2009) *Maximum likelihood estimator for Hidden Markov Models in continuous time* in *Statistical Inference for Stochastic Processes*, Vol. 12, No. 2, pp. 139-163

[9] Diaconis, P. and Freedman, D. (1999) *Iterated Random Functions*, in *SIAM Review*, Vol. 41, No. 1, pp. 45-76

[10] Douc,R., Moulines, E., Olsson, J. and van Handel, R. (2010) *Consistency of the maximum likelihood estimator for general hidden Markov models*, preprint

[11] Doukhan, P. (1994) *Mixing: Properties and Examples*, Springer-Verlag, New York

[12] Gerencsér, L. (1989) *On a Class of Mixing Processes* in Stochastics, Vol. 26, pp. 165-191

[13] Gerencsér, L. (1991) *Almost Sure Exponential Stability of Random Linear Differential Equations* in Stochastics and Stochastics Reports, Vol.36, pp. 91-107

[14] Gerencsér, L. (1994) *Fixed gain off-line estimators of ARMA parameters* in Journal of Mathematical Systems, Estimation and Control, Vol. 4, No. 2, pp. 1-26

[15] Gerencsér, L. and Molnár-Sáska, G., (2003) *Estimation and Strong Approximation of Hidden Markov Models* in Lecture Notes in Control and Information Sciences, Springer, Vol. 294, pp.313-320

[16] Gerencsér, L. and Molnár-Sáska, G., (2004) *Change Detection of Hidden Markov Models*, in Proceedings of the 43th IEEE Conference on Decision and Control, pp. 1754-1758

[17] Gerencsér, L., Michaletzky, G. and Molnár-Sáska, G. (2007) *An improved bound for the exponential stability of predictive filters of Hidden Markov Models* in Communications in Information and Systems Vol. 7, No. 2, pp. 133-152

[18] Gerencsér, L. and Molnár-Sáska, G. (2007) *Identification of Hidden Markov Models - Uniform LLN-s*, in Modeling, Estimation and Control, Festschrift in Honor of Giorgio Picci on the Occasion of his Sixty-Fifth Birthday, Lecture Notes in Control and Information Sciences, Springer, No. 364, pp. 135-149

[19] Gerencsér, L. and Prosdocimi, C. (2010) *Input-output properties of the Hinkley detector*, submitted

[20] Hinkley, D.V. (1971) *Inference about the Change-Point from Cumulative Sum Tests* in Biometrika, **58**, 3, 509-523

[21] Le Gland, F. and Mevel, L. (2000) *Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models* in Math. Control Signals and Systems, Vol. 13, pp. 63-93

[22] Leroux, B.G. (1992) *Maximum Likelihood estimation for Hidden Markov Models* in Stochastic Processes and their Applications, Vol. 40, pp. 127-143

[23] Lorden, G. (1971) *Procedures for Reacting to a Change in Distribution* in The Annals Mathematical Statistics, **42**, 6, 1897-1908

[24] Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability*, Springer Verlag, London

[25] Page, E.S. (1954) *Continuous Inspection Schemes* in Biometrika, **41**, 1/2, 100-115

[26] Panjer, H.H. and Willmot, G.E. (1992) *Insurance Risk Models*, Society of Actuaries