New methods for the statistical analysis of Hidden Markov Models

László Gerencsér, Gábor Molnár-Sáska
Computer and Automation Institute
Hungarian Academy of Sciences
13-17 Kende u., Budapest 1111, Hungary
gerencser@sztaki.hu, molnar@math.bme.hu
György Michaletzky
University of Eötvös Loránd
1/c Pázmány Péter, Budapest 1117, Hungary
michgy@cs.elte.hu
Gábor Tusnády
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
13-15. Reáltanoda u., Budapest 1053, Hungary
tusnady@math-inst.hu
Zsuzsanna Vágó
Computer and Automation Institute
Hungarian Academy of Sciences
13-17 Kende u., Budapest 1111, Hungary
vago@oplab.sztaki.hu

Abstract

The estimation of Hidden Markov Models has attracted a lot of attention recently, see [19], [22], [23]. The purpose of this paper is to lay the foundation for a new approach for the analysis of the maximum-likelihood estimation of HMM-s, using representation of HMM-s due to [7]. A useful connection between the estimation theory of HMM-s and linear stochastic systems is established via the theory of L-mixing processes developed in [11]. The results are potentially useful for deriving strong approximation results, which are in turn applicable to analyze adaptive predictors and change detection methods.

Keywords: Hidden Markov Models, stochastic systems, random transformations, Doeblin-condition, L-mixing processes, maximum-likelihood estimation.

1 Introduction

Hidden Markov Models have become a basic tool for modeling stochastic systems with a wide range of applicability. For a general introduction see [25]. The estimation of the dynamics of a Hidden Markov Model is a basic problem in applications. A key element in the statistical analysis of HMM-s is a strong law of large numbers for the log-likelihood function. In previous works geometric ergodicity of Markov chains and the subadditive ergodic theorem were used, see [19], [22], [23]. An alternative tool that has been widely used in linear system identification is theory of L-mixing processes. The relevance of this theory will be established in this paper using a random-transformation representation for Markov-processes (see [17], [7], [15]). The advantage of this approach is that, under suitable conditions a more precise characterization of the estimation error-process can be obtained, which, in turn, is crucial for the analysis of the performance of adaptive prediction, see [14]. The paper is an extension of [15]. The main advance is the extension of Theorem 5.1 to unbounded functions.

2 Hidden Markov Models

We consider Hidden Markov Models with a general state space $\mathcal{X}$ and an observation or read-out space $\mathcal{Y}$. Both are assumed to be Polish spaces, i. e. they are complete, separable metric spaces. The pair (X_n, Y_n) is a Hidden Markov process if (X_n) is a homogenous Markov chain, with state space $\mathcal{X}$ and the observations Y_n are conditionally independent and identically distributed given (X_n).

If $\mathcal{X}$ and $\mathcal{Y}$ are discrete then we have with $X^n = (X_n, \ldots, X_1)$, $Y^n = (Y_n, \ldots, Y_1)$

$$P(Y^n|X^n) = \prod_{i=0}^{n} P(Y_i = y_i|X_i = x_i).$$

In this case we will use the following notations

$$P(Y_k = y|X_k = x) = b^x(y), \quad B(y) = \mathrm{diag}(b^i(y)).$$

Let Q be the transition matrix of the unobserved Markov process (X_n), i.e. $Q_{ij} = P(X_{n+1} = j|X_n = i)$. Let the predictive filter be defined by

$$p^j_{n+1} = P(X_{n+1} = j|Y_n, \dots Y_0).$$

Writing $p_{n+1} = (p^1_{n+1}, \dots, p^N_{n+1})^T$, the filter process satisfies the Baum-equation

$$p_{n+1} = \pi(Q^* B(Y_n) p_n). \tag{1}$$

where π is the normalizing operator, to ensure that p_{n+1} is a probability vector, see [1]. Here $p^j_0 = P(X_0 = j)$, but in practice we take an arbitrary probability vector q as initial condition, and the solution of the Baum-equation will be denoted by $p_n(q)$. A key property of the Baum-equation is its exponential stability with respect to the initial condition. This has been established in [19] for continuous read-outs, but here we state it for finite-state finite read-out HMM-s.

Theorem 2.1 *Assume that for some positive integer m we have $Q^m > 0$ and $b^x(y) > 0$ for all x, y. Then*

$$||p_n(q) - p_n(q')||_{TV} \leq C(1-\delta)^n ||q - q'||,$$

where $|| \quad ||_{TV}$ denotes the norm in total variation.

This basic property of the prediction filter will be used to introduce the concept of exponential stability abstracted and will be used to derive general mixing properties of the extended process (X_n, Y_n, p_n).

3 Representation of Markov processes

Let the state space of a Markov-chain be a Polish space $\mathcal{X}$, and let $\mathcal{M} : \mathcal{X} \longrightarrow \mathcal{X}$ be the space of Borel-measurable mappings. Assume that $\mathcal{M}$ is equipped with a σ-algebra of its subsets. We assume that the sets defined for any fixed $x \varepsilon \mathcal{X}$ and $G \varepsilon \mathcal{B}(\mathcal{X})$, where $\mathcal{B}(\mathcal{X})$ is the σ-algebra of Borel sets in $\mathcal{X}$, $\{f : \quad f x \varepsilon G\}$ are mesaureable. Let m be a probability measure on this σ-algebra. Let T_n be a sequence of i.i.d mappings according to m. Let X_0 be independent of (T_n). Then the process (X_n) defined by $X_0 \varepsilon \mathcal{X}$, $X_{n+1} = T_{n+1} X_n$ is Markov.

Conversely let (X_n) be a Markov-process on a Polish-space $\mathcal{X}$ with transition probabilities $P(x, G)$, $x \varepsilon \mathcal{X}$, $G \varepsilon \mathcal{B}(\mathcal{X})$, then we can find a representation, in other words a measure m on $\mathcal{M}$, satisfying

$$P(x, G) = m\{T : Tx \varepsilon G\}.$$

For the proof see [17]. Later on we will "identify" m with P.

Next we are going to introduce the notion of Doeblin-condition (see [5]):

Definition 3.1 *If there exists an integer $m \geq 1$ such that $P^m(x, A) \geq \delta \nu(A)$ is valid for all $x \varepsilon \mathcal{X}$ and $A \varepsilon \mathcal{B}(\mathcal{X})$ with some probability measure ν, then we say that the Doeblin-condition is satisfied.*

Lemma 3.1 *Let (X_n) be a Markov chain. The Doeblin condition is valid if and only if there exists a sequence of i.i.d. mappings (T_n) such that $P(T_m \dots T_1 \varepsilon \Gamma_c) \geq \delta$, where Γ_c is the set of constant mappings.*

Proof. See [5] ∎

Theorem 3.1 *Assume that the Doeblin condition holds for a Markov chain (X_n). Then there exists an invariant distribution π, and*

$$|P^n(x, A) - \pi(A)| \leq (1-\delta)^n \quad \text{for} \quad \forall A \varepsilon \mathcal{B}(\mathcal{X}).$$

Proof. See [5]. ∎

Now let (X_n, Y_n) be a Hidden Markov process and assume that the state space $\mathcal{X}$ and the observed space $\mathcal{Y}$ are Polish.

Lemma 3.2 *Assume that the Doeblin condition holds for the Markov chain (X_n). Then the Doeblin condition holds for (X_n, Y_n) as well.*

Proof. Let T_n be the representation of the Markov chain as in Lemma 3.1. It means there exists a sequence of i.i.d. mappings (T_n) such that $X_{n+1} = T_{n+1} X_n$ with $P(T_n \varepsilon \Gamma_c) \geq \delta > 0$, and (T_n) is independent from the starting point X_0.

Let $P(x, G)$ be the read-out transition kernel of the original Markov chain (X_n), where $x \varepsilon \mathcal{X}$ and $G \varepsilon \mathcal{Y}$. By Lemma 3.1 there is a probability measure m on the space of Borel mappings $\mathcal{X} \longrightarrow \mathcal{Y}$ for which $P(x, G) = m\{U : Ux \varepsilon G\}$.

With the notation $Y_n = U_n X_n$ we get

$$Y_{n+1} = U_{n+1} T_{n+1} X_n.$$

Obviously if $T_n \varepsilon \Gamma_c(\mathcal{X} \to \mathcal{X})$, then $U_n T_n \varepsilon \Gamma_c(\mathcal{X} \to \mathcal{Y})$, and thus the lemma is proved. ∎

4 Markov chains and L-mixing processes

Now we are going to introduce a class of processes called L-mixing processes which have been used extensively in the statistical analysis of linear stochastic systems, see [11].

Definition 4.1 *A stochastic process (X_n) $(n \geq 0)$ taking its values in an Euclidean space is M-bounded if for all $q \geq 1$*

$$M_q = \sup_{n \geq 0} E^{1/q} \|X_n\|^q < \infty.$$

Let $(\mathcal{F}_n)$ and $(\mathcal{F}_n^+)$ be two sequences of monoton increasing and monoton decreasing σ-algebras, respectively such that $\mathcal{F}_n$ and $\mathcal{F}_n^+$ are independent for all n.

Definition 4.2 *A stochastic process (X_n) taking its values in a finite-dimensional Euclidean space is L-mixing, if it is M-bounded and with*

$$\gamma_q(\tau) = \sup_{n \geq \tau} E^{1/q} \|X_n - E(X_n | \mathcal{F}_{n-\tau}^+)\|^q$$

$$\Gamma(q) = \sum_{\tau=0}^{\infty} \gamma_q(\tau) < \infty.$$

holds.

The following lemma is useful in checking whether a process is L-mixing or not.

Lemma 4.1 *Let X be a random variable as above with $E\|X\|^q < \infty$ for all q, and let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra and η is a $\mathcal{G}$ measurable random variable. Then we have*

$$E^{1/q} \|X - E(X|\mathcal{G})\|^q \leq 2E^{1/q} \|X - \eta\|^q.$$

The following proposition shows the importance of the L-mixing processes.

Proposition 4.1 *Let (X_n) be a Markov chain with state space X, where X is a Polish space, and assume that the Doeblin condition is valid for $m = 1$. Furthermore let $g : X \longrightarrow \mathbb{R}$ be a bounded, measurable function. Then $g(X_n)$ is an L-mixing process.*

Proof. Let $n > m$ and $n - m = \tau$. To approximate the process $g(X_n)$, let

$$\mathcal{F}_n = \sigma\{X_0, T_k : k \leq n\}$$

and

$$\mathcal{F}_n^+ = \sigma\{T_k : k \geq n+1\}.$$

First of all we approximate X_n by $X_{n,m}^+$, where

$$X_{n,m}^+ = T_n \dots T_{m+1} X^*, \tag{2}$$

and X^* is a constant and independent of (T_k). Obviously $X_{n,m}^+$ is $\mathcal{F}_m^+$ measurable. Furthermore $P(X_n \neq X_{n,m}^+)$ is majorated by

$$P(T_k \notin \Gamma_c \text{for } m+1 \leq k \leq n) \leq (1-\delta)^{n-m}.$$

Now

$$E^{1/q} \|g(X_n) - g(X_{n,m}^+)\|^q \leq 2KP^{1/q}(X_n \neq X_{n,m}^+).$$

The last expression is majorated by $2K(1-\delta)^{\frac{n-m}{q}}$, where K is an upper bound for $|g|$. Due to Lemma 4.1

$$\gamma_q(\tau) \leq 4K(1-\delta)^{\frac{\tau}{q}}$$

holds, and the right hand side of the last inequality is summable, thus the statement is proved. ■

5 Exponentially stable random mappings

Now we formulate a general concept of exponential stability motivated by Theorem 2.1. Let X be an arbitrary abstract set, $\mathcal{Z}$ be an Euclidean space and $f : X \times \mathcal{Z} \longrightarrow \mathcal{Z}$ be a Borel-measurable function. Consider the recursion

$$z_{n+1} = f(x_n, z_n), \quad z_0 = \xi, \tag{3}$$

$x_n \varepsilon X$ arbitrary, and the solution be denoted by $z_n(\xi)$.

Definition 5.1 *The mapping f is uniformly exponentially stable if for every sequence $(x_n) \varepsilon X$*

$$\|z_n(\xi) - z_n(\xi')\| \leq C(1-\rho)^n \|\xi - \xi'\|,$$

where $C > 0, 1 > \rho > 0$ are independent of the sequence (x_n).

Under reasonable technical conditions this condition is satisfied for the Baum-equation and its derivatives, see [19]. Let $z_n = z(n,m,\xi)$ denote the solution of (3) initialized at $z_m = \xi$ with $m \leq n$.

We need the following lemma that is proved by direct calculation.

Lemma 5.1 *For the solution of (3) we have*

$$z_n = \xi + \sum_{m=0}^{n-1} (z(n, m+1, f(x_m, \xi)) - z(n, m+1, \xi)).$$

Define the process (Z_n) by

$$Z_{n+1} = f(X_n, Z_n), \quad Z_0 = \xi, \tag{4}$$

where (X_n) is a Markov-chain. To prove M-boundedness of (Z_n) we impose the condition that $E\|Z_1\|^q < \infty$ for any fixed $Z_0 = \xi$. This is expressed in the following condition:

Condition 5.1 *Assume for all $\xi \varepsilon \mathcal{Z}$ and for any $q \geq 1$*

$$\int_X \|f(x, \xi)\|^q d\pi(x) \leq K_1(\xi),$$

where π is the stationary distribution of (X_n).

Condition 5.2 *Let the distribution of X_0 be π_0. Then*

$$\frac{d\pi_0}{d\pi} \le C_1.$$

Condition 5.2 and 5.1 imply $E\|f(X_n,\xi)\|^q \le K_1(\xi)C_1$.

Lemma 5.2 *Let the mapping $f(x,z)$ be uniformly exponentially stable, and let Condition 5.1 and 5.2 hold. Then the process Z_n defined by (4) is M-bounded.*

The proof is obtained using Lemma 5.1.

Theorem 5.1 *Consider the process (X_n,Z_n), where (X_n) satisfies the Doeblin-condition with $m=1$, and Z_n is defined by (4) with a uniformly exponentially stable mapping f. Assume that X_0 is independent of $\{T_n\}$, $n\ge 1$, and Conditions 5.1 and 5.2 hold. Furthermore let $g(x,z)$ be a bounded, measurable function on $\mathcal{X}\times\mathcal{Z}$ such that is Lipschitz-continuous in z for every x with an x-independent Lipschitz constant L. Then*

$$v_n = g(x_n,z_n)$$

is an L-mixing process.

Proof. The process $g(x_n,z_n)$ is obviously M-bounded. Now let $n>m$, $\tau=n-m$, $\mathcal{F}_n,\mathcal{F}_n^+$, and $X_{n,m}^+$ be the same as in Proposition 4.1, such that the distribution of X^* is stationary. Let an approximation of Z_n be defined by

$$Z_{k+1,m}^+ = f(X_{k,m}^+,Z_{k,m}^+), \quad \text{where} \quad Z_{m,m}^+ = z^* \quad \text{is a constant.}$$

Obviously, $Z_{n,m}^+$ is $\mathcal{F}_m^+$-measurable. Let $m'=n-[\frac{\tau}{2}]$ and let B denote the event no coupling occurs in the interval $[m',m]$:

$$B=\{\omega:\ \text{for} \quad m<k\le m' \quad T_k\notin\Gamma_c\}$$

It is easy to see that for $q\ge 1$

$$E^{\frac{1}{q}}\|g(X_n,Z_n)-g(X_{n,m}^+,Z_{n,m}^+)\|^q \le$$

$$2KP^{\frac{1}{q}}(B)+LC(1-\rho)^{[\frac{\tau}{2}]}E^{\frac{1}{q}}\|Z_{m',m}^+-Z_{m'}\|^q$$

where $\|g(x,z)\|\le K$. Now by Lemma 5.2 $Z_{m'}$ and $Z_{m',m}^+$ are M-bounded, hence the claim of the theorem follows. ∎

6 The case of unbounded g

In this section we consider an extension of Theorem 5.1 for *unbounded* g. The process (X_n,Z_n) is Markov. A representation of this Markov process is

$$(x,z)\longrightarrow (Tx,f(x,z)). \tag{5}$$

Lemma 6.1 *Assume that f is uniformly exponentially stable and Condition 5.1 holds. Then the process (X_n,Z_n) has a stationary distribution, say μ*

Proof. In Theorem 3.1 it is shown that X_n has stationary distribution. Define X_{-n} as the limit

$$X_{-n}=\lim_k T_{-n}\circ\cdots\circ T_{-n-k}\eta,$$

Denote the mapping $f(x_n,\cdot):\mathcal{Z}\longrightarrow\mathcal{Z}$ by f_{x_n} and let be

$$Z_0^*=\lim_n f_{X_0}\circ f_{X_{-1}}\circ\cdots\circ f_{X_{-n}}\xi.$$

We have to prove that the limit exists. Take a realization of (X_{-n}) denoted by (x_{-n}). Taking into account the exponential stability of f and Lemma 5.2 it is easy to see that $f_{x_0}\circ f_{x_{-1}}\circ\cdots\circ f_{x_{-n}}\xi$ is a Cauchy sequence in L_q, hence it converges. Now it is easy to see that the distribution of (X_0^*,Z_0^*) is stationary. ∎

Let the distribution of the process (X_0^*,Z_0^*) be denoted by μ, and the distribution of (X_n,Z_n) by μ_n.

Condition 6.1 *The Radon-Nikodym derivate of μ_0 w.r.t. μ is bounded, say*

$$\frac{d\mu_0}{d\mu}\le K.$$

Condition 6.2 *We have*

$$\int_{\mathcal{X}\times\mathcal{Z}} \|g(x,z)\|^q d\mu(x,z)\le M.$$

Condition 6.1 implies that

$$\frac{d\mu_n}{d\mu}\le K \quad \text{for all} \quad n$$

and we get that Condition 6.1 and 6.2 imply

$$E\|g(X_n,Z_n)\|^q\le KM.$$

Condition 6.3 *Let the initial value ξ be random with $E^{\frac{1}{q}}\|\xi\|^q<\infty$ and distribution $\nu(\xi)$, and let it be independent of (X_n). Then*

$$\int_{\mathcal{X}\times\mathcal{Z}} \|f(x,\xi)\|^q d\pi(x)d\nu(\xi)<\infty.$$

Condition 6.4 *If ξ has the stationary distribution of (Z_n), then*

$$E\|\xi\|^q<\infty \quad \textit{for all} \quad q.$$

Theorem 6.1 *Consider the process (X_n, Z_n), where X_n satisfies the Doeblin-condition with $m = 1$, and let (Z_n) be defined by (4) with a uniformly exponentially stable mapping f. Let $Z_0 = \xi$ be a random variable with $E^{\frac{1}{q}} \|\xi\|^q < \infty$ and let X_0 be independent of (T_n). Let Condition 5.1, 6.1, 6.3 and 6.4 hold. Let $g(x,z)$ be a measurable function of $\mathcal{X} \times \mathcal{Z}$ that is Lipschitz-continuous in z for every x with Lipschitz constant L. Assume that Condition 6.2 is satisfied. Then*

$$v_n = g(x_n, z_n)$$

is an L-mixing process.

Proof. The proof is analogous to the proof of Theorem 5.1. The key difference is that $2KP^{\frac{1}{q}}(B)$ will be replaced by

$$E^{\frac{1}{qs}} \|g(X_n, Z_n)\|^{qs} + E^{\frac{1}{qs}} \|g(X_n^+, Z_{n,m}^+)\|^{qs}.$$

Due to Condition 6.1, 6.2, 6.3, 6.4 the latter expression is exponentially small, which implies the proof. ∎

7 Estimation of Hidden Markov Models

This section gives a brief outline of the maximum likelihood estimation of Hidden Markov Models. Consider the classical case when $\mathcal{X}$ and $\mathcal{Y}$ are finite. With the notation $p_n^i = P(X_n = i | Y_{n-1}, \ldots, Y_0)$ we have

$$p_{n+1} = \pi(Q^T B(Y_n) p_n) = f(Y_n, p_n).$$

We can write $\log p(y_{n-1}, \ldots y_0, \theta)$ as

$$\sum_{k=1}^{n-1} \log p(y_k | y_{k-1}, \ldots y_0, \theta) + \log p(y_0, \theta).$$

Here the k-th term is

$$\log \sum_x b^x(y_k) P(x | y_{k-1}, \ldots, y_0, \theta) = \log \sum_x b^x(y_k) p_k^x.$$

Now write

$$g(y,p) = \log \sum_x b^x(y) p^x. \tag{6}$$

Thus we get

$$\log p(y_n, \ldots, y_0, \Theta) = \frac{1}{N} \sum_{k=1}^{N} g(y_k, p_k).$$

Theorem 6.1 is applicable under suitable conditions, and then the arguments of [13] yield the following result:

Theorem 7.1 *Assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite and $Q^m > 0$ for some positive integer m and $b^x(y) > 0$ for all x, y. Let $\hat{\theta}_N$ be the ML estimate of θ^*. Then $\hat{\theta}_N - \theta^*$ can be written as*

$$-(R^*)^{-1} \frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial \theta} \log p(Y_n | Y_{n-1}, \ldots, Y_0, \theta^*) + \delta_n,$$

where $\delta_n = O_M(N^{-1})$, and R^* is the Fisher-information matrix.

A key point here is that the error term is $O_M(N^{-1})$. This ensures that all basic limit theorems, that are known for the dominant term, which is a martingale, are also valid for $\hat{\theta}_N - \theta^*$.

8 Change detection of HMM-processes

The negative of the log-likelihood can be interpreted as a codelength, modulo a constant, which is obtained when encoding the data sequence $(y_N, \ldots, y_1)$ with a prescribed accuracy, using the assumed joint density $p(y_N, \ldots, y_0; \theta)$. This interpretation of the likelihood is a central idea of the theory of stochastic complexity, see [24]. Let

$$C_n(y_n; \theta) \stackrel{\Delta}{=} -\log p(y_n | y_{n-1}, \ldots, y_0; \theta),$$

A key observation of the theory of the stochastic complexity is that for the mean loss in the codelength we have

$$E(C_n(y_n, \widehat{\theta}) - C_n(y_n, \theta^*)) \cong \frac{1}{2} \mathrm{tr} I(\theta^*) S,$$

where $I(\theta^*)$ is the Fisher information matrix.

A useful device for change detection is the use of forgetting. Let $\widehat{\theta} = \widehat{\theta}_n(\lambda)$ be the estimator of θ^* obtained by minimizing

$$\sum_{n=1}^{N} (1-\lambda)^{N-n} \lambda \log p(y_n | y_{n-1}, \ldots, y_0; \theta),$$

with $0 < \lambda < 1$ is the forgetting factor. Using the techniques developed in [12] it can be shown that under the conditions of Theorem 7.1 with $p = \dim \theta$

$$E(C_n(y_n, \widehat{\theta}_{n-1}(\lambda)) - C_n(y_n, \theta^*) \cong \frac{1}{2} p \lambda.$$

We are now in a position to apply the techniques developed in [2] for change point detection of ARMA processes. Define

$$C_N(\lambda) = \sum_{n=1}^{N} -\log q_n,$$

where $q_n = p(y_n | y_{n-1}, \ldots, y_0, \widehat{\theta}_{n-1})$. Then using (7) and the law of large numbers we have with $0 < \lambda_1 < \lambda_2$:

$$\delta_N = C_N(\lambda_1) - C_N(\lambda_2)/N \cong \frac{1}{2} p(\lambda_1 - \lambda_2).$$

The faster the forgetting is, the more we loose in the performance of the encoding procedure.

Now if a change does occur at some τ, then the procedure using forgetting factor λ_2 will adapt faster, hence locally (in time) we will have a better encoding performance. This will bring us to the following application of the Hinkley detector: let

$$\delta_N^* = \min_{n \le N} \delta_n,$$

and let $\varepsilon > 0$ be a fixed threshold. Whenever

$$\delta_N - \delta_N^* > \varepsilon$$

we generate an alarm. This on-line detector has been extensively tested for ARMA processes. Simulation studies for HMM-s are under way. A number of interesting theoretical issues concerning the properties of the above detector have been solved, see [2].

The above techniques can be used to detect changes in so called ARCH or GARCH processes playing an important role in mathematical finance. These processes introduced by Engle in [10] and Bollerslev in [6] are used to model the log-return of a stock-price process. For the best recent results on the estimation of GARCH processes see Berkes et al. [4]

9 Acknowledgement

Three authors acknowledge the support of the National Research Foundation of Hungary (OTKA) under Grant no. T 032932.

References

[1] L. E. Baum, T. Petrie, Statistical inference for probabilistic functions of finite state Markov chains, Ann. Math. Stat. 37 (1966) 1559-1563.

[2] J. Baikovicius and L. Gerencsér. Change point detection in a stochastic complexity framework. In *Proc. of the 29-th IEEE CDC*, volume 6, pages 3554–3555, 1990.

[3] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*. Springer-Verlag, Berlin, 1990.

[4] I. Berkes, J. Horváth, and P. Kokoszka. GARCH processes: structure and estimation, Manuscript, 2002.

[5] R. Bhattacharya, E. C, Waymire, An Approach to the Existence of Unique Invariant Probabilities for Markov Processes, 1999

[6] T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[7] V. S. Borkar, On white noise representations in stochastic realization theory, SIAM J. Control and Optim. 31 (1993) 1093-1102.

[8] R. Douc, C. Matias, Asymptotics of the maximum likelihood estimator for general hidden Markov models, Bernoulli 7(3) (2001) 381-420.

[9] R. J. Elliott, J. van der Hoek, An application of hidden Markov models to asset allocation problems, Finance Stochast. 1. (1997) 229-238.

[10] R.F. Engle. Autoregressive Conditional Heteroskedasticity with Estimates of The Variance of the UK Inflation. *Econometrica*, 50:987–1008, 1982.

[11] L. Gerencsér, On a Class of Mixing Processes, Stochastics 26. 165-191.

[12] L. Gerencsér. Predictive stochastic complexity associated with fixed gain estimators. In C. Praagman J.W. Nieuewenhuis and H.L. Trentelman, editors, *Proceedings of the 2nd European Control Conference, Groningen*, pages 1673–1677, 1993.

[13] L. Gerencsér, Rate of convergence of recursive estimators, SIAM J. Control and Optimization, Vol 30, No. 5 (1992) 1200-1227.

[14] L. Gerencsér, On the martingale approximation of the estimation error of ARMA parameters, Szstems & Control Letters 15 (1990) 417-423. North-Holland

[15] L. Gerencsér, G. Molnár-Sáska, A New Method for the Analysis of Hidden Markov Model Estimates, IFAC Congress, Barcelona, 2002 jul.

[16] S. Karlin, H. M. Taylor, Stochastic Processes, Academic Press, New York

[17] Y. Kifer, Ergodic Theory of Random Transformation, Progress in Probability and Statistics Vol 10, 1986

[18] F. Le Gland, L. Mevel, Basic Properties of the Projective Product with Application to Products of Column-Allowable Nonnegative Matrices, Math. Control Signals Systems 13. (2000) 41-62.

[19] F. Le Gland, L. Mevel, Exponential Forgetting and Geometric Ergodi city in Hidden Markov Models, Math. Control Signals Systems 13. (2000) 63-93.

[20] F. Le Gland, Stability and Approximation of Nonlinear Filters: an Information Theoretic Approach, Proceedings of the 38th Conference on Decision & Control, Phoenix, Arizona (1999)

[21] F. Le Gland, L. Mevel, Asymptotic Properties of the MLE in Hidden Markov Models, Proceedings of the 4th European Control Conference, Bruxelles 1997

[22] B. G. Leroux, Maximum-likelihood estimation for hidden Markov models, Stochastic Processes and their Applications 40 (1992) 127-143. North-Holland

[23] L. Mevel, L. Finesso, Convergence Rates of the Maximum Likelihood Estimator of Hidden Markov Models,2000

[24] J. Rissanen. *Stochastic complexity in statistical inquiry*. World Scientific Publisher, 1989.

[25] J. H. van Schuppen, Lecture Notes on Stochastic Systems, Manuscript

[26] Zs. Vágó ,L. Gerencsér, Molnár-Sáska, G., Change detection of stochastic volatility processes, Bachelier Congress, Crete, 2002 jun.