RECURSIVE SMOOTHERS FOR HIDDEN DISCRETE-TIME MARKOV CHAINS

LAKHDAR AGGOUN

Received 22 September 2004 and in revised form 4 February 2005

We consider a discrete-time Markov chain observed through another Markov chain. The proposed model extends models discussed by Elliott et al. (1995). We propose improved recursive formulae to update smoothed estimates of processes related to the model. These recursive estimates are used to update the parameter of the model via the expectation maximization (EM) algorithm.

1. Introduction

Hidden Markov chains have been the subject of extensive studies, see the books [1, 2] and the references therein. Of particular interest are the discret-time, finite-state hidden Markov models.

In this paper, using the same techniques as in [3], we propose results that improve the finite-dimensional smoothers of functionals of a partially observed discrete-time Markov chain. The model itself extends models discussed in [2]. The proposed formulae for updating these quantities are recursive. Therefore, recalculation of all backward estimates is not required in the implementation of the EM algorithm.

This paper is organized as follows. In Section 2, we introduce the model. In Section 3, a new probability measure under which all processes are independent is defined and a recursive filter for the state is derived. The main results of this paper are in Section 4 where recursive smoothers are derived.

2. Model dynamics

A system is considered, whose state is described by a finite-state, homogeneous, discrete-time Markov chain X_k, $k \in \mathbb{N}$. We suppose that X_0 is given, or its distribution is known. If the state space of X_k has N elements, it can be identified without loss of generality, with the set

$$S_X = \{e_1, \dots, e_N\}, \tag{2.1}$$

where e_i are unit vectors in $\mathbb{R}^N$ with unity as the ith element and zeros elsewhere.

Journal of Applied Mathematics and Stochastic Analysis 2005:3 (2005) 345–351
DOI: 10.1155/JAMSA.2005.345

Write $\mathcal{F}_k^0 = \sigma\{X_0,\dots,X_k\}$, for the σ-field generated by $X_0,\dots,X_k$, and $\{\mathcal{F}_k\}$ for the complete filtration generated by the $\mathcal{F}_k^0$; this augments $\mathcal{F}_k^0$ by including all subsets of events of probability zero. The Markov property implies here that

$$P(X_{k+1} = e_j \mid \mathcal{F}_k) = P(X_{k+1} = e_j \mid X_k). \tag{2.2}$$

Write $a_{ji} = P(X_{k+1} = e_j \mid X_k = e_i)$, $A = (a_{ji}) \in \mathbb{R}^{N\times N}$.

Define $V_{k+1} := X_{k+1} - AX_k$ so that

$$X_{k+1} = AX_k + V_{k+1}. \tag{2.3}$$

$\{V_k\}$, $k \in \mathbb{N}$, is a sequence of martingale increments.

The state process X is not observed directly. We observe a second Markov chain Y on the same state space as X but with probability transitions perturbated by X. More precisely, suppose that

$$P(Y_{k+1} = e_s \mid \mathcal{G}_k \vee \sigma\{X_{k+1}\}) = P(Y_{k+1} = e_s \mid Y_k, X_{k+1}), \tag{2.4}$$

where $\{\mathcal{G}_k\}$ is the complete filtration generated by X and Y.

Write

$$b_{s,ri} = P(Y_{k+1} = e_s \mid Y_k = e_r, X_{k+1} = e_i), \tag{2.5}$$

and $B = \{b_{s,ri}\}$, $1 \le s, r, i \le N$. Note that $\sum_{s=1}^{M} b_{s,ri} = 1$. We immediately have the following representation for Y:

$$Y_{k+1} = B\ Y_k \otimes X_{k+1} + W_{k+1}, \tag{2.6}$$

where W_k, $k \in \mathbb{N}$, is a sequence of martingale increments.

Let $\{\mathcal{Y}_k\}$ be the complete filtration generated by Y.

Our objective here is to seek recursive filters and smoothers for the states of the Markov chain X, the number of jumps from one state to another for the occupation time of a state, and for a process related to the observations.

3. An unnormalized finite-dimensional recursive filter for the state

What we wish to do now is starting with a probability measure $\overline{P}$ on $(\Omega, \bigvee_{n=1}^{\infty} \mathcal{G}_n)$ such that

(1) the process X is a finite-state Markov chain with transition matrix A;

(2) $\{Y_k\}$, $k \in \mathbb{N}$, is a sequence of *i.i.d.* random variables and $\overline{P}(Y_{k+1} = e_r \mid \mathcal{G}_k \vee \sigma\{X_{k+1}\}) = \overline{P}(Y_{k+1} = e_r) = 1/M$.

We will now construct a new measure P on $(\Omega, \bigvee_{n=1}^{\infty} \mathcal{G}_n)$ such that under P, $E[Y_{k+1} \mid \mathcal{G}_k \vee \sigma\{X_{k+1}\}] = BY_k \otimes X_{k+1}$. Write

$$\lambda_\ell = \prod_{s,r,i=1}^{N} (Nb_{s,ri})^{\langle Y_\ell, e_s\rangle \langle Y_{\ell-1}, e_r\rangle \langle X_\ell, e_i\rangle}, \quad \ell \in \mathbb{N}, \tag{3.1}$$

$$\Lambda_k = \prod_{\ell=1}^{k} \lambda_\ell, \tag{3.2}$$

where $b_{s,ri}$ is the probability transition defined in (2.5).

With the above definitions, $\overline{E}[\lambda_{k+1} \mid \mathcal{G}_k] = 1$. Now set $(dP/d\overline{P})\mid_{\mathcal{G}_k} = \Lambda_k$. (The existence of P follows from Kolmogorov's extension theorem.)

Recall that for a $\mathcal{G}$-adapted sequence $\{\phi_k\}$,

$$E[\phi_k \mid \mathcal{Y}_k] = \frac{\overline{E}[\Lambda_k \phi_k \mid \mathcal{Y}_k]}{\overline{E}[\Lambda_k \mid \mathcal{Y}_k]}. \tag{3.3}$$

Write $q_k(e_{\mathfrak{m}})$, $1 \le t \le N$, $k \in \mathbb{N}$, for the unnormalized, conditional probability distribution such that

$$\overline{E}[\Lambda_k \langle X_k, e_{\mathfrak{m}}\rangle \mid \mathcal{Y}_k] = q_k(e_{\mathfrak{m}}). \tag{3.4}$$

Now $\sum_{i=1}^{N} \langle X_k, e_i\rangle = 1$, so

$$\sum_{i=1}^{N} q_k(e_i) = \overline{E}\left[\Lambda_k \sum_{i=1}^{N} \langle X_k, e_i\rangle \mid \mathcal{Y}_k\right] = \overline{E}[\Lambda_k \mid \mathcal{Y}_k]. \tag{3.5}$$

Therefore, the normalized conditional probability distribution

$$p_k(e_{\mathfrak{m}}) = E[\langle X_k, e_{\mathfrak{m}}\rangle \mid \mathcal{Y}_k] \tag{3.6}$$

is given by

$$p_k(e_{\mathfrak{m}}) = \frac{q_k(e_{\mathfrak{m}})}{\sum_{j=1}^{k} q_k(e_j)}. \tag{3.7}$$

To simplify the notation, we write

$$\begin{aligned} c_{\mathfrak{m}}(Y_k, Y_{k-1}) &= \prod_{s,r=1}^{N} (Nb_{s,r\mathfrak{m}})^{\langle Y_k, e_s\rangle \langle Y_{k-1}, e_r\rangle}, \\ \underline{c}(Y_k, Y_{k-1}) &= (c_1(Y_k, Y_{k-1}), \ldots, c_N(Y_k, Y_{k-1}))'. \end{aligned} \tag{3.8}$$

Theorem 3.1. *For $k \in \mathbb{N}$, the recursive filter for the unnormalized estimates of the states is given by*

$$q_k = \operatorname{diag}[\underline{c}(Y_k, Y_{k-1})] A q_{k-1}. \tag{3.9}$$

Proof. In view of (3.1), (3.2), (2.3), and the notation in (3.8),

$$\begin{aligned}\overline{E}[\Lambda_k\langle X_k,e_{\mathfrak{m}}\rangle \mid \mathcal{Y}_k] &= \prod_{s,r=1}^{N}(Nb_{s,r\mathfrak{m}})^{\langle Y_k,e_s\rangle\langle Y_{k-1},e_r\rangle}\sum_{i=1}^{N}\overline{E}[\Lambda_{k-1}\langle X_{k-1},e_i\rangle\langle Ae_i,e_{\mathfrak{m}}\rangle \mid \mathcal{Y}_k]\\ &= c_{\mathfrak{m}}(Y_k,Y_{k-1})\sum_{i=1}^{N}a_{\mathfrak{m}i}\langle q_{k-1},e_i\rangle,\end{aligned} \tag{3.10}$$

and

$$\begin{aligned}\overline{E}[\Lambda_k X_k \mid \mathcal{Y}_k] &= \sum_{\mathfrak{m}=1}^{N}\overline{E}[\Lambda_k\langle X_k e_{\mathfrak{m}}\rangle \mid \mathcal{Y}_k]e_{\mathfrak{m}}\\ &= \sum_{\mathfrak{m}=1}^{N}\sum_{i=1}^{N}a_{\mathfrak{m}i}c_{\mathfrak{m}}(Y_kY_{k-1})\langle q_{k-1},e_i\rangle e_{\mathfrak{m}}\\ &= \operatorname{diag}[\underline{c}(Y_k,Y_{k-1})]Aq_{k-1},\end{aligned} \tag{3.11}$$

which finishes the proof. □

4. Recursive smoothers

We emphasize again that these improved recursive formulae to update smoothed estimates are used to update the parameters of the model via the EM algorithm.

Theorem 4.1. *For $k > m$, the unnormalized smoothed estimate $\overline{E}[\Lambda_k X_m \mid \mathcal{Y}_k] \triangleq \gamma_{m,k}$ is given by*

$$\gamma_{m,k} = \operatorname{diag}[q_m]v_m. \tag{4.1}$$

Proof. Write $\prod_{\ell=m+1}^{k}\lambda_\ell \triangleq \Lambda_{m+1,k}$.

$$\begin{aligned}\overline{E}[\Lambda_k\langle X_m,e_i\rangle \mid \mathcal{Y}_k] &= \overline{E}[\Lambda_m\langle X_m,e_i\rangle\Lambda_{m+1,k} \mid \mathcal{Y}_k]\\ &= \overline{E}[\Lambda_m\langle X_m,e_i\rangle\overline{E}[\Lambda_{m+1,k} \mid \mathcal{Y}_k\vee\mathcal{F}_m] \mid \mathcal{Y}_k]\\ &= \overline{E}[\Lambda_m\langle X_m,e_i\rangle\overline{E}[\Lambda_{m+1,k} \mid \mathcal{Y}_k\vee\mathcal{F}_m] \mid \mathcal{Y}_k]\\ &= \overline{E}[\Lambda_m\langle X_m,e_i\rangle\overline{E}[\Lambda_{m+1,k} \mid \mathcal{Y}_k\vee\{X_m=e_i\}] \mid \mathcal{Y}_k]\\ &= \overline{E}[\Lambda_m\langle X_m,e_i\rangle \mid \mathcal{Y}_m]\overline{E}[\Lambda_{m+1,k} \mid \mathcal{Y}_k\vee\{X_m=e_i\}]\\ &\triangleq \langle q_m,e_i\rangle\langle v_m,e_i\rangle,\end{aligned} \tag{4.2}$$

where

$$v_m = (\overline{E}[\Lambda_{m+1,k} \mid \mathcal{Y}_k\vee\{X_m=e_1\}],\ldots,\overline{E}[\Lambda_{m+1,k} \mid \mathcal{Y}_k\vee\{X_m=e_N\}])'. \tag{4.3}$$

Therefore,

$$\begin{aligned}\overline{E}[\Lambda_k X_m \mid \mathcal{Y}_k] &= \sum_{i=1}^{N} e_i \overline{E}[\Lambda_k \langle X_m, e_i\rangle \mid \mathcal{Y}_k] \\ &= \sum_{i=1}^{N} \langle q_m, e_i\rangle \langle v_m, e_i\rangle e_i \\ &= \operatorname{diag}[q_m] v_m.\end{aligned} \tag{4.4}$$

□

The same argument shows that the following lemma holds.

Lemma 4.2. *The process v satisfies the backward dynamics*

$$v_m = A^* \operatorname{diag}[\underline{c}(Y_{m+1}, Y_m)] v_{m+1}; \qquad v_k = (1,\dots,1) \in \mathbb{R}^N. \tag{4.5}$$

Here A^* is the matrix transpose of A.

4.1. Recursive smoother for the number of jumps. The number of jumps from state e_r to state e_s in time k is given by

$$\mathcal{J}_k^{rs} = \sum_{\ell=1}^{k} \langle X_{\ell-1}, e_r\rangle \langle X_\ell, e_s\rangle. \tag{4.6}$$

Theorem 4.3. *Write $\sigma(\mathcal{J}_k^{rs}) = \overline{E}[\Lambda_k \mathcal{J}_k^{rs} \mid \mathcal{Y}_k]$.*

$$\sigma(\mathcal{J}_k^{rs}) = \sum_{\ell=1}^{k} a_{sr} \langle q_{\ell-1}, e_r\rangle \langle v_{\ell-1}, e_r\rangle. \tag{4.7}$$

Proof.

$$\begin{aligned}\overline{E}[\Lambda_k \mathcal{J}_k^{rs} \mid \mathcal{Y}_k] &= \sum_{\ell=1}^{k} \overline{E}[\langle X_{\ell-1}, e_r\rangle \langle X_\ell, e_s\rangle \Lambda_k \mid \mathcal{Y}_k] \\ &= \sum_{\ell=1}^{k} \overline{E}[\langle X_{\ell-1}, e_r\rangle \langle A X_{\ell-1}, e_s\rangle \Lambda_k \mid \mathcal{Y}_k] \\ &= \sum_{\ell=1}^{k} a_{sr} \overline{E}[\langle X_{\ell-1}, e_r\rangle \Lambda_k \mid \mathcal{Y}_k] \\ &= a_{sr} \sum_{\ell=1}^{k} \overline{E}[\Lambda_{\ell-1} \langle X_{\ell-1}, e_r\rangle \overline{E}[\Lambda_{\ell,k} \mid \mathcal{Y}_k \vee \{X_{\ell-1} = e_r\}] \mid \mathcal{Y}_k] \\ &= a_{sr} \sum_{\ell=1}^{k} \langle q_{\ell-1}, e_r\rangle \langle v_{\ell-1}, e_r\rangle,\end{aligned} \tag{4.8}$$

which finishes the proof. □

Lemma 4.4.

$$\sigma(\mathcal{J}_{k+1}^{rs}) = \Gamma_k' A^* \operatorname{diag}[\underline{c}(Y_{k+1}, Y_k)] \cdot \underline{1} + a_{sr}\langle q_k, e_r\rangle, \tag{4.9}$$

where $\underline{1} = (1,\dots,1) \in \mathbb{R}^N$ *and*

$$\Gamma_k' \triangleq a_{sr} \sum_{\ell=1}^{k} \langle q_{\ell-1}, e_r\rangle e_r' A^* \operatorname{diag}[\underline{c}(Y_\ell, Y_{\ell-1})] \cdots A^* \operatorname{diag}[\underline{c}(Y_{k-1}, Y_k)]. \tag{4.10}$$

Proof. Using the backward recursion (4.5),

$$\begin{aligned}\sigma(\mathcal{J}_k^{rs}) &= a_{sr} \sum_{\ell=1}^{k} \langle q_{\ell-1}, e_r\rangle e_r' A^* \operatorname{diag}[\underline{c}(Y_\ell, Y_{\ell-1})] \cdots A^* \operatorname{diag}[\underline{c}(Y_{k+1}, Y_k)] \cdot \underline{1} \\ &= \Gamma_k' \cdot \underline{1}.\end{aligned} \tag{4.11}$$

Also note that

$$\Gamma_{k+1}' = \Gamma_k' A^* \operatorname{diag}[\underline{c}(Y_{k+1}, Y_k)] + a_{sr}\langle q_k, e_r\rangle e_r. \tag{4.12}$$

Therefore,

$$\sigma(\mathcal{J}_{k+1}^{rs}) = \Gamma_{k+1}' \cdot \underline{1} = \Gamma_k' A^* \operatorname{diag}[\underline{c}(Y_{k+1}, Y_k)] \cdot \underline{1} + a_{sr}\langle q_k, e_r\rangle, \tag{4.13}$$

and the result follows. □

4.2. Recursive smoother for the occupation time. The number of occasions up to time k for which the Markov chain X has been in state e_r, $1 \le r \le N$, is

$$\mathbb{O}_k^r = \sum_{\ell=0}^{k} \langle X_\ell, e_r\rangle. \tag{4.14}$$

Lemma 4.5. *Write* $\sigma(\mathbb{O}_k^r) = \overline{E}[\Lambda_k \mathbb{O}_k^r \mid \mathcal{Y}_k]$.

$$\sigma(\mathbb{O}_k^r) = \sum_{\ell=0}^{k} \langle q_\ell, e_r\rangle \langle v_\ell, e_r\rangle, \tag{4.15}$$

$$\sigma(\mathbb{O}_{k+1}^r) = \Sigma_k' A^* \operatorname{diag}[\underline{c}(Y_{k+1}, Y_k)] \cdot \underline{1} + \langle q_{k+1}, e_r\rangle, \tag{4.16}$$

where

$$\Sigma_k' \triangleq \sum_{\ell=1}^{k} \langle q_\ell, e_r\rangle e_r' A^* \operatorname{diag}[\underline{c}(Y_\ell, Y_{\ell-1})] \cdots A^* \operatorname{diag}[\underline{c}(Y_{k-1}, Y_k)], \tag{4.17}$$

and

$$\Sigma_{k+1}' = \Sigma_k' A^* \operatorname{diag}[\underline{c}(Y_{k+1}, Y_k)] + \langle q_{k+1}, e_r\rangle e_r. \tag{4.18}$$

4.3. Recursive smoother for state-to-observation transitions. The parameters estimation of our model requires estimates and smoothers of the process

$$\mathcal{T}_k^{rs\mathrm{m}} = \sum_{\ell=1}^{k} \langle X_{\ell-1}, e_r \rangle \langle Y_{\ell-1}, e_s \rangle \langle Y_\ell, e_{\mathrm{m}} \rangle. \tag{4.19}$$

Lemma 4.6. *Write* $\sigma(\mathcal{T}_k^{rs\mathrm{m}}) = \overline{E}[\Lambda_k \mathcal{T}_k^{rs} \mid \mathcal{Y}_k]$.

$$\sigma(\mathcal{T}_k^{rs}) = \sum_{\ell=1}^{k} \langle q_{\ell-1}, e_r \rangle \langle v_{\ell-1}, e_r \rangle \langle Y_{\ell-1}, e_s \rangle \langle Y_\ell, e_{\mathrm{m}} \rangle, \tag{4.20}$$

$$\sigma(\mathcal{T}_{k+1}^{rs\mathrm{m}}) = \Phi_k' A^* \operatorname{diag}[\underline{c}(Y_{k+1})] \cdot \underline{1} + \langle q_k, e_r \rangle \langle Y_{k+1}, e_s \rangle, \tag{4.21}$$

where

$$\Phi_k' \triangleq \sum_{\ell=1}^{k} \langle q_{\ell-1}, e_r \rangle \langle Y_{\ell-1}, e_s \rangle \langle Y_\ell, e_{\mathrm{m}} \rangle e_r' A^* \operatorname{diag}[\underline{c}(Y_\ell)] \cdots A^* \operatorname{diag}[\underline{c}(Y_k)], \tag{4.22}$$

and

$$\Phi_{k+1}' = \Phi_k' A^* \operatorname{diag}[\underline{c}(Y_{k+1})] + \langle q_k, e_r \rangle \langle Y_{\ell-1}, e_s \rangle \langle Y_\ell, e_{\mathrm{m}} \rangle e_r. \tag{4.23}$$

References

[1] L. Aggoun and R. J. Elliott, *Measure Theory and Filtering: Introduction and Applications*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2004.

[2] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Applications of Mathematics (New York), vol. 29, Springer-Verlag, New York, 1995.

[3] R. J. Elliott and W. P. Malcolm, *Improved smoother dynamics for discrete-time HMM parameter estimation*, Proceeding of the 40th IEEE Conference on Decision and Control, Florida, 2001.

Lakhdar Aggoun: Department of Mathematics and Statistics, Sultan Qaboos University, P.O. Box 36, Al-Khodh 123, Muscat, Oman

E-mail address: laggoun@squ.edu.om