Communications on Stochastic Analysis
Vol. 9, No. 3 (2015) 367-374
Serials Publications
www.serialspublications.com

HIDDEN MARKOV CHANGE POINT ESTIMATION

ROBERT J. ELLIOTT* AND SEBASTIAN ELLIOTT

Abstract. A hidden Markov model is considered where the dynamics of the hidden process change at a random 'change point' . In principle this gives rise to a non-linear filter but closed form recursive estimates are obtained for the conditional distribution of the hidden process and of .

1. Introduction

Hidden Markov models have found many applications from speech processing to regime switching dynamics in financial models. An early description is given in the paper [4] by Rabiner and Huang and a fuller treatment can be found in the book "Hidden Markov Models: Estimation and Control" by Elliott, Aggoun and Moore. Recent results can be found in [3].

In this paper we consider the situation where a discrete time Markov chain X is observed indirectly through a second Markov chain Y. However, the dynamics of X undergo a change at a random time . Given the observed process Y, filtered recursive estimates for the conditional distribution of X and the change point are derived. This paper uses the expectation maximization, EM algorithm as discussed in [1] and [2].

2. Chain Dynamics

Suppose $X = \{X_k, k = 0, 1, ...\}$ is a discrete time finite state Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$. Without lose of generality, the state space can be taken to be the set of unit vectors $\{e_1, e_2, ...e_N\}$ in R^N, where N is the number of elements of the state space and $e_i = (0, 0, ..., 0, 1, 0, ..., 0)'$ is a standard basis vector in R^N.

Suppose $\tau \in \{1, 2, ...\}$ is a random time and

$$P(\tau = k) = p_k \geq 0.$$

Write $F_k = P(\tau > k) = \sum_{l=k+1}^{\infty} p_l$.

The random time τ represents a random time at which there is a change in the dynamics of the chain X.

Received 2015-8-27; Communicated by the editors.
2010 *Mathematics Subject Classification.* 60J10; 60J22.
Key words and phrases. Hidden Markov chain, change points, estimation.
* Professor Robert J. Elliott wishes to thank the ARO for support under project W911NF-15-1-0220.

In fact, if $k \leq \tau$, suppose

$$\begin{aligned} a^1_{ji} &= P(X_{k+1} = e_j | X_k = e_i) \\ &= P(X_1 = e_j | X_0 = e_i). \end{aligned}$$

Write A^1 for the matrix $(a^1_{ji}, 1 \leq i, j \leq N)$.
If $k > \tau$, suppose

$$a^2_{ji} = P(X_{k+1} = e_j | X_k = e_i),$$

and write A^2 for the matrix $(a^2_{ji}, 1 \leq i, j \leq N)$.

Write $\bar{Z}_k = I_{k \geq \tau}$.

The state space of $\bar{Z}$ will be mapped onto the unit vectors $\bar{e}_1 = (1, 0)$, $\bar{e}_2 = (0, 1)$ by considering the process

$$Z_k = (1 - \bar{Z}_k)\bar{e}_1 + \bar{Z}_k \bar{e}_2 \in R^2.$$

Define the filtrations

$$\begin{aligned} \mathcal{F}^X_k &= \sigma\{X_i, i \leq k\}, \\ \mathcal{F}^Z_k &= \sigma\{Z_i, i \leq k\}, \\ \mathcal{F}_k &= \mathcal{F}^X_k \vee \mathcal{F}^Z_k. \end{aligned}$$

Lemma 2.1. *Suppose* $A_k = A^1\langle Z_k, \bar{e}_1\rangle + A^2\langle Z_k, \bar{e}_2\rangle$, *then* $X_{k+1} = A_k X_k + M_{k+1} \in R^N$, *where* M *is a sequence of Martingale increments, that is* $E[M_{k+1} | \mathcal{F}_k] = 0 \in R^N$.

Proof.

$$\begin{aligned} E[M_{k+1} | \mathcal{F}_k] &= E[X_{k+1} - A_k X_k | \mathcal{F}^X_k \vee \mathcal{F}^Z_k] \\ &= E[X_{k+1} | X_k, Z_k] - (A^1\langle Z_k, \bar{e}_1\rangle + \\ &\quad A^2\langle Z_k, \bar{e}_2\rangle) X_k \\ &= E[X_{k+1}(\langle Z_k, \bar{e}_1\rangle + \langle Z_k, \bar{e}_2\rangle) | X_k, Z_k] - \\ &\quad (A^1\langle Z_k, \bar{e}_1\rangle + A^2\langle Z_k, \bar{e}_2\rangle) X_k \\ &= (A^1\langle Z_k, \bar{e}_1\rangle + A^2\langle Z_k, \bar{e}_2\rangle) X_k - \\ &\quad (A^1\langle Z_k, \bar{e}_1\rangle + A^2\langle Z_k, \bar{e}_2\rangle) X_k \\ &= 0. \end{aligned}$$

□

TRANSITIONS OF Z:
Note $P(Z_{k+1} = \bar{e}_1 | \mathcal{F}^Z_k) = P(Z_{k+1} = \bar{e}_1 | Z_k)$. Now

$$\begin{aligned} P(Z_{k+1} = \bar{e}_1 | Z_k = \bar{e}_1) &= P(\tau > k+1 | \tau > k) = \frac{F_{k+1}}{F_k}, \\ P(Z_{k+1} = \bar{e}_2 | Z_k = \bar{e}_1) &= P(\tau = k+1 | \tau > k) = \frac{p_{k+1}}{F_k}, \\ P(Z_{k+1} = \bar{e}_1 | Z_k = \bar{e}_2) &= P(\tau > k+1 | \tau \leq k) = 0, \\ P(Z_{k+1} = \bar{e}_2 | Z_k = \bar{e}_2) &= P(\tau \leq k+1 | \tau \leq k) = 1. \end{aligned}$$

Lemma 2.2. *Write* $F_k = \begin{pmatrix} \frac{F_{k+1}}{F_k} & 0 \\ \frac{p_{k+1}}{F_k} & 1 \end{pmatrix}$. *Then* $Z_{k+1} = F_k Z_k + N_{k+1} \in R^2$, *where* $E[N_{k+1}|\mathcal{F}_k^Z] = 0 \in R^2$.

3. Observations.

The chain X is not observed directly. Rather, there is another finite state process Y which is a 'noisy' observation of X.

Suppose the finite state space of Y is identified with the unit vectors where $\{f_1, f_2, ..., f_M\}$ of R^M, $f_j = (0, 0, ..., 1, ..., 0)' \in R^N$. We can have $M > N$, $M < N$ or $M = N$.

Suppose $c_{j,i} = P(Y_k = f_j | X_k = e_i) \geq 0$. Note $\sum_{j=1}^M c_{ji} = 1$.

Write $C = (c_{ji}), 1 \leq i \leq N, 1 \leq j \leq M, \mathcal{F}_k^Y = \sigma\{Y_j, 0 \leq j \leq k\}$.

Lemma 3.1. $Y_k = CX_k + L_k \in R^M$, *where* $E[L_k|\mathcal{F}_k^X] = 0 \in R^M$.

Proof.

$$\begin{aligned} E[L_k|\mathcal{F}_k^X] &= E[Y_k - CX_k|\mathcal{F}_k^X] \\ &= E[Y_k - CX_k|X_k] \\ &= E[Y_k|X_k] - CX_k \\ &= \sum_{j=1}^M \sum_{i=1}^N E[Y_k\langle Y_k, f_j\rangle\langle X_k, e_i\rangle|X_k] - CX_k \\ &= \sum_{j=1}^M \sum_{i=1}^N E[\langle Y_k, f_j\rangle\langle X_k, e_i\rangle|X_k] f_j - CX_k \\ &= \sum_{j=1}^M \sum_{i=1}^N \langle X_k, e_i\rangle c_{ji} f_j - CX_k \\ &= 0 \in R^M. \end{aligned}$$

□

JOINT DISTRIBUTIONS:

Lemma 3.2.

$$\begin{aligned} E[X_{k+1} \otimes Z_{k+1}|X_k, Z_k] &= A^1 X_k \otimes \begin{pmatrix} \frac{F_{k+1}}{F_k} \\ \frac{p_{k+1}}{F_k} \end{pmatrix} \langle Z_k, \bar{e}_1\rangle \\ &\quad + A^2 X_k \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \langle Z_k, \bar{e}_2\rangle. \end{aligned} \tag{3.1}$$

Proof.

$$
\begin{aligned}
&E[X_{k+1}\otimes Z_{k+1}|X_k,Z_k] \\
&= E[E[X_{k+1}\otimes Z_{k+1}|X_{k+1},X_k,Z_k]|X_k,Z_k] \\
&= E[X_{k+1}\otimes E[Z_{k+1}|Z_k]|X_k,Z_k] \\
&= E[X_{k+1}\otimes E[Z_{k+1}(\langle Z_k,\bar{e}_1\rangle+\langle Z_k,\bar{e}_2\rangle)|Z_k]|X_k,Z_k] \\
&= E[X_{k+1}\otimes \begin{pmatrix}\frac{F_{k+1}}{F_k}\\ \frac{p_{k+1}}{F_k}\end{pmatrix}\langle Z_k,\bar{e}_1\rangle|X_k,Z_k] \\
&\quad + E[X_{k+1}\otimes\begin{pmatrix}0\\1\end{pmatrix}\langle Z_k,\bar{e}_2\rangle|X_k,Z_k] \\
&= A^1X_k\otimes\begin{pmatrix}\frac{F_{k+1}}{F_k}\\ \frac{p_{k+1}}{F_k}\end{pmatrix}\langle Z_k,\bar{e}_1\rangle + A^2X_k\otimes\begin{pmatrix}0\\1\end{pmatrix}\langle Z_k,\bar{e}_2\rangle.
\end{aligned}
\tag{3.2}
$$

□

4. Reference Probability.

The above dynamics of X, Z, Y are under the 'real world' probability P. Suppose we have another probability $\bar{P}$ under which X and Z have the same dynamics as above, but under which Y is a process independent of X and Z and under which Y is a sequence of independent, identity distributed random variables with

$$\bar{P}(Y_k=f_j|\mathcal{F}_k^X\vee\mathcal{F}_k^Z\vee\mathcal{F}_{k-1}^Y)=\bar{P}(Y_k=f_j)=1/M.$$

Lemma 4.1. *Let*

$$\lambda_k = M\sum_{j=1}^{M}\langle CX_k,f_j\rangle\langle Y_k,f_j\rangle, \tag{4.1}$$

$$\Lambda_k = \prod_{i=0}^{k}\lambda_i. \tag{4.2}$$

The probability P can be defined by setting $\frac{dP}{d\bar{P}}|_{\mathcal{G}_k}=\Lambda_k$, where $\mathcal{G}_k=\mathcal{F}_k^X\vee\mathcal{F}_k^Y\vee\mathcal{F}_k^Z$. Then under P the dynamics of X and Z are unchanged and

$$P(Y_k=f_j|X_k=e_i)=c_{ji}.$$

That is, under P

$$Y_k=CX_k+L_k,$$

where $E[L_k|\mathcal{F}_k^X]=0\in R^M$.

Proof. Consider first

$$\begin{aligned}\bar{E}[\lambda_k|\mathcal{G}_{k-1}\vee X_k] &= M\sum_{j=1}^{M}\bar{E}[\langle CX_k, f_j\rangle\langle Y_k, f_j\rangle|X_k]\\ &= M\sum_{j=1}^{M}\langle CX_k, f_j\rangle\bar{E}[\langle Y_k, f_j\rangle]\\ &= \sum_{j=1}^{M}\sum_{i=1}^{N}c_{ji}\langle X_k, e_i\rangle\\ &= 1.\end{aligned}$$

From Bayes' theorem, (see [1]), with $\frac{dP}{d\bar{P}}|_{\mathcal{G}_k} = \Lambda_k$,

$$\begin{aligned}&E[\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee X_k]\\ &= \frac{\bar{E}[\Lambda_k\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee X_k]}{\bar{E}[\Lambda_k|\mathcal{G}_{k-1}\vee X_k]}\\ &= \frac{\Lambda_{k-1}\bar{E}[\lambda_k\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee X_k]}{\Lambda_{k-1}\bar{E}[\lambda_k|\mathcal{G}_{k-1}\vee X_k]}\\ &= M\bar{E}[(\sum_{r=1}^{M}\langle CX_k, f_r\rangle\langle Y_k, f_r\rangle)\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee X_k]\\ &= M\bar{E}[\langle CX_k, f_j\rangle\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee X_k]\\ &= \langle CX_k, f_j\rangle.\end{aligned}$$

Therefore, if $X_k = e_i$,

$$P(Y_k = f_j|X_k = e_i) = E[\langle Y_k, f_j\rangle|X_k = e_i] = c_{ji}.$$

□

5. Filters.

We wish to estimate both X and Z, given the noisy observations Y. This is required under the 'real world' probability P. That is, we wish to determine $E[X_k\otimes Z_k|\mathcal{F}_k^Y]\in R^{N\times 2}$.

However, it is easier to work under the 'reference probability' $\bar{P}$ for which the dynamics of X and Z remain unchanged but the Y sequence is i.i.d. with $P(Y_k = f_j) = 1/M$ for all k and j.

Using Bayes' rule,

$$E[X_k\otimes Z_k|\mathcal{F}_k^Y] = \frac{\bar{E}[\Lambda_k X_k\otimes Z_k|\mathcal{F}_k^Y]}{\bar{E}[\Lambda_k|\mathcal{F}_k^Y]}.$$

Write

$$\begin{aligned}q_k &:= \bar{E}[\Lambda_k X_k\otimes Z_k|\mathcal{F}_k^Y]\in R^{N\times 2},\\ q_k^1 &:= \bar{E}[\Lambda_k X_k\langle Z_k, \bar{e}_1\rangle|\mathcal{F}_k^Y],\\ q_k^2 &:= \bar{E}[\Lambda_k X_k\langle Z_k, \bar{e}_2\rangle|\mathcal{F}_k^Y].\end{aligned}$$

for unnormalized conditional expectations given the observations $\mathcal{F}_k^Y$ to time k. Then

$$q_k = (q_k^1, q_k^2).$$

Note that, with $\mathbf{1}$ denoting the vector of $1's$ in R^N or R^2:

$$\mathbf{1}' \cdot X_k \otimes Z_k \cdot \mathbf{1} = 1.$$

Therefore,

$$\mathbf{1}' \cdot q_k \cdot \mathbf{1} = \bar{E}[\mathbf{1}' \cdot \Lambda_k X_k \otimes Z_k \cdot \mathbf{1} | \mathcal{F}_k^Y] = \bar{E}[\Lambda_k | \mathcal{F}_k^Y].$$

Consequently, if we know q_k, the denominator $\bar{E}[\Lambda_k | \mathcal{F}_k^Y]$ is just the sum of all components of q_k.

The filter is now a recursive estimate for q_k given the observations Y.

Theorem 5.1.

$$q_{k+1}^1 = M \sum_{j=1}^{M} \sum_{i=1}^{N} \langle A^1 q_k^1, e_i \rangle \frac{F_{k+1}}{F_k} c_{ji} e_i \langle Y_{k+1}, f_j \rangle, r \tag{5.1}$$

$$q_{k+1}^2 = M \sum_{j=1}^{M} \sum_{i=1}^{N} [\langle A^1 q_k^1, e_i \rangle \frac{p_{k+1}}{F_k} + \langle A^2 q_k^2, e_i \rangle] c_{ji} e_i \langle Y_{k+1}, f_j \rangle. \tag{5.2}$$

Proof.

$$\begin{aligned}
& q_{k+1}^1 \\
= \; & \bar{E}[\Lambda_k \lambda_{k+1} X_{k+1} \langle Z_{k+1}, \bar{e}_1 \rangle | \mathcal{F}_{k+1}^Y] \\
= \; & M \sum_{j=1}^{M} \bar{E}[\Lambda_k \langle C X_{k+1}, f_j \rangle X_{k+1} \langle Z_{k+1}, \bar{e}_1 \rangle | \mathcal{F}_{k+1}^Y] \langle Y_{k+1}, f_j \rangle \\
= \; & M \sum_{j=1}^{M} \sum_{i=1}^{N} \bar{E}[\Lambda_k \langle X_{k+1}, e_i \rangle \langle Z_{k+1}, \bar{e}_1 \rangle | \mathcal{F}_{k+1}^Y] c_{ji} e_i \langle Y_{k+1}, f_j \rangle \\
= \; & M \sum_{j=1}^{M} \sum_{i=1}^{N} \sum_{r=1}^{2} \bar{E}[\Lambda_k \langle X_{k+1}, e_i \rangle \langle Z_{k+1}, \bar{e}_1 \rangle \langle Z_k, \bar{e}_r \rangle | \mathcal{F}_{k+1}^Y] \\
& \times c_{ji} e_i \langle Y_{k+1}, f_j \rangle \\
= \; & M \sum_{j=1}^{M} \sum_{i=1}^{N} \langle A^1 q_k^1, e_i \rangle \frac{F_{k+1}}{F_k} c_{ji} e_i \langle Y_{k+1}, f_j \rangle.
\end{aligned}$$

For q^2:

$$\begin{aligned}
&q^2_{k+1}\\
&= \bar{E}[\Lambda_k \lambda_{k+1} X_{k+1} \langle Z_{k+1}, \bar{e}_2\rangle | \mathcal{F}^Y_{k+1}]\\
&= M \sum_{j=1}^{M} \bar{E}[\Lambda_k \langle C X_{k+1}, f_j\rangle X_{k+1} \langle Z_{k+1}, \bar{e}_2\rangle | \mathcal{F}^Y_{k+1}] \langle Y_{k+1}, f_j\rangle\\
&= M \sum_{j=1}^{M} \sum_{i=1}^{N} \bar{E}[\Lambda_k \langle X_{k+1}, e_i\rangle \langle Z_{k+1}, \bar{e}_2\rangle | \mathcal{F}^Y_{k+1}] c_{ji} e_i \langle Y_{k+1}, f_j\rangle\\
&= M \sum_{j=1}^{M} \sum_{i=1}^{N} \sum_{r=1}^{2} \bar{E}[\Lambda_k \langle X_{k+1}, e_i\rangle \langle Z_{k+1}, \bar{e}_2\rangle \langle Z_k, \bar{e}_r\rangle | \mathcal{F}^Y_{k+1}] \cdot\\
&\quad c_{ji} e_i \langle Y_{k+1}, f_j\rangle\\
&= M \sum_{j=1}^{M} \sum_{i=1}^{N} [\langle A^1 q^1_k, e_i\rangle \frac{p_{k+1}}{F_k} c_{ji} e_i \langle Y_{k+1}, f_j\rangle +\\
&\quad \langle A^2 q^2_k, e_i\rangle c_{ji} e_i \langle Y_{k+1}, f_j\rangle].
\end{aligned}$$

□

Remark 5.2. The filter is initialized by assuming the change point τ has not occured, so $Z_0 = \bar{e}_1$, and by taking an initial distribution $p_0 = q_0$ for X_0.

Corollary 5.3. *The normalized conditional distributions of X_k and Z_k are then given by*

$$p^1_k = E[X_k \langle Z_k, \bar{e}_1\rangle | \mathcal{F}^Y_k] = \frac{q^1_k}{\mathbf{1}' \cdot q_k \cdot \mathbf{1}}, \tag{5.3}$$

$$p^2_k = E[X_k \langle Z_k, \bar{e}_2\rangle | \mathcal{F}^Y_k] = \frac{q^2_k}{\mathbf{1}' \cdot q_k \cdot \mathbf{1}}. \tag{5.4}$$

Given the observations, the conditional probability of the change point τ is then

$$P(\tau > k) = E[\langle Z_k, \bar{e}_1\rangle | \mathcal{F}^Y_k] = \mathbf{1}' \cdot q^1_k. \tag{5.5}$$

Remark 5.4. Recall M is the number of elements in the state space of the observation process Y. The right hand sides of the recursions (5.5) and (5.6) of Theorem 5.1 involve a factor M. However, this will cancel when we consider the normalized forms in (5.7) and (5.8). Consequently, equation (5.5) and (5.6) can be modified to give recursions for unnormalized quantities $\bar{q}^1_k$, $\bar{q}^2_k$ as:

$$\bar{q}^1_{k+1} = \sum_{j=1}^{M} \sum_{i=1}^{N} \langle A^1 \bar{q}^1_k, e_i\rangle \frac{F_{k+1}}{F_k} c_{ji} e_i \langle Y_{k+1}, f_j\rangle, \tag{5.6}$$

$$\bar{q}^2_{k+1} = \sum_{j=1}^{M} \sum_{i=1}^{N} [\langle A^1 \bar{q}^1_k, e_i\rangle \frac{p_{k+1}}{F_k} + \langle A^2 \bar{q}^2_k, e_i\rangle] c_{ji} e_i \langle Y_{k+1}, f_j\rangle. \tag{5.7}$$

Again the initialization is $Z_0 = \bar{e}_1$ and $\bar{q}_0 = p_0 \in R^N$.

6. A Viterbi Recursion.

As noted in our earlier papers, the Viterbi filter is given by replacing the expected value by a maximum likelihood. That is, the Viterbi estimation is given by a sequence of vectors

$$\begin{aligned} q_k^{*1} &= [q_k^{*1}(1), q_k^{*1}(2), ..., q_k^{*1}(N)]', \\ q_k^{*2} &= [q_k^{*2}(1), q_k^{*2}(2), ..., q_k^{*2}(N)]'. \end{aligned}$$

defined recursively by $Z_0 = \bar{e}_1$, $q_0^{*1} = p_0$, $q_0^{*2} = \mathbf{0} \in R^N$, and

$$q_{k+1}^{*1}(i) := \langle A^1 q_k^{*1}, e_i \rangle \frac{F_{k+1}}{F_k} \max_j (c_{ji} \langle Y_{k+1}, f_j \rangle), \tag{6.1}$$

$$q_{k+1}^{*2}(i) := [\langle A^1 q_k^{*1}, e_i \rangle \frac{p_{k+1}}{F_k} + \langle A^2 q_k^{*2}, e_i \rangle] \max_j (c_{ji} \langle Y_{k+1}, f_j \rangle). \tag{6.2}$$

7. Conclusion.

Hidden Markov chains, that is Markov chains observed indirectly through the observations of a second Markov chain, have been extensively studied. For a bibliography see the book [1] by Aggoun, Elliott and Moore.

In this paper we have considered a hidden Markov chain X whose dynamics undergo a change at a random time τ. Given an observed process filtered estimates for the conditional distribution of X and the change point time τ are derived.

References

1. Aggoun, L., Elliott, R. J. and Moore, J. B.: *Hidden Markov models: estimation and control*, Springer - Verlag, erlin Heidelberg - New York, 1997.
2. Dembo, A. and Zeitouni, O.: Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm, *Stoch. Proc. Appl.* **23** (1986), no. 1, 91–113.
3. Elliott, R. J. and Malcolm, W. P.: Data- recursive smoother formulae for paritally observed discrete-time Markov chains, *Stoch. Anal. Appl.* **24** (2006), no. 3, 579–597.
4. Rabiner, L. R. and Juang, B. H.: An introduction to Hidden Markov Models, *IEEE ASSP Magazine*, (January 1986), 4–16.

Robert J. Elliott: Haskayne School of Busiless, University of Calgary, Calgary, AB T2N1N4, Canada

E-mail address: relliott@ucalgary.edu

Sebastian Elliott: Elliott Stochastics LLC. 31 Varsity Estates Park NW, Calgary, AB Canada

E-mail address: ellio02@gmail.com