Journal of Stochastic Analysis

Volume 2 | Number 4 Article 3

December 2021

Recursive and Viterbi Estimation for Semi-Markov Chains

Robert J. Elliott
University of South Australia, Campus Central - City West, GPO Box 2471, Robert.Elliott@unisa.edu.au

W. P. Malcolm
Defence Science and Technology Group Australia, Russell Offices Canberra, Constitution Ave, Russell ACT 2600, Australia, Paul.Malcolm2@defence.gov.au

Journal of Stochastic Analysis
Vol. 2, No. 4 (2021) Article 3 (16 pages)
DOI: 10.31390/josa.2.4.03

RECURSIVE AND VITERBI ESTIMATION FOR SEMI-MARKOV CHAINS

ROBERT ELLIOTT* AND W. P. MALCOLM

Abstract. Recursive and Viterbi filters and smoothers are found for a semi-Markov chain observed in Gaussian noise. As is well known, the sojourn times of all states in first-order time-homogeneous Markov chain are necessarily geometrically distributed. A semi-Markov chain can have one, more than one, or all state sojourns distributed by general probability distributions. In this article the filters and smoothers presented are all exact, that is, they are not based upon any approximations to the semi-Markov chain dynamics.

1. Introduction

In this paper we consider the estimation of a partially observed finite-state discrete-time semi-Markov process $X = \{X_k\}_{k\geq 0}$. Such models are often referred to as a Hidden Semi-Markov Model (HsMM). We suppose X is observed in Gaussian noise. Exact filters and smoothers are obtained for X and its related quantities. Our exact estimation schemes are not based upon any approximations to the dynamics of a semi-Markov process. Exact schemes are important for various reasons. Firstly, if an exact estimation scheme can be written down in an implementable closed form, then it should be investigated as a matter of course. Exact estimation schemes also provide a valuable benchmark against which any approximate schemes may be measured in respect of performance, or estimator properties. Viterbi filters and smoothers are also obtained for a HsMM. Recursive estimation schemes are given for model parameter estimation.
The two main contributions of this note are

(1) explicit recursive formulae for filters and smoothers.
(2) Viterbi versions of these formulae.

Any discrete-time first-order time-homogeneous Markov chain (by definition) has geometrically distributed sojourns for each of its states. Semi-Markov chains can have general distributions for their state-sojourn times. Their estimation in filtering, smoothing and Viterbi forms is an important problem.

Some related expressions can be found in [9]. The results below are more direct and clearer. Earlier references on semi-Markov processes include the books by Koski [7], Barbu and Limnios [2], and van der Hoek and Elliott [9]. References

Received 2021-7-10; Accepted 2021-8-29; Communicated by the editors.
2010 *Mathematics Subject Classification.* 60G35; 60K15; 62L12.
Key words and phrases. Discrete-time, semi-Markov models, Viterbi algorithm, filters, smoothers, martingales, change of probability measure.
* Corresponding author.

on filtering include Yu [10], Krishnamurthy, Moore and Chung [8] and Elliott, Limnios and Swishchuk [5] and Elliott and Malcolm [6].

2. Stochastic Dynamics

To establish the framework for our results we must repeat some definitions of [6]. All processes are defined on a probability space $(\Omega, \mathcal{F}, P)$.

Our process of interest is a finite-state discrete-time, stochastic process $X = \{X_k\}_{k\geq 0}$ with arbitrary state sojourn distributions.

Without loss of generality the state space for the process X can be identified with the set of unit vectors

$$S \coloneqq \{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_N\}$$

$$\boldsymbol{e}_i \coloneqq (0, \dots, 0, 1, 0, \dots, 0)' \in \mathbb{R}^N.$$

We also write $m \in \{1, 2, 3, \dots\}$ for state sojourns.

Notation 2.1. The initial state $X_0 \in S$, is taken as given, or, its probability distribution $p_0 = (p_0^1, p_0^2, \dots, p_0^N)' \in \mathbb{R}^N$, is known. The processes we are interested in will change state at discrete random times τ_n. State transitions at these times are of the type $\boldsymbol{e}_i \to \boldsymbol{e}_j$, with $i \neq j$. We set $\tau_0 \coloneqq 0$. Successive jump event times form a strictly increasing sequence $\tau_0 < \tau_1 < \tau_2 < \tau_3 \dots$. Write $\mathcal{F}_k \coloneqq \sigma\{X_u\, , u \leq k\}$ and $\mathcal{F} = \{\mathcal{F}_u\}_{u\geq 0}$ for the filtration generated by X.

We now define a time-homogeneous semi-Markov chain.

Definition 2.2. A stochastic process X is a time-homogeneous semi-Markov process if

$$\begin{aligned} &P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j,\, \tau_{n+1} - \tau_n = m \mid \mathcal{F}_{\tau_n}\big) \\ &\qquad = P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j,\, \tau_{n+1} - \tau_n = m \mid X_{\tau_n} = \boldsymbol{e}_i\big). \end{aligned} \tag{2.1}$$

This can be factorized as

$$\begin{aligned} q\big(\boldsymbol{e}_j, \boldsymbol{e}_i, m\big) &\coloneqq P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j,\, \tau_{n+1} - \tau_n = m \mid X_{\tau_n} = \boldsymbol{e}_i\big) \\ &= P\big(\tau_{n+1} - \tau_n = m \mid X_{\tau_n} = \boldsymbol{e}_i\big) P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j \mid \tau_{n+1} - \tau_n = m, X_{\tau_n} = \boldsymbol{e}_i\big) \\ &= \pi_i(m) p_{j,i}(m), \text{ say.} \end{aligned}$$

Here

$$\begin{aligned} \pi_i(m) &\coloneqq P\big(\tau_{n+1} - \tau_n = m \mid X_{\tau_n} = \boldsymbol{e}_i\big) \quad \text{and} \\ p_{j,i}(m) &\coloneqq P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j \mid \tau_{n+1} - \tau_n = m, X_{\tau_n} = \boldsymbol{e}_i\big). \end{aligned}$$

Notation 2.3. Write

$$\begin{aligned} G_i(k) &\coloneqq P\big(\tau_{n+1} - \tau_n \leq k \mid X_{\tau_n} = \boldsymbol{e}_i\big) = \sum_{m=1}^{k} \pi_i(m), \\ F_i(k) &\coloneqq P\big(\tau_{n+1} - \tau_n > k \mid X_{\tau_n} = \boldsymbol{e}_i\big) = 1 - G_i(k). \end{aligned}$$

Then

$$\begin{aligned} P\big(\tau_{n+1} = \tau_n + k \mid X_{\tau_n+k-1} = X_{\tau_n} = \boldsymbol{e}_i\big) \\ &= P\big(\tau_{n+1} = \tau_n + k \mid \tau_{n+1} > \tau_n + k - 1, X_{\tau_n} = \boldsymbol{e}_i\big) \\ &= \frac{\pi_i(k)}{F_i(k-1)} \,. \end{aligned}$$

Write $\Delta^i(k) \coloneqq \dfrac{\pi_i(k)}{F_i(k-1)}$.

Definition 2.4. For each index i, $1 \le i \le N$, we define the recursive process h^i_k by

$$\begin{aligned} h^i_0 &\coloneqq \big\langle X_0, \boldsymbol{e}_i \big\rangle \in \{0,1\}. \\ h^i_k &\coloneqq \big\langle X_k, \boldsymbol{e}_i \big\rangle + \langle X_k, \boldsymbol{e}_i \rangle \big\langle X_k, X_{k-1} \big\rangle \, h^i_{k-1}, \qquad k = 1, 2, \dots . \end{aligned}$$

The h^i processes are non-zero only at times when $X = \boldsymbol{e}_i$. The process h^i returns the cumulative time spent in state $\boldsymbol{e}_i$. If

$$h_k = \sum_{i=1}^{N} h^i_k$$

then $h_0 = 1$ and

$$h_k = 1 + \langle X_k, X_{k-1} \rangle h_{k-1}.$$

The process h_k measures the amount of time since the last transition event. This process is never zero.

2.1. Transition-Event Probabilities.

Lemma 2.5. *Suppose $i \ne j$, $\;1 \le i,\, j \le N$. Then*

$$P\big(X_{k+1} = \boldsymbol{e}_j \mid X_k = \boldsymbol{e}_i,\, h_k\big) = p_{j,i}\big(h^i_k\big)\Delta^i\big(h^i_k\big).$$

Proof. The conditioning event fixes $X_k = \boldsymbol{e}_i$, so $h_k = h^i_k$. Also $\tau_n = k - h_k + 1$, so $X_k = X_{\tau_n} = \boldsymbol{e}_i$. Further, if a state transition occurs at time $k+1$, then $\tau_{n+1} = k + 1$. Consequently

$$\begin{aligned} P\big(X_{k+1} = \boldsymbol{e}_j \mid X_k = \boldsymbol{e}_i,\, h^i_k\big) &= P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j \mid X_{\tau_n} = e_i \text{ and} \tau_{n+1} = \tau_n + h^i_k\big) \\ &= p_{j,i}\big(h^i_k\big)\Delta^i\big(h_k\big). \end{aligned}$$

□

Remark 2.6. We are assuming there is a jump from $\boldsymbol{e}_i$ to a different $\boldsymbol{e}_j$, $i \ne j$, at time $k+1$. So, $\displaystyle\sum_{\substack{j=1 \\ j \ne i}}^{N} p_{j,i}(k+1) = 1$.

Corollary 2.7. *Under the same hypotheses,*

$$
\begin{aligned}
P\big(X_{k+1} = \boldsymbol{e}_i \mid X_k = e_i, h_k^i\big) &= 1 - \Delta^i\big(h_k^i\big) \\
&= 1 - \Big(\Delta^i\big(h_k^i\big) \sum_{\substack{j=1 \\ j\neq i}}^{N} p_{j,i}\big(h_k^i\big)\Big) \\
&= 1 - \sum_{\substack{j=1 \\ j\neq i}}^{N} \big(p_{j,i}(h_k^i)\Delta^i(h_k^i)\big)\,.
\end{aligned}
$$

Notation 2.8. For $k = 1, 2, \dots,$ write $A(k)$ for the $N \times N$ matrix with entries

$$
\begin{aligned}
a_{i,i}(k) &= 1 - \Delta^i(k) \\
a_{j,i}(k) &= p_{j,i}(k)\Delta^i(k).
\end{aligned}
$$

Example 2.9. Then for $N = 3$

$$
A(k) = \begin{bmatrix} a_{1,1}(k) & a_{1,2}(k) & a_{1,3}(k) \\ a_{2,1}(k) & a_{2,2}(k) & a_{2,3}(k) \\ a_{3,1}(k) & a_{3,2}(k) & a_{2,3}(k) \end{bmatrix}
$$

$$
= \begin{bmatrix} 1 - \Delta^1(k) & p_{1,2}(k)\Delta^2(k) & p_{1,3}(k)\Delta^3(k) \\ p_{2,1}(k)\Delta^1(k) & 1 - \Delta^2(k) & p_{2,3}(k)\Delta^3(k) \\ p_{3,1}(k)\Delta^1(k) & p_{3,2}(k)\Delta^2(k) & 1 - \Delta^3(k) \end{bmatrix}.
$$

Notation 2.10. Define the matrices:

$$
\Pi(k) := \big(p_{i,j}(k),\, 1 \le i,\, j \le N\big)
$$

where $p_{i,i}(k) = -1$ and $p_{j,i}(k) = P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j \mid \tau_{n+1} - \tau_n = k,\ X_{\tau_n} = \boldsymbol{e}_i\big)$, for $i \neq j$. Write

$$
D(k) := \operatorname{diag}\big(\Delta^1(k), \Delta^2(k), \dots, \Delta^N(k)\big).
$$

Then

$$
A(k) = I + \Pi(k)D(k),
$$

where I is the $N \times N$ identity matrix.

For the case when $N = 3$

$$
\Pi(k) = \begin{bmatrix} -1 & p_{1,2}(k) & p_{1,3}(k) \\ p_{2,1}(k) & -1 & p_{2,3}(k) \\ p_{3,1}(k) & p_{3,2}(k) & -1 \end{bmatrix} \qquad D(k) = \begin{bmatrix} \Delta^1(k) & 0 & 0 \\ 0 & \Delta^2(k) & 0 \\ 0 & 0 & \Delta^3(k) \end{bmatrix}.
$$

This decomposition nicely separates the probabilities of when the jump occurs and where it goes.

A key result is the following representation of the semi-Markov chain X.

Theorem 2.11. *The semi-Markov chain X has the following semi-martingale dynamics:*

$$X_{k+1} = A(h_k)X_k + M_{k+1} \in \mathbb{R}^N.$$

Here M_{k+1} is a martingale increment:

$$E\big[M_{k+1} \mid X_k, h_k\big] = \mathbf{0} \in \mathbb{R}^N.$$

Proof. We need only remark that we can write h_k in $A(h_k)$ rather than the different occupation times $h_k^1, \dots, h_k^N$. This is because the components of X_k are in effect indicator functions. If $X_k = \boldsymbol{e}_i$ the product $A(h_k)\boldsymbol{e}_i$ selects the i^{th} column of $A_k(h_k)$ and the h_k will be that for h_k^i. □

3. An Exact Recursive Filter for the State

In this section, we recall the result of Section 9 of [6]. It will then be extended.

We use the notation of Section 2. That is, X is a semi-Markov chain with finite state space $S = \big\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_N\big\}$, $\boldsymbol{e}_i \in \mathbb{R}^N$. The observation process is $y = \big\{y_0, y_1, \dots, y_k, \dots\big\}$ where $y_i \in \mathbb{R}$. The extension to vector observation processes is immediate. We suppose

$$y_k = c(X_k) + d(X_k)w_k$$

where $\{w_k,\, k = 0, 1, 2, \dots\}$ is a sequence of i.i.d. $N(0,1)$ random variables. As $X_k \in S$ there are vectors $\boldsymbol{c} = (c_1, c_2, \dots, c_N)' \in \mathbb{R}^N$, $\boldsymbol{d} = (d_1, d_2, \dots, d_N)' \in \mathbb{R}^N$ such that

$$c(X_k) = \big\langle \boldsymbol{c}, X_k \big\rangle \quad \text{and} \quad d(X_k) = \big\langle \boldsymbol{d}, X_k \big\rangle\,.$$

We suppose $d_i > 0$ for $i = 1, \dots, N$. We suppose there is a second 'reference' probability measure $\overline{P}$ under which X remains a semi-Markov chain with the same dynamics but $y = \{y_0, y_1, \dots\}$ is a sequence of i.i.d. $N(0,1)$ random variables.

We now construct the original probability P under which X is a semi-Markov chain with dynamics

$$X_{k+1} = A(h_k)X_k + M_{k+1} \in \mathbb{R}^N$$

and the process $w = \{w_0, w_1, \dots\}$ is a sequence of i.i.d $N(0,1)$ random variables where

$$w_k = \frac{y_k - \langle \boldsymbol{c}, X_k \rangle}{\langle \boldsymbol{d}, X_k \rangle}\,.$$

Definition 3.1. For $k = 0, 1, 2, \dots$ write

$$\lambda_k(y_k) = \frac{\phi\big((y_k - \langle \boldsymbol{c}, X_k \rangle)/\langle \boldsymbol{d}, X_k \rangle\big)}{\langle \boldsymbol{d}, X_k \rangle \phi(y_k)}$$

where $\phi(x)$ is the $N(0,1)$ density

$$\frac{1}{\sqrt{2\pi}}\ \exp\ \big(-x^2/2\big).$$

Set

$$\Lambda_{0,k} = \prod_{\ell - 0}^{k} \lambda_k\,.$$

Recall

$$\mathcal{F}_k = \sigma\big\{X_0, X_1, \ldots, X_k\big\}, \quad \mathcal{Y}_k = \sigma\big\{y_0, y_1, \ldots, y_k\big\}$$

and write

$$\mathcal{G}_k = \sigma\big\{X_0, \ldots, X_k, y_0, \ldots, y_k\big\}.$$

The related filtrations are

$$\{\mathcal{F}_k\},\ \{\mathcal{Y}_k\} \quad \text{and} \quad \{\mathcal{G}_k\}.$$

The original 'real world' probability is defined in terms of $\overline{P}$ by setting

$$\frac{dP}{d\overline{P}}\Big|_{\mathcal{G}_k} = \Lambda_{0,k}\,.$$

Lemma 3.2. *Under P, X is a semi-Markov chain with dynamics*

$$X_{k+1} = A(h_k)X_k + M_{k+1}$$

and $\{w_k, k = 0, 1, \ldots\}$ is a sequence of i.i.d. $N(0,1)$ *random variables where*

$$w_k = \big(y_k - \langle \boldsymbol{c}, X_k\rangle / \langle \boldsymbol{d}, X_k\rangle\big).$$

That, is under P,

$$y_k = \langle \boldsymbol{c}, X_k\rangle + \langle \boldsymbol{d}, X_k\rangle w_k.$$

For a proof see [3].

Remark 3.3. Suppose $F : \mathbb{N} \to \mathbb{R}$ is an arbitrary function.

We wish to determine for any $i \in \{1, 2, \ldots, N\}$

$$E\big[\big\langle X_k, \boldsymbol{e}_i\big\rangle F(h_k^i) \mid \mathcal{Y}_k\big].$$

Using Bayes' theorem as in [3] this is

$$E\big[\big\langle X_k, \boldsymbol{e}_i\big\rangle F(h_k^i) \mid \mathcal{Y}_k\big] = \frac{\overline{E}\big[\Lambda_k \big\langle X_k, \boldsymbol{e}_i\big\rangle F(h_k^i) \mid \mathcal{Y}_k\big]}{\overline{E}\big[\Lambda_k \mid \mathcal{Y}_k\big]}\,.$$

Suppose there are unnormalized probabilities $\gamma_k^i(m)$ so that

$$\overline{E}\big[\Lambda_k \big\langle X_k, \boldsymbol{e}_i\big\rangle F(h_k^i) \mid \mathcal{Y}_k\big] = \sum_{m=1}^{\infty} F(m)\gamma_k^i(m).$$

However, $h_k^i \le k+1$ so $\gamma_k^i(m) = 0$ for $m > k+1$ and the sum is only up to $k+1$.

Write

$$\lambda_k^i(y_k) = \frac{\phi\big((y_k - c_i)/d_i\big)}{d_i \phi(y_k)}\,.$$

We now obtain recursions for the γ.

Theorem 3.4. *For $m = 1$ and with A as in Notation* (2.8)

$$\gamma_k^i(1) = \lambda_k^i(y_k) \sum_{\substack{j=1 \\ j \ne i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\gamma_{k-1}^j(m)\,.$$

For $1 < m \le k+1$

$$\gamma_k^i(m) = \lambda_k^i(y_k) a_{i,i}(m-1)\gamma_{k-1}^i(m-1)\,.$$

Proof. Suppose $i \in \{1, 2, \dots, N\}$ and $F : \mathbb{N} \to \mathbb{R}$ is an arbitrary function. Then

$$\begin{aligned}
\overline{E}\Big[\Lambda_k \langle X_k, \boldsymbol{e}_i\rangle F(h^i_k) \mid \mathcal{Y}_k\Big] &= \sum_{m=1}^{k+1} F(m)\gamma^i_k(m) \\
&= \overline{E}\Big[\Lambda_{k-1}\lambda_k(y_k)\langle X_k, \boldsymbol{e}_i\rangle F\big(\langle X_k, \boldsymbol{e}_i\rangle + \langle X_k, \boldsymbol{e}_i\rangle \langle X_{k-1}, \boldsymbol{e}_i\rangle h^i_{k-1}\big) \mid \mathcal{Y}_k\Big] \\
&= \lambda^i_k(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_k, \boldsymbol{e}_i\rangle F\big(1 + \langle X_{k-1}, \boldsymbol{e}_i\rangle h^i_{k-1}\big) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda^i_k(y_k) \sum_{j=1}^{N} \overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle F\big(1 + \langle X_{k-1}, \boldsymbol{e}_i\rangle h^i_{k-1}\big) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda^i_k(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} \overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle F(1) \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad + \lambda^i_k(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle\langle X_k, \boldsymbol{e}_i\rangle F(1 + h^i_{k-1}) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda^i_k(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} \overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle \, \langle A\big(h^j_{k-1}\big)X_{k-1}, \boldsymbol{e}_i\rangle F(1) \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad + \lambda^i_k(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle \, \langle A\big(h^i_{k-1}\big)X_{k-1}, \boldsymbol{e}_i\rangle F\big(1 + h^i_{k-1}\big) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda^i_k(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} F(1)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle \, a_{i,j}\big(h^j_{k-1}\big) \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad\quad + \lambda^i_k(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle a_{i,i}\big(h^i_{k-1}\big)F\big(1 + h^i_{k-1}\big) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda^i_k(y_k)F(1) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\gamma^j_{k-1}(m) + \lambda^i_k(y_k) \sum_{m=1}^{k+1} a_{i,i}(m)F(1+m)\gamma^i_{k-1}(m) \\
&= \lambda^i_k(y_k)F(1) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\gamma^j_{k-1}(m) + \lambda^i_k(y_k) \sum_{m=2}^{k+1} a_{i,i}(m-1)F(m)\gamma^i_{k-1}(m-1) \\
&= \lambda^i_k(y_k)F(1) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\gamma^j_{k-1}(m) + \lambda^i_k(y_k) \sum_{m=2}^{k+1} a_{i,i}(m-1)F(m)\gamma^i_{k-1}(m-1).
\end{aligned}$$

F is an arbitrary function. First choose F so that $F(1) = 1$ and $F(m) = 0$ if $m \neq 1$. Then from (3.1)

$$\gamma^i_k(1) = \lambda^i_k(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\gamma^j_{k-1}(m)\,.$$

This is the recursion in k for $\gamma_k^i(1)$, the unnormalized conditional probability given $\mathcal{Y}_k$ that at time k $h_k^i(X_k) = 1$ and $X_k = \boldsymbol{e}_i$. Now choose F so that $F(m) = 1$ for some $m > 1$ and $F(m) = 0$ otherwise. Then from (3.1)

$$\gamma_k^i(m) = \lambda_k^i(y_k) a_{i,i}(m-1)\gamma_{k-1}^i(m-1).$$

This is the recursion in k for $\gamma_{k-1}^i(m)$, the unnormalized conditional probability given $\mathcal{Y}_k$ that, at time k, $h_k^i = m$ and $X_k = \boldsymbol{e}_i$. □

Remark 3.5. By definition

$$\overline{E}\big[\Lambda_k \big\langle X_k, \boldsymbol{e}_i \big\rangle F(h_k^i) \mid \mathcal{Y}_k\big] = \sum_{m=1}^{k+1} F(m)\, \gamma_k^i(m).$$

Taking $F(m) = 1$ for all m gives

$$\overline{E}\big[\Lambda_k \big\langle X_k, \boldsymbol{e}_i \big\rangle \mid \mathcal{Y}_k\big] = \sum_{m=1}^{k+1} \gamma_k^i(m).$$

As $\sum_{i=1}^N \big\langle X_k, \boldsymbol{e}_i \big\rangle = 1$, $\overline{E}\big[\Lambda_k \mid \mathcal{Y}_k\big] = \sum_{i=1}^N \sum_{m=1}^{k+1} \gamma_k^i(m)$,
so $E\big[\big\langle X_k, \boldsymbol{e}_i \big\rangle \mid \mathcal{Y}_k\big] = \dfrac{\sum_{m=1}^{k+1} \gamma_k^i(m)}{\sum_{i=1}^N \sum_{m=1}^{k+1} \gamma_k^i(m)}$.

Taking $F(m) = m$ for all m gives,

$$\overline{E}\big[\Lambda_k \big\langle X_k, \boldsymbol{e}_i \big\rangle h_k^i \mid \mathcal{Y}_k\big] = \sum_{m=1}^{k+1} m\, \gamma_k^i(m)$$

and

$$\overline{E}\big[\Lambda_k h_k^i \mid \mathcal{Y}_k\big] = \sum_{i=1}^N \sum_{m=1}^{k+1} m\, \gamma_k^i(m)$$

so

$$E\big[h_k^i \mid \mathcal{Y}_k\big] = \frac{\sum_{i=1}^N \sum_{m=1}^{k+1} m\, \gamma_k^i(m)}{\sum_{i=1}^N \sum_{m=1}^{k+1} \gamma_k^i(m)}.$$

4. Recursive Estimates

To estimate parameters of our model, we require recursive filters for quantities of the form:

$$N_k^{p,q} := \sum_{\ell=1}^{k} \big\langle X_{\ell-1}, \boldsymbol{e}_p \big\rangle \big\langle X_\ell, \boldsymbol{e}_q \big\rangle.$$

This gives the number of jumps from state $\boldsymbol{e}_p$ to $\boldsymbol{e}_q$ up to time k. The quantity

$$J_k^p := \sum_{\ell=1}^{k} \big\langle X_{\ell-1}, \boldsymbol{e}_p \big\rangle$$

gives the amount of time spent in state $\boldsymbol{e}_p$ up to time $k-1$. We also need estimates of sums of the form

$$G_k^p \coloneqq \sum_{\ell=1}^{k} f(y_\ell)\,\langle X_{\ell-1}, \boldsymbol{e}_p\rangle,$$

where f is a measurable function. As in [3] and [6], we first consider recursive estimates of the related $\mathbb{R}^N$-valued quantities

$$N_k^{p,q} X_k,\ J_k^p X_k \quad \text{and} \quad G_k^p X_k\,.$$

With $\mathbf{1} \coloneqq (1,1,\ldots,1)' \in \mathbb{R}^N$, for all k

$$\langle X_k, \mathbf{1}\rangle = \sum_{i=1}^{N} \langle X_k, e_i\rangle = 1.$$

Knowing, say, $E\big[N_k^{p,q} X_k \mid \mathcal{Y}_k\big] \in \mathbb{R}^N$ we immediately obtain

$$\big\langle E\big[N_k^{p,q} X_k \mid \mathcal{Y}_k\big], \mathbf{1}\big\rangle = E\big[N_k^{p,q}\,\langle X_k, \mathbf{1}\rangle \mid \mathcal{Y}_k\big] = E\big[N_k^{p,q} \mid \mathcal{Y}_k\big].$$

As in the previous section we have for any function $F : \mathbb{N} \to \mathbb{R}$ that

$$E\big[N_k^{p,q}\langle X_k, \boldsymbol{e}_i\rangle\, F(h_k^i) \mid \mathcal{Y}_k\big] = \frac{\overline{E}[\Lambda_k N_k^{p,q}\langle X_k, \boldsymbol{e}_i\rangle F(h_k^i) \mid \mathcal{Y}_k\big]}{\overline{E}\big[\Lambda_k \mid \mathcal{Y}_k\big]}.$$

As in [9], estimates for say $p_{i,j}$ are given by

$$\frac{E\big[N_k^{i,j} \mid \mathcal{Y}_k\big]}{E\big[J_k^i \mid \mathcal{Y}_k\big]} = \frac{\overline{E}\big[\Lambda_k N_k^{i,j} \mid \mathcal{Y}_k\big]}{\overline{E}\big[\Lambda_k J_k^i \mid \mathcal{Y}_k\big]}.$$

Suppose there are unnormalized probabilities $\delta_k^i(m,p,q)$ such that

$$\overline{E}\big[\Lambda_k N_k^{p,q}\langle X_k, \boldsymbol{e}_i\rangle F(h_k^i) \mid \mathcal{Y}_k\big] = \sum_{m=1}^{k+1} F(m)\delta_k^i(m,p,q).$$

We suppose for the moment that p, q are fixed and write $\delta_k^i(m,p,q)$ as $\delta_k^i(m)$. A recursion will be obtained for $\delta_k^i(m)$.

Theorem 4.1. *For $m = 1$ and $i \neq q$,* $\delta_k^i(1) = \lambda_k^i(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\delta_{k-1}^j(m).$

For $m = 1$ and $i = q$,

$$\delta_k^q(1) = \lambda_k^q(y_k)\Big[\sum_{\substack{j=1\\ j\neq p,\, j\neq q}}^{N} \sum_{m=1}^{k+1} a_{q,j}(m)\delta_{k-1}^j(m) + \sum_{m=1}^{k+1} a_{q,p}(m)\delta_{k-1}^p(m) + \sum_{m=1}^{k+1} a_{q,p}(m)\gamma_{k-1}^p(m)\Big].$$

For $m \neq 1$, the cases $i \neq q$ and $i = q$ are, respectively,

$$\delta_k^i(m) = \lambda_k^i(y_k)a_{i,i}(m-1)\delta_{k-1}^i(m-1), \quad \delta_k^q(m) = \lambda_k^q(y_k)a_{q,q}(m-1)\delta_{k-1}^q(m-1).$$

Proof. In this proof we first evaluate the expectation $E\big[\Lambda_k N_k^{p,q}\langle X_k, \boldsymbol{e}_i\rangle F(h_k^i) \mid \mathcal{Y}_k\big]$ in a general setting and thereafter specialise the outcome into scenarios; 1) $i \neq q$ and 2) $i = q$.

$$
\begin{aligned}
\overline{E}\big[\Lambda_k N_k^{p,q}\langle X_k, \boldsymbol{e}_i\rangle F(h_k^i) \mid \mathcal{Y}_k\big] &= \sum_{m=1}^{k+1} F(m)\delta_k^i(m) \\
&= \overline{E}\Big[\Lambda_{k-1}\lambda_k^i(y_k)\langle X_k, \boldsymbol{e}_i\rangle N_k^{p,q} F\big(\langle X_k, \boldsymbol{e}_i\rangle + \langle X_k, \boldsymbol{e}_i\rangle\langle X_{k-1}, \boldsymbol{e}_i\rangle h_{k-1}^i\big) \mid \mathcal{Y}_k\Big] \\
&= \lambda_k^i(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_k, \boldsymbol{e}_i\rangle N_k^{p,q} F\big(1 + \langle X_{k-1}, \boldsymbol{e}_i\rangle h_{k-1}^i\big) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda_k^i(y_k)\sum_{j=1}^{N}\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle \\
&\qquad \times\Big(N_{k-1}^{p,q} + \langle X_{k-1}, \boldsymbol{e}_p\rangle\langle X_k, \boldsymbol{e}_q\rangle\Big) F\big(1 + \langle X_{k-1}, \boldsymbol{e}_i\rangle h_{k-1}^i\big) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda_k^i(y_k)\sum_{\substack{j=1\\ j\neq i}}^{N}\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle\Big(N_{k-1}^{p,q} \\
&\qquad + \langle X_{k-1}, \boldsymbol{e}_p\rangle\langle X_k, \boldsymbol{e}_q\rangle\Big) F(1) \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad\quad + \lambda_k^i(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle\langle X_k, \boldsymbol{e}_i\rangle\Big(N_{k-1}^{p,q} \\
&\qquad\qquad + \langle X_{k-1}, \boldsymbol{e}_p\rangle\langle X_k, \boldsymbol{e}_q\rangle\Big) F(1 + h_{k-1}^i) \mid \mathcal{Y}_{k-1}\Big].
\end{aligned} \tag{4.1}
$$

Case 1, taking $i \neq q$, the last line of (4.1) becomes

$$
\begin{aligned}
&= \lambda_k^i(y_k)\sum_{\substack{j=1\\ j\neq i}}^{N}\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle N_{k-1}^{p,q} F(1) \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad + \lambda_k^i(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle\langle X_k, \boldsymbol{e}_i\rangle N_{k-1}^{p,q} F(1 + h_{k-1}^i) \mid \mathcal{Y}_k\Big] \\
&= \lambda_k^i(y_k)F(1)\sum_{\substack{j=1\\ j\neq i}}^{N}\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle a_{i,j}\big(h_{k-1}^i\big) N_{k-1}^{p,q} \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad + \lambda_k^i(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle a_{i,i}\big(h_{k-1}^i\big) F(1 + h_{k-1}^i) N_{k-1}^{p,q} \mid \mathcal{Y}_{k-1}\Big]
\end{aligned}
$$

$$= \lambda_k^i(y_k) F(1) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\delta_{k-1}^j(m) + \lambda_k^i(y_k) \sum_{m=1}^{k+1} a_{i,i}(m) F(1+m)\delta_{k-1}^i(m). \tag{4.2}$$

Case 2, suppose now that $i = q$, then the last line of (4.1) becomes

$$\begin{aligned}
&= \lambda_k^q(y_k) \sum_{j=1}^{N} \overline{E}\Big[\Lambda_{k-1}\big\langle X_{k-1}, \boldsymbol{e}_j\big\rangle\big\langle X_k, \boldsymbol{e}_q\big\rangle\big(N_{k-1}^{p,q} + \langle X_{k-1}, e_p\rangle\langle X_k, e_q\rangle\big) \\
&\qquad \times F\big(1 + \big\langle X_{k-1}, \boldsymbol{e}_q\big\rangle h_{k-1}^q\big) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda_k^q(y_k) \sum_{\substack{j=1\\ j\neq q}}^{N} \overline{E}\Big[\Lambda_{k-1}\big\langle X_{k-1}, \boldsymbol{e}_j\big\rangle\big\langle X_k, \boldsymbol{e}_q\big\rangle \\
&\quad \times \Big(N_{k-1}^{p,q} + \big\langle X_{k-1}, \boldsymbol{e}_p\big\rangle\big\langle X_k, \boldsymbol{e}_q\big\rangle\Big) F\big(1 + \big\langle X_{k-1}, \boldsymbol{e}_q\big\rangle h_{k-1}^q\big) \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad + \lambda_k^q(y_k)\overline{E}\Big[\Lambda_{k-1}\big\langle X_{k-1}, \boldsymbol{e}_q\big\rangle \big\langle X_k, \boldsymbol{e}_q\big\rangle \\
&\qquad\quad \times \big(N_{k-1}^{p,q} + \big\langle X_{k-1}, \boldsymbol{e}_q\big\rangle\big\langle X_k, \boldsymbol{e}_q\big\rangle\big) F(1 + h_{k-1}^q) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda_k^q(y_k) \sum_{\substack{j=1\\ j\neq p,\, j\neq q}}^{N} \overline{E}\Big[\Lambda_{k-1}\big\langle X_{k-1}, \boldsymbol{e}_j\big\rangle\big\langle X_k, \boldsymbol{e}_q\big\rangle N_{k-1}^{p,q} F(1) \mid \mathcal{Y}_{k-1}\Big] \\
&\quad + \lambda_k^q(y_k)\overline{E}\Big[\Lambda_{k-1}\big\langle X_{k-1}, \boldsymbol{e}_p\big\rangle\big\langle X_k, \boldsymbol{e}_q\big\rangle (N_{k-1}^{p,q} + 1) F(1) \mid \mathcal{Y}_{k-1}\Big] \\
&\qquad + \lambda_k^q(y_k)\overline{E}\Big[\Lambda_{k-1}\big\langle X_{k-1}, \boldsymbol{e}_q\big\rangle\big\langle X_k, \boldsymbol{e}_q\big\rangle N_{k-1}^{p,q} F(1 + h_{k-1}^q) \mid \mathcal{Y}_{k-1}\Big] \\
&= \lambda_k^q(y_k) F(1) \sum_{\substack{j=1\\ j\neq p,\, j\neq q}}^{N} \sum_{m=1}^{k+1} a_{q,j}(m)\delta_{k-1}^j(m) + \lambda_k^q(y_k) F(1) \sum_{m=1}^{k+1} a_{q,q}(m)\delta_{k-1}^p(m) \\
&\quad + \lambda_k^q(y_k) F(1) \sum_{m=1}^{k+1} a_{q,p}(m)\gamma_{k-1}^p(m) + \lambda_k^q(y_k) \sum_{m=1}^{k+1} a_{q,q}(m)\delta_{k-1}^q(m) F(1+m).
\end{aligned} \tag{4.3}$$

As in Theorem 3.4 we now take $F(1) = 1$ and $F(m) = 0$ if $m \neq 1$. Then, from the outcome taking $i \neq q$:

$$\delta_k^i(1) = \lambda_k^i(y_k) \sum_{\substack{j=1 \\ j\neq i}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\delta_{k-1}^j(m).$$

With $i = q$ and $F(1) = 1$, $F(m) = 0$ if $m \neq 1$:

$$\delta_k^q(1) = \lambda_k^q(y_k)\Bigg[\sum_{\substack{j=1 \\ j\neq p,\, j\neq q}}^{N} \sum_{m=1}^{k+1} a_{q,j}(m)\delta_{k-1}^j(m) + \sum_{k=1}^{k+1} a_{q,p}(m)\delta_{k-1}^p(m) + \sum_{m=1}^{k+1} a_{q,p}(m)\gamma_{k-1}^p(m)\Bigg].$$

We now consider a function F such that for some $m \neq 1$ $F(m) = 1$ and $\forall\, n \in \{1, 2, \dots, k+1\} \setminus \{m\}$, $F(n) = 0$. Then from equation 4.3 with $i = q$: $\delta_k^i(m) = \lambda_k^i(y_k)a_{i,i}(m-1)\delta_{k-1}^i(m-1)$ and also with $i \neq q$, $\delta_k^q(m) = \lambda_k^q(y_k)a_{q,q}(m-1)\delta_{k-1}^q(m-1)$. □

Remark 4.2. This result gives a recursive estimate for $\delta^{\,i}(m, p, q)$. As noted above, these can be used to determine estimates for $N_k^{p,q}$.

We now consider estimates for $G_k^q := \sum_{\ell=1}^{k} f(y_\ell)\langle X_{\ell-1}, \boldsymbol{e}_q\rangle$. As before, suppose there are probabilities such that for $i \in \{1, 2, \dots, N\}$ and any $F : \mathbb{N} \to \mathbb{R}$

$$\overline{E}\big[\Lambda_k G_k^q \langle X_k, \boldsymbol{e}_i\rangle F(h_k^i) \mid \mathcal{Y}_k\big] = \sum_{m=1}^{k+1} F(m)\nu_k^i(m).$$

We then have the following recursions:

Theorem 4.3. *For $m = 1$,*

$$\nu_k^i(1) = \lambda_k^i(y_k)\Big[\sum_{\substack{j=1 \\ j\neq i, j\neq q}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\nu_{k-1}^j(m) + \sum_{m=1}^{k+1} a_{i,q}(m)\nu_{k-1}^q(m) + f(y_k) \sum_{m=1}^{k+1} a_{i,q}(m)\gamma_{k-1}^q(m)\Big].$$

For $m \neq 1$:

$$\nu_k^i(m) = \lambda_k^i(y_k)\big[a_{i,i}(m-1)\nu_{k-1}^i(m-1) + f(y_k)a_{i,q}(m-1)\gamma_{k-1}^q(m-1)\langle \boldsymbol{e}_i, \boldsymbol{e}_q\rangle\big].$$

Proof.

$$\overline{E}\big[\Lambda_k G_k^q \langle X_k, \boldsymbol{e}_i\rangle F(h_k^i) \mid \mathcal{Y}_k\big] = \sum_{m=1}^{k+1} F(m)\nu_k^i(m)$$

$$= \lambda_k^i(y_k) \sum_{\substack{j=1 \\ j\neq i}}^{N} \overline{E}\big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle\big(G_{k-1}^q + f(y_k)\langle X_{k-1}, e_q\rangle\big)F(1) \mid \mathcal{Y}_k\big]$$

$$+\lambda_k^i(y_k)\overline{E}\Big[\Lambda_{k-1}, \langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle\big(G_{k-1}^q + f(y_k)\langle X_{k-1}, \boldsymbol{e}_q\rangle\big)F(1+h_{k-1}^i) \mid \mathcal{Y}_{k-1}\Big]$$

$$= \lambda_k^i(y_k) \sum_{\substack{j=1 \\ j\neq i, j\neq q}}^{N} \overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle G_{k-1}^q F(1) \mid \mathcal{Y}_{k-1}\Big]$$

$$+ \lambda_k^i(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_q\rangle\langle X_k, \boldsymbol{e}_i\rangle\big(G_{k-1}^q + f(y_k)\big)F(1) \mid \mathcal{Y}_{k-1}\Big]$$

$$+ \lambda_k^i(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle\langle X_k, \boldsymbol{e}_i\rangle G_{k-1}^q F(1+h_{k-1}^i) \mid \mathcal{Y}_k\Big]\big(1 - \langle \boldsymbol{e}_i, \boldsymbol{e}_q\rangle\big)$$

$$+ \lambda_k^q(y_k)\overline{E}\Big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_q\rangle\langle X_k, \boldsymbol{e}_q\rangle\big(G_{k-1}^q + f(y_k)\big)F(1+h_{k-1}^q) \mid \mathcal{Y}_k\Big]\langle \boldsymbol{e}_i, \boldsymbol{e}_q\rangle$$

$$= \lambda_k^i(y_k)F(1) \sum_{\substack{j=1 \\ j\neq i, j\neq q}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\nu_{k-1}^j(m) + \lambda_k^i(y_k)F(1)\sum_{m=1}^{k+1} a_{i,q}(m)\nu_{k-1}^q(m)$$

$$+ \lambda_k^i(y_k)f(y_k)F(1)\sum_{m=1}^{k+1} a_{i,q}(m)\gamma_{k-1}^q(m)$$

$$+ \lambda_k^i(y_k)\sum_{m=1}^{k+1} a_{i,i}(m)\nu_{k-1}^i(m)F(m+1)(1 - \langle \boldsymbol{e}_i, \boldsymbol{e}_q\rangle)$$

$$+ \lambda_k^i(y_k)\sum_{m=1}^{k+1} a_{q,q}(m)F(m+1)\nu_{k-1}^q(m)\langle e_i, e_q\rangle$$

$$+ \lambda_k^i(y_k)f(y_k)\sum_{m=1}^{\infty} a_{q,q}(m)F(m+1)\gamma_{k-1}^q(m)\langle \boldsymbol{e}_i, \boldsymbol{e}_q\rangle\,.$$

Taking $F(1) = 0$ and $F(m) = 0$ if $m \neq 1$

$$\nu_k^i(1) = \lambda_k^i(y_k)\Big[\sum_{\substack{j=1 \\ j\neq i,\, j\neq q}}^{N} \sum_{m=1}^{k+1} a_{i,j}(m)\nu_{k-1}^j(m) + \sum_{m=1}^{k+1} a_{i,q}(m)\nu_{k-1}^q(m)$$

$$+ f(y_k)\sum_{m=1}^{k+1} a_{i,q}(m)\gamma_{m-1}^q(m)\Big].$$

With $m \neq 1$, $F(m) = 1$ and again, $\forall\, n \in \{1, 2, \ldots, k+1\} \setminus \{m\}$, $F(n) = 0$,

$$\nu_k^i(m) = \lambda_k^i(y_k)\Big[a_{i,i}(m-1)\nu_{k-1}^i(m-1) + f(y_k)a_{i,q}(m-1)\gamma_{k-1}^q(m-1)\big\langle \boldsymbol{e}_i, \boldsymbol{e}_q\big\rangle\Big].$$

□

Remark 4.4. This provides recursive estimates for G_k^q which arise in the estimation of the model (parameters). See [3]. Taking $f(y_k) = 1$ for all y_k we see that G_k^q reduces to J_k^p, the occupation time up to time $k-1$, so the results of 4.3 give recursions for estimates related to the occupation times.

5. Smoothers

Suppose $0 \le k \le T$ and we have observed $\big\{y_0, y_1, \ldots, y_T\big\}$. We wish to find the smoothed estimate $E\big[X_k F(h_k) \mid \mathcal{Y}_T\big]$. With

$$\Lambda_{k+1,T} \coloneqq \prod_{\ell=k+1}^{T} \lambda_\ell$$

using Bayes' theorem, we have

$$E\big[X_k F(h_k) \mid \mathcal{Y}_T\big] = \frac{\overline{E}\big[\Lambda_{0,T} X_k F(h_k) \mid \mathcal{Y}_T\big]}{\overline{E}\big[\Lambda_{0,T} \mid \mathcal{Y}_T\big]}\,.$$

Now $\Lambda_{0,T} = \lambda_{0,k}\Lambda_{k+1,T}$ so

$$\overline{E}\big[\Lambda_{0,T} X_k F(h_k) \mid \mathcal{Y}_T\big] = \overline{E}\big[\lambda_{0,k} X_k F(h_k) \overline{E}\big[\Lambda_{k+1,T} \mid \mathcal{Y}_T, \mathcal{F}_k\big]\mathcal{Y}_T\big].$$

We shall find a backward recursion for

$$\overline{E}[\Lambda_{k+1,T} \mid \mathcal{Y}_T, \mathcal{F}_k\big].$$

The process (X_k, h_k) is Markov. Suppose there are probabilities ρ such that

$$\overline{E}[\Lambda_{k+1,T} \mid \mathcal{Y}_T, x_k = \boldsymbol{e}_i, h_k^i = m\big] = \rho_{k,T}^i(m).$$

Theorem 5.1. *For $i \in \{1, 2, \ldots, N\}$ and $m \in \{1, 2, \ldots\}$*

$$\rho_{k,T}^i(m) = a_{i,i}(m+1)\rho_{k+1,T}(m+1)\lambda_{k+1}^i(y_{k+1})$$
$$\rho_{k,T}^i(1) + \sum_{\substack{j=1\\ j\neq i}}^{N} a_{j,i}(m)\rho_{k+1}^j(m)\lambda_{k+1}^i(y_{k+1}).$$

Proof.

$$\begin{aligned}\rho_{k,T}^i(m) &= \overline{E}\big[\Lambda_{k+1,T} \mid \mathcal{Y}_T, X_k = e_i, h_k^i = m\big]\\ &= \sum_{n=1}^{k+2}\sum_{j=1}^{N} E\big[\big\langle x_{k+1}, \boldsymbol{e}_j\big\rangle \boldsymbol{I}(h_{k+1} = n)\Lambda_{k+2,T} \mid \mathcal{Y}_T, X_k = \boldsymbol{e}_i, h_k^i = m\big]\lambda_{k+1}^j(y_k).\end{aligned}$$

In the second sum, if $j = i$, we have

$$\rho_{k,T}^i(m) = a_{i,i}(m+1)\rho_{k+1,T}^i(m+1)\lambda_{k+1}^i(y_{k+1})$$

and if $j \neq i$:

$$\rho^i_{k,T}(1) = \sum_{\substack{j=1\\ j\neq i}}^{N} a_{j,i}(m)\rho^j_{k+1,T}(m)\lambda^j_{k+1}(y_{k+1}).$$

□

6. Viterbi Estimates

The basic idea of the Viterbi estimates is that expectation is replaced by maximum likelihood. Consequently, in the above results summations are replaced by maximizations. Therefore, we obtain approximate unnormalised state estimates.

From Theorem 3.4 γ^* is determined recursively by

$$\gamma^i_k(1)^* = \lambda^i_k(y_k)\max_{j,\ j\neq i}\max_m a_{i,j}(m)\gamma^i_{k-1}(m)^*$$

and for $1 < m \leq k+1$

$$\gamma^i_k(m)^* = \lambda^i_k(y_k)a_{i,i}(m-1)\gamma^i_{k-1}(m-1)^*.$$

For the number of jumps, from Theorem 4.1 for $i \neq p$

$$\delta^i_k(1)^* = \lambda^i_k(y_k)\max_{j,\ j\neq i}\max_m a_{i,j}(m)\delta^i_{k-1}(m)^*,$$

for $i = p$

$$\begin{aligned}\sigma^p_k(1)^* = \lambda^p_k(y_k)\Big[&\max_{\substack{j\\ j\neq p,\ j\neq q}}\max_m a_{i,j}(m)\delta^j_{k-1}(m)^* \\ &+ \max_m a_{p,q}(m)\delta^q_{k-1}(m)^* + a_{p,q}(n)\gamma^q_{k-1}(m)^*\Big],\end{aligned}$$

for $m \neq 1$: $\delta^i_k(m)^* = \lambda^i_k(y_k)a_{i,i}(m-1)\delta^i_{k-1}(m-1)^*$.
For the G process, from Theorem 4.3,

$$\begin{aligned}\nu^i_k(1)^* = \lambda^i_k(y_k)\Big[&\max_{\substack{j\\ j\neq i, j\neq q}} a_{i,j}(m)\nu^j_{k-1}(m)^* + \max_m a_{i,q}(m)\nu^q_{k-1}(m)^* \\ &+ f(y_k)\max_m a_{i,q}(m)\gamma^q_{k-1}(m)^*\Big].\end{aligned}$$

For $m \neq 1$:

$$\begin{aligned}\nu^i_k(m)^* = \lambda^i_k(y_k)\big[&a_{i,i}(m-1)\nu^i_{k-1}(m-1)^* \\ &+ f(y_k)a_{i,q}(m-1)\gamma^i_{k-1}(m-1)^*\langle \boldsymbol{e}_i, \boldsymbol{e}_q\rangle\big].\end{aligned}$$

For Smoothers, from Theorem 5.1, when $m \neq 1$

$$\rho^i_{k,T}(m)^* = a_{i,i}(m+1)\rho^i_{k+1,T}(m+1)^*\lambda^i_{k+1}(y_{k+1})$$

and

$$\rho^i_{k,T}(1)^* = \max_{\substack{j\\ j\neq i}} a_{j,i}(m)\rho^j_{k+1,T}(m)^*\lambda^j_{k+1}(y_{k+1}).$$

7. Conclusion

Explicit expressions have been obtained for filtered and smoothed recursive estimates of a finite state semi-Markov chain observed in Gaussian noise. Viterbi versions of the formulae are also derived.

Acknowledgment. This research is supported the Natural Science and Engineering Research Council of Canada (NSERC) and by the Australian Research Council (ARC).

References

1. Abdullah, K. M.: *Application of hidden semi-Markov models in regime switching time series models*, PhD Thesis, University of South Australia, 2017.
2. Barbu, V. S. and Limnios, N.: *Semi-Markov Chains and Hidden Semi-Markov Models Towards Applications*, Lecture Notes in Statistics. Springer Verlag, 2008, Number 191.
3. Elliott, R. J., Aggoun, L., and Moore, J. B.: *Hidden Markov Models: Estimation and Control*, Applications of Mathematics, Springer Verlag, 1995, no.29.
4. Elliott, R. J. and Deng, Jia: "A Viterbi smoother for discrete state space model", Systems and Control Letters, Volume 58, Issue 6, june 2009, pp. 400–405
5. Elliott, R. J., Limnios, N., and Swishchuk, A.: "Filtering hidden semi-Markov chains", *Statistics and Probability Letters*, vol. 83, 2007–2014, 2018.
6. Elliott, R. J. and Malcolm, W. P.: "New representations for a semi-Markov chain and related filters", *Journal of Stochastic Analysis*, vol. 2, no. 1, 2021, D01 10.31390/josa.2.1.08.
7. Koski, T.: *Hidden Markov Models for Bio-informatics*, Kluwer Academic Publishers 2001.
8. Krishnamurthy, V., Moore, J. B., and Chung, S.-H.: "On hidden fractal model signal processing", *Signal Processing*, Volume 24, pp. 177–192, 1991.
9. van der Hoek, J. and Elliott, R. J.: *Introduction to Hidden Semi-Markov Models*. London Mathematical Society Lecture Notes, Cambridge University Press, 2018, Number 224.
10. Yu, S.-Z.: "Hidden semi-Markov models," *Artificial Intelligence*, Volume 174, Number 2, pp. 215–243, February 2010.

Robert J. Elliott: UNISA Business, University of South Australia, Campus Central - City West, GPO Box 2471, Australia and University of Calgary, Calgary T2N 1 N4, Canada

Email address: Robert.Elliott@unisa.edu.au and relliott@ucalgary.ca

URL: https://people.unisa.edu.au/Robert.Elliott

URL: http://www.ucalgary.ca/~relliott

W. P. Malcolm: Defence Science and Technology Group Australia, Russell Offices Canberra, Constitution Ave, Russell ACT 2600, Australia

Email address: Paul.Malcolm2@defence.gov.au