Journal of Stochastic Analysis

Volume 2 | Number 1 Article 8

March 2021

New Representations for a Semi-Markov Chain and Related Filters

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

Recommended Citation
Elliott, Robert J. and Malcolm, W. P. (2021) "New Representations for a Semi-Markov Chain and Related Filters," *Journal of Stochastic Analysis*: Vol. 2 : No. 1 , Article 8.
DOI: 10.31390/josa.2.1.08
Available at: https://digitalcommons.lsu.edu/josa/vol2/iss1/8

Journal of Stochastic Analysis
Vol. 2, No. 1 (2021) Article 8 (16 pages)
DOI: 10.31390/josa.2.1.08

NEW REPRESENTATIONS FOR A SEMI-MARKOV CHAIN AND RELATED FILTERS

ROBERT ELLIOTT* AND W. P. MALCOLM

Abstract. In this article we investigate estimation for a partially observed semi-Markov chain, or a Hidden semi-Markov Model (HsMM). We derive semimartingale dynamics for a semi-Markov chain and give them in a new vector form which explicitly exhibits the times at which jump-events occur and the probabilities of state transitions. However, the most important result is the new vector lattice state-space representation for a general finite-state, discrete-time semi-Markov chain. On this space the semi-Markov chain and its occupation times are a Markov process with dynamics described by finite matrices. These representations are new. Finite dimensional recursive filters are derived for a HsMM.

1. Introduction

Semi-Markov chains are related to renewal processes and have been used in applications since their introduction over 60 years ago. Their general occupation-time distributions offer a far richer class of models than standard Markov chains. The two main contributions of this paper are;

(1) a new vector state-space representation for a general finite-state semi-Markov chain which exhibits it as a Markov chain,
(2) the consequent extension to semi-Markov chains of the filtering, smoothing and estimation results,
(3) the matrix and vector semimartingale dynamics for the semi-Markov chain.

Earlier references on semi-Markov processes include the books by Koski [8], Barbu and Limnios [2], and van der Hoek and Elliott [10]. References on filtering include, Krishnamurthy, Moore and Chung [9] and Elliott, Limnios and Swishchuk [5]. Filters for Markov modulated time series were obtained in the PhD Thesis [1]. The matrix representation in this paper of the dynamics is new.

2. Stochastic Dynamics

All processes are defined on a probability space $\big(\Omega, \mathcal{F}, P\big)$. Our process of interest is a semi-Markov chain $X = \big\{X_k,\, k = 0, 1, 2, \dots\big\}$ with arbitrary state sojourn distributions. As is now standard the finite state space for the process

Received 2020-11-29; Accepted 2021-2-9; Communicated by the editors.
2010 *Mathematics Subject Classification.* 60G35; 60K15.
Key words and phrases. Discrete-time, semi-Markov models, filters, smoothers, martingales, change of probability measure, filter-based (online) and smoother-based (offline) EM algorithms.
* Corresponding author.

X is identified with the set of unit vectors $S = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_N\}$, where $\boldsymbol{e}_i := (0, \dots, 0, 1, 0, \dots, 0)' \in \mathbb{R}^N$. We also write $m \in \{1, 2, 3, \dots\}$ exclusively as a time index for state sojourns.

Notation 2.1. The initial state $X_0 \in S$, is taken as given, or its probability distribution $p_0 = (p_0^1, p_0^2, \dots, p_0^N)' \in [0,1]^N$ is known. The chain will change state at random discrete 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 $\tau_0 := 0$. Successive jump event times form a strictly increasing sequence $\tau_0 < \tau_1 < \tau_2 < \tau_3 \dots$. Write $\mathcal{F}_k := \sigma\{X_u, u \leq k\}$ and $\mathbb{F} = \{\mathcal{F}_u\}_{u \geq 0}$ for the filtration generated by X.

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

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

$$P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j,\, \tau_{n+1} - \tau_n = m \mid \mathcal{F}_{\tau_n}\big) = P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j,\, \tau_{n+1} - \tau_n = m \mid X_{\tau_n} = \boldsymbol{e}_i\big).$$

If $X_{\tau_n} = \boldsymbol{e}_i$ we write this as $q\big(\boldsymbol{e}_j, \boldsymbol{e}_i, m\big)$.

This can be factorized as

$$P\big(\tau_{n+1} - \tau_n = m \mid X_{\tau_{n+1}} = \boldsymbol{e}_j,\, X_{\tau_n} = \boldsymbol{e}_i\big)\, P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j | X_{\tau_n} = \boldsymbol{e}_i\big) = f_{j,i}(m) p_{j,i},$$

say. Here

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

Consequently

$$q\big(\boldsymbol{e}_i, \boldsymbol{e}_j, m\big) = f_{j,i}\big(m\big) p_{j,i}\,. \tag{2.1}$$

We can also consider the factorization

$$\begin{aligned} P(X_{\tau_{n+1}} = \boldsymbol{e}_j, \tau_{n+1} - \tau_n = m | X_{\tau_n} = \boldsymbol{e}_i) &= P(\tau_{n+1} - \tau_n = m | X_{\tau_n} = \boldsymbol{e}_i) \times \\ &\qquad P(X_{\tau_{n+1}} = \boldsymbol{e}_j | \tau_{n+1} - \tau_n = m, X_{\tau_n} = \boldsymbol{e}_i) \\ &= \pi_i(m) p_{j,i}(m), \text{ say.} \end{aligned} \tag{2.2}$$

Here

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

Approximations 2.3. If $f_{j,i}(m)$ does not depend upon $\boldsymbol{e}_j$ we can write

$$\begin{aligned} P(\tau_{n+1} - \tau_n = m \mid X_{\tau_{n+1}} = \boldsymbol{e}_j,\, X_{\tau_n} = \boldsymbol{e}_i) &= P(\tau_{n+1} - \tau_n = m \mid X_{\tau_n} = \boldsymbol{e}_i) \\ &= \pi_i(m). \end{aligned} \tag{2.3}$$

That is, for each i, $1 \leq i \leq N$, $\big\{\pi_i(m),\, m = 1, 2, 3, \dots\big\}$ is a probability distribution on the positive integers. Then under this simplification

$$q\big(\boldsymbol{e}_j, \boldsymbol{e}_i, m\big) = \pi_i\big(m\big) p_{j,i}\,. \tag{2.4}$$

Note that, as we assumed X is homogeneous in time, all these probabilities are independent of n. If $p_{j,i}(m)$ does not depend upon m then from (2.2) we again have:

$$q\big(\boldsymbol{e}_j, \boldsymbol{e}_i, m\big) = \pi_i(m) p_{j,i}\,. \tag{2.5}$$

The approximation given by equation (2.4) or equation (2.5) is that used by Ferguson [7] However, in this paper we shall not discuss any approximations but use the general decomposition

$$q\big(\boldsymbol{e}_j, \boldsymbol{e}_i, m\big) = \pi_i(m) p_{j,i}(m). \tag{2.6}$$

Notation 2.4. Write

$$G_i(m) \coloneqq P\big(\tau_{n+1} - \tau_n \le m \mid X_{\tau_n} = \boldsymbol{e}_i\big) = \sum_{\ell=1}^{m} \pi_i(\ell),$$
$$F_i(m) \coloneqq P\big(\tau_{n+1} - \tau_n > m \mid X_{\tau_n} = \boldsymbol{e}_i\big) = 1 - G_i(m).$$

We now provide the conditional probability for a state-transition to occur at the next discrete time. This probability plays an important role in subsequent calculations and is denoted by $\Delta^i(m)$. Given some discrete-time k, write τ_n for the most recent transition-event time prior to k, (or at k), that is, $\tau_n \coloneqq \max_\ell\{\tau_\ell \le k\}$. Further, suppose that for some m, $X_{\tau_n+m-1} = \boldsymbol{e}_i$. The probability of a transition-event occuring at the next time $\tau_n + m$ is

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

This result is from the definition of conditional probability.
Write $A \coloneqq \big\{\tau_{n+1} = \tau_n + m\big\}$, $B \coloneqq \big\{\tau_{n+1} > \tau_n + m - 1\big\}$ and $C \coloneqq \big\{\tau_n = \boldsymbol{e}_i\big\}$. Then

$$\begin{aligned} P(\tau_{n+1} = \tau_n + m \mid X_{\tau_n+m-1} = X_{\tau_n} = \boldsymbol{e}_i) \\ &= P\big(A \mid B \cap C\big) \\ &= \frac{P\big(A \cap B \mid C\big)}{P\big(B \mid C\big)}, \end{aligned}$$

(but $A \cap B = A$ as $A \subset B$, so it equals)

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

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

Definition 2.5. For each index i, $1 \le i \le N$, we define the recursive process $h^i_k \coloneqq \big\langle X_k, \boldsymbol{e}_i \big\rangle + \langle X_k, \boldsymbol{e}_i\rangle \langle X_k, X_{k-1}\rangle\, h^i_{k-1}$, with $h^i_0 \coloneqq \big\langle X_0, \boldsymbol{e}_i\big\rangle \in \{0, 1\}$. 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 $h_k = \sum_{i=1}^{N} h_k^i$ 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 2.6. *Suppose $i \neq j$, $1 \leq i, j \leq N$. Then $P\big(X_{k+1} = \boldsymbol{e}_j | X_k = \boldsymbol{e}_i,\, h_k^i\big) = p_{j,i}\big(h_k^i\big)\Delta^i\big(h_k^i\big)$.*

Proof. Write $A \coloneqq \big\{X_{k+1} = \boldsymbol{e}_j\big\}$, $B' \coloneqq \big\{\tau_{n+1} - \tau_n = h_k^i\big\}$,
$B'' \coloneqq \big\{\tau_{n+1} > \tau_n + h_k^i - 1\big\}$ and $C \coloneqq \big\{X_{\tau_n} = \boldsymbol{e}_i = X_k\big\}$. Then

$$
\begin{aligned}
P\big(X_{k+1} = \boldsymbol{e}_j \mid X_k = \boldsymbol{e}_i, h_k^i\big) &= P\big(A \cap B' \mid B'' \cap C\big) \\
&= \frac{P\big(A \cap B' \cap B'' \cap C\big)}{P\big(B'' \cap C\big)} = \frac{P\big(A \cap B' \cap C\big)}{P\big(B'' \cap C\big)} \\
&= P\big(A \mid B' \cap C\big)\frac{P\big(B' \cap C\big)}{P\big(B'' \cap C\big)}
\end{aligned}
$$

(as $B' \cap B'' = B'$)

$$
\begin{aligned}
&= p_{j,i}\big(h_k^i\big)\frac{\pi^i\big(h_k^i\big)}{F_i\big(h_k^i - 1\big)} \\
&= p_{j,i}\big(h_k^i\big)\Delta^i\big(h_k^i\big).
\end{aligned}
$$

□

Remark 2.7. We are assuming there is a jump from $\boldsymbol{e}_i$ to a different $\boldsymbol{e}_j$, $i \neq j$, at time $k+1$. So, $\sum_{\substack{j=1 \\ j \neq i}}^{N} p_{j,i}(k+1) = 1$.

Corollary 2.8. *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 2.9. For $m = 1, 2, \ldots$, write $A(m)$ for the $N \times N$ matrix with entries $a_{i,i}(m) = 1 - \Delta^i(m)$ and $a_{j,i}(m) = p_{j,i}(m)\Delta^i(m)$.

Example 2.10. Then for $N = 3$ and some $h_k = m$,

$$
A(m) \coloneqq \begin{bmatrix} 1 - \Delta^1(m) & p_{1,2}(m)\Delta^2(m) & p_{1,3}(m)\Delta^3(m) \\ p_{2,1}(m)\Delta^1(m) & 1 - \Delta^2(m) & p_{2,3}(m)\Delta^3(m) \\ p_{3,1}(m)\Delta^1(m) & p_{3,2}(m)\Delta^2(m) & 1 - \Delta^3(m) \end{bmatrix} .
$$

Notation 2.11. Define the matrices: $\Pi(m) := (p_{i,j}(m), 1 \le i, j \le N)$ where $p_{i,i}(m) = -1$ and $p_{j,i}(m) = P\big(X_{\tau_{n+1}} = \boldsymbol{e}_j \mid \tau_{n+1} - \tau_n = m,\ X_{\tau_n} = \boldsymbol{e}_i\big)$, for $i \neq j$.
Write $D(m) := \operatorname{diag}\big(\Delta^1(m), \Delta^2(m), \dots, \Delta^N(m)\big)$.
Then $A(m) = I + \Pi(m)D(m)$, where I is the $N \times N$ identity matrix.

For the case when $N = 3$,

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

and so $A(m) = I + \Pi(m)D(m)$. 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 2.12. *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. For $i \neq j$ $E\big[\langle X_{k+1}, \boldsymbol{e}_j\rangle \mid X_k = \boldsymbol{e}_i, h_k^i\big] = P\big(X_{k+1} = \boldsymbol{e}_j \mid X_k = \boldsymbol{e}_i, h_k^i\big) = a_{j,i}\big(h_k^(i)\big)$ from Lemma 2.6 and the definition of $a_{j,i}\big(h_k^i\big)$. For the transition $X_k = \boldsymbol{e}_i \to X_{k+1} = \boldsymbol{e}_i$,
$E\big[\langle X_{k+1}, \boldsymbol{e}_i\rangle \mid X_k = \boldsymbol{e}_i, h_k^i\big] = P\big(X_{k+1} = \boldsymbol{e}_i \mid X_k = \boldsymbol{e}_i, h_k^i\big) = a_{i,i}\big(h_k^(i)\big)$ from Corollary 2.8 and the definition of $a_{i,i}\big(h_k^i\big)$. So $E\big[X_{k+1} \mid X_k, h_k\big] = A\big(h_k\big)X_k \in \mathbb{R}^N$ and

$$\begin{aligned} E\big[M_{k+1} \mid X_k, h_k\big] &= E\big[X_{k+1} - A\big(h_k\big)X_k \mid X_k, h_k\big] \\ &= \mathbf{0} \in \mathbb{R}^N. \end{aligned}$$

That is M_{k+1} is a (vector) martingale increment. □

3. Lattice-based State-Space Dynamics

In this section we describe a countably infinite state space for a general semi-Markov chain. In this state space the process (X, h) is in fact a Markov chain. This property is known but the matrix representations are new.

3.1. Lattice-based State-Space. The complete description of the state of our semi-Markov chain X at time k is given by the state of the chain $X_k \in \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_N\}$ and the number of time steps h_k the chain has been in that state since the last jump. To simplify the discussion here we suppose $N = 3$ so $S = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$.
A state space $\overline{S}$ for the chain $\overline{X}_k := (X_k, h_k)$ can be identified with countably many copies of S as follows: Elements of $\overline{S}$ can be thought of as infinite column vectors so, for example,

$$\begin{aligned} (\boldsymbol{e}_1, 1) \quad &\text{corresponds to} \quad \big(\underbrace{1, 0, 0}_{h=1} \mid 0, 0, 0 \mid 0, \cdots\big)' \text{ and} \\ (\boldsymbol{e}_2, \ell) \quad &\text{corresponds to} \quad \big(0, 0, 0 \mid \cdots \mid \underbrace{0, 1, 0}_{h_k=\ell} \mid 0, \cdots\big)' \end{aligned}$$

with $\boldsymbol{e}_2 = (0,1,0)'$ in the ℓ^{th} block. As a basis of unit vectors for this $\overline{X} = (X, h)$ process we take unit vectors $\boldsymbol{e}_{i,n}$, $1 \le i \le 3, \quad n = 1, 2, \ldots$. Here the i denotes the state in $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ and the n corresponds to the sojourn time in state $\boldsymbol{e}_i$ since the last jump at τ_n . Recall $S = \big\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\big\}$, so $\boldsymbol{e}_{i,n}$ is in the n^{th} block of $\big(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\big)$. Write $\overline{S} = \{e_{i,n}\,,\ 1 \le i \le 3,\ n = 1, 2, \ldots\}$. There is a map from $\mathbb{R}^3$ to $\mathbb{R}^{3\times\mathbb{N}}$ given by
$T : (\alpha_1, \alpha_2, \alpha_3)' \to \big((\alpha_1, \alpha_2, \alpha_3), (\alpha_1, \alpha_2, \alpha_3), \cdots\big)'$. With I_N the $N\times N$ unit matrix this is given by the $\mathbb{N} \times N$ matrix: $T = \big(I_N, I_N, I_N, \ldots\big)'$. The adjoint of this is a map from $\mathbb{R}^{N\times\mathbb{N}}$ to $\mathbb{R}^N$ given by $T^* = (I_N,\, I_N,\, I_N,\, \ldots)$.

3.2. State Transition Events. Note the counter $h_k = h_k(X_k)$ starts at 1, the first time X jumps to a new state. With the above notation $(\boldsymbol{e}_i, r) = \boldsymbol{e}_{i,r} \to (\boldsymbol{e}_i, r+1) = \boldsymbol{e}_{i,(r+1)}$ with probability $\big(1 - \Delta^i(r)\big)$, or $(\boldsymbol{e}_i, r) = \boldsymbol{e}_{i,r} \to \ (\boldsymbol{e}_j, 1) = \boldsymbol{e}_{j,i},\ j \ne i$, with probability $p_{j,i}(r)\Delta^i(r)$. For example, suppose at time 0 the chain is in state $(\boldsymbol{e}_1, 1) = \boldsymbol{e}_{1,1} = \big(1,0,0 \mid 0,0,0 \mid \cdots\big)'$. This can become either $\boldsymbol{e}_{1,2} = (\boldsymbol{e}_1, 2) = (0,0,0 \mid 1,0,0 \mid 0,\cdots)'$ with probability $\big(1 - \Delta^1(1)\big)$, or $\boldsymbol{e}_{2,1} = \big(\boldsymbol{e}_2, 1) = (0,1,0|0,0,0|0,0,\cdots\big)'$ with probability $p_{2,1}(r)\Delta^1(1)$, or
$\boldsymbol{e}_{3,1} = \big(\boldsymbol{e}_3, 1) = (0,0,1 \mid 0,0,0 \mid 0,\cdots\big)'$ with probability $p_{3,1}(r)\Delta^1(1)$. There is then an infinite matrix C which describes these transitions.

3.3. Dynamics for $\overline{X}_k := \big(X_k, h_k\big)$. In the $N = 3$ state case and for some value of $m \in \{1, 2, \ldots\}$, write

$$\Pi(m) = \begin{bmatrix} 0 & p_{1,2}(m)\Delta^2(m) & p_{1,3}(m)\Delta^3(m) \\ p_{2,1}(m)\Delta^1(m) & 0 & p_{2,3}(m)\Delta^3(m) \\ p_{3,1}(i)\Delta^1(m) & p_{3,2}(m)\Delta^2(m) & 0 \end{bmatrix}$$

and $D(m) = \operatorname{diag}\big\{1 - \Delta^1(m),\ 1 - \Delta^2(m),\ 1 - \Delta^3(m)\big\}$. With $\mathbf{0}$ representing the 3×3 zero matrix

$$C = \begin{bmatrix} \Pi(1) & \Pi(2) & \Pi(3) & \cdots \\ D(1) & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & D(2) & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{3.1}$$

If we write the enlarged vectors as $\overline{X}_k$ then the semi-martingale dynamics of the Markov chain can be written as $\overline{X}_{k+1} = C\overline{X}_k + \overline{M}_{k+1} \in \overline{S}$. This gives $E\big[\,\overline{X}_{k+1} \mid \overline{X}_k\,\big] = C\overline{X}_k$ and $E\big[\,\overline{X}_{k+1} \mid \overline{X}_0\,\big] = C^{k+1}\,\overline{X}_0$. At time $k \in \{0, 1, 2, \ldots\}$ the sojourn time h^i_k cannot be more than $k+1$ and the next possible value of h^i_k is $k+2$. Consequently the size of C at time k is at most $(k+2)N \times (k+1)N$. For example, at time 0 the C matrix has the form $C = \begin{bmatrix} \Pi(1) \\ D(1) \end{bmatrix}$. At time 1 the C matrix has the form

$$C = \begin{bmatrix} \Pi(1) & \Pi(2) \\ D(1) & \mathbf{0} \\ \mathbf{0} & D(2) \end{bmatrix} \quad \text{and so on.} \tag{3.2}$$

Consequently, at any finite time the C matrix is finite. Also the state space of $\overline{X}$ at time k only has $(k+1)N$ elements. The size of the state space and the corresponding matrices will also remain finite if the sojourn distributions all have finite support.

4. Observation Dynamics

The filtering results of [6] are now adapted to this situation. Note that if $\overline{X}_k \in \overline{S}$ then $T^* \overline{X}_k = X_k \in S$. We suppose the Markov chain $\overline{X}$ is not observed directly. Instead there is an observation sequence $y = \{y_0, y_1, \dots, y_k, \dots\}$ where

$$y_k = c(X_k) + d(X_k) w_k \,. \tag{4.1}$$

The observations are of $X_k = T^* \overline{X}_k$ rather than $\overline{X}_k$. $\{w_k,\, k = 0, 1, 2, \}$ is a sequence of i.i.d. $N(0,1)$ random variables. $c(\cdot)$ and $d(\cdot)$ are known real valued functions. Note that any real function $g(X_k)$ takes only the finite number of values $g(\boldsymbol{e}_1), g(\boldsymbol{e}_2), \dots, g(\boldsymbol{e}_N)$. Write $g_k = g(\boldsymbol{e}_k)$ and $\boldsymbol{g} = (g_1, g_2, \dots, g_N)' \in \mathbb{R}^N$. Then $g(X_k) = \langle \boldsymbol{g}, X_k \rangle$. Consequently there are vectors $\boldsymbol{c} = (c_1, c_2, \dots, c_N)$, $\boldsymbol{d} = (d_1, d_2, \dots, d_N)$ such that $c(X_k) = \langle \boldsymbol{c}, X_k \rangle$ and $d(X_k) = \langle \boldsymbol{d}, X_k \rangle$. We suppose $d_k > 0$ for $k = 1, \dots, N$.

Remark 4.1. We suppose the observation process y is scalar-valued. The extension to a vector-valued y is straight forward.

5. Finite-Dimensional Recursive Filters

5.1. Change of Probability Measure Formulation. We suppose there is a second 'reference' probability measure , $\overline{P}$, under which 1.) the process $\overline{X}$ is still a Markov chain with dynamics $\overline{X}_{k+1} = C\overline{X}_k + \overline{M}_k$ and 2.) the process $y = \{y_0, y_1, \dots\}$ is a sequence of i.i.d. $N(0,1)$ random variables. From $\overline{P}$ we now construct the original probability P under which; 1.) the process $X = T^* \overline{X}$ is a semi-Markov chain with dynamics as above so $X_{k+1} = A\big(h(k)\big) X_k + M_{k+1}$ and 2.) The process $w = (w_0, w_1, \dots)$ is a sequence of i.i.d. $N(0,1)$ random variables where $w_k = \dfrac{y_k - \langle \boldsymbol{c}, X_k \rangle}{\langle \boldsymbol{d}, X_k \rangle}$.

Definition 5.1. For $k = 0, 1, 2, \dots$ write $\lambda_k := \dfrac{\phi\Big((y_k - \langle \boldsymbol{c}, X_k \rangle)/\langle \boldsymbol{d}, X_k \rangle\Big)}{\langle \boldsymbol{d}, X_k \rangle \phi(y_k)}$, where $\phi(x)$ is the $N(0,1)$ density $\frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2} x^2$, and

$$\Lambda_{0,k} := \prod_{\ell=0}^{k} \lambda_\ell \,. \tag{5.1}$$

Recall $\mathcal{F}_k = \sigma\{X_0, X_1, \dots, X_k\}$ and write $\mathcal{Y}_k = \sigma\{y_0, y_1, \dots, y_k\}$ and $\mathcal{G}_k = \sigma\{X_0, \dots, X_k, y_0, \dots, y_k\}$. We consider the related filtrations $\{\mathcal{F}_k\}$, $\{\mathcal{Y}_k\}$ and $\{\mathcal{G}_k\}$.

Definition 5.2. The original 'real world' probability P is defined in terms of $\overline{P}$ by setting

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

We can then prove

Lemma 5.3. *Under P X is a semi-Markov chain with dynamics* $X_{k+1} = A\big(h(k)\big)X_k + M_{k+1}$ *and* $\big\{w_k,\, k = 0, 1, \dots\big\}$ *is a sequence of i.i.d.* $N(0,1)$ *random variables where* $w_k = (y_k - \big\langle \boldsymbol{c}, X_k\big\rangle)/\big\langle \boldsymbol{d}, X_k\big\rangle$.

That is, under P $y_k = \langle c, X_k\rangle + \langle d, X_k\rangle\, w_k$.

Proof. For a proof see [4]. □

Recall from §3.1, that the chain $\overline{X}$ has dynamics $\overline{X}_{k+1} = C\overline{X}_k + \overline{M}_{k+1} \in \overline{S}$. We suppose, as in §4, that the observation process is $y_k = c(X_k) + d(X_k)w_k$, where $X_k = T^*\,\overline{X}_k$. As above

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

Write $\lambda_k^i = \dfrac{\phi\Big((y_k - c_i)/d_i\Big)}{d_i\phi(y_k)}$ and $\Lambda_{0,k} = \prod_{\ell=0}^{k} \lambda_\ell$. However, for any $n = 1, 2, \dots,$ $T^*\boldsymbol{e}_{i,n} = \boldsymbol{e}_i$ so, for example, $\big\langle \boldsymbol{c}, \boldsymbol{e}_i\big\rangle = c_i = \big\langle \boldsymbol{c}, T^*\boldsymbol{e}_{i,n}\big\rangle = \big\langle T\boldsymbol{c}, \boldsymbol{e}_{i,n}\big\rangle$ and λ_k can be written in terms of the full state $\overline{X}_k$:

$$\lambda_k = \frac{\phi\Big(\big(y_k - \big\langle T\boldsymbol{c}, \overline{X}_k\big\rangle\big)/\big\langle T\boldsymbol{d}, \overline{X}_k\big\rangle\Big)}{\big\langle T\boldsymbol{d}, \overline{X}_k\big\rangle \phi(y_k)} .$$

5.2. A Finite-dimensional recursive filter for X. Write $\gamma_k = \overline{E}\big[\, \Lambda_k \overline{X}_k \mid \mathcal{Y}_k \,\big]$ for the unnormalized conditional expected value of $\overline{X}_k$ given the observations $\mathcal{Y}_k$ to time k. Again suppose $N = 3$, write $\Gamma_3(y_{k+1}) \coloneqq \operatorname{diag}\big\{\lambda_{k+1}^1(y_{k+1}),\, \lambda_{k+1}^2(y_{k+1}),\, \lambda_{k+1}^3(y_{k+1})\big\}$ and $\Gamma(y_{k_1}) \coloneqq \operatorname{diag}\big\{\Gamma_3(y_{k+1}),\, \Gamma_3(y_{k+1}),\, \Gamma_3(y_{k+1}), \dots\big\}$. We then have the recursion.

Theorem 5.4. $\gamma_{k+1} = \Gamma(y_{k+1})\, C\, \gamma_k$ *with* γ_0 *given by* X_0*, or its probability distribution.*

Proof.

$$\begin{aligned}
\gamma_{k+1} &= \overline{E}\big[\Lambda_{k+1}\overline{X}_{k+1} \mid \mathcal{Y}_{k+1}\big] \\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{N} \overline{E}\big[\Lambda_k\lambda_{k+1}\big\langle\, \overline{X}_{k+1}\,,\, \boldsymbol{e}_{i,n}\big\rangle \mid \mathcal{Y}_{k+1}\big]\boldsymbol{e}_{i,n} \\
&= (y_{k+1})\sum_{n=1}^{\infty}\sum_{i=1}^{N} \lambda_{k+1}^i(y_{k+1})\overline{E}\big[\Lambda_k\big\langle C\,\overline{X}_k\,,\, \boldsymbol{e}_{i,n}\big\rangle \mid \mathcal{Y}_k\big]\boldsymbol{e}_{i,n} \\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{N} \lambda_{k+1}^i(y_{k+1})\langle C\gamma_k\,,\, \boldsymbol{e}_{i,j}\rangle\boldsymbol{e}_{i,n} = \Gamma(y_{k+1})\, C\, \gamma_k\, .
\end{aligned}$$

□

Remark 5.5. As noted above, at any finite time, or if the sojourn distributions have finite support, the matrices C are of finite dimension.

6. Parameter Estimation

Recall that with $a_{i,i}(m) = 1 - \Delta^i(m)$, $a_{j,i}(m) = p_{j,i}(m)\Delta^i(m)$ and $N = 3$, then $D(m) = \operatorname{diag}\{a_{1,1}(m), a_{2,2}(m), a_{3,3}(d)\}$, $\Pi(m) = \mathbf{0} + \sum_{\substack{i,j\in\mathcal{M}\\ i\neq j}} a_{i,j}(m)$ and

$$C = \begin{bmatrix} \Pi(1) & \Pi(2) & \Pi(3) & \dots \\ D(1) & 0 & 0 & \dots \\ 0 & D(2) & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

The dynamics of the chain $\overline{X}_k \in \overline{S}$ are given by $\overline{X}_{k+1} = C\,\overline{X}_k + \overline{M}_{k+1} \in \overline{S}$. With $X_k \coloneqq T^*\,\overline{X}_k$ the observation process is given by

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

and for some $\boldsymbol{e}_{i,n}$, this is

$$\begin{aligned} y_k &= \big\langle \boldsymbol{c}, T^*\boldsymbol{e}_{i,n} \big\rangle + \big\langle \boldsymbol{d}, T^*\boldsymbol{e}_{i,n} \big\rangle\, w_k \\ &= \big\langle T\boldsymbol{c}, \boldsymbol{e}_{i,n} \big\rangle + \big\langle T\boldsymbol{d}, \boldsymbol{e}_{i,n} \big\rangle\, w_k, \quad \text{for } i \in \{1,2,\dots,N\} \text{ and } n \in \{1,2,\dots\}. \end{aligned}$$

We wish to estimate the parameters of the model, that is the $\boldsymbol{c}, \boldsymbol{d} \in \mathbb{R}^N$ and the $a_{j,i}(k)$, $1 \le i,\ j \le N$, $i \neq j$. Note we need only estimate the off-diagonal elements $a_{j,i}(k)$ of the matrices $\Pi(k)$.
Now $N_k^{j,i}(m) = \sum_{\ell=1}^k \big\langle \overline{X}_{\ell-1}, \boldsymbol{e}_{i,m} \big\rangle \big\langle \overline{X}_\ell, \boldsymbol{e}_{j,i} \big\rangle$ gives the number of jumps from state $\boldsymbol{e}_{im}$ to state $\boldsymbol{e}_{j,1}$ up to time k. $J_k^i(m) = \sum_{\ell=1}^k \big\langle \overline{X}_{\ell-1}, \boldsymbol{e}_{i,m} \big\rangle$ gives the amount of time spent in state $\boldsymbol{e}_{i,m}$ up to time k. We also need estimates for sums of the form

$$G_k^i = \sum_{\ell=1}^k f(y_\ell) \big\langle X_{\ell-1}, \boldsymbol{e}_i \big\rangle = \sum_{\ell=1}^k \sum_{m=1}^l f(y_\ell) \big\langle X_{\ell-1}, T^*\boldsymbol{e}_{i,m} \big\rangle.$$

Here the function $f(\cdot)$ is any bounded mapping. As in [6] we first consider the unnormalized vector estimate $\sigma\big(N_k^{ji}(m)\overline{X}_k\big) \coloneqq \overline{E}\left[\Lambda_k N_k^{j,i}(m)\,\overline{X}_k \mid \mathcal{Y}_k\right]$, A recursion for this quantity is given by:

Lemma 6.1.

$$\sigma\big(N_k^{j,i}(m)\,\overline{X}_k\big) = \Gamma(y_{k+1}) C \sigma\big(N_k^{j,i}(m) X_k\big) + a_{j,i}(m) \big\langle \gamma_k, \boldsymbol{e}_{i,m} \big\rangle \boldsymbol{e}_{j,i}$$

where γ_k is determined by Theorem 5.4.

Proof. Suppose γ_k and $\sigma\big(N_k^{j,i}(m)X_k\big)$ have been determined. Then

$$\begin{aligned}&\overline{E}\left[\Lambda_{k+1}N_{k+1}^{j,i}(m)\,\overline{X}_{k+1}\mid \mathcal{Y}_{k+1}\right]\\ &\quad=\overline{E}\left[\Lambda_k\lambda_{k+1}\big(N_k^{j,i}(n)+\big\langle\,\overline{X}_{k+1},\boldsymbol{e}_{j,i}\big\rangle\big\langle\,\overline{X}_k,\boldsymbol{e}_{i,m}\big\rangle\big)\,\overline{X}_{k+1}\mid\mathcal{Y}_{k+1}\right]\\ &\quad\text{(similarly to Theorem 5.4 this is)}\\ &\quad=\Gamma(y_{k+1})\sum_{p,q=1}^{m}\overline{E}\Big[\Lambda_kN_k^{j,i}(m)\langle C\,\overline{X}_k,\boldsymbol{e}_{p,q}\rangle|\mathcal{Y}_{k+1}\Big]\boldsymbol{e}_{p,q}+\\ &\quad\overline{E}\Big[\Lambda_k\lambda_{k+1}\,\big\langle\,\overline{X}_{k+1},\boldsymbol{e}_{j,i}\big\rangle\,\big\langle\,\overline{X}_k,\boldsymbol{e}_{i,m}\big\rangle\mid\mathcal{Y}_{k+1}\Big]\boldsymbol{e}_j\,.\end{aligned}$$

The result follows. □

Similarly we can establish:

Lemma 6.2. *With* $\sigma\big(J_k^i(m)\,\overline{X}_k\big)=\overline{E}\left[\Lambda_kJ_k^i(m)\,\overline{X}_k\mid\mathcal{Y}_k\right]$

$$\sigma\big(J_{k+1}^i(m)\,\overline{X}_{k+1}\big)=\Gamma(y_{k+1})C\sigma\big(J_k^i(m)\,\overline{X}_k\big)+\big\langle\gamma_k,\boldsymbol{e}_{i,m}\big\rangle\,\Gamma(y_{k+1})C\boldsymbol{e}_{i,m}\,.$$

Proof.

$$\begin{aligned}&\overline{E}\left[\Lambda_{k+1}J_{k+1}^i(m)\,\overline{X}_{k+1}\mid\mathcal{Y}_{k+1}\right]\\ &\qquad=\overline{E}\left[\Lambda_k\lambda_{k+1}\big(J_k^i(m)+\big\langle\,\overline{X}_k,\boldsymbol{e}_{i,m}\big\rangle\big)\,\overline{X}_{k+1}\mid\mathcal{Y}_{k+1}\right]\\ &\qquad=\Gamma(y_{k+1})C\sigma\big(J_k^i(m)\,\overline{X}_k\big)+\big\langle\gamma_k,\boldsymbol{e}_{i,m}\big\rangle\,\Gamma(y_{k+1})C\boldsymbol{e}_{i,m}\,.\end{aligned}$$

□

In general, with $\sigma(G_k^i\,\overline{X}_k)=\overline{E}\left[\Lambda_kG_k^i\,\overline{X}_k\mid\mathcal{Y}_k\right]$ we have

Lemma 6.3.

$$\sigma\big(G_{k+1}^i\overline{X}_{k+1}\big)=\Gamma\big(y_{k+1}\big)C\sigma\big(G_k^i\,\overline{X}_k\big)+f(y_{k+1})\sum_{m=1}^{k}\big\langle\gamma_k,\boldsymbol{e}_{i,m}\big\rangle C\boldsymbol{e}_{i,m}\,.$$

Remark 6.4. Now $\big\langle\,\overline{X}_k,\mathbf{1}\big\rangle=1$ for all k, where $\mathbf{1}$ is an infinite column vector of 1s. Therefore, for example, $\big\langle\sigma\big(N_k^{j,i}(m)\,\overline{X}_k,\mathbf{1}\big\rangle$ gives an unnormalized estimate for $\sigma\big(N_k^{j,i}(m)\big)$.

In turn these provide estimates such as $\widehat{a}_{j,i}(m)=\dfrac{\sigma\big(N_k^{j,i}(m)\big)}{\sigma\big(J_k^i(m)\big)}$ for $i\neq j$, and for the other parameters of the model as in [4].

7. Smoothers

Suppose $0\le k\le T$ and we have observed $\big\{y_0,y_1,\dots,y_T\big\}$. We wish to find $E\big[X_k\mid\mathcal{Y}_T\big]$. Write $\Lambda_{k+1,T}=\prod_{\ell=k+1}^T\lambda_\ell$. Using Bayes' theorem again, (see [4]) and the reference measure of §5.1.

$$E\big[\,\overline{X}_k\mid\mathcal{Y}_T\big]=\frac{\overline{E}\big[\Lambda_{0,T}\overline{X}_k\mid\mathcal{Y}_T\big]}{\overline{E}\big[\Lambda_{0,T}\mid\mathcal{Y}_T\big]}\,.$$

Now $\Lambda_{0,t} = \Lambda_{0,k}\Lambda_{k+1,T}$ and $\overline{E}\left[\Lambda_{0,T}\overline{X}_k \mid \mathcal{Y}_T\right] = \overline{E}\left[\Lambda_{0,k}X_k\overline{E}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_T, \mathcal{F}_k\right]\mathcal{Y}_T\right]$. However, $\overline{X}$ is Markov so $\overline{E}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_T, \mathcal{F}_k\right] = \overline{E}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_T, \overline{X}_k\right]$.

Definition 7.1. For $1 \le i \le N$ and $n \in \{1, 2, \dots, k+1\}$ write $v^i_{T,T}(n) = 1$ and $v^i_{k,T}(n) = \overline{E}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_T\, , \overline{X}_k = \boldsymbol{e}_{i,n}\right]$. Set

$$v_{k,T} \coloneqq \Big(v^1_{k,t}(1)\dots v^N_{k,T}(1) \mid v^1_{k,T}(2)\dots v^N_{k,T}(2) \mid \times \cdots \mid v^1(k+1)\dots v^N_{k,T}(k+1)\Big).$$

Theorem 7.2. *The process v satisfies the backward dynamics*

$$v_{k,T} = C^*\,\Gamma\big(y_{k+1}\big)\, v_{k,T}, \ \textit{with}\ v_{T,T} = (1, 1, \dots, 1)' \in \mathbb{R}^{(k+1)N}. \tag{7.1}$$

Proof. For $\boldsymbol{e}_{i,n} \in \overline{S}$ consider

$$\begin{aligned}
&\big\langle v_{k,T}, \boldsymbol{e}_{i,n}\big\rangle = v^i_{k,T}(n)\\
&\quad= \overline{E}\Big[\Lambda_{k+2,T}, \lambda_{k+1} \mid \mathcal{Y}_T, \overline{X}_k = \boldsymbol{e}_{i,n}\Big]\\
&\quad= \sum_{m=1}^{k+1}\sum_{i=1}^{N} \overline{E}\Big[\big\langle \overline{X}_{k+1}, \boldsymbol{e}_{j,m}\big\rangle \Lambda_{k+2,T} \mid \mathcal{Y}_T, \overline{X}_k = \boldsymbol{e}_{i,n}\Big]\lambda^j_{k+1}(y_{k+1})\\
&\quad= \sum_{m=1}^{k+1}\sum_{i=1}^{N} \overline{E}\Big[\langle \overline{X}_{k+1}, \boldsymbol{e}_{j,m}\rangle \overline{E}\left[\Lambda_{k+2,T} \ \mid \mathcal{Y}_T, \overline{X}_{k+1} = \boldsymbol{e}_{j,m}, \overline{X}_k = \boldsymbol{e}_{i,n}\right] \Big|\\
&\qquad\quad \mathcal{Y}_T, \overline{X}_k = \boldsymbol{e}_{i,n}\Big]\lambda^j_{k+1}(y_{k+1})\\
&\quad= \sum_{m=1}^{k+1}\sum_{i=1}^{N} \overline{E}\Big[\langle \overline{X}_{k+1}, \boldsymbol{e}_{j,m}\rangle\, \langle v_{k+1,T}, \boldsymbol{e}_{j,m}\rangle | \mathcal{Y}_T, \overline{X}_k = \boldsymbol{e}_{i,n}\Big]\lambda^j_{k+1}(y_{k+1})\\
&\quad= \sum_{m=1}^{k+1}\sum_{i=1}^{N} C_{jm,in}(k)\, \langle v_{k+1,T}, \boldsymbol{e}_{j,m}\rangle\, \lambda^j_{k+1}(y_{k+1}).
\end{aligned}$$

and the result follows. □

Theorem 7.3. *An unnormalized smoothed estimate for $\overline{X}_k$ given observations $\big\{y_0, y_1, \dots y_T\big\}$ is*

$$q_{k,T} = \overline{E}\left[\Lambda_{0,T}\overline{X}_k \mid \mathcal{Y}_T\,\right] = \operatorname{diag}\lambda_k \cdot v_{k,T}\,.$$

Proof.

$$\begin{aligned}
\overline{E}\left[\Lambda_{0,T}\langle \overline{X}_k, \boldsymbol{e}_{i,n}\rangle \mid \mathcal{Y}_T\,\right] &= \overline{E}\Big[\Lambda_{0,k}\,\big\langle \overline{X}_k, \boldsymbol{e}_{i,n}\big\rangle\, \overline{E}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_T, \overline{X}_k = \boldsymbol{e}_{i,n}\right]\mathcal{Y}_T\Big]\\
&= \langle \gamma_k, \boldsymbol{e}_{i,n}\big\rangle \big\langle v_{k,T}, \boldsymbol{e}_{i,n}\big\rangle\,.
\end{aligned}$$

Therefore

$$\overline{E}\left[\Lambda_{0,T}\overline{X}_k \mid \mathcal{Y}_T\right] = \sum_{m=1}^{k+1}\sum_{i=1}^{N} \langle \gamma_k, \boldsymbol{e}_{i,n}\rangle\, \langle v_{k,T}, \boldsymbol{e}_{i,n}\rangle\, \boldsymbol{e}_{i,n} = \operatorname{diag}\gamma_k \cdot v_{k,T}\,.$$

□

8. Simulation Study

8.1. Example Stochastic Dynamics. The indirectly observed X-process we consider is a three-state semi Markov chain with distinct classes for its sojourn distributions. The distributions for the states $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ are, respectively, a Dirac distribution with full mass on 5, a finite distribution on the the natural numbers $\{2,3,5\}$ with corresponding probabilities $\{0.3, 0.5, 0.2\}$ and a geometric distribution with parameters 0.35. The (column stochastic) transition matrix used to simulate an embedded Markov chain, (from which we construct a semi Markov chain realisation) has the form $\begin{bmatrix} 0 & 3/10 & 1/4 \\ 1/3 & 0 & 3/4 \\ 2/3 & 7/10 & 0 \end{bmatrix}$. The initial distribution used for X_0 was uniform across the state space. The $\Delta^i(h)$ probabilities for the model we describe here are listed in Table 1. Given that the sojourns distributions for

Table 1. **End-of-Sojourn Probabilities**

Sojourn :	h=1	h=2	h=3	h=4	h=5	h=6	$\cdots$
$\Delta^1(h)$	0	0	0	0	1	0	$\cdots$
$\Delta^2(h)$	0	0.3	0.71	0	1	0	$\cdots$
$\Delta^3(h)$	0.35	0.35	0.35	0.35	0.35	0.35	$\cdots$

the states $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are both finite, the corresponding support for $\Delta^1(h)$ and $\Delta^2(h)$ is also finite. However, the state $\boldsymbol{e}_3$ has a geometric sojourn, this means $\Delta^3(h)$ is constant on $\mathbb{N}$. The observation dynamics used were given above by equation (4.1), with parameter values $c(\boldsymbol{e}_1) = -1$, $c(\boldsymbol{e}_2) = 1$ and $c(\boldsymbol{e}_3) = 2$, each determined by the state of X at time k, and $d(\boldsymbol{e}_1) = 1.2$, $d(\boldsymbol{e}_2) = 0.4$ and $d(\boldsymbol{e}_3) = 0.2$, also determined by the state of X at time k. The inclusion of one or more geometric sojourns in a semi Markov model (or indeed any other candidate sojourn distribution defined on $\mathbb{N}$) means that the matrix C defined at (3.1) will be an infinite matrix. Consequently a suitable truncation of the matrix C must be used. For the simulation study we assumed that the maximum realized state sojourn (for the geometric distribution) was no more $h_{\text{Max}} = 50$, (for a geometric distribution parameter of 0.35, the event that $h > 50$ has measure approximately equal to 4.42250e-10). Consequently our C matrix had dimensions 150×150.

8.2. Results. The recursive filter given in Theorem 5.4 generates a sequence of unnormalized probabilities distributions $\{\gamma_\ell\}_{\ell \geq 0}$. These unnormalized probabilities are joint distributions for the random variables $X \in \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ and $h \in \{1, 2, \dots, h_{\text{Max}}\}$. The corresponding normalized estimated densities for X and for h are easily recovered from the normalized version of γ_k by marginalisation. In our example we compute Maximum a Posteriori (MAP) estimates from these densities. In Figure 1 we show a realization of the observed process $\{y_\ell\}_{\ell \geq 0}$, the partially observed semi Markov process X and the filtered estimate of X. For clarity, the independent variable on these plots is marked only at the embedded chain transition times, $\tau_0, \tau_1, \ldots$. Similarly, in Figure 2, we plot the exact h-process and the MAP estimates of h marginalized from each γ_k. Comparing the estimates of X in Figure 1 with the estimates of h in Figure 2, we can see that the expected dependence between X and h is clear, as errors in these estimators appear in the same time regions, for example $k \in \{7, 8, 9, 10\}$ and in $k \in \{28, 29, 30\}$. It is encouraging that at the times following these regions the filter has recovered.

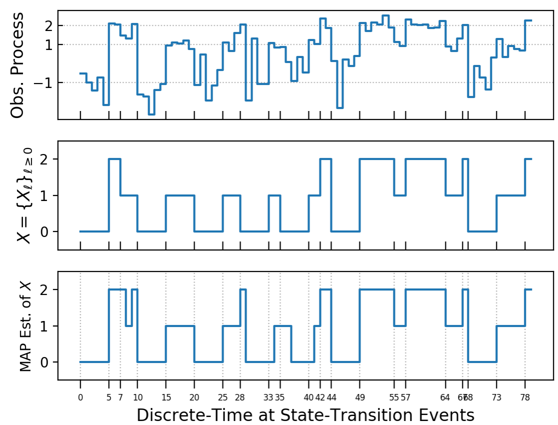

Figure 1. The uppermost plot shows the observation process. The middle plot shows the exact X process. The bottom plot is the filtered estimate of X.

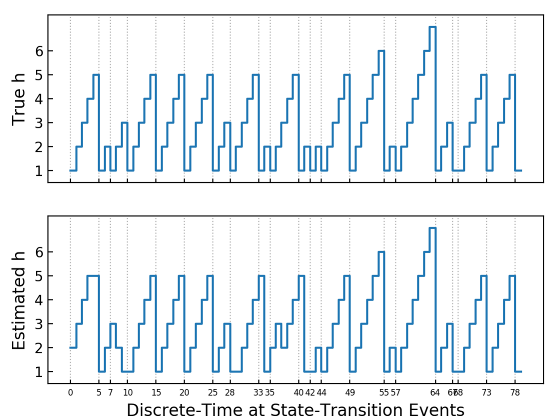

Figure 2. The uppermost plot is the exact h-process. The bottom plot in this figure is the filtered estimate of h.

9. An Exact Filter for the State and State-Sojourn Time

In this section we provide a second, direct, derivation for the recursive filter of Theorem 5.4. Given the semi-Markov chain X and the observation process y of Section 4 we wish to obtain joint conditional estimates of X_k and h_k given $\mathcal{Y}_k = \sigma\{y_0, y_1, \dots, y_k\}$. Suppose $F : \{1, 2, \dots\} \to \mathbb{R}$ is an arbitrary function. We consider $E\big[\langle X_k, \boldsymbol{e}_i\rangle\, F\big(h_k^i\big) \mid \mathcal{Y}_k\big]$ for any $i \in \{1, 2, \dots, N\}$. We wish to find $E\big[\big\langle X_k, \boldsymbol{e}_i\big\rangle F\big(h_k^i\big) \mid \mathcal{Y}_k\big]$. Using the Bayes' rule of [4], this equals $\dfrac{\overline{E}\big[\Lambda_k\big\langle X_k, \boldsymbol{e}_i\big\rangle F\big(h_k^i\big) \mid \mathcal{Y}_k\big]}{\overline{E}\big[\Lambda_k \mid \mathcal{Y}_k\big]}$. The denominator here is derived from the numerator by taking $F = 1$ and summing over i.

Notation 9.1. Suppose there are unnormalized probabilities $\gamma_k^i(n)$ such that

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

However, as noted in Section 3.3, $h_k^i \le k+1$ so $\gamma_k^i(n) = 0$ for $n > k+1$ and the sum here is only up to $n = k+1$. As in section 5 write $\lambda_k^i(y_k) = \dfrac{\phi\big((y_k - c_i)/d_i\big)}{d_i\phi(y_k)}$.

We shall obtain the following recursions for the γ

Theorem 9.2. *For* $n = 1$

$$\gamma_k^i(1) = \lambda_k^i(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{n=1}^{k+1} a_{i,j}(n)\gamma_{k-1}^j(n)\,. \tag{9.1}$$

For $1 < n \le k+1$

$$\gamma_k^i(n) = \lambda_k^i(y_k)a_{i,i}(n-1)\gamma_{k-1}^i(n-1)\,. \tag{9.2}$$

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

$$\begin{aligned}
&\overline{E}[\Lambda_k\langle X_k, \boldsymbol{e}_i\rangle F(h_k^i) \mid \mathcal{Y}_k] = \sum_{n=1}^{k+1} F(n)\gamma_k^i(n)\\
&= \overline{E}\big[\Lambda_{k-1}\lambda_k\langle X_k, \boldsymbol{e}_i\rangle F\big(\langle X_k, \boldsymbol{e}_i\rangle\langle X_k, \boldsymbol{e}_i\rangle\,\langle X_{k-1}, \boldsymbol{e}_i\rangle h_{k-1}^i\big) \mid \mathcal{Y}_k\big]\\
&= \lambda_k^i(y_k)\overline{E}\big[\Lambda_{k-1}\langle X_k, \boldsymbol{e}_i\rangle F\big(1 + \langle X_{k-1}, \boldsymbol{e}_i\rangle h_{k-1}^i\big) \mid \mathcal{Y}_{k-1}\big]\\
&= \lambda_k^i(y_k) \sum_{j=1}^{N} \overline{E}\big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle F\big(1 + \langle X_{k-1}, \boldsymbol{e}_i\rangle h_{k-1}^i\big) \mid \mathcal{Y}_{k-l}\big]\\
&= \lambda_k^i(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} \overline{E}\big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle\langle X_k, \boldsymbol{e}_i\rangle F(1) \mid \mathcal{Y}_{k-1}\big]\\
&\quad + \lambda_k^i(y_k)\overline{E}\left[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle\langle X_k, \boldsymbol{e}_i\rangle F(1 + h_{k-1}^i) \mid \mathcal{Y}_{k-1}\right]
\end{aligned}$$

$$= \lambda_k^i(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} \overline{E}[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle \, \langle A\big(h_{k-1}^j\big)X_{k-1}, \boldsymbol{e}_i\rangle \times$$
$$F(1) \mid \mathcal{Y}_{k-1}\big] + \lambda_k^i(y_k)\overline{E}\left[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle \times\right.$$
$$\langle A\big(h_{k-1}^i\big)X_{k-1}, \boldsymbol{e}_i\rangle F\big(1 + h_{k-1}^i\big) \mid \mathcal{Y}_k\big]$$
$$= \lambda_k^i(y_k) \sum_{\substack{j=1\\ j\neq i}}^{N} F(1)\overline{E}\big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle \, a_{ij}\big(h_{k-1}^j\big) \mid \mathcal{Y}_{k-1}\big]$$
$$+ \lambda_k^i(y_k)\overline{E}\left[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle a_{i,i}\big(h_{k-1}^i\big)F\big(1 + h_{k-1}^i\big) \mid \mathcal{Y}_{k-1}\right]$$
$$= \lambda_k^i(y_k)F(1) \sum_{\substack{j=1\\ j\neq i}}^{n} \overline{E}\big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_j\rangle a_{i,j}\big(h_{k-1}^j\big) \mid \mathcal{Y}_{k-1}\big]$$
$$+ \lambda_k^i(y_k)\overline{E}\big[\Lambda_{k-1}\langle X_{k-1}, \boldsymbol{e}_i\rangle a_{i,i}\big(h_{k-1}^i\big)F\big(1 + h_{k-1}^i\big) \mid \mathcal{Y}_{k-1}\big]$$
$$= \lambda_k^i(y_k)F(1) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{n=1}^{k+1} a_{i,j}(n)\gamma_{k-1}^j(n) + \lambda_k^i(y_k) \sum_{n=1}^{k+1} a_{ii}(n)F(1+n)\gamma_{k-1}^i(n)$$
$$= \lambda_k^i(y_k)F(1) \sum_{\substack{j=1\\ j\neq i}}^{N} \sum_{n=1}^{k+1} a_{i,j}(n)\gamma_{k-1}^j(n) + \lambda_k^i(y_k) \sum_{m=2}^{k+1} a_{i,i}(m-1)F(m)\gamma_{k-1}^i(m-1)$$
$$= \lambda_k^i(y_k)F(1) \sum_{\substack{j=1\\ j\neq i}}^{N} a_{i,j}(n)\gamma_{k-1}^j(n) + \lambda_k^i(y_k) \sum_{n=2}^{k+1} a_{i,i}(n-1)F(n)\gamma_{k-1}^i(n-1). \tag{9.3}$$

Now F is an arbitrary function $F : \{1, 2, \dots\} \to \mathbb{R}$. Consider an F such that $F(1) = 1$ and $F(n) = 0$ if $n \neq 1$. Then from (9.3)

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

This is the recursion for $\gamma_k^i(1)$, the unnormalized conditional probability given $\mathcal{Y}_k$ that at time k $h_k^i(X_k) = 1$ and $x_k = \boldsymbol{e}$. Now consider another F which is such that $F(m) = 1$ for some $m > 1$ and $F(m) = 0$ otherwise. Then from (A.3) and (5.4):

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

This is the recursion for $\gamma_,^i(k-1)$, the unnormalized conditional probability given $\mathcal{Y}_k$ that, at time k, $h_k^i = m$ and $X_k = \boldsymbol{e}_i$. This provides a coordinate-wise version of Theorem 5.4. □

Remark 9.3. Note that, as in the earlier results, the recursions only involve finite sums.

Acknowledgment. Robert Elliott would like to acknowledge the support of the Natural Science and Engineering Research Council of Canada (NSERC). W. P. Malcolm would like to gratefully acknowledge J. D. Ferguson.

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*, Springer Verlag, Lecture Notes in Statistics Number 191, 2008.
3. Cinlar, E.: *Introduction to Stochastic Processes*, Prentice-Hall Inc., Chapter 10, 1975.
4. Elliott, R. J. and Aggoun, L. and Moore, J. B.: *Hidden Markov Models, Estimation and Control*, Springer Verlag, Applications in Mathematics Number 29, 3rd corrected printing, 2008.
5. Elliott, R. J. and Limnios, N. and Swishchuk, A.: Filtering hidden semi-Markov chains, *Statistics and Probability Letters* **83** (2018), 2007–2014.
6. Elliott, R. J. and Malcolm, W. P.: Data recursive smoother formulae for partially observed discrete time Markov chains, *Stochastic Analysis and Applications* **24** (2016), 579–597.
7. Ferguson, J. D.: *Symposium on the application of hidden Markov models to text and speech*, Institute for Defense Analyses, 1980.
8. Koski, T.:*Hidden Markov models for bioinformatics*, Kluwer Academic Publishers, 2001.
9. Krishnamurthy, V. and Moore, J. B. and Chung, H.: On hidden fractal model signal processing, *IEEE Transactions on Signal Processing* **24** (1991), 177–192.
10. van er Hoek, J. and Elliott, R. J.: *Introduction to Semi-Markov Models*, Cambridge University Press, London Mathematical Society Lecture Note Series, Number 445, 2018.

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

E-mail address: Robert.Elliott@unisa.edu.au and relliott@ucalgary.ca

URL: https://people.unisa.edu.au/Robert.Elliott and http://www.ucalgary.ca/~relliott

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

E-mail address: Paul.Malcolm2@defence.gov.au