APPENDIX B

Some proofs

In this appendix we present proofs of five results, shown here in boxes, that are used principally in Section 4.1. None of these results is surprising given the structure of an HMM. Indeed a more intuitive, and less laborious, way to establish such properties is to invoke the separation properties of the directed graph of the model. More precisely, if one can establish that the sets of random variables $\mathbf{A}$ and $\mathbf{B}$ in a directed graphical model are 'd-separated' by the set $\mathbf{C}$, it will then follow that $\mathbf{A}$ and $\mathbf{B}$ are conditionally independent given $\mathbf{C}$; see Pearl (2000, pp. 16–18) or Bishop (2006, pp. 378 and 619). An account of the properties of HMMs that is similar to the approach we follow here is provided by Koski (2001, Chapter 13). We present the results for the case in which the random variables X_t are discrete. Analogous results hold in the continuous case.

B.1 A factorization needed for the forward probabilities

The first purpose of the appendix is to establish the following result, which we use in Section 4.1.1 in order to interpret $\alpha_t(i)$ as the forward probability $\Pr(\mathbf{X}^{(t)}, C_t = i)$. (Recall that $\alpha_t(i)$ was defined as the i th element of $\boldsymbol{\alpha}_t = \boldsymbol{\delta}\mathbf{P}(x_1)\prod_{s=2}^{t}\boldsymbol{\Gamma}\mathbf{P}(x_s)$: see Equation (2.15).)

For positive integers t:

$$\Pr(\mathbf{X}^{(t+1)}, C_t, C_{t+1}) = \Pr(\mathbf{X}^{(t)}, C_t)\ \Pr(C_{t+1} \mid C_t)\ \Pr(X_{t+1} \mid C_{t+1}). \tag{B.1}$$

Throughout this appendix we assume that

$$\Pr(C_t \mid \mathbf{C}^{(t-1)}) = \Pr(C_t \mid C_{t-1})$$

and

$$\Pr(X_t \mid \mathbf{X}^{(t-1)}, \mathbf{C}^{(t)}) = \Pr(X_t \mid C_t).$$

In addition, we assume that these (and other) conditional probabilities are defined, in which case the probabilities that appear as denominators in what follows are strictly positive. The model may be represented, as usual, by the directed graph in Figure B.1.

The tool we use throughout this appendix is the following factorization for the joint distribution of the set of random variables V_i in a directed

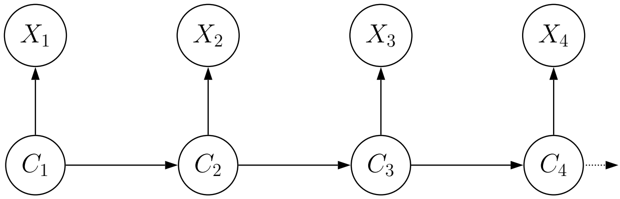

Figure B.1 *Directed graph of basic HMM.*

graphical model, which appeared earlier as Equation (2.5):

$$\Pr(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} \Pr(V_i \mid \mathrm{pa}(V_i)), \tag{B.2}$$

where $\mathrm{pa}(V_i)$ denotes all the parents of V_i in the set $V_1, V_2, \ldots, V_n$. In our model, the only parent of X_k is C_k, and (for $k = 2, 3, \ldots$) the only parent of C_k is C_{k-1}; C_1 has no parent. The joint distribution of $\mathbf{X}^{(t)}$ and $\mathbf{C}^{(t)}$, for instance, is therefore given by

$$\Pr(\mathbf{X}^{(t)}, \mathbf{C}^{(t)}) = \Pr(C_1) \prod_{k=2}^{t} \Pr(C_k \mid C_{k-1}) \prod_{k=1}^{t} \Pr(X_k \mid C_k). \tag{B.3}$$

In order to prove Equation (B.1), note that Equation (B.3) and the analogous expression for $\Pr(\mathbf{X}^{(t+1)}, \mathbf{C}^{(t+1)})$ imply that

$$\Pr(\mathbf{X}^{(t+1)}, \mathbf{C}^{(t+1)}) = \Pr(C_{t+1} \mid C_t)\ \Pr(X_{t+1} \mid C_{t+1})\ \Pr(\mathbf{X}^{(t)}, \mathbf{C}^{(t)}).$$

Now sum over $\mathbf{C}^{(t-1)}$; the result is Equation (B.1). □

Furthermore, (B.1) can be generalized as follows.

For any (integer) $T \geq t + 1$:

$$\Pr(\mathbf{X}_1^T, C_t, C_{t+1}) = \Pr(\mathbf{X}_1^t, C_t)\ \Pr(C_{t+1} \mid C_t)\ \Pr(\mathbf{X}_{t+1}^T \mid C_{t+1}). \tag{B.4}$$

(Recall the notation $\mathbf{X}_a^b = (X_a, X_{a+1}, \ldots, X_b)$.) Briefly, the proof of (B.4) proceeds as follows. First write $\Pr(\mathbf{X}_1^T, \mathbf{C}_1^T)$ as

$$\Pr(C_1) \prod_{k=2}^{T} \Pr(C_k \mid C_{k-1}) \prod_{k=1}^{T} \Pr(X_k \mid C_k),$$

then split each of the two products into $k \le t$ and $k \ge t+1$. Use the fact that

$$\Pr(\mathbf{X}_{t+1}^T, \mathbf{C}_{t+1}^T) = \Pr(C_{t+1}) \prod_{k=t+2}^{T} \Pr(C_k \mid C_{k-1}) \prod_{k=t+1}^{T} \Pr(X_k \mid C_k),$$

and sum $\Pr(\mathbf{X}_1^T, \mathbf{C}_1^T)$ over $\mathbf{C}_{t+2}^T$ and $\mathbf{C}_1^{t-1}$. □

B.2 Two results needed for the backward probabilities

In this section we establish the two results used in Section 4.1.2 in order to interpret $\beta_t(i)$ as the backward probability $\Pr(\mathbf{X}_{t+1}^T \mid C_t = i)$.

The first of these is that,

for $t = 0, 1, \ldots, T-1$,

$$\Pr(\mathbf{X}_{t+1}^T \mid C_{t+1}) = \Pr(X_{t+1} \mid C_{t+1}) \Pr(\mathbf{X}_{t+2}^T \mid C_{t+1}). \tag{B.5}$$

This is established by noting that

$$\begin{aligned} & \Pr(\mathbf{X}_{t+1}^T, \mathbf{C}_{t+1}^T) \\ = \;& \Pr(X_{t+1} \mid C_{t+1}) \left(\Pr(C_{t+1}) \prod_{k=t+2}^{T} \Pr(C_k \mid C_{k-1}) \prod_{k=t+2}^{T} \Pr(X_k \mid C_k) \right) \\ = \;& \Pr(X_{t+1} \mid C_{t+1}) \Pr(\mathbf{X}_{t+2}^T, \mathbf{C}_{t+1}^T), \end{aligned}$$

and then summing over $\mathbf{C}_{t+2}^T$ and dividing by $\Pr(C_{t+1})$. □

The second result is that,

for $t = 1, 2, \ldots, T-1$,

$$\Pr(\mathbf{X}_{t+1}^T \mid C_{t+1}) = \Pr(\mathbf{X}_{t+1}^T \mid C_t, C_{t+1}). \tag{B.6}$$

This we prove as follows. The right-hand side of Equation (B.6) is

$$\frac{1}{\Pr(C_t, C_{t+1})} \sum_{\mathbf{C}_{t+2}^T} \Pr(\mathbf{X}_{t+1}^T, \mathbf{C}_t^T),$$

which by (B.2) reduces to

$$\sum_{\mathbf{C}_{t+2}^T} \prod_{k=t+2}^{T} \Pr(C_k \mid C_{k-1}) \prod_{k=t+1}^{T} \Pr(X_k \mid C_k).$$

The left-hand side is

$$\frac{1}{\Pr(C_{t+1})} \sum_{\mathbf{C}_{t+2}^T} \Pr(\mathbf{X}_{t+1}^T, \mathbf{C}_{t+1}^T),$$

which reduces to the same expression. □

B.3 Conditional independence of $\mathbf{X}_1^t$ and $\mathbf{X}_{t+1}^T$

Here we establish the conditional independence of $\mathbf{X}_1^t$ and $\mathbf{X}_{t+1}^T$ given C_t, used in Section 4.1.3 to link the forward and backward probabilities to the probabilities $\Pr(\mathbf{X}^{(T)} = \mathbf{x}^{(T)}, C_t = i)$. That is, we show that,

for $t = 1, 2, \ldots, T-1$,

$$\Pr(\mathbf{X}_1^T \mid C_t) = \Pr(\mathbf{X}_1^t \mid C_t)\ \Pr(\mathbf{X}_{t+1}^T \mid C_t). \tag{B.7}$$

To prove this, first note that

$$\Pr(\mathbf{X}_1^T, \mathbf{C}_1^T) = \Pr(\mathbf{X}_1^t, \mathbf{C}_1^t) \frac{1}{\Pr(C_t)} \Pr(\mathbf{X}_{t+1}^T, \mathbf{C}_t^T),$$

which follows by repeated application of Equation (B.2). Then sum over $\mathbf{C}_1^{t-1}$ and $\mathbf{C}_{t+1}^T$. This yields

$$\Pr(\mathbf{X}_1^T, C_t) = \Pr(\mathbf{X}_1^t, C_t) \frac{1}{\Pr(C_t)} \Pr(\mathbf{X}_{t+1}^T, C_t),$$

from which the result is immediate. □