Eye movement analysis with switching hidden Markov models: Appendix

Tim Chuk[1], Antoni B. Chan[2], Shinsuke Shimojo[3], and Janet H. Hsiao[1]
[1] *Department of Psychology, University of Hong Kong*
[2] *Department of Computer Science, City University of Hong Kong*
[3] *Department of Biology and Biological Engineering, California Institute of Technology*

This is the appendix for the paper "Eye movement analysis with switching hidden Markov models", which appears in *Behavior Research Methods* (2020) 52:1026-1043.

A. Derivation of Switching HMM

The switching HMM is an HMM with an additional high-level state sequence that selects the currently active transition matrix. Formally, at time t for the nth sequence, let $z_{n,t} = \{1, \cdots, K\}$ be the hidden-state variable, $s_{n,t} = \{1, \cdots, S\}$ be the high-level state variable, and $x_{n,t}$ be the observation. The high-level state variable and hidden state variable are both 1st-order Markov chains. The transition matrix for the hidden state variable $z_{n,t}$ depends on the current high-level state $s_{n,t}$,

$$p(\mathbf{s}_n) = p(s_{n,1}) \prod_{t=2}^{\tau_n} p(s_{n,t}|s_{n,t-1}), \tag{1}$$

$$p(\mathbf{z}_n|\mathbf{s}_n) = p(z_{n,1}|s_{n,1}) \prod_{t=2}^{\tau_n} p(z_{n,t}|z_{n,t-1}, s_{n,t}) \tag{2}$$

where the individual probability distributions are

$$\text{initial high-level state:} \qquad p(s_{n,1} = j) = \rho_j \tag{3}$$

$$\text{high-level transition probability:} \qquad p(s_{n,t} = j'|s_{n,t-1} = j) = b_{j,j'} \tag{4}$$

$$\text{initial state:} \qquad p(z_{n,1} = k|s_{n,1} = j) = \pi_k^{(j)} \tag{5}$$

$$\text{transition probability:} \qquad p(z_{n,t} = k'|z_{n,t-1} = k, s_{n,t} = j) = a_{k,k'}^{(j)}. \tag{6}$$

Note that we will always use j for the high-level state (s), and k for the lower hidden state (z). In some cases, we may want to set $\rho_1 = 1$ and $\rho_j = 0$ for $j \neq 1$, which will force the SHMM to always start in the same high-level state.

The emission densities are Gaussian, and only depend on the current lower hidden state (i.e., are shared among high-level states),

$$\text{observation likelihood:} \qquad p(\mathbf{x}_{n,t}|z_{n,t} = k) = \mathcal{N}(\mathbf{x}_{n,t}|\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}), \tag{7}$$

where $(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k)$ are the mean and precision matrix of a Gaussian.

Finally, the joint probability model is

$$p(\mathbf{X}, \mathbf{Z}, \mathbf{S}) = \prod_{n=1}^{D} \left[p(s_{n,1}) p(z_{n,1}|s_{n,1}) p(x_{n,1}|z_{n,1}) \prod_{t=2}^{\tau_n} p(s_{n,t}|s_{n,t-1}) p(z_{n,t}|z_{n,t-1}, s_{n,t}) p(x_{n,t}|z_{n,t}). \right] \tag{8}$$

A.1 Equivalence to standard HMM

The switching HMM can be turned into a standard HMM by combining the high-level state variable and hidden state variable into a single hidden state variable. This can be seen by looking at their joint distribution,

$$p(\mathbf{z}_n, \mathbf{s}_n) = p(\mathbf{z}_n|\mathbf{s}_n) p(\mathbf{s}_n) \tag{9}$$

$$= p(z_{n,1}|s_{n,1}) p(s_{n,1}) \prod_{t=2}^{\tau_n} p(z_{n,t}|z_{n,t-1}, s_{n,t}) p(s_{n,t}|s_{n,t-1}) \tag{10}$$

$$= p(s_{n,1}, z_{n,1}) \prod_{t=2}^{\tau_n} p(s_{n,t}, z_{n,t}|s_{n,t-1}, z_{n,t-1}), \tag{11}$$

where

$$\text{initial state:} \quad p(s_{n,1} = j, z_{n,1} = k) = \rho_j \pi_k^{(j)}, \tag{12}$$

$$\text{transition probability:} \quad p(s_{n,t} = j', z_{n,t} = k'|s_{n,t-1} = j, z_{n,t-1} = k) = b_{j,j'} a_{k,k'}^{(j')}. \tag{13}$$

Hence, an equivalent HMM can be formed by defining an augmented set of hidden states $\tilde{z}_{n,t}$ that takes a state value pair (j, k), where j is the high-level state and k is the low-level hidden state. The transition matrix and initial state probability take a special form,

$$p(\tilde{z}_{n,t} = (j', k')|\tilde{z}_{n,t-1} = (j, k)) = \tilde{a}_{(j,k),(j',k')} = b_{j,j'} a_{k,k'}^{(j')}, \tag{14}$$

$$p(\tilde{z}_{n,1} = (j, k)) = \tilde{\pi}_{(j,k)} = \rho_j \pi_k^{(j)}. \tag{15}$$

Note that this is equivalent to defining an HMM with SK hidden states, where the pair (j, k) is mapped to a single index via $\tilde{k} = k + (j-1)S$. The transition matrix is a block matrix

$$\tilde{\mathbf{A}} = \begin{bmatrix} b_{1,1}\mathbf{A}^{(1)} & b_{1,2}\mathbf{A}^{(2)} & \cdots \\ b_{2,1}\mathbf{A}^{(1)} & b_{2,2}\mathbf{A}^{(2)} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{16}$$

where $\mathbf{A}^{(j)} = [a_{k,k'}^{(j)}]_{k,k'}$ is the transition matrix for high-level state j, and the initial state probabilities are

$$\tilde{\boldsymbol{\pi}} = \begin{bmatrix} \rho_1 \boldsymbol{\pi}^{(1)} \\ \rho_2 \boldsymbol{\pi}^{(2)} \\ \vdots \end{bmatrix}, \tag{17}$$

where $\boldsymbol{\pi}^{(j)} = [\pi_1^{(j)}, \cdots, \pi_K^{(j)}]^T$ is the prior state probability vector for high-level state j.

Finally, the Gaussian emissions are shared among high-levle states, i.e., do not depend on the high-level level state value j,

$$p(\mathbf{x}_{n,t}|\tilde{z}_{n,t} = (j,k)) = \mathcal{N}(\mathbf{x}_{n,t}|\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}). \tag{18}$$

A.2 Parameter estimation with EM

The SHMM parameters can be estimated by modifying the EM algorithm for HMMs. Here we drop the tilde notation for z to reduce clutter. Define the indicator variables $z_{n,t,(j,k)}$ which is 1 if and only if $z_{n,t} = (j,k)$.

A.2.1 Complete data likelihood and $\mathcal{Q}$ function

Using the indicator variable trick, the complete data log-likelihood is

$$\log p(\mathbf{X}, \mathbf{Z}) = \log p(\mathbf{Z}) + \log p(\mathbf{X}|\mathbf{Z}) \tag{19}$$

$$= \sum_{n=1}^{N} \left[\log p(z_{n,1}) + \sum_{t=2}^{\tau} \log p(z_{n,t}|z_{n,t-1}) + \sum_{t=1}^{\tau} \log p(\mathbf{x}_{n,t}|z_{n,t}) \right] \tag{20}$$

$$= \sum_{n=1}^{N} \Bigg[\sum_{(j,k)} z_{n,1,(j,k)} \log \tilde{\pi}_{(j,k)} + \sum_{t=2}^{\tau} \sum_{(j,k)} \sum_{(j',k')} z_{n,t,(j',k')} z_{n,t-1,(j,k)} \log \tilde{a}_{(j,k),(j',k')} + \sum_{t=1}^{\tau} \sum_{(j,k)} z_{n,t,(j,k)} \log p(\mathbf{x}_{n,t}|z_{n,t} = (j,k)) \Bigg] \tag{21}$$

Taking the conditional expectation, we obtain the $\mathcal{Q}$ function for the EM algorithm,

$$\mathcal{Q} = \sum_{n=1}^{N} \sum_{(j,k)} r_{n,1,(j,k)} \log \tilde{\pi}_{(j,k)} + \sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{(j,k)} \sum_{(j',k')} \gamma_{n,t,(j,k),(j',k')} \log \tilde{a}_{(j,k),(j',k')} + \sum_{n=1}^{N} \sum_{t=1}^{\tau} \sum_{(j,k)} r_{n,t,(j,k)} \log p(\mathbf{x}_{n,t}|z_{n,t} = (j,k)), \tag{22}$$

where $r_{n,t,(j,k)}$ is the responsibility of observing the Gaussian for state (j,k), and $\gamma_{n,t,(j,k),(j',k')}$ is the transition responsibility between (j,k) to (j',k'),

$$r_{n,t,(j,k)} = \mathbb{E}_{z_{n,t}|\mathbf{X}}[z_{n,t,(j,k)}] = p(z_{n,t} = (j,k)|\mathbf{X}) \tag{23}$$

$$\gamma_{n,t,(j,k),(j',k')} = \mathbb{E}_{z_{n,t},z_{n,t-1}|\mathbf{X}}[z_{n,t-1,(j,k)} z_{n,t,(j',k')}] = p(z_{n,t-1} = (j,k), z_{n,t} = (j',k')|\mathbf{X}). \tag{24}$$

A.2.2 E-Step

In the E-step, the responsibilities in (23) and (24) are calculated using the standard forward-backward algorithm for HMMs with the augmented transition matrix, initial state vector, and emission densities given in (16), (17), and (18).

A.2.3 M-Step

The M-step is similar to standard HMMs, except now we need to consider that the parameters have a particular form to make it equivalent to an SHMM.

Looking at the prior term in the $\mathcal{Q}$ function,

$$\mathcal{Q} = \sum_{n=1}^{N} \sum_{(j,k)} r_{n,1,(j,k)} \log \tilde{\pi}_{(j,k)} = \sum_{n=1}^{N} \sum_{(j,k)} r_{n,1,(j,k)} \log \rho_j \pi_k^{(j)} \tag{25}$$

$$= \sum_{n=1}^{N} \sum_{(j,k)} r_{n,1,(j,k)} \log \rho_j + \sum_{n=1}^{N} \sum_{(j,k)} r_{n,1,(j,k)} \log \pi_k^{(j)} \tag{26}$$

$$= \sum_{j} \left(\sum_{n=1}^{N} \sum_{k} r_{n,1,(j,k)} \right) \log \rho_j + \sum_{j} \sum_{k} \left(\sum_{n=1}^{N} r_{n,1,(j,k)} \right) \log \pi_k^{(j)}. \tag{27}$$

Define the summed responsibilities,

$$N_{1,(j,k)} = \sum_{n=1}^{N} r_{n,1,(j,k)}, \tag{28}$$

$$L_{1,j} = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{n,1,(j,k)} = \sum_{k=1}^{K} N_{1,(j,k)}. \tag{29}$$

Maximizing with respect to ρ_j and $\pi_k^{(j)}$, and noting that they must sum to 1 (over j and k respectively), yields the parameter updates,

$$\rho_j = \frac{L_{1,j}}{\sum_{j'=1}^{S} L_{1,j'}}, \tag{30}$$

$$\pi_k^{(j)} = \frac{N_{1,(j,k)}}{\sum_{k'=1}^{K} N_{1,(j,k')}} = \frac{N_{1,(j,k)}}{L_{1,j}}. \tag{31}$$

Looking at the transition matrix in the $\mathcal{Q}$ function,

$$\mathcal{Q} = \sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{(j,k)} \sum_{(j',k')} \gamma_{n,t,(j,k),(j',k')} \log \tilde{a}_{(j,k),(j',k')} \tag{32}$$

$$= \sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{(j,k)} \sum_{(j',k')} \gamma_{n,t,(j,k),(j',k')} \log b_{j,j'} a_{k,k'}^{(j')} \tag{33}$$

$$= \sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{(j,k)} \sum_{(j',k')} \gamma_{n,t,(j,k),(j',k')} \log b_{j,j'} + \sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{(j,k)} \sum_{(j',k')} \gamma_{n,t,(j,k),(j',k')} \log a_{k,k'}^{(j')} \tag{34}$$

$$= \sum_{j} \sum_{j'} \left(\sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{k} \sum_{k'} \gamma_{n,t,(j,k),(j',k')} \right) \log b_{j,j'} + \sum_{j'} \sum_{k} \sum_{k'} \left(\sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{j} \gamma_{n,t,(j,k),(j',k')} \right) \log a_{k,k'}^{(j')} \tag{35}$$

Define the summed responsibilities,

$$O_{j,j'} = \sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{k=1}^{K} \sum_{k'=1}^{K} \gamma_{n,t,(j,k),(j',k')}, \tag{36}$$

$$M_{k,(j',k')} = \sum_{n=1}^{N} \sum_{t=2}^{\tau} \sum_{j=1}^{S} \gamma_{n,t,(j,k),(j',k')}. \tag{37}$$

Maximizing the $\mathcal{Q}$ function w.r.t. the parameters gives

$$b_{j,j'} = \frac{O_{j,j'}}{\sum_{l=1}^{S} O_{j,l}}, \tag{38}$$

$$a_{k,k'}^{(j')} = \frac{M_{k,(j',k')}}{\sum_{l=1}^{K} M_{k,(j',l)}}. \tag{39}$$

Looking at the Gaussian term in the $\mathcal{Q}$ function,

$$\mathcal{Q} = \sum_{n=1}^{N} \sum_{t=1}^{\tau} \sum_{(j,k)} r_{n,t,(j,k)} \log p(\mathbf{x}_{n,t} | z_{n,t} = (j,k)) \tag{40}$$

$$= \sum_{n=1}^{N} \sum_{t=1}^{\tau} \sum_{(j,k)} r_{n,t,(j,k)} \mathcal{N}(\mathbf{x}_{n,t} | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}) \tag{41}$$

$$= \sum_{k} \sum_{n=1}^{N} \sum_{t=1}^{\tau} \left(\sum_{j} r_{n,t,(j,k)} \right) \mathcal{N}(\mathbf{x}_{n,t} | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}), \tag{42}$$

which is the same form as standard HMM but with the responsibilities summed over j first. Defining

$$\hat{r}_{n,t,k} = \sum_{j=1}^{S} r_{n,t,(j,k)}, \tag{43}$$

$$N_k = \sum_{n=1}^{N} \sum_{t=1}^{\tau} \hat{r}_{n,t,k}, \tag{44}$$

the parameter updates are

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \sum_{t=1}^{\tau} \hat{r}_{n,t,k} \mathbf{x}_{n,t}, \tag{45}$$

$$\boldsymbol{\Lambda}_k^{-1} = \frac{1}{N_k} \sum_{n=1}^{N} \sum_{t=1}^{\tau} \hat{r}_{n,t,k} (\mathbf{x}_{n,t} - \boldsymbol{\mu}_k)(\mathbf{x}_{n,t} - \boldsymbol{\mu}_k)^T. \tag{46}$$