Hidden Markov Models: Exercises

Xavier Didelot

Exercise 1

Let X be a random variable distributed as a δ_1, δ_2 mixture of two distributions with expectations μ_1, μ_2, and variances σ_1^2 and σ_2^2, respectively, where $\delta_1 + \delta_2 = 1$.

1. Show that $\mathrm{E}(x) = \delta_1\mu_1 + \delta_2\mu_2$
2. Show that $\mathrm{Var}(x) = \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1\delta_2(\mu_1 - \mu_2)^2$
3. Show that a mixture of two Poisson distributions is overdispersed, that is $\mathrm{Var}(x) > \mathrm{E}(x)$
4. Generalize to a mixture of $m \geq 2$ Poisson distributions

Answer

1. Let C denote the random variable that performs the mixing, ie $C = 1$ with probability δ_1 and $C = 2$ with probability δ_2. Let X_1 and X_2 denote the random variables associated with each component, so that $\mathrm{E}(X_1) = \mu_1$, $\mathrm{E}(X_2) = \mu_2$, $\mathrm{Var}(X_1) = \sigma_1^2$ and $\mathrm{Var}(X_2) = \sigma_2^2$. We have:

$$\mathrm{E}(X) = \int xp(X = x)dx = \int x(p(X = x|C = 1)p(C = 1) + p(X = x|C = 2)p(C = 2))dx$$
$$= \delta_1 \int xp(X_1 = x)dx + \delta_2 \int xp(X_2 = x)dx = \delta_1\mu_1 + \delta_2\mu_2$$

2. Likewise we have:

$$\mathrm{E}(X^2) = \delta_1\mathrm{E}(X_1^2) + \delta_2\mathrm{E}(X_2^2)$$

   and therefore:

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1\delta_2(\mu_1 - \mu_2)^2$$

3. In this case, we have $\mu_1 = \sigma_1^2 = \lambda_1$ and $\mu_2 = \sigma_2^2 = \lambda_2$. Therefore:

$$\mathrm{Var}(X) = \delta_1\lambda_1 + \delta_2\lambda_2 + \delta_1\delta_2(\lambda_1 - \lambda_2)$$
$$\mathrm{E}(X) = \delta_1\lambda_1 + \delta_2\lambda_2$$

4. Proof by induction, using the fact that a mixture of three Poissons is a mixture of one Poisson with a mixture of two Poissons, etc.

Exercise 2

Consider a stationary two-state Markov chain with transition probability matrix given by:

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

1. Show that the stationary distribution is:

$$(\delta_1, \delta_2) = \frac{1}{\gamma_{12} + \gamma_{21}}(\gamma_{21}, \gamma_{12})$$

2. Consider the case
$$\mathbf{\Gamma} = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$
The following two sequences of observations are assumed to be generated by this Markov chain:

Sequence A: 1 1 1 2 2 1
Sequence B: 2 1 1 2 1 1

Compute the probability of each of the sequences. Note that each sequence contains the same number of ones and twos. Why are these sequences not equally probable?

Answer

1. Stationarity implies $\boldsymbol{\delta\Gamma} = \boldsymbol{\delta}$ so that:
$$\delta_1\gamma_{11} + \delta_2\gamma_{21} = \delta_1 \text{ and } \delta_1\gamma_{21} + \delta_2\gamma_{22} = \delta_2$$
The stationary distribution is found by replacing one of the equations in the system with $\delta_1 + \delta_2 = 1$ and then solving for $\boldsymbol{\delta}$. So we have:
$$\delta_1\gamma_{11} + \delta_2\gamma_{21} = \delta_1 \text{ and } \delta_1 + \delta_2 = 1$$
and therefore:
$$\delta_1 = \frac{\gamma_{21}}{\gamma_{12} + \gamma_{21}} \text{ and } \delta_2 = \frac{\gamma_{12}}{\gamma_{12} + \gamma_{21}}$$

2. The probability of sequence A is $\gamma_{11}\gamma_{11}\gamma_{12}\gamma_{22}\gamma_{21} = 0.01296$. The probability of sequence B is $\gamma_{21}\gamma_{11}\gamma_{12}\gamma_{21}\gamma_{11} = 0.0396$. The sequences have different probability because they consist of different transitions between the states of the Markov process. In other words, the ordering of the sequences matters, unlike an independent mixture model.

Exercise 3

Consider a stationary two-state Poisson-HMM with parameters
$$\mathbf{\Gamma} = \begin{pmatrix} 0.1 & 0.9 \\ 0.4 & 0.6 \end{pmatrix} \text{ and } \boldsymbol{\lambda} = (1, 3)$$
We observe the data $X = (X_1, X_2, X_3) = (0, 2, 1)$.

1. Compute the probability of the observation by considering all possible sequences of state of the Markov chain that could have occurred.
2. Compute the probability of the observation using the relevant matrix form of the HMM likelihood equation.

Answer

1. First compute the stationary distribution $\boldsymbol{\delta} = \left(\frac{4}{13}, \frac{9}{13}\right)$. Then we consider all eight possible state sequences that can occur in a two-state Markov chain in three steps:

i	j	k	$p_i(0)$	$p_j(2)$	$p_k(1)$	δ_i	γ_{ij}	γ_{jk}	product
1	1	1	0.3679	0.1839	0.3679	$\frac{4}{13}$	$\frac{1}{10}$	$\frac{1}{10}$	$7.66 \cdot 10^{-5}$
1	1	2	0.3679	0.1839	0.1494	$\frac{4}{13}$	$\frac{1}{10}$	$\frac{9}{10}$	$2.80 \cdot 10^{-4}$
1	2	1	0.3679	0.2240	0.3679	$\frac{4}{13}$	$\frac{9}{10}$	$\frac{4}{10}$	$3.36 \cdot 10^{-3}$
1	2	2	0.3679	0.2240	0.1494	$\frac{4}{13}$	$\frac{9}{10}$	$\frac{6}{10}$	$2.05 \cdot 10^{-3}$
2	1	1	0.0498	0.1839	0.3679	$\frac{9}{13}$	$\frac{4}{10}$	$\frac{1}{10}$	$9.33 \cdot 10^{-5}$
2	1	2	0.0498	0.1839	0.1494	$\frac{9}{13}$	$\frac{4}{10}$	$\frac{9}{10}$	$3.41 \cdot 10^{-4}$
2	2	1	0.0498	0.2240	0.3679	$\frac{9}{13}$	$\frac{6}{10}$	$\frac{4}{10}$	$6.82 \cdot 10^{-4}$
2	2	2	0.0498	0.2240	0.1494	$\frac{9}{13}$	$\frac{6}{10}$	$\frac{6}{10}$	$4.15 \cdot 10^{-4}$
Total probability									0.0073

2. The likelihood equation is:

$$p(X_1 = 0, X_2 = 2, X_3 = 1) = \boldsymbol{\delta} \boldsymbol{P}(0)\boldsymbol{\Gamma}\boldsymbol{P}(2)\boldsymbol{\Gamma}\boldsymbol{P}(1)\boldsymbol{1}' \text{ with } \boldsymbol{P}(x) = \begin{pmatrix} \lambda_1^x e^{-\lambda_1}/x! & 0 \\ 0 & \lambda_2^x e^{-\lambda_2}/x! \end{pmatrix}$$

   Carrying out the calculations one obtains:

$$p(X_1 = 0, X_2 = 2, X_3 = 1) = \boldsymbol{\delta} \boldsymbol{P}(0)\boldsymbol{\Gamma}\boldsymbol{P}(2)\boldsymbol{\Gamma}\boldsymbol{P}(1)\boldsymbol{1}' = 0.0073$$

Exercise 4

Consider a stationary two-state Poisson-HMM with parameters

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.1 & 0.9 \\ 0.4 & 0.6 \end{pmatrix} \text{ and } \boldsymbol{\lambda} = (1, 3)$$

We observe the data $X = (X_1, X_3) = (0, 1)$. Note that X_2 is missing data.

1. Compute the probability of the observation by considering all possible sequences of state of the Markov chain that could have occurred.
2. Compute the probability of the observation using the relevant matrix form of the HMM likelihood equation.

Answer

1. First compute the stationary distribution $\boldsymbol{\delta} = \left(\frac{4}{13}, \frac{9}{13}\right)$. Then we consider all eight possible state sequences that can occur in a two-state Markov chain in three steps:

i	j	k	$p_i(0)$	$p_k(1)$	δ_i	γ_{ij}	γ_{jk}	product
1	1	1	0.3679	0.3679	$\frac{4}{13}$	$\frac{1}{10}$	$\frac{1}{10}$	$4.16 \cdot 10^{-4}$
1	1	2	0.3679	0.1494	$\frac{4}{13}$	$\frac{1}{10}$	$\frac{9}{10}$	$1.52 \cdot 10^{-3}$
1	2	1	0.3679	0.3679	$\frac{4}{13}$	$\frac{9}{10}$	$\frac{4}{10}$	$1.50 \cdot 10^{-2}$
1	2	2	0.3679	0.1494	$\frac{4}{13}$	$\frac{9}{10}$	$\frac{6}{10}$	$9.13 \cdot 10^{-3}$
2	1	1	0.0498	0.3679	$\frac{9}{13}$	$\frac{4}{10}$	$\frac{1}{10}$	$5.07 \cdot 10^{-4}$
2	1	2	0.0498	0.1494	$\frac{9}{13}$	$\frac{4}{10}$	$\frac{9}{10}$	$1.85 \cdot 10^{-3}$
2	2	1	0.0498	0.3679	$\frac{9}{13}$	$\frac{6}{10}$	$\frac{4}{10}$	$3.04 \cdot 10^{-3}$
2	2	2	0.0498	0.1494	$\frac{9}{13}$	$\frac{6}{10}$	$\frac{6}{10}$	$1.85 \cdot 10^{-3}$
Total probability								0.0333

2. The likelihood equation is:

$$p(X_1 = 0, X_3 = 1) = \boldsymbol{\delta} \boldsymbol{P}(0)\boldsymbol{\Gamma}^2\boldsymbol{P}(1)\boldsymbol{1}' \text{ with } \boldsymbol{P}(x) = \begin{pmatrix} \lambda_1^x e^{-\lambda_1}/x! & 0 \\ 0 & \lambda_2^x e^{-\lambda_2}/x! \end{pmatrix}$$

   Carrying out the calculations one obtains:

$$p(X_1 = 0, X_3 = 1) = \boldsymbol{\delta} \boldsymbol{P}(0)\boldsymbol{\Gamma}^2\boldsymbol{P}(1)\boldsymbol{1}' = 0.0333$$

Exercise 5

Show that the general expression for the likelihood of a stationary HMM simplifies into the expression for the likelihood of an independent mixture model if the transition matrix $\mathbf{\Gamma}$ is such that the current state does not depend on the previous one.

Answer

For a stationary HMM, the general expression for the likelihood is:

$$L_T = p(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}) = \boldsymbol{\delta} \boldsymbol{P}(x_1) \mathbf{\Gamma} \boldsymbol{P}(x_2) \mathbf{\Gamma} \boldsymbol{P}(x_3) ... \mathbf{\Gamma} \boldsymbol{P}(x_T) \mathbf{1'}$$

To make the current HMM state independent of the previous one, we need:

$$\mathbf{\Gamma} = \begin{bmatrix} \delta_1 & \delta_2 & \dots & \delta_m \\ \delta_1 & \delta_2 & \dots & \delta_m \\ \vdots & \vdots & \ddots & \delta_m \\ \delta_1 & \delta_2 & \dots & \delta_m \end{bmatrix}$$

We expand the HMM likelihood equation from right to left, starting on the right with $\mathbf{1'}$. The product $\boldsymbol{P}(x)\mathbf{1'}$ is the column vector with i-th term equal to $p_i(x)$. The product $\mathbf{\Gamma} \boldsymbol{P}(x)\mathbf{1'}$ is therefore equal to $\sum_{i=1}^{m} \delta_i p_i(x)$ times $\mathbf{1'}$. We repeat for each x_j for j from T down to 2 to obtain:

$$L_T = \boldsymbol{\delta} \boldsymbol{P}(x_1) \mathbf{1'} \prod_{j=2}^{T} \sum_{i=1}^{m} \delta_i p_i(x_j)$$

and since $\boldsymbol{\delta} \boldsymbol{P}(x_1) \mathbf{1'} = \sum_{i=1}^{m} \delta_i p_i(x_1)$ we conclude that:

$$L_T = \prod_{j=1}^{T} \sum_{i=1}^{m} \delta_i p_i(x_j)$$

This is indeed the likelihood of an independent mixture model.

Exercise 6

Simplify the general expression for the likelihood of a stationary HMM in the case where for every state i and every observation x we have $p_i(x) = p(x)$.

Answer

For a stationary HMM, the general expression for the likelihood is:

$$L_T = p(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}) = \boldsymbol{\delta} \boldsymbol{P}(x_1) \mathbf{\Gamma} \boldsymbol{P}(x_2) \mathbf{\Gamma} \boldsymbol{P}(x_3) ... \mathbf{\Gamma} \boldsymbol{P}(x_T) \mathbf{1'}$$

Since for every i and x we have $p_i(x) = p(x)$ we deduce $\boldsymbol{P}(x) = p(x)\boldsymbol{I}$, and therefore:

$$L_T = \left(\prod_{j=1}^{T} p(x_j) \right) \boldsymbol{\delta} \mathbf{\Gamma}^T \mathbf{1'} = \prod_{j=1}^{T} p(x_j)$$

This is the likelihood of iid observations from distribution $p(x)$.

Exercise 7

Consider a stationary m-state Poisson-HMM $\{X_t\}$ with transition probability matrix $\mathbf{\Gamma}$ and state-dependent means $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_m)$. Let $\boldsymbol{\delta}$ be the stationary distribution of the Markov chain. Let $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$.

1. Show that $\text{E}(X_t) = \boldsymbol{\delta}\boldsymbol{\lambda}'$
2. Show that $\text{E}(X_t^2) = \boldsymbol{\delta}\mathbf{\Lambda}\boldsymbol{\lambda}' + \boldsymbol{\delta}\boldsymbol{\lambda}'$
3. Show that $\text{Var}(X_t) = \boldsymbol{\delta}\mathbf{\Lambda}\boldsymbol{\lambda}' + \boldsymbol{\delta}\boldsymbol{\lambda}' - (\boldsymbol{\delta}\boldsymbol{\lambda}')^2$
4. Show that X_t is overdispersed, ie $\text{Var}(X_t) > \text{E}(X_t)$
5. Show that $\text{E}(X_t X_{t+k}) = \boldsymbol{\delta}\mathbf{\Lambda}\mathbf{\Gamma}^k\boldsymbol{\lambda}'$
6. Show that $\text{Corr}(X_t, X_{t+k}) = \frac{\boldsymbol{\delta}\mathbf{\Lambda}\mathbf{\Gamma}^k\boldsymbol{\lambda}' - (\boldsymbol{\delta}\boldsymbol{\lambda}')^2}{\boldsymbol{\delta}\mathbf{\Lambda}\boldsymbol{\lambda}' + \boldsymbol{\delta}\boldsymbol{\lambda}' - (\boldsymbol{\delta}\boldsymbol{\lambda}')^2}$

Answer

1.
$$\text{E}(X_t) = \sum_{i=1}^{m} \text{E}(X_t|C_t = i)p(C_t = i) = \sum_{i=1}^{m} \delta_i\lambda_i = \boldsymbol{\delta}\boldsymbol{\lambda}'$$

2.
$$\text{E}(X_t^2) = \sum_{i=1}^{m} \text{E}(X_t^2|C_t = i)p(C_t = i) = \sum_{i=1}^{m} \delta_i(\lambda_i^2 + \lambda_i) = \boldsymbol{\delta}\mathbf{\Lambda}\boldsymbol{\lambda}' + \boldsymbol{\delta}\boldsymbol{\lambda}'$$

3.
$$\text{Var}(X_t) = \text{E}(X_t^2) - \text{E}(X_t)^2 = \boldsymbol{\delta}\mathbf{\Lambda}\boldsymbol{\lambda}' + \boldsymbol{\delta}\boldsymbol{\lambda}' - (\boldsymbol{\delta}\boldsymbol{\lambda}')^2$$

4.
$$\text{Var}(X_t) - \text{E}(X_t) = \boldsymbol{\delta}\mathbf{\Lambda}\boldsymbol{\lambda}' - (\boldsymbol{\delta}\boldsymbol{\lambda}')^2 = \sum_{i=1}^{m} (\delta_i\lambda_i^2)(1 - \delta_i) > 0$$

5.
$$\text{E}(X_t X_{t+k}) = \sum_{i=1}^{m}\sum_{j=1}^{m} \text{E}(X_t X_{t+k}|C_t = i, C_{t+k} = j)p(C_t = i, C_{t+k} = j)$$

   Since X_t and X_{t+k} are independent given C_t and C_{t+k} we have:

$$\text{E}(X_t X_{t+k}) = \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i\lambda_j\delta_i\gamma_{i,j}(k) = \boldsymbol{\delta}\mathbf{\Lambda}\mathbf{\Gamma}^k\boldsymbol{\lambda}'$$

   Note that in the last step we replace the double sum with a product of matrices, exactly as we did in the lectures for the bivariate distribution and for the likelihood calculation.

6.
$$\text{Corr}(X_t, X_{t+k}) = \frac{\text{E}(X_t X_{t+k}) - \text{E}(X_t)\text{E}(X_{t+k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+k})}} = \frac{\boldsymbol{\delta}\mathbf{\Lambda}\mathbf{\Gamma}^k\boldsymbol{\lambda}' - (\boldsymbol{\delta}\boldsymbol{\lambda}')^2}{\boldsymbol{\delta}\mathbf{\Lambda}\boldsymbol{\lambda}' + \boldsymbol{\delta}\boldsymbol{\lambda}' - (\boldsymbol{\delta}\boldsymbol{\lambda}')^2}$$

Exercise 8

Consider a stationary Bernouilli-HMM with transition probability matrix $\mathbf{\Gamma} = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$, probability of emission of a one in state 1 equal to $p_1 = 1/4$, probability of emission of a one in state 2 equal to $p_2 = 3/4$. We observe the data $x_1 = 1$ and $x_2 = 1$.

1. Write down the emission probability matrix $\boldsymbol{P} = \boldsymbol{P}(x = 1)$
2. Find the stationary distribution $\boldsymbol{\delta}$
3. Calculate the forward vector $\boldsymbol{\alpha}_1$, then $\boldsymbol{\alpha}_2$ and deduce the likelihood L_T using the foward algorithm
4. Calculate the backward vector $\boldsymbol{\beta}_1'$ and deduce the likelihood L_T using the backward algorithm
5. Calculate L_T using $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$

Answer

1.

$$\boldsymbol{P} = \begin{pmatrix} 1/4 & 0 \\ 0 & 3/4 \end{pmatrix}$$

2.

$$\boldsymbol{\delta} = (1/2, 1/2)$$

3.

$$\boldsymbol{\alpha}_1 = \boldsymbol{\delta}\boldsymbol{P} = (1/8, 3/8)$$
$$\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1\mathbf{\Gamma}\boldsymbol{P} = (3/16, 5/16)\boldsymbol{P} = (3/64, 9/64)$$
$$L_T = 9/32$$

4.

$$\boldsymbol{\beta}_1 = (\mathbf{\Gamma}\boldsymbol{P}\mathbf{1}')' = (3/8, 5/8)$$
$$L_T = \boldsymbol{\delta}\boldsymbol{P}\boldsymbol{\beta}_1' = (1/8, 3/8)\begin{pmatrix} 3/8 \\ 5/8 \end{pmatrix} = 9/32$$

5.

$$L_T = \boldsymbol{\alpha}_1\boldsymbol{\beta}_1' = (1/8, 3/8)\begin{pmatrix} 3/8 \\ 5/8 \end{pmatrix} = 9/32$$

Exercise 9

1. Show that state prediction can be performed for a given HMM model using the formula:

   $$p(C_{T+h} = i|\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}) = \boldsymbol{\alpha}_T\mathbf{\Gamma}^h\boldsymbol{e}_i'/L_T$$

   where $\boldsymbol{e}_i$ is a vector of zeros except for a one at the i-th position.
2. Deduce that forecasting can be performed using the formula:

   $$p(X_{T+h} = x|\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}) = \boldsymbol{\alpha}_T\mathbf{\Gamma}^h\boldsymbol{P}(x)\mathbf{1}'/L_T$$

3. Prove the forecasting formula again, using a ratio of likelihoods.

Answer

1.

$$p(C_{T+h}=i|\boldsymbol{X}^{(T)})=\sum_{j=1}^{m}p(C_{T+h}=i|C_T=j,\boldsymbol{X}^{(T)})p(C_T=j|\boldsymbol{X}^{(T)})$$

We already know that:

$$p(C_T=j|\boldsymbol{X}^{(T)})=p(C_T=j,\boldsymbol{X}^{(T)})/p(\boldsymbol{X}^{(T)})=\alpha_T(j)/L_T$$

Furthermore:

$$p(C_{T+h}=i|C_T=j,\boldsymbol{X}^{(T)})=p(C_{T+h}=i|C_T=j)=\boldsymbol{\Gamma}^h(j,i)$$

Therefore:

$$p(C_{T+h}=i|\boldsymbol{X}^{(T)})=\sum_{j=1}^{m}\boldsymbol{\Gamma}^h(j,i)\alpha_T(j)/L_T=\boldsymbol{\alpha}_T\boldsymbol{\Gamma}^h\boldsymbol{e}_i'/L_T$$

2.

$$p(X_{T+h}=x|\boldsymbol{X}^{(T)}=\boldsymbol{x}^{(T)})=\sum_{i=1}^{m}p(C_{T+h}=i|\boldsymbol{X}^{(T)})p(X_{T+h}=x|C_{T+h}=i)=\sum_{i=1}^{m}\boldsymbol{\alpha}_T\boldsymbol{\Gamma}^h\boldsymbol{e}_i'p_i(x)/L_T$$

$$=\boldsymbol{\alpha}_T\boldsymbol{\Gamma}^h\boldsymbol{P}(x)\mathbf{1}'/L_T$$

3.

$$p(X_{T+h}=x|\boldsymbol{X}^{(T)}=\boldsymbol{x}^{(T)})=\frac{p(X_{T+h}=x,\boldsymbol{X}^{(T)}=\boldsymbol{x}^{(T)})}{p(\boldsymbol{X}^{(T)}=\boldsymbol{x}^{(T)})}$$

$$=p(\boldsymbol{X}^{(T)}=\boldsymbol{x}^{(T)})=\boldsymbol{u}(1)\boldsymbol{P}(x_1)\boldsymbol{\Gamma}\boldsymbol{P}(x_2)...\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\boldsymbol{\Gamma}^h\boldsymbol{P}(x)\mathbf{1}'/L_T=\boldsymbol{\alpha}_T\boldsymbol{\Gamma}^h\boldsymbol{P}(x)\mathbf{1}'/L_T$$

Exercise 10

Let $\{X_t\}$ be a second-order HMM, based on a stationary second-order Markov chain $\{C_t\}$ with m states. We use the following definitions:

$$u(i,j)=p(C_{t-1}=i,C_t=j)$$

$$p_i(x_t)=p(X_t=x_t|C_t=i)$$

$$\gamma(i,j,k)=p(C_t=k|C_{t-1}=j,C_{t-2}=i)$$

$$v_t(i,j|\boldsymbol{x}^{(t)})=p(\boldsymbol{X}^{(t)}=\boldsymbol{x}^{(t)},C_{t-1}=i,C_t=j)$$

1. Write down $v_2(j,k|\boldsymbol{x}^{(2)})$ in terms of the definitions above.
2. Show that for integers $t\geq 3$ we have:

$$v_t(j,k|\boldsymbol{x}^{(t)})=\left(\sum_{i=1}^{m}v_{l-1}(i,j|\boldsymbol{x}^{(t-1)})\gamma(i,j,k)\right)p_k(x_t)$$

3. Show how the recursion above can be used to calculate the likelihood.
4. What is the computational effort required to calculate the likelihood?

Answer

1. $$v_2(i,j|\boldsymbol{x}^{(2)}) = u(i,j)p_i(x_1)p_j(x_2)$$

2. $$v_t(j,k|\boldsymbol{x}^{(t)}) = p(\boldsymbol{x}^{(t)}, C_{t-1}=j, C_t=k) = \sum_{i=1}^{m} p(\boldsymbol{x}^{(t)}, C_{t-2}=i, C_{t-1}=j, C_t=k)$$
$$= \sum_{i=1}^{m} p(\boldsymbol{x}^{(t-1)}, C_{t-2}=i, C_{t-1}=j) p(X_t=x_t, C_t=k|\boldsymbol{X}^{(t-1)}=\boldsymbol{x}^{(t-1)}, C_{t-2}=i, C_{t-1}=j)$$
$$= \sum_{i=1}^{m} v_{l-1}(i,j|\boldsymbol{x}^{(t-1)}) p(C_t=k|C_{t-2}=i, C_{t-1}=j) p(X_t=x_t|C_t=k) = \sum_{i=1}^{m} v_{l-1}(i,j|\boldsymbol{x}^{(t-1)})\gamma(i,j,k)p_k(x_t)$$

3. We use the recursion above to calculate the $v_t(j,k|\boldsymbol{x}^(t))$ values for increasing values of t up to $t=T$, and then calculate the likelihood as follows:
$$L_T = p(\boldsymbol{X}^{(t)} = \boldsymbol{x}^{(t)}) = \sum_{i=1}^{m}\sum_{j=1}^{m} p(\boldsymbol{X}^{(t)} = \boldsymbol{x}^{(t)}, C_T=j, C_{T-1}=i) = \sum_{i=1}^{m}\sum_{j=1}^{m} v_T(i,j)$$

4. At each step $t = 3..T$, we need to calculate $v_t(j,k|\boldsymbol{x}^(t))$ for all $j = 1..m$ and all $k = 1..m$, so m^2 terms. Each such calculation involves a summation over m states, so that the complexity is $O(Tm^3)$.