

Probabilistic Modeling and Reasoning

Homework — 8

Nikolaos Liouliakis (AID25001)
Vasileios-Efraim Tsavalia (AID25006)

MSc in Artificial Intelligence and Data Analytics
University of Macedonia

Supervisor: Professor Dimitris Christou-Varsakelis

February 2025

Problem 1

We consider a Hidden Markov Model (HMM) with 3 states ($M = 3$) and 2 output symbols. The transition matrix A , emission matrix B , and initial state probabilities a are given by:

$$A = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.3 & 0.6 & 0.0 \\ 0.2 & 0.4 & 1.0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.7 & 0.4 & 0.8 \\ 0.3 & 0.6 & 0.2 \end{bmatrix}$$

$$a = (0.9, 0.1, 0.0)^T$$

Given the observed sequence $v_{1:3} = (1, 2, 1)$, we solve the following:

1. Compute $p(v_{1:3})$ using the forward algorithm

The forward probabilities α_t are computed iteratively as follows:

$$\alpha_1 = a \odot B_{v_1, \cdot}^T$$

$$\alpha_t = (A \cdot \alpha_{t-1}) \odot B_{v_t, \cdot}$$

where \odot denotes the Hadamard (element-wise) product.

Step-by-Step Computation

For $t = 1$:

Since $v_1 = 1$, we use the first row of B :

$$\alpha_1 = a \odot B_{1, \cdot}^T$$

$$B_{1, \cdot}^T = \begin{bmatrix} 0.7 \\ 0.4 \\ 0.8 \end{bmatrix}$$

$$\alpha_1 = \begin{bmatrix} 0.9 \\ 0.1 \\ 0.0 \end{bmatrix} \odot \begin{bmatrix} 0.7 \\ 0.4 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 0.9 \cdot 0.7 \\ 0.1 \cdot 0.4 \\ 0.0 \cdot 0.8 \end{bmatrix} = \begin{bmatrix} 0.63 \\ 0.04 \\ 0.0 \end{bmatrix}$$

For $t = 2$:

Using $v_2 = 2$, we take the second row of B :

$$\alpha_2 = (A \cdot \alpha_1) \odot B_{2, \cdot}^T$$

$$B_{2, \cdot}^T = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.2 \end{bmatrix}$$

Compute:

$$A \cdot \alpha_1 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.3 & 0.6 & 0.0 \\ 0.2 & 0.4 & 1.0 \end{bmatrix} \begin{bmatrix} 0.63 \\ 0.04 \\ 0.0 \end{bmatrix} = \begin{bmatrix} 0.315 \\ 0.213 \\ 0.142 \end{bmatrix}$$

$$\alpha_2 = \begin{bmatrix} 0.315 \\ 0.213 \\ 0.142 \end{bmatrix} \odot \begin{bmatrix} 0.3 \\ 0.6 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.0945 \\ 0.1278 \\ 0.0284 \end{bmatrix}$$

For $t = 3$:

Since $v_3 = 1$, we use the first row of B :

$$\alpha_3 = (A \cdot \alpha_2) \odot B_{1,}^T$$

$$A \cdot \alpha_2 = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.3 & 0.6 & 0.0 \\ 0.2 & 0.4 & 1.0 \end{bmatrix} \begin{bmatrix} 0.0945 \\ 0.1278 \\ 0.0284 \end{bmatrix} = \begin{bmatrix} 0.0473 \\ 0.1050 \\ 0.0984 \end{bmatrix}$$

$$\alpha_3 = \begin{bmatrix} 0.0473 \\ 0.1050 \\ 0.0984 \end{bmatrix} \odot \begin{bmatrix} 0.7 \\ 0.4 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 0.0331 \\ 0.0420 \\ 0.0787 \end{bmatrix}$$

Final probability of the observation sequence:

$$p(v_{1:3}) = \sum_i \alpha_3(i) = 0.0331 + 0.0420 + 0.0787 = 0.1538$$

2. Compute $p(h_1|v_{1:3})$

Using Bayes' theorem:

$$p(h_1|v_{1:3}) = \frac{p(h_1, v_{1:3})}{p(v_{1:3})}$$

where $p(h_1, v_{1:3})$ is by definition

$$\alpha_3$$

therefore:

$$p(h_1|v_{1:3}) = \begin{bmatrix} 0.0331 \\ 0.0420 \\ 0.0787 \end{bmatrix} \cdot \frac{1}{0.1538} = \begin{bmatrix} 0.2150 \\ 0.2731 \\ 0.5119 \end{bmatrix}$$

3. Find the most probable hidden state sequence

$$\arg \max_{h_{1:3}} p(h_{1:3}|v_{1:3})$$

Using the definition, the most probable state sequence is:

$$h_{1:3} = [3, 3, 3]$$

which was verified with the script HW8.P1.m.

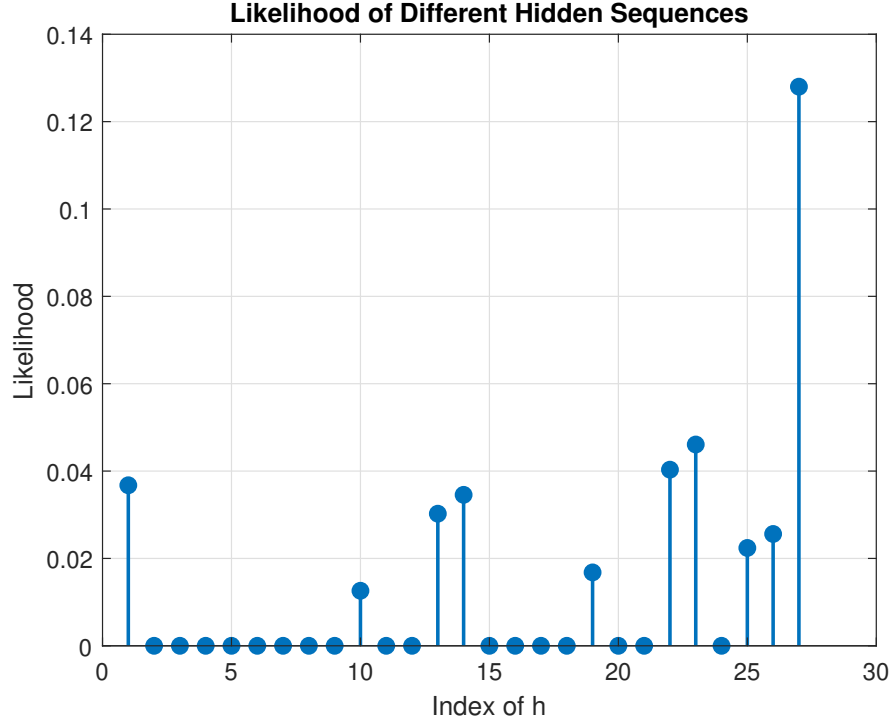


Figure 1: Stem plot for all possible states of $h_{1:3}$

Note that for many state sequences $h_{1:3}$, the likelihood is zero. This is expected because, as observed in A , state 2 cannot transition to state 1, and similarly, state 3 cannot transition to either state 1 or state 2.

Problem 2

Suppose the HMM transition matrix A and emission matrix B are initialized to uniform values:

$$A_{ij} = \frac{1}{N}, \quad \forall i, j$$

where N is the number of hidden states, ensuring that A is a valid stochastic matrix. Similarly, let

$$B_{jk} = \frac{1}{M}, \quad \forall j, k$$

where M is the number of observation symbols.

E-Step: The EM algorithm computes expected counts based on current parameters. Since all transition and emission probabilities are equal, the forward and backward probabilities computed during the E-step will also be uniform across states.

M-Step: The parameter A is updated as:

$$A_{ij}^{\text{new}} = \frac{\sum_t \gamma_t(i, j)}{\sum_t \gamma_t(i)}$$

where $\gamma_t(i, j)$ is the expected count of transition $i \rightarrow j$. Since the initialization is uniform, the expected counts remain uniform, and thus the updates do not change the structure of A and B .

Conclusion: Since the EM updates preserve the uniformity of the matrices, the algorithm fails to update parameters meaningfully. The algorithm stagnates, leading to failure in learning a useful model.

Problem 3

1. Probability of Sequence under p_{new}

The probability of the sequence $S = A, A, G, T, A, C, T, T, A, C, C, T, A, C, G, C$ under the transition matrix p_{new} is given by:

$$P(S|p_{\text{new}}) = p(h_1) \prod_{t=1}^{15} p_{\text{new}}(S_{t+1}|S_t)$$

where $p(h_1)$ is the initial probability of being in the first state. Given that it is uniformly distributed among A, C, G, T :

$$p(h_1) = \frac{1}{4}$$

To compute the probability of the sequence under p_{new} , the function `sequence_prob` was written, which is called from the `HW8_P3.m` file resulting in:

$$P(S|p_{\text{new}}) = \frac{1}{4} \prod_{t=1}^{15} p_{\text{new}}(S_{t+1}|S_t) = 8.1781 \cdot 10^{-13}$$

2. Probability under q_{new} and Likelihood Comparison

Similarly, we compute:

$$P(S|q_{\text{new}}) = \frac{1}{4} \prod_{t=1}^{15} q_{\text{new}}(S_{t+1}|S_t) = 8.6147 \cdot 10^{-24}$$

Comparing $P(S|p_{\text{new}})$ and $P(S|q_{\text{new}})$, we determine that S has a higher likelihood under p_{new} as expected, because using the transition matrix p_{new} is more probable the next state to be bigger (the numbers in the sequence increase) whereas with q_{new} it is more probable the next state to be smaller.

3. Maximum Likelihood Estimation with MixMarkov

The function `generate_markov_samples` was created to:

- Generate 100 sequences of length 16 from p_{new} .
- Generate 100 sequences of length 16 from q_{new} .

While using the function `MixMarkov.m` some errors occurred, which prevented from getting the final results.

4. Hidden Markov Model and Most Likely Sequence

Given the emission distribution:

$$p(v = i|h = j) = \begin{cases} 0.7 & \text{if } i = j \\ 0.1 & \text{if } i \neq j \end{cases}$$

we adapted `demoHMMInferenceSimple.m` to compute the most likely hidden sequences $h_1^{1:16}$ under both transition matrices. The preferred hidden sequence was determined based on likelihood comparisons. The majority of the time the preferred hidden sequence was from p_{new} .

Problem 4

We are tasked with proving that a suitable value of M such that

$$\frac{p^*(x)}{q(x)} \leq M$$

is

$$M = e^{1 + \frac{\pi^2}{2\sigma^2}} \sqrt{2\pi\sigma^2}.$$

1. Define $p^*(x)$ and $q(x)$

The target unnormalized probability density is given as:

$$p^*(x) = e^{\sin(x)}.$$

The proposal distribution $q(x)$ is a Gaussian:

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

2. Compute $\frac{p^*(x)}{q(x)}$

The ratio of the unnormalized target density to the proposal density is:

$$\frac{p^*(x)}{q(x)} = \sqrt{2\pi\sigma^2} e^{\sin(x)} \exp\left(\frac{x^2}{2\sigma^2}\right).$$

3. Find an upper bound of $\frac{p^*(x)}{q(x)}$

To find M , we need an upper bound of $\frac{p^*(x)}{q(x)}$ over all $x \in [-\pi, \pi]$.

- The term $e^{\sin(x)}$ achieves its maximum when $\sin(x) = 1$, which occurs at $x = \frac{\pi}{2}$. Hence:

$$\max e^{\sin(x)} = e^1.$$

- The term $\exp\left(\frac{x^2}{2\sigma^2}\right)$ achieves its maximum when x^2 is maximal in the range $[-\pi, \pi]$, which occurs at $x = \pm\pi$. Hence:

$$\max \exp\left(\frac{x^2}{2\sigma^2}\right) = \exp\left(\frac{\pi^2}{2\sigma^2}\right).$$

Thus, an upper bound of $\frac{p^*(x)}{q(x)}$ is:

$$M = \sqrt{2\pi\sigma^2} \cdot e^1 \cdot \exp\left(\frac{\pi^2}{2\sigma^2}\right).$$

Simplify:

$$M = e^{1 + \frac{\pi^2}{2\sigma^2}} \sqrt{2\pi\sigma^2}.$$

4. Conclusion

We have shown that the suitable value of M is:

$$M = e^{1 + \frac{\pi^2}{2\sigma^2}} \sqrt{2\pi\sigma^2}.$$

Which ensures that $\frac{p^*(x)}{q(x)} \leq M$ for all $x \in [-\pi, \pi]$.

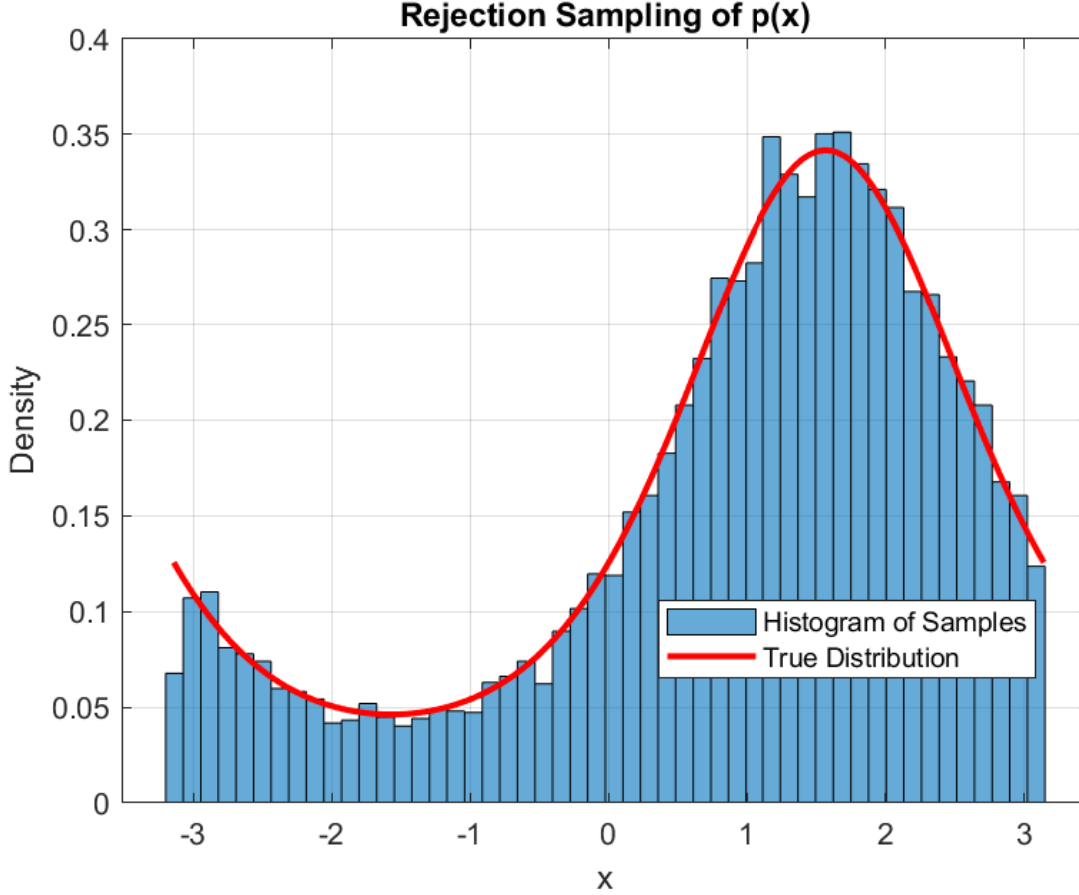


Figure 2: Histogram of 10000 samples drawn from $p(x)$

Problem 5

The original joint distribution is given by

$$p(x_1, x_2, x_3, x_4, x_5, x_6) = p(x_1) p(x_2) p(x_3 | x_1, x_2) p(x_4 | x_3) p(x_5 | x_3) p(x_6 | x_4, x_5). \quad (1)$$

When we fix (or observe) $x_5 = x_5^*$, the conditional distribution over the remaining variables is

$$\begin{aligned} p(x_1, x_2, x_3, x_4, x_6 | x_5 = x_5^*) &= \frac{p(x_1, x_2, x_3, x_4, x_5^*, x_6)}{p(x_5^*)} \\ &= \frac{p(x_1) p(x_2) p(x_3 | x_1, x_2) p(x_4 | x_3) p(x_5^* | x_3) p(x_6 | x_4, x_5^*)}{p(x_5^*)}. \end{aligned} \quad (2)$$

Notice that the likelihood term $p(x_5^* | x_3)$ plays a role in the posterior over x_3 .

Ancestral Sampling Procedure

To sample from $p(x_1, x_2, x_3, x_4, x_6 \mid x_5 = x_5^*)$, follow these steps:

1. **Sample x_1 :** Draw $x_1 \sim p(x_1)$.
2. **Sample x_2 :** Draw $x_2 \sim p(x_2)$.
3. **Sample x_3 :** Draw x_3 from

$$p(x_3 \mid x_1, x_2, x_5^*) \propto p(x_3 \mid x_1, x_2) p(x_5^* \mid x_3).$$

4. **Sample x_4 :** Draw $x_4 \sim p(x_4 \mid x_3)$.
5. **Sample x_6 :** Draw $x_6 \sim p(x_6 \mid x_4, x_5^*)$.