

Assignment 1

Per Emil Hammarlund, Albert Öst

2019-04-15

Contents

Exercise 1	2
Exercise 2	3
Exercise 3	4
Calculation of $\mathbb{E}[X]$	4
Calculation of $\text{var}_X[X]$	4
Verification of implemented code	4
Exercise 4	5
Exercise 5	6
Exercise 6	7
Exercise 7	9

Exercise 1

For the markov chain:

$$q = \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}, A = \begin{pmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{pmatrix}$$

The state distribution for the first time step $t = 1$ is:

$$P(S_1 = 1) = q(1) = 0.75$$

$$P(S_1 = 2) = q(2) = 0.25$$

For the second time step $t = 2$, the distribution of states is:

$$\begin{aligned} P(S_2 = 1) &= \sum_{j=1}^2 P(S_1 = j)P(S_2 = 1|S_1 = j) \\ &= P(S_1 = 1)P(S_2 = 1|S_1 = 1) + P(S_1 = 2)P(S_2 = 1|S_1 = 2) \\ &= q(1)A(1, 1) + q(2)A(2, 1) = 0.75 \cdot 0.99 + 0.25 \cdot 0.03 = 0.75 \end{aligned}$$

$$\begin{aligned} P(S_2 = 2) &= \sum_{j=1}^2 P(S_1 = j)P(S_2 = 2|S_1 = j) \\ &= P(S_1 = 1)P(S_2 = 2|S_1 = 1) + P(S_1 = 2)P(S_2 = 2|S_1 = 2) \\ &= q(1)A(1, 2) + q(2)A(2, 2) = 0.75 \cdot 0.01 + 0.25 \cdot 0.97 = 0.25 \end{aligned}$$

Which again, is the same distribution as the first time step. And since the distribution of a given time step will always be calculated by marginalizing out the previous time step from the joint $P(S_t, S_{t-1})$ for each of the states in the current time step. The same calculations will be performed $\forall t \in \{1, 2, 3, \dots\}$ and thus, $P(S_t = j)$ is constant for all timesteps.

Exercise 2

The distribution of states for 10000 samples from the given markov chain takes the following form:

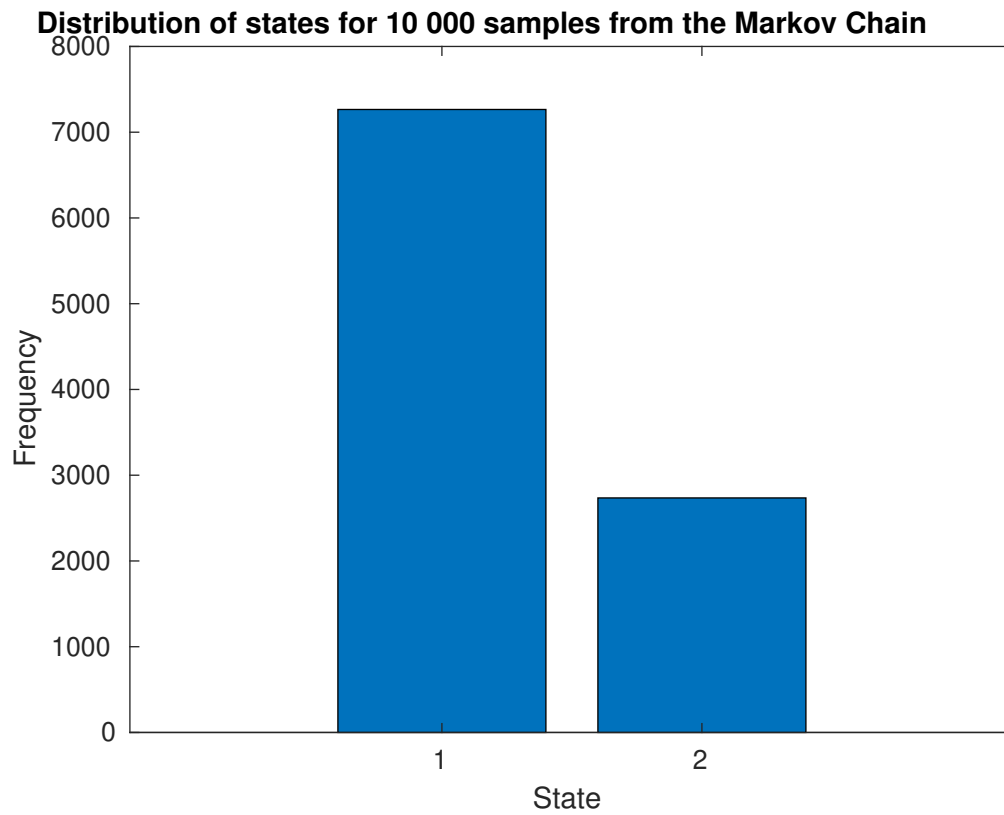


Figure 1: Distribution of states over 10 000 samples

With the frequencies:

State	Frequency	Relative Frequency
1	7265	0.7625
2	2735	0.2735

Exercise 3

Calculation of $\mathbb{E}[X]$

$$\begin{aligned}\mu_X &= \mathbb{E}[X] = \mathbb{E}_Z[\mathbb{E}_X[X|Z]] = \mathbb{E}_Z[\mathbb{E}_X[b_Z(X)]] = \sum_{j=1}^2 \mathbb{E}_X[X|Z=j]P(Z=j) \\ &= \mathbb{E}_X[b_1(x)]P(Z=1) + \mathbb{E}[b_2(x)]P(Z=2) \\ &= \{\text{since the first order moment of a normal distribution is its mean}\} \\ &= 0 \cdot 0.75 + 3 \cdot 0.25 = \frac{3}{4}\end{aligned}$$

Calculation of $\text{var}_X[X]$

$$\begin{aligned}\text{var}[X] &= \mathbb{E}_Z[\text{var}_X[X|Z]] + \text{var}_Z[\mathbb{E}_X[X|Z]] = \mathbb{E}_Z[\text{var}_X[X|Z]] + \mathbb{E}_Z[(\mathbb{E}_X[X|Z])^2] - (\mathbb{E}[X])^2 \\ &= \sum_{j=1}^2 \text{var}_X[X|Z=j]P(Z=j) + \sum_{j=1}^2 (\mathbb{E}_X[X|Z=j])^2 P(Z=j) - \left(\frac{3}{4}\right)^2 \\ &= \sum_{j=1}^2 (\text{var}_X[X|Z=j] + (\mathbb{E}_X[X|Z=j])^2) P(Z=j) - \left(\frac{3}{4}\right)^2 \\ &= (\text{var}_X[X|Z=1] + (\mathbb{E}_X[X|Z=1])^2)P(Z=1) + (\text{var}_X[X|Z=2] + (\mathbb{E}_X[X|Z=2])^2)P(Z=2) - \frac{9}{16} \\ &= (1 + 0^2) \cdot 0.75 + (4 + 9) \cdot 0.25 - \frac{9}{16} = \frac{3}{4} + \frac{13}{4} - \frac{9}{16} = \frac{12}{16} + \frac{52}{16} - \frac{9}{16} = \frac{55}{16} = 3.4375\end{aligned}$$

Verification of implemented code

When generating 10 000 samples from the HMM rand function, the following mean and variance was observed:

mean	variance
0.7831	3.5241

Exercise 4

The samples from 500 contiguous samples X_t took the following form:

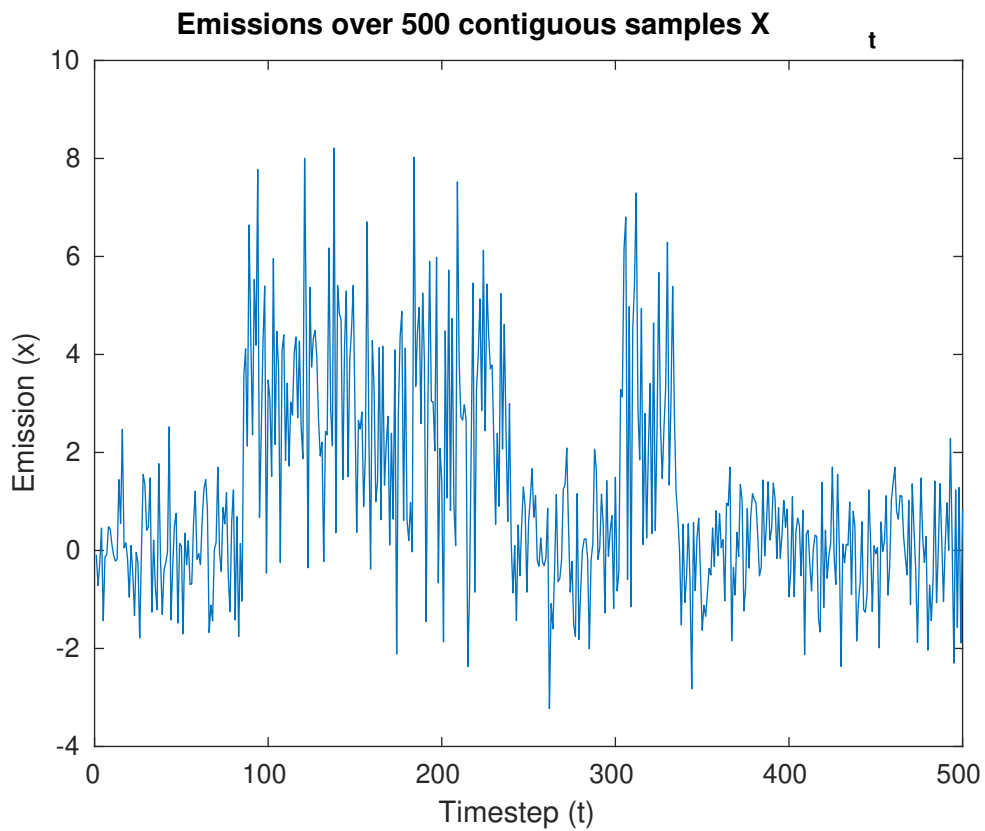


Figure 2: Samples from 500 contiguous samples from the hmm rand function

As can be seen in figure 2 above, the distribution that is generating emissions from the hidden markov model (i.e. the hidden state) tends to stick in the same hidden state rather than transition between the states. This makes sense since staying in the same hidden state is far more likely than transitioning.

Exercise 5

The observed emissions from the hidden markov model took the following form:

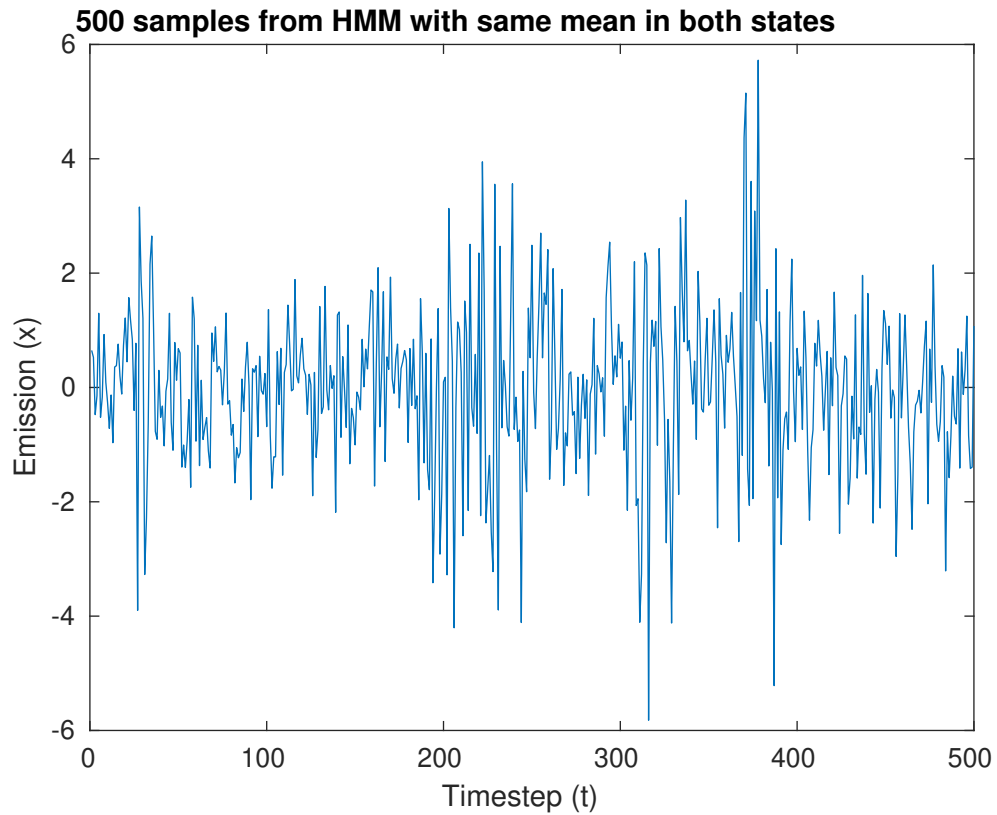


Figure 3: Samples from HMM with $\mu_1 = \mu_2 = 0$

As can be seen in figure 3 above, it is much more difficult to differentiate the transition between the two distributions and thus by extension know what hidden state the observations are coming from. It is of course, possible to approximate what hidden states have generated the emissions, but due to the similarity between the two distributions. It will not be as accurate.

Exercise 6

The finite hidden markov model which can be found in equation 5.17 in the course book was constructed using the following code:

```
mc = MarkovChain([1 ;0], [0.7 0.25 0.05; 0.1 0.8 0.1])
h = HMM(mc, [GaussD('Mean', 0, 'StDev', 1), GaussD('Mean', 3, 'StDev', 2)])
```

And the rand function was queried for 500 samples 10 000 times and the length of the observation sequence was saved for each try using:

```
num_obs = zeros(10000, 1);
for i=1:10000
    num_obs(i) = size(h.rand(500), 2);
end
fprintf(sprintf('The observations sequence had an average length of %f.\n', mean(num_obs)));
```

Which gave an average length of 12.886700.

This was then compared to the distribution of the expected length of observations generated from the code below:

```
lenDistr = zeros(100, 1);
for i=1:100
    lenDistr(i) = sum(mc.InitialProb' * mc.TransitionProb(:, 1:2)^i) * i;
end
lenDistr = lenDistr ./ sum(lenDistr);
plot(lenDistr);
hold on
xlabel('Number of timesteps in transeint states');
ylabel('Probability');
title('Distribution over first 100 timesteps of expected sequence length');
hold off
```

The distribution of the expected length of the observation sequence generated from the code above took the following form:

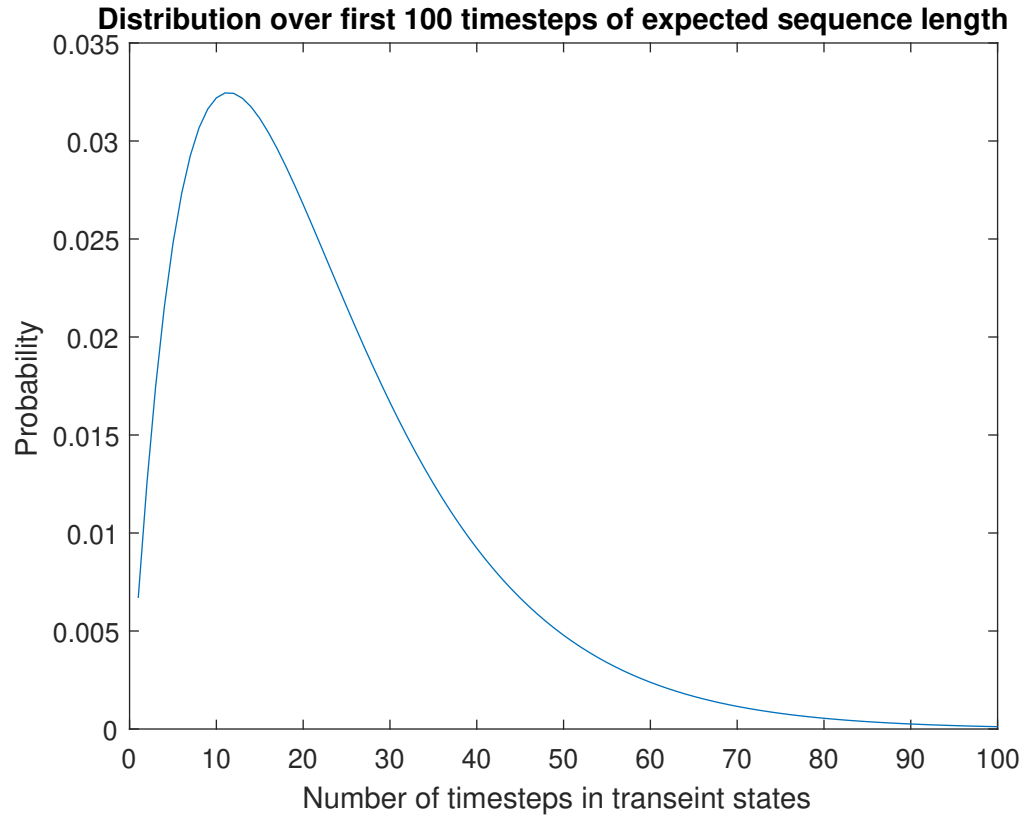


Figure 4: Distribution of expected sequence length

From which it can be observed that the highest density is around 12-13, which is in the same neighborhood as our average sequence length.

Exercise 7

The following hidden markov model was constructed:

```
g1 = GaussD('Mean', [3; 3], 'Covariance', [0.01 0; 0 0.01]);
g2 = GaussD('Mean', [0; 0], 'Covariance', [0.01 0.01; 0.01 0.01]);
mc = MarkovChain([0.75; 0.25], [0.99 0.01; 0.03 0.97]);
h = HMM(mc, [g1, g2]);
```

And 100 samples was drawn from it and visualized using the following code:

```
h = HMM(mc, [g1, g2]);
emissions = h.rand(1000000);
hist3(emissions, 'CdataMode', 'auto', 'Nbins', [100, 100]);
view(2);
colorbar;
hold on
xlabel('dimension 1');
ylabel('dimension 2');
title('Visualisation of emissions from HMM');
hold off
```

the distribution took the following form:

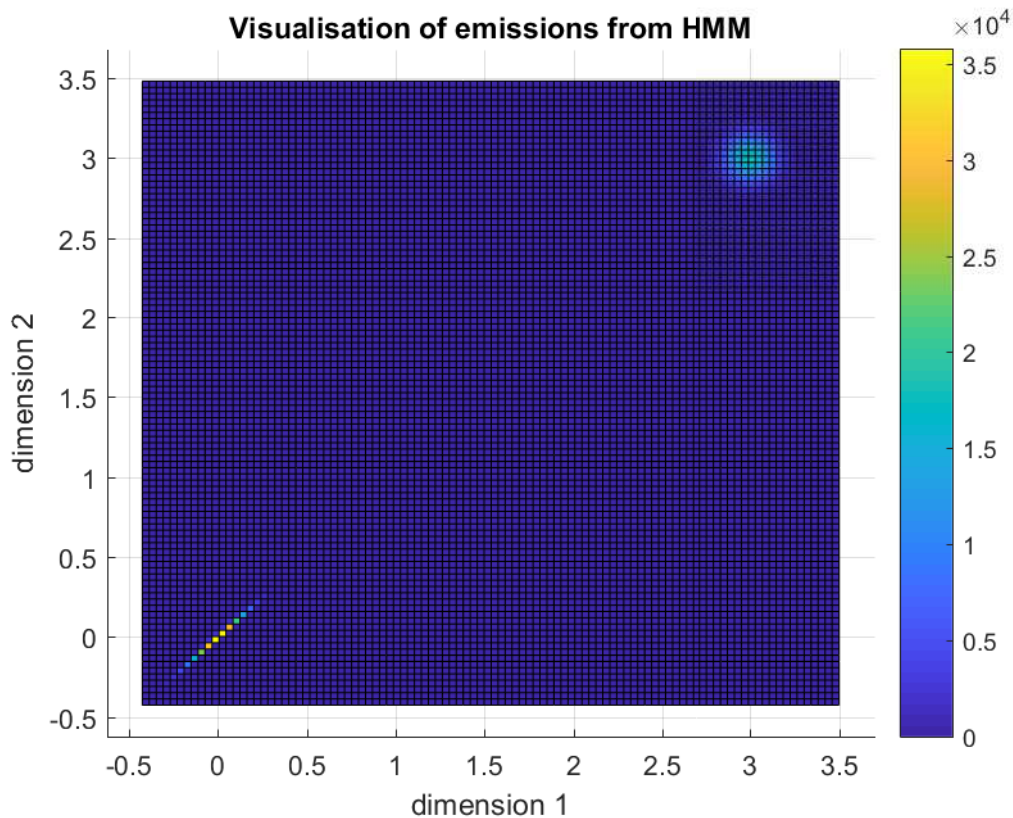


Figure 5: 100 samples from HMM with multivariate gaussian distributions

From which it is clear that distribution 1 has no correlation between dimensions and that distribution 2 has a strong correlation between dimensions.