# EQ2341 Pattern Recognition and Machine Learning Assignment 1

Yihan Wang, Yichen Yang

April 10, 2019

## 1 Introduction

The task in this Assignment is to discover how an HMM can generate a sequence with a kind of structured randomness that is typical of many real life signals[1]. HMM model is widespread used in many applications, through this assignment, we'll get a deeper understanding on how HMM model generates a signal source by both mathematical deduction and computer simulation.

## 2 Part I: HMM Random Source

Note: This part corresponds to question 1,2,3 in A1.1

Random function: See Matlab code in the attachment:

## 3 Part II: Verify the Markov Chain and HMM Sources

### 3.1 State probability calculation

Note: This part corresponds to question 1,2,3 in A1.2
Assuming a HMM model with the following parameters: Initial state distribution:

$$q = \begin{cases} 0.75 & S = 1 \\ 0.25 & S = 2 \end{cases}$$

This q-distribution implies that at the beginning, the probability of state 1 equals 0.75 and that of state 2 equals 0.25. Transition matrix:

$$A = \begin{bmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{bmatrix}$$

This transition matrix implies the probability of transition from the current state $S_i$ to the next one $S_{i+1}$, or $P(S_{i+1}|S_i)$. Using the condition above, and note that the state distribution is binary, we could calculate the state probability at each time:
$P_1(S = 1) = 0.75; P_1(S = 2) = 0.25$
$P_2(S = 1) = P_1(S = 1)P(S_2 = 1|S_1 = 1) + P_1(S = 2)P(S_2 = 1|S_1 = 2)$
$P_2(S = 2) = 1 - P_2(S = 1)$

For any instance i, we have

$$P_i(S=1) = P_{i-1}(S=1)P(S_i=1|S_{i-1}=1) + P_{i-1}(S=2)P(S_i=1|S_{i-1}=2) \quad (1)$$

$$P_i(S=2) = 1 - P_i(S=1) \quad (2)$$

Then, we assume this infinite Malkov Chain is convergent, which implies at time goes to infinity, $P_i(S=1) = P_{i-1}(S=1)$, plugging it in equation (1) and (2), we deduce that: $\lim_{i\to\infty} P_i(S=1) = 0.75$; $\lim_{i\to\infty} P_i(S=2) = 0.25$
This conclusion corresponds to the simulation result in Matlab program.

## 3.2 Mean and variance calculation

Note: This part corresponds to question 1,2,3 in A1.2
Next we need to calculate the statistic property of the observation $X_t$. Given the same model above with observation??? parameter:

$$B = \begin{bmatrix} b_1(x), & S=1 \\ b_2(x), & S=2 \end{bmatrix}$$

Where $b_1(x)$ $N(0,1)$ and $b_2(x)$ $N(3,4)$. This variable tells us that the PDF on the corresponding observation $P(X_i=x|S_i=1) = b_1(x)$ and $P(X_i=x|S_i=2) = b_2(x)$
The expectation $u_X = E(X) = E_S(E_X(X|S))$
$u_X = P(S=1)E_X(X|S=1) + P(S=2)E_X(X|S=2)$
$u_X = P(S=1)E_x(b_1(x)) + P(S=2)E_x(b_2(x))$
$u_X = 0.75$
The variance:

$$var(X) = E_S(var_X(X|S)) + var_S(E_X(X|S)) \quad (3)$$

$$E_S(var_X(X|S)) = P(S=1)var_X(X|S=1) + P(S=2)var_X(X|S=2) \quad (4)$$

Before getting into the next term, note that the random variable $T = E_X(X|S)$ complies to binary distribution with the following probability:

| $T=E(X\|S)$ | $E_X(X\|S=1)$ | $E_X(X\|S=2)$ |
|---|---|---|
| Probability | P(S=1) | P(S=2) |

Plugging the values in, we derive the distribution of T as the following:

| $T=E(X\|S)$ | 0 | 3 |
|---|---|---|
| Probability | 0.75 | 0.25 |

Then we calculate the variance of $T = E_X(X|S)$

$$var_S(E_X(X|S) = var(T) = E(T^2) - E^2(T) = 0.25 * 9 - 0.75^2 \quad (5)$$

Finally, plugging (4) and (5) into (3), we derive that $var(X) = 3.4375$
Similarly, the theoretical result corresponds to simulation result.

## 3.3 HMM realization plot

Note: This part corresponds to question 4,5 in A1.2
The plot is shown in 1. Firstly, we initialize HMM model by setting the following parameters:

$$A = \begin{bmatrix} 0.97 & 0.03 \\ 0.04 & 0.96 \end{bmatrix}$$

,

$$q = \begin{cases} 0.5 & S = 1 \\ 0.5 & S = 2 \end{cases}$$

,

$$B = \begin{bmatrix} b_1(x), & S = 1 \\ b_2(x), & S = 2 \end{bmatrix}$$

Then, we set $b_1(x)$ $N(0, 0.01)$ and $b_2(x) = N(3, 0.04)$, which means the observation functions have different means. After simulation, we plot the state sequence S in respect to T in Figure 3, and the observation X in respect to T in Figure 1. From these 2 graphs, we know that the state variation is only dependent on transition matrix, and that **if the observation parameters have conspicuous difference in mean (compared with their variance), we can identify the state through the observation.**
Simultaneously, we set $b_1(x)$ $N(0, 0.01)$ and $b_2(x) = N(0, 1)$, which means the observation functions are zero-mean but different variance. Similarly, we plot Figure 2 and 4, where we can see the observation mean remains the same in spite of the states. From these 2 graphs, we know that **even with the same mean value, when the observation variance in one state is far away from that of the other, we can still identify the state through observation, but it doesn't hold the water when the observation variances are close to each other.**
Under these 2 conditions, we can see that the observation is clearer when we have different observation means, and that the smaller the observation is, the more accurately we can identify the corresponding states. Also, by setting a larger T, we generalize that the state probability will converge as T increases.

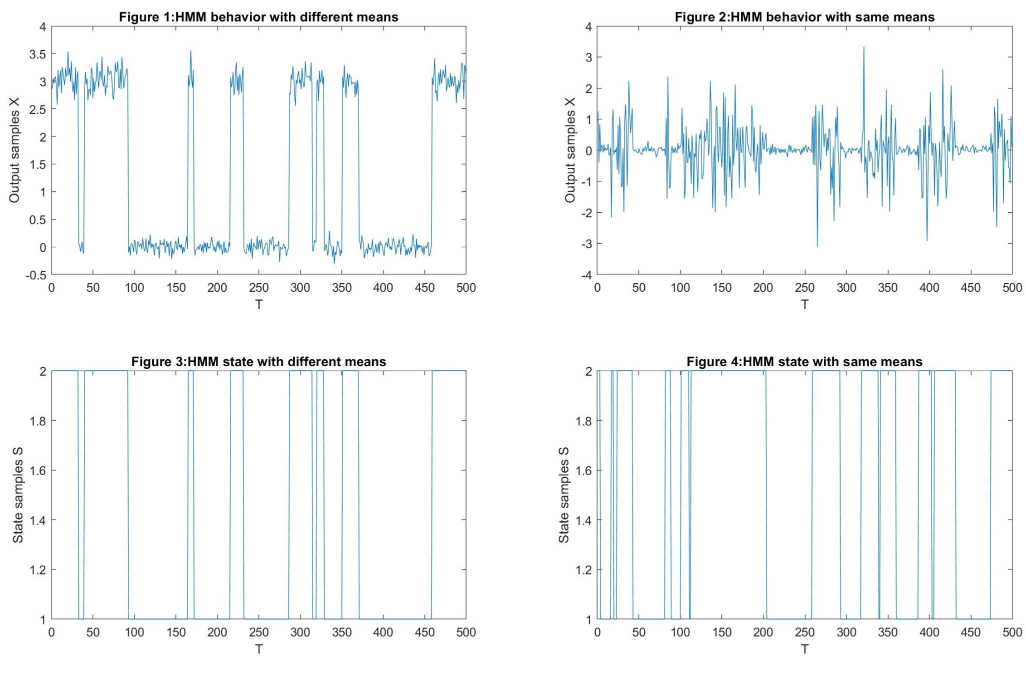


Figure 1: HMM realization plot

## 3.4 Finite-duration HMM Denition

Note: This part corresponds to question 6 in A1.2
When the row of transition matrix A is smaller than the column of that ( Alternatively,

row(A)¡column(A) ), we claim that this HMM Definition is finite. Due to the smaller row, there's a lack in transition from the final state to the previous one, and that state marks the end of this Malkov chain. (For example, when the transition matrix has a size of 4*3, that means there's no transition from the state 4 to any other state, so when this process reaches state 4, it won't go to any other state and the process suspends.)
From the code "", we can also see that even we require it to generate a sequence with the length of 500, the actual sequence is smaller than that in size.

### 3.5 Vector-valued HMM Denition

Note: This part corresponds to question 7 in A1.2
When we define a vector-valued observation, what we observe is a set of vectors instead of values, and the vector's element can absolutely be correlated. In 2, we can see the different and correlated output vectors, by simulating a large number of samples. (Note the distribution of this 2-d graph looks like an ellipse, showing that there's a correlation between the 2 elements.)

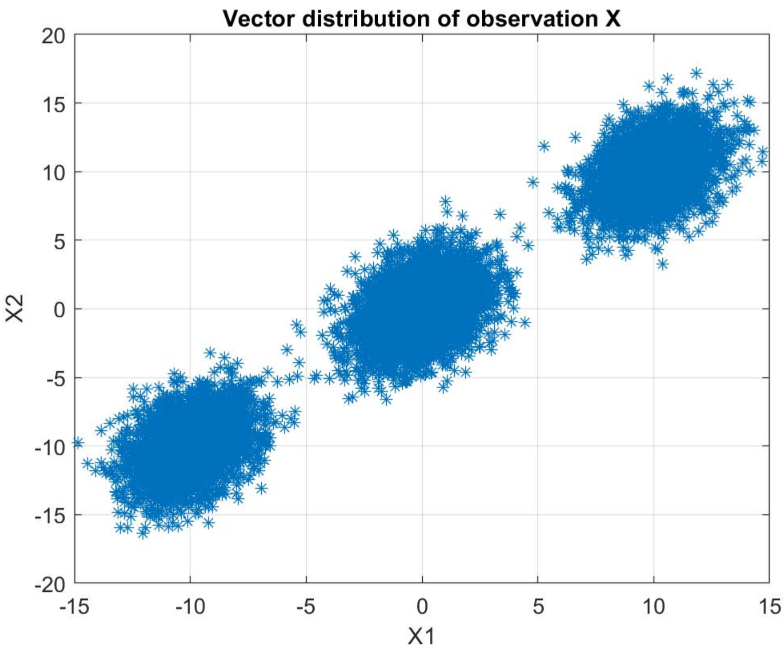


Figure 2: Vector observation plot

## 4 Conclusion

Through this assignment, we understand how the HMM model is generated and also the feature of state and observation. By calculation and simulation, we get a deeper understanding of how the parameter B, A and q can affect the model, and how the finite length HMM model is defined and its features. Also, we have a rough concept of identifying states from the observation. Hoping to learn more interesting things in the following projects!

# Appendix

# References

[1] Pattern Recognition Fundamental Theory and Exercise Problems, ARNE LEIJON and GUSTAV EJE HENTER, 2015

## MatLab code

See attachments.
Section A.1.2: Question 2 and 3: 'Q_2_and_3.m"
Question 4 and 5: "Q_4_and_5.m"
Question 6 : "Q_6.m"
Question 7 : "Q_7.m"