Hidden Markov Chains

Marco, Nicolás, Sebastián and César

November 2, 2019

Index

Reminder: What a Markov chain is?

- A Markov chain is a stochastic process with a particular characteristic, must be "memory-less" (i.e the probability of future actions are not dependent upon the steps that led up to the present state.). This property is known as the **Markov property**.
- Markov chains are very useful, we can find very often phenomena that satisfy the Markov property. This fact make Markov chains very useful in various application scenarios.
- Additionally, Markov chains allow us modelling sequential processes in a simple but effective way.

Reminder: What a Markov chain is?

A simple example...

- Bag of balls without replacement: **is NOT** a Markov process

ex1_bag_of_balls.png

Figure: example: probability of getting a blue ball / without replacement

- Bag of balls with replacement **is** a Markov process

ex1_bag_of_balls_wrep.png

Figure: example: probability of getting a blue ball / with replacement

Markov chains: Definition

Markov Chain: definition

A Markov chain is a sequence $X_{0:T} = (X_0, X_1, \ldots, X_T)$ of random variables taking values in some finite set $\mathcal{X}$ that satisfies the rule of conditional independence:

$$P\left(X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0\right) = P\left(X_{t+1} = x_{t+1} \mid X_t = x_t\right)$$

For this case, we will study time-homogeneous Markov chains (i.e. the probability of any state transition is independent of time).

$$A_{i,j} := P\left(X_{t+1} = j \mid X_t = i\right) \; i, j \in \mathcal{X}$$

However, the general case of Markov chains allows for time-inhomogeneous Markov chains (as time goes on, the probability of moving from one state to another may change).

Transition matrices

The movement among states are defined by a **transition matrix**, A_t. Particularly, this matrix contains the information on the probability of transitioning between states.

$$A_{i,j} := P\left(X_{t+1} = j \mid X_t = i\right) \; i, j \in \mathcal{X}$$

Each row of the matrix is a probability vector, and the sum of its entries is 1.

For $x_0 \in \mathcal{X}$, let $\mu_{x0} = P\left(X_0 = x_0\right)$. Then, the joint promabiloty mass function of $X_{1:T}$ in terms of $\left(A_{i,j}\right)_{i,j\in\mathcal{X}}$ and $\left(\mu_i\right)_{i\in\mathcal{X}}$ is as follows:

$$p\left(x_{0:T}\right) := P\left(X_0 = x_0, \ldots, X_T = x_t\right)$$

$=\mathrm{P}(X_0 = x_0) \prod_{t=1}^{T} P\left(X_t = x_t \mid X_{t-1} = x_{t-1}\right)$
$=\mu_{x0} \prod_{t=1}^{T} A_{x_{t-1},x_t}$

Hidden Markov Chains : Motivation

- A **Hidden Markov Model** is a statistical model which studies a system assumed as a Markov process including hidden (unobservable) states.
- In the simple Markov chain model, the state of the system is directly visible to the observer. On the other hand, HMM assume that the data observed is not the actual state of the model but instead is generated by underlying hidden states.
- Each state has a probability distribution over the possible outputs.
- HMM are often used to model temporal data.

Definition

HMM :Definition

Let $X_{0:T} = (X_0, X_1, \ldots, X_T)$ be a homogeneous Markov chain taking values in $\mathcal{X}$ with and associated transition matrix (A_{ij}). Additionally, there exits another sequence of random variables $Y_{0:T} = (Y_1, \ldots, Y_T)$ taking values in $\mathcal{Y}$, known as the observation space.

We assume that the random variables of sequence $Y_{0:T}$ are independent conditional on the state sequence $X_{0:T}$, which is equivalent to:

$$P(Y_1 = y_1, \ldots, Y_T = y_T \mid X_0 = x_0, \ldots, X_T = x_t) = \prod_{t=1}^{T} P(Y_t = y_t \mid X_t = x_t$$

if we are considering an homogeneous HMM, then we have an **emission probability** mass function:

$$g_x(y) := P(Y_t = y \mid X_t = x)$$

More Definitions

Using the emission probabilities to connect the observation space and the hidden space and introducing the transition matrix that governs the movements among hidden states, we can define the joint probability of the hidden states and observations as follows:

$$P\left(X_{0:T} = x_{0:T}, Y_{0:T} = y_{0:T}\right) = \mu_{x_0} \prod_{t=1}^{T} g_{x_t}\left(y_t\right) A_{x_{t-1},x_t}$$

Example

Where is the frog at the beginning?

lightgray 1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

Table: Hidden Mark-frog chain example - Part I

Example - Transition probabilities (Matrix A)

Where can the frog jump?

	Level to						
	lightgray	1	2	3	4	5	6
Level from	1	0.4	0.6				
	2	0.3	0.4	0.6			
	3		0.3	0.4	0.6		
	4			0.3	0.4	0.6	
	5				0.3	0.4	0.6
	6	0.6				0.3	0.4

Table: Hidden Mark-frog chain example - Part II

Example - Emission probabilities (Matrix B)

What is the probability to detect the frog?

Movement sensor

	Detection	No detection
1	0.9	0.1
2	0.5	0.5
3	0.1	0.9
4	0	1
5	0	1
6	0	1

Table: Hidden Mark-frog chain example - Part III

Example - Observations

After 14 times, this are the sensor's results:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0	0	0	1	1	0	0	0	0	1	1	0	1

Table: Hidden Mark-frog chain example - Part IV

Using filtering algorithm

Forward Algorithm initialization equation

$$\alpha_1(i) = \pi_i \cdot b_i(O)$$

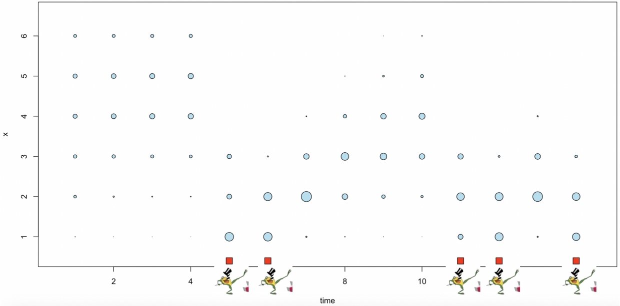
Forward Algorithm recursion equation

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i)\alpha_{ij}b_{ij}(O)$$

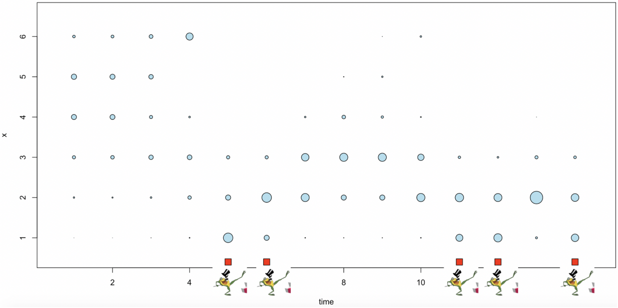
Termination

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

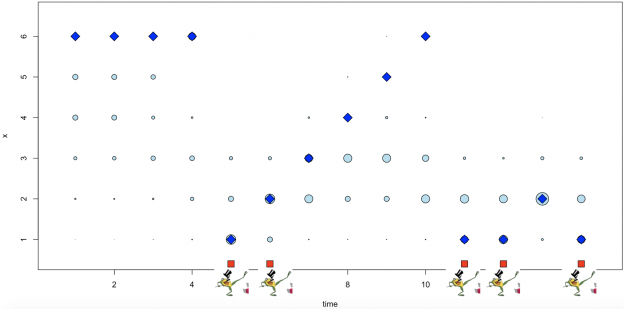
Distribution of X (filtering)



Distribution of X (smoothing)



Decoding of X



Formal Filtering

We are interested in the conditional probability mass function $p\left(x_t \mid y_{1:t}\right)$ of the state X_t given the data observed up to time t .

$$p(x_t|y_{1:t}) = \frac{p(x_t, y_{1:t})}{\sum_{x_t' \in \mathcal{X}} p(x_t', y_{1:t})}$$

Formal Filtering

We will derive a recursion for:

$$
\begin{aligned}
p(x_t, y_{t:T}) &= \sum_{x_{t-1}\in\mathcal{X}} p(x_t, x_{t-1}, y_t, y_{1:t-1}) \\
&= \sum_{x_t\in\mathcal{X}} p(y_t|x_t, x_{t-1}, y_{1:t-1})p(x_t|x_{t-1}, y_{1:t-1})p(x_{t-1}, y_{1:t-1}) \\
&= p(y_t|x_t)\sum_{x_t\in\mathcal{X}} p(x_t|x_{t-1})p(x_{t-1}, y_{1:t-1})
\end{aligned}
$$

If we define $\alpha(t) := p(x_t, y_{1:t})$, then we have a forward recursion for $t = 1, \ldots, T$, $x_t \in \mathcal{X}$:

$$
\alpha_t(x_t) = p(y_t|x_t)\sum_{x_t\in\mathcal{X}} p(x_t|x_{t-1})\alpha_{t-1}(x_{t-1});\;\; \alpha_0(x_0) = 0
$$

Formal Filtering and likelihood

So, we compute a filtering using:
For $t = 1, \ldots, T$, $x \in X$:

$$\alpha_t(x_t) = p(y_t \mid x_t) \sum_{x_{t-1} \in \mathcal{X}} p(x_t \mid x_{t-1}) \, \alpha_{t-1}(x_{t-1})$$

Formal Filtering and likelihood

The filtering pmf is obtained by normalizing $\alpha_t(x_t)$ as:

$$p(x_t|y_{1:t}) = \frac{p(x_t, y_{1:t})}{p(y_{1:t})} = \frac{\alpha_t(x_t)}{\sum_{x\in\mathcal{X}} \alpha_t(x)}$$

The likelihood term $p(y_{1:t-1})$ can be computed from the α-recursion:

$$p(y_{1:T}) = \sum_{x\in\mathcal{X}} \alpha_T(x)$$

Predict

$$p(x_t|y_{1:t-1}) = \sum_{x_{t-1}\in\mathcal{X}} p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})$$

Update

$$p(x_t|y_{1:t}) = \frac{g_{x_t}(y_t)p(x_t|y_{1:t})}{\sum_{x'_t\in\mathcal{X}} g_{x'_t}(y_t)p(x'_t|y_{1:t})}$$

Considerations in Filtering

- The proposed recursion may suffer from numerical underflow/overflow, as α_t may become very small or very large for large t.
- The proposed recursion may suffer from numerical underflow/overflow, as α_t may become very small or very large for large t.
- To avoid this, we can normalize α_t, or propagate the filtering *pmf* $p\left(x_t \mid y_{1:t}\right)$ instead of α_t, using the following two-step predict-update recursion:

$$p\left(x_t \mid y_{1:t-1}\right) = \sum_{x_{t-1} \in \mathcal{X}} p\left(x_t \mid x_{t-1}\right) p\left(x_{t-1} \mid y_{1:t-1}\right) \qquad \textit{Predict}$$

$$p\left(x_t \mid y_{1:t}\right) = \frac{g_{x_t}\left(y_t\right) p\left(x_t \mid y_{1:t-1}\right)}{\sum_{x_t' \in \mathcal{X}} g_{x_t'}\left(y_t\right) p\left(x_t' \mid y_{1:t-1}\right)} \qquad \textit{Update}$$