

IEOR 165 – Lecture 13

Markov Processes

1 Definition

A Markov process is a process in which the probability of being in a future state conditioned on the present state and past states is equal to the probability of being in a future state conditioned only on the present state. There are certain key features of Markov processes that can be used to classify different models:

1.1 State Space

The space to which the states of the model belong can be used to classify different models. The state describes the current configuration of the system, and it captures the important aspects of the system. The fact that the future evolution of a Markov process only depends on the current state is important because it means that for these classes of models the state completely characterizes the configuration of the system. In other words, there are no hidden variables that influence the evolution of the system. The notion of hidden states is particularly relevant in physics to quantum mechanics, where the question of whether the stochastic nature of quantum mechanics is due to hidden variables has been deeply studied.

Discrete state spaces are those in which the states are represented by a discrete set of objects. For example, we could describe a doctor in a hospital with a discrete state space in which some states could correspond to

- Meeting Patient
- Performing Procedure
- Idling
- Administration

Depending on the level of analysis, we might have more or less states to describe the system.

Continuous state spaces are those in which the states are represented by a continuous set of objects. An example is position of a ball we have thrown in space. The position of the ball can occupy different continuous values.

Hybrid state spaces are those in which the states are represented by a combination of discrete and continuous sets. For example, consider a bouncing ball. We need to use a discrete state to describe if the ball is in the air or if the ball is bouncing on the ground. Furthermore, we need continuous states to describe the position and velocity of the ball.

1.2 Initial Conditions

To define the evolution of a Markov process, we need to specify an *initial condition*, which represent the states of the system at the “start”. The initial condition can be stochastic or deterministic. For instance, we can specify a distribution of states the system “starts” at.

1.3 Time

Two common classes of models are discrete-time and continuous-time models. In most Markov processes, time is a *privileged variable*, meaning that it is interpreted as a clock that keeps track of the duration between events; however, this is not the case in general. For instance, special and general relativity models from physics drop time from this privileged position in a precise way.

Discrete-time models are those in which time increases in discrete increments. Starting from $t = 0$, time advances as $t = t + 1$, and actions in this class of models occur at every increment. Because of the discrete nature of time in this class of models, the time variable t is often used as an index for the state-variables. So if, say, the state is $x \in \mathbb{R}^p$, then the value of the state at time t is denoted x_t or $x[t]$. Because of the discrete nature of time, another convention is to use the variables n or k to denote time.

Continuous-time models are those in which time continuously increases. Starting from $t = 0$, time advances as $\frac{d}{dt}t = 1$, and actions in this class of models can occur at any point in time. Because of the continuous nature of time in this class of models, the time variable is used as the input into a function describing the states. So if the state is x , then the value of the state at time t is denoted as $x(t)$.

These are not the only classes of models for temporal evolution. For instance, hybrid-time models are those in which time increases in both discrete and continuous increments. These types of models occur when describing certain semi-autonomous, robotic, or embedded systems.

2 Discrete Time Markov Chains

A Markov chain is a Markov process in which the state space is discrete. Vertices are used as an abstraction for different quantities/states of the system. For instance, we can use vertices to represent the number of people waiting in the queue for a service system. The time in this class of models can be continuous, discrete, or hybrid.

Because of the discrete nature of the state space, the system is represented by a weighted directed graph $G = (V, E)$, where the vertices $v_i \in V$ represent states of the system and the edges $e_{ij} \in E$ denotes an edge from v_i going towards v_j . Furthermore, every vertex v_i has a self-loop, meaning an edge $e_{ii} \in E$ for all $v_i \in V$. By convention, these self-loops are not drawn because

they have a fundamentally different characteristic.

The edges $e_{ij} \in E$ for $i \neq j$ have strictly positive weights w_{ij} . For a discrete-time Markov chain, the weight w_{ij} will represent the probability of a transition to state v_j occurring when the system is currently at state v_i . In particular, if edge e_{ij} has weight w_{ij} then this means that we have the following probability of transitioning states for all $v_i, v_j \in V$ (even for $v_i = v_j$):

$$\mathbb{P}[x_{t+1} = v_j | x_t = v_i] = w_{ij}.$$

2.1 MLE Estimator

Suppose we have data that consists of $x_t \in V$ for $t = 0, \dots, n$. The question we would like to answer is how can we estimate the w_{ij} weights? One possible approach is maximum likelihood estimation. Before we define the likelihood, it is useful to define the following values

$$N_{ij} = \sum_{t=0}^{n-1} \mathbf{1}(x_t = v_i \text{ and } x_{t+1} = v_j).$$

These values count the number of transitions between vertices v_i to vertices v_j that were observed in the measured data. Next, note that the likelihood is given by

$$\begin{aligned} \mathbb{P}(x_t \text{ for } t = 0, \dots, n) &= \mathbb{P}(x_0) \cdot \mathbb{P}[x_1 | x_0] \cdot \mathbb{P}[x_2 | x_1, x_0] \cdots \mathbb{P}[x_n | x_{n-1}, \dots, x_0] \\ &= \mathbb{P}(x_0) \cdot \mathbb{P}[x_1 | x_0] \cdot \mathbb{P}[x_2 | x_1] \cdots \mathbb{P}[x_n | x_{n-1}] \\ &= \mathbb{P}(x_0) \cdot \prod_{t=0}^{n-1} \mathbb{P}[x_{t+1} | x_t] \\ &= \mathbb{P}(x_0) \cdot \prod_{v_i} \prod_{v_j} w_{ij}^{N_{ij}} \end{aligned}$$

We have used the Markov property in going from the first to second line, and the third line comes by definition of the N_{ij} values. Consequently, the MLE is given by

$$\max \left\{ \mathbb{P}(x_0) \cdot \prod_{v_i} \prod_{v_j} w_{ij}^{N_{ij}} \mid \sum_{v_j} w_{ij} = 1, \forall v_i \right\}.$$

The solution does not depend on $\mathbb{P}(x_0)$, and so we can drop this term. To simplify the calculation, we can take the negative log-likelihood. As a result, we solve

$$\min \left\{ - \sum_{v_i} \sum_{v_j} N_{ij} \log w_{ij} \mid \sum_{v_j} w_{ij} = 1, \forall v_i \right\}.$$

The KKT conditions are

$$\begin{aligned} -N_{ij}/w_{ij} + \lambda_i &= 0, \forall v_i, v_j \\ \sum_{v_j} w_{ij} &= 1, \forall v_i \end{aligned}$$

This means that $w_{ij} = N_{ij}/\lambda_i$. And so, $\sum_{v_j} N_{ij}/\lambda_i = 1$ implies that $\lambda_i = \sum_{v_j} N_{ij}$. Consequently, our estimate is

$$\hat{w}_{ij} = \frac{N_{ij}}{\sum_{v_j} N_{ij}}.$$

2.2 Example: Weather Forecasting

Q: A model for weather forecasting that works surprisingly well is to assume the weather on a given day only depends on the weather on the previous day. This model is then a discrete time Markov chain if we restrict the categories of weather to: sunny, rainy or cloudy. Suppose the data comparing the weather “today” and “tomorrow” for a given location can be summarized as follows:

	tomorrow sunny	tomorrow rainy	tomorrow cloudy
today sunny	286	123	165
today rainy	128	333	256
today cloudy	245	255	189

Find the maximum likelihood estimates of the transition probabilities.

A: We first construct a graph with three vertices, and we assign v_1 to “sunny”, v_2 to “rainy”, and v_3 to “cloudy”. Next, note we use the notation p_{ij} to represent the transition probability from v_i to v_j . (This notation is slightly different from above, where we used w_{ij} to represent this equanity. Then, the MLE is given by:

$$\begin{aligned}
\hat{p}_{11} &= \frac{n_{11}}{n_{11} + n_{12} + n_{13}} = \frac{286}{286 + 123 + 165} = \frac{286}{574} = 0.498 \\
\hat{p}_{12} &= \frac{n_{12}}{n_{11} + n_{12} + n_{13}} = \frac{123}{286 + 123 + 165} = \frac{123}{574} = 0.214 \\
\hat{p}_{13} &= \frac{n_{13}}{n_{11} + n_{12} + n_{13}} = \frac{165}{286 + 123 + 165} = \frac{165}{574} = 0.288 \\
\hat{p}_{21} &= \frac{n_{21}}{n_{21} + n_{22} + n_{23}} = \frac{128}{128 + 333 + 256} = \frac{128}{717} = 0.179 \\
\hat{p}_{22} &= \frac{n_{22}}{n_{21} + n_{22} + n_{23}} = \frac{333}{128 + 333 + 256} = \frac{333}{717} = 0.464 \\
\hat{p}_{23} &= \frac{n_{23}}{n_{21} + n_{22} + n_{23}} = \frac{256}{128 + 333 + 256} = \frac{256}{717} = 0.357 \\
\hat{p}_{31} &= \frac{n_{31}}{n_{31} + n_{32} + n_{33}} = \frac{245}{245 + 255 + 189} = \frac{245}{689} = 0.356 \\
\hat{p}_{32} &= \frac{n_{32}}{n_{31} + n_{32} + n_{33}} = \frac{255}{245 + 255 + 189} = \frac{255}{689} = 0.370 \\
\hat{p}_{33} &= \frac{n_{33}}{n_{31} + n_{32} + n_{33}} = \frac{189}{245 + 255 + 189} = \frac{189}{689} = 0.274
\end{aligned}$$