Probability and Stochastic Processes
Lecture 22: Discrete Time Markov Processes

Prof. Washim Uddin Mondal
Department of Electrical Engineering
Indian Institute of Technology Kanpur

Conditional Independence

Definition

Let X, Y, Z be three discrete random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The random variables X, Z are called conditionally independent given Y if the following holds $\forall x, y, z \in \mathbb{R}$ whenever $\mathbf{P}(Y = y) > 0$.

$$\mathbf{P}(X = x, Z = z | Y = y) = \mathbf{P}(X = x | Y = y)\mathbf{P}(Z = z | Y = y)$$

- The above condition is equivalent to the following.

$$\mathbf{P}(X = x, Y = y, Z = z)\mathbf{P}(Y = y) = \mathbf{P}(X = x, Y = y)\mathbf{P}(Y = y, Z = z), \ \forall x, y, z \in \mathbb{R}$$

- The above condition does not require $\mathbf{P}(Y = y) > 0$.

Conditional Independence

Example

Suppose there are two biased coins whose probabilities of occurrence of a head in a toss are 0.3 and 0.4, respectively. One of these coins is chosen uniformly at random and tossed twice independently. Let the random variables X_1, X_2 denote outcomes of these tosses (0: Tail, 1: Head), and $Y \in \{1, 2\}$ denote the index of the chosen coin.

(a) Argue that X_1, X_2 are conditionally independent given Y.

(b) Show that X_1, X_2 are not independent.

Solution to Part (a)

- Given a coin, the outcomes of the two tosses are independent.

Conditional Independence

Solution to Part (b)

- Let the underlying probability measure is $\mathbf{P}$. Note that $\mathbf{P}(Y=1) = \mathbf{P}(Y=2) = 0.5$.
- Moreover, $\mathbf{P}(X_1 = 1, X_2 = 1|Y=1) = (0.3)^2 = 0.09$
- Also, $\mathbf{P}(X_1 = 1, X_2 = 1|Y=2) = (0.4)^2 = 0.16$.
- Therefore, $\mathbf{P}(X_1 = 1, X_2 = 1) = \sum_{y=1}^{2} \mathbf{P}(Y=y)\mathbf{P}(X_1 = 1, X_2 = 1|Y=y) = 0.125$.
- Note that $\mathbf{P}(X_1 = 1) = \sum_{y=1}^{2} \mathbf{P}(Y=y)\mathbf{P}(X_1 = 1|Y=y) = 0.35$. Similarly, $\mathbf{P}(X_2 = 1) = 0.35$. Since, $0.35 \times 0.35 \neq 0.125$, we conclude the result.

Note

Conditional independence does NOT imply independence, in general.

Conditional Independence

Proposition 22.1

Let X, Y, Z be discrete random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. If X, Z are conditionally independent given Y, then $\mathbf{P}(Z = z | X = x, Y = y) = \mathbf{P}(Z = z | Y = y)$, $\forall x, y, z \in \mathbb{R}$ whenever $\mathbf{P}(X = x, Y = y) > 0$. The converse is also true.

Proof of Proposition 22.1

$$\mathbf{P}(Z = z | X = x, Y = y) = \frac{\mathbf{P}(X = x, Y = y, Z = z)}{\mathbf{P}(X = x, Y = y)}$$
$$= \frac{\mathbf{P}(Y = y, Z = z)}{\mathbf{P}(Y = y)} = \mathbf{P}(Z = z | Y = y)$$

Conditional Independence

Definition

Let (X, Y, Z) be jointly continuous random variables. We say that X, Z are conditionally independent given Y if $f_{X,Z|Y}(x, z|y) = f_{X|Y}(x|y)f_{Z|Y}(z|y)$, $\forall x, y, z \in \mathbb{R}$ whenever $f_Y(y) > 0$.

Equivalent Definitions

If (X, Y, Z) are jointly continuous random variables, we say that (X, Z) are conditionally independent given Y if the following holds $\forall x, y, z \in \mathbb{R}$.

(a) $f_{X,Y,Z}(x, y, z)f_Y(y) = f_{X,Y}(x, y)f_{Y,Z}(y, z)$.

(b) $f_{Z|X,Y}(z|x, y) = f_{Z|Y}(z|y)$, whenever $f_{X,Y}(x, y) > 0$.

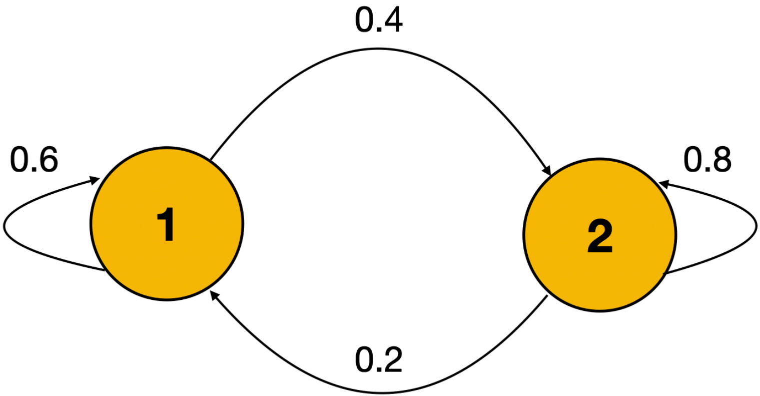
Markov Chain

Motivating Example

Assume that the weather of a certain region can be either sunny or cloudy. Let the random variable $X_n \in \{1, 2\}$ denotes the weather of the nth day $(1 :$ cloudy and $2 :$ sunny). Clearly, $\{X_n\}_{n \in \mathbb{N}}$ is a discrete time random process. Assume, additionally, that if the nth day is sunny, then the $(n+1)$th day is also sunny with a probability of 0.8. On the other hand, if the nth day is cloudy, then the $(n+1)$th day is also cloudy with a probability of 0.6. Note the following for this simple weather model.

- X_n completely specifies the statistics of X_{n+1}.
- The knowledge of the history $(X_1, \cdots, X_{n-1})$ does not help better predicting the future X_{n+1} if the current weather X_n is known.

Markov Chain



Markov Chain

Definition

Let X_n be a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ that takes values in $\mathcal{S}$, $\forall n \in \mathbb{N}$. We say that $\{X_n\}_{n\in\mathbb{N}}$ is a discrete-time Markov chain (DTMC) with state space $\mathcal{S}$ if $\forall n, m \in \mathbb{N}$ either of the following equivalent conditions holds.

(a) $(X_1, \cdots, X_{n-1})$ and X_{n+1} are conditionally independent given X_n.

(b) $(X_1, \cdots, X_{n-1})$ and $(X_{n+1}, \cdots, X_{n+m})$ are conditionally independent given X_n.

Notes

- The state space $\mathcal{S}$ can be either countable or uncountable.
- For simplicity, we take $\mathcal{S} = \{1, \cdots, S\}$ in this course.

Markov Chain

Notes

- The conditions (a) and (b) can be mathematically expressed as:

$$
\begin{aligned}
&\mathbf{P}(X_{n+1} = x_{n+1} | X_n = x_n, \cdots, X_1 = x_1) = \mathbf{P}(X_{n+1} = x_{n+1} | X_n = x_n) \\
&\mathbf{P}(X_{n+1} = x_{n+1}, \cdots, X_{n+m} = x_{n+m} | X_n = x_n, \cdots, X_1 = x_1) \\
&\qquad = \mathbf{P}(X_{n+1} = x_{n+1}, \cdots, X_{n+m} = x_{n+m} | X_n = x_n)
\end{aligned}
$$

  where $x_1, \cdots, x_{n+m} \in \mathcal{S}$.
- Although both conditions are equivalent, condition (a) is easier to check in practice.

Markov Chain

Notes

- Markov chains are *almost memoryless* since the future states depend only on the current state, not on the history.
- $\mathbf{P}(X_{n+1} = j | X_n = i)$ is called the (one-step) transition probability from state i to state j.
- The above term, in general, may depend on n as well as on i, j.
- If the above transition probability solely depends on i, j, and not on n, then $\{X_n\}_{n \in \mathbb{N}}$ is called a *time-homogeneous* DTMC.
- This course primarily discusses time-homogeneous DTMCs.

Markov Chain

Notes

- For time-homogeneous DTMC, we denote $\mathbf{P}(X_{n+1} = j | X_n = i) = p_{ij}$.
- $P = (p_{ij})_{i,j \in \mathcal{S}}$ is called the transition probability matrix of the underlying DTMC. If $\mathcal{S} = \{1, \cdots, S\}$, the transition probability matrix takes the following form.

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1S} \\ \vdots & \vdots & \vdots \\ p_{S1} & \cdots & p_{SS} \end{pmatrix}$$

- P is row stochastic since each of its entries is non-negative, and each row of P sums up to 1, i.e., $\sum_{j=1}^{S} p_{ij} = 1, \forall i \in \{1, \cdots, S\}$.

Markov Chain

Example

For the simple weather model stated before, identify the state space and write the transition probability matrix.

Solution

- The state space is $\mathcal{S} = \{1, 2\}$. Moreover, the transition probability matrix is:

$$P = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

- P_{ij} is the transition probability from state i to state j, where $(i, j) \in \{1, 2\}^2$.

Markov Chain

Proposition 22.2

Let $\{X_n\}_{n\in\mathbb{N}}$ be a time-homogeneous DTMC with state space $\mathcal{S} = \{1, \cdots, S\}$, and transition probability matrix P. If the underlying probability measure is $\mathbf{P}$, and $\pi^{(n)} = [\pi_1^{(n)}, \cdots, \pi_S^{(n)}]$, where $\pi_i^{(n)} = \mathbf{P}(X_n = i)$, then the following holds, $\forall n \in \mathbb{N}$.

$$\pi^{(n+1)} = \pi^{(n)} P = \pi^{(1)} P^n$$

Proof of Proposition 22.2

$$\mathbf{P}(X_{n+1} = j) = \sum_{i=1}^{S} \mathbf{P}(X_{n+1} = j | X_n = i)\mathbf{P}(X_n = i)$$

Markov Chain

Example

For the simple weather model stated before, if the weather is sunny with probability 0.3 on day 1, what is the probability that day 3 is sunny?

Solution

- Clearly, $\pi^{(1)} = [\mathbf{P}(X_1 = 1),\ \mathbf{P}(X_1 = 2)] = [0.7, 0.3]$. Hence,

$$\pi^{(3)} = \begin{pmatrix} 0.7 & 0.3 \end{pmatrix} \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}^2 = \begin{pmatrix} 0.7 & 0.3 \end{pmatrix} \begin{pmatrix} 0.44 & 0.56 \\ 0.28 & 0.72 \end{pmatrix} = \begin{pmatrix} 0.392 & 0.608 \end{pmatrix}$$

- Required probability is 0.608.

Markov Chain

Notes

- What do the entries of P^n represent?
- If, initially, the system is at state i, i.e., $\pi_i^{(1)} = 1$ and $\pi_k^{(1)} = 0, \forall k \neq i$, then

$$\pi^{(n+1)} = \pi^{(1)} P^n = (P^n)_i$$

- Hence, $(P^n)_{ij} = \pi_j^{(n+1)} = \mathbf{P}(X_{n+1} = j)$.
- $(P^n)_{ij}$ is the n-step transition probability from state i to state j.
- One can show that P^n is also row stochastic $\forall n \in \mathbb{N}$.

Ergodicity

Definition

Let $\{X_n\}_{n\in\mathbb{N}}$ be a DTMC with a finite state space $\mathcal{S}$, and a transition probability matrix P. If, for some $n \in \mathbb{N}$, $(P^n)_{ij} > 0$, $\forall i, j \in \mathcal{S}$, then $\{X_n\}_{n\in\mathbb{N}}$ is said to be ergodic.

Proposition 22.3

Let $\{X_n\}_{n\in\mathbb{N}}$ be an ergodic DTMC with a finite state space $\mathcal{S} = \{1, \cdots, S\}$ and a transition probability P. The following results hold.

(a) $\lim_{n\to\infty}(P^n)_{ij} = \pi_j$ exists and is independent of i, $\forall i, j \in \mathcal{S}$.

(b) $\pi_j \geq 0$, $\forall j \in \mathcal{S}$, and $\sum_{j=1}^{S} \pi_j = 1$.

(c) $\pi = \pi P$ where $\pi = [\pi_1, \cdots, \pi_S]$.

Ergodicity

Notes

- Note that π is a valid PMF over the state space $\mathcal{S}$.
- π is called the stationary distribution of the underlying ergodic DTMC.
- Note that, for ergodic DTMC, the limiting value of P^n has the following form.

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi_1 & \cdots & \pi_S \\ \vdots & \vdots & \vdots \\ \pi_1 & \cdots & \pi_S \end{pmatrix}$$

- In other words, the rows of $\lim_{n\to\infty} P^n$ are identical.

Ergodicity

Example

Find the stationary distribution of the simple weather model stated before.

Solution

- The underlying DTMC is ergodic since all entries of P are non-zero.
- Let $\pi = (\pi_1, \pi_2)$ be the stationary distribution. We have,

$$\begin{aligned}
\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} &= \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} \\
&= \begin{pmatrix} 0.6\pi_1 + 0.2\pi_2 & 0.4\pi_1 + 0.8\pi_2 \end{pmatrix}
\end{aligned}$$

Ergodicity

Solution (Continued)

- We get $\pi_1 = 0.6\pi_1 + 0.2\pi_2$, which implies that $\pi_2 = 2\pi_1$.
- Also, $\pi_1 + \pi_2 = 1$, which leads to $\pi = (\frac{1}{3}, \frac{2}{3})$.
- One can also obtain π by computing $\lim_{n\to\infty} P^n$. Note that

$$P^5 = \begin{pmatrix} 0.34016 & 0.65984 \\ 0.32992 & 0.67008 \end{pmatrix}, \quad P^{10} = \begin{pmatrix} 0.3334 & 0.6666 \\ 0.3333 & 0.6667 \end{pmatrix}$$

- Each row of P^n converges towards $\pi = (\frac{1}{3}, \frac{2}{3})$ as $n \to \infty$.