Markov Chains

Kshitij Tomar

July 14, 2024

IID (Type of Distribution)

A simple example of a random variable, which you have already experienced, are situations such as rolling a dice or flipping a coin. These systems, doing these tasks repeatedly, are considered $i.i.d$. That is, they are independent and identical distributions. Let X represent of the outcome of a standard dice roll.

$$P(X = 1) = P(X = 2) = P(X = 3) = \cdots P(X = 6) = \frac{1}{6}.$$

In specific, rolling a dice is also uniformly distributed, which means that picking any of the possible outcomes is equally likely. Now let's suppose that we roll another standard dice. Let Y be that outcome. X and Y are i.i.d. We know how systems like these behave, however they present an unrealistic model of most actual systems we want to study.

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
    - Weather evolves over time

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
   - Weather evolves over time
2. Weather tomorrow *depends* on the weather in the past couple of days

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
    - Weather evolves over time
2. Weather tomorrow *depends* on the weather in the past couple of days
3. Weather patterns must exhibit *some* probabilistic behavior

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
   - Weather evolves over time
2. Weather tomorrow *depends* on the weather in the past couple of days
3. Weather patterns must exhibit *some* probabilistic behavior
   - There must be some non-zero probability of it being sunny tomorrow

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
   - Weather evolves over time
2. Weather tomorrow *depends* on the weather in the past couple of days
3. Weather patterns must exhibit *some* probabilistic behavior
   - There must be some non-zero probability of it being sunny tomorrow
   - The total probability of all choices must sum to one

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
   - Weather evolves over time
2. Weather tomorrow *depends* on the weather in the past couple of days
3. Weather patterns must exhibit *some* probabilistic behavior
   - There must be some non-zero probability of it being sunny tomorrow
   - The total probability of all choices must sum to one
   - Because there has to be some weather tomorrow

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
    - Weather evolves over time
2. Weather tomorrow *depends* on the weather in the past couple of days
3. Weather patterns must exhibit *some* probabilistic behavior
    - There must be some non-zero probability of it being sunny tomorrow
    - The total probability of all choices must sum to one
    - Because there has to be some weather tomorrow
    - Probability of Sunny tomorrow is impacted by Rain today

The Problem with IID

Lets suppose we are working at a weather station and we aim to built a long term probability model to predict the weather. Note that

1. Weather tomorrow is **not** independent to Weather today
   - Weather evolves over time
2. Weather tomorrow *depends* on the weather in the past couple of days
3. Weather patterns must exhibit *some* probabilistic behavior
   - There must be some non-zero probability of it being sunny tomorrow
   - The total probability of all choices must sum to one
   - Because there has to be some weather tomorrow
   - Probability of Sunny tomorrow is impacted by Rain today

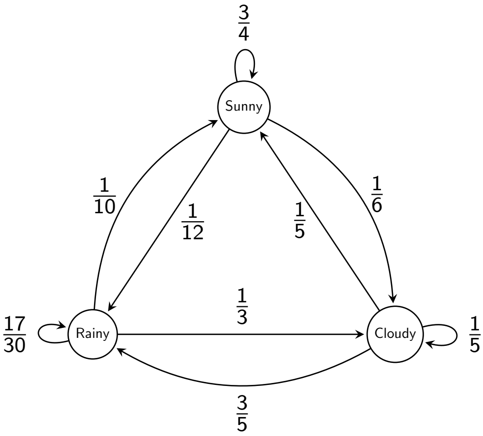
For the sake of simplicity, let's assume that the weather tomorrow only depends on the weather today. Since we are also at a weather station, we can keep track of the weather patterns as they emerge and collect samples from this unknown probability distribution.

The Solution (Its not the graph)

Let's now suppose that we spent a month collecting data at our weather station about the possible weather conditions of the sky: Sunny, Cloudy, and Rainy. Drawing out the results from the lab.

The Solution (Its not the graph)

Let's now suppose that we spent a month collecting data at our weather station about the possible weather conditions of the sky: Sunny, Cloudy, and Rainy. Drawing out the results from the lab.



What does this tell us?

Lets assume that these are the probabilities that describe transitions from one weather (state) to another. Instead of drawing a graph every time, these probabilities may be succinctly written into a transition matrix. Let P be this matrix.

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ \frac{1}{10} & \frac{1}{3} & \frac{17}{30} \end{pmatrix}$$

What does this tell us?

Lets assume that these are the probabilities that describe transitions from one weather (state) to another. Instead of drawing a graph every time, these probabilities may be succinctly written into a transition matrix. Let P be this matrix.

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ \frac{1}{10} & \frac{1}{3} & \frac{17}{30} \end{pmatrix}$$

For any entry, it represents the movement from a state to another state. For example, the entry $P(0,1) = \frac{1}{6}$ represents the probability of moving from a sunny weather today to a cloudy weather tomorrow. For reference, Sunny $= 0$, Cloudy $= 1$, and Rainy $= 2$.

What does this tell us?

Lets assume that these are the probabilities that describe transitions from one weather (state) to another. Instead of drawing a graph every time, these probabilities may be succinctly written into a transition matrix. Let P be this matrix.

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ \frac{1}{10} & \frac{1}{3} & \frac{17}{30} \end{pmatrix}$$

For any entry, it represents the movement from a state to another state. For example, the entry $P(0,1) = \frac{1}{6}$ represents the probability of moving from a sunny weather today to a cloudy weather tomorrow. For reference, Sunny $= 0$, Cloudy $= 1$, and Rainy $= 2$.
Notice all rows have to sum to 1!

What does this tell us?

Lets assume that these are the probabilities that describe transitions from one weather (state) to another. Instead of drawing a graph every time, these probabilities may be succinctly written into a transition matrix. Let P be this matrix.

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ \frac{1}{10} & \frac{1}{3} & \frac{17}{30} \end{pmatrix}$$

For any entry, it represents the movement from a state to another state. For example, the entry $P(0,1) = \frac{1}{6}$ represents the probability of moving from a sunny weather today to a cloudy weather tomorrow. For reference, Sunny $= 0$, Cloudy $= 1$, and Rainy $= 2$.

Notice all rows have to sum to 1! So what?

So What?

To see the power of the transition matrix, consider the probability we want for the Sunny weather in the upcoming two days if today was Sunny. Let S represent a Sunny weather, C represent a cloudy weather and R represent a rainy weather. Let $\chi = \{S, C, R\}$. Then, weather on day $t \geq 0$ can be represented as X_t

$$\mathbb{P}(X_t = S \mid X_{t-2} = S) = \sum_{x \in \chi} P(X_t = S \mid X_{t-1} = x)\, P(X_{t-1} = x \mid X_{t-2} = S).$$

$$\mathbb{P}(X_t = S \mid X_{t-2} = S) = P^2(S, S).$$

This can be done iterative for any integer number of times t.

So...we can do ~~Magic~~ Math

We can start with an arbitrary selection out of our possible choices. Let's suppose that we pick $x \in \chi$. Note that the starting weather could be determined through some probability distribution or it could be picked arbitrarily. Now we define a probability distribution μ so that $\mu_t(x) = \frac{1}{t}\#\{X_q = x : q \leq t\}$ for all $x \in \chi$. As we sample more according to our transition matrix P, our probability distribution will get closer to the actual long-term probability distribution.

So...we can do ~~Magic~~ Math

We can start with an arbitrary selection out of our possible choices. Let's suppose that we pick $x \in \chi$. Note that the starting weather could be determined through some probability distribution or it could be picked arbitrarily. Now we define a probability distribution μ so that $\mu_t(x) = \frac{1}{t}\#\{X_q = x : q \leq t\}$ for all $x \in \chi$. As we sample more according to our transition matrix P, our probability distribution will get closer to the actual long-term probability distribution.
Notice that $\mu_{t+1}(x) = \mu_t P$ for all $t \geq 0$.

So...we can do ~~Magic~~ Math

We can start with an arbitrary selection out of our possible choices. Let's suppose that we pick $x \in \chi$. Note that the starting weather could be determined through some probability distribution or it could be picked arbitrarily. Now we define a probability distribution μ so that $\mu_t(x) = \frac{1}{t}\#\{X_q = x : q \leq t\}$ for all $x \in \chi$. As we sample more according to our transition matrix P, our probability distribution will get closer to the actual long-term probability distribution.
Notice that $\mu_{t+1}(x) = \mu_t P$ for all $t \geq 0$.
We are interested in the limiting distribution, which is taking $t \to \infty$. As we travel further and further out into the chain, this makes it so that our initial arbitrary decision has negligible effect over the entire distribution.
All limiting distributions π are stationary in the sense that $\pi = \pi P$.

So...we can do ~~Magic~~ Math

We can start with an arbitrary selection out of our possible choices. Let's suppose that we pick $x \in \chi$. Note that the starting weather could be determined through some probability distribution or it could be picked arbitrarily. Now we define a probability distribution μ so that $\mu_t(x) = \frac{1}{t}\#\{X_q = x : q \leq t\}$ for all $x \in \chi$. As we sample more according to our transition matrix P, our probability distribution will get closer to the actual long-term probability distribution.

Notice that $\mu_{t+1}(x) = \mu_t P$ for all $t \geq 0$.

We are interested in the limiting distribution, which is taking $t \to \infty$. As we travel further and further out into the chain, this makes it so that our initial arbitrary decision has negligible effect over the entire distribution.

All limiting distributions π are stationary in the sense that $\pi = \pi P$.

So formally...

Markov Chains

Definition

Let χ be a finite state space. A sequence of random variables (X_t) is called a **Markov Chain**, where $t \in A$ is an indexing set and $X_t \in \chi$ for all t, when for all $x, y \in \chi$ and $t \in A$,

$$\mathbb{P}(X_t = y \mid X_{t-1} = x, X_{t-2}, \ldots, X_0) = \mathbb{P}(X_t = y \mid X_{t-1} = x).$$

Stationary Distributions

Definition

Let χ be a finite state space. A **transition matrix** P of the Markov Chain (X_t) on χ is a $|\chi| \times |\chi|$ matrix such that for all $x, y \in \chi$

$$\mathbb{P}(X_t = y \mid X_{t-1} = x) = P(x, y)$$

$P(x, y)$ is known as the **transition probability** of moving from a state x to a state y.

Definition

For a Markov Chain (X_t) on a finite state space χ, a distribution π is a **stationary distribution** if $\pi P = \pi$ where P is the transition matrix.

Why this works

Theorem (Ergodic Theorem)

Let $f : \chi \to \mathbb{R}$ be a function on the finite state space χ. If (X_i) is an irreducible chain with stationary distribution π, then for any starting distribution μ

$$\mathbb{P}\left(\lim_{t\to\infty} \frac{1}{t}\sum_{i=0}^{t-1} f(X_i) = \mathbb{E}\left[f\right]_\pi\right) = 1.$$

Theorem (Convergence Theorem)

Suppose P is an irreducible and aperiodic transition matrix, with stationary distribution π on a finite state space χ. Then for all $t > 0$, there exist constants $a \in (0,1)$ and $C > 0$ such that

$$\max_{x\in\chi}\{d(P^t(x,\cdot),\pi)_{TV}\} \leq Ca^t.$$

Applications Mentioned in Paper

Applications Mentioned in Paper

1. Markov Chain Monte Carlo: Creating a Markov chain to find a transition matrix such that a given probability distribution is stationary with respect to that transition matrix

Applications Mentioned in Paper

1. Markov Chain Monte Carlo: Creating a Markov chain to find a transition matrix such that a given probability distribution is stationary with respect to that transition matrix
2. Monte Carlo Integration: Approximating the area of higher dimension functions using Monte Carlo Simulation techniques

Applications Mentioned in Paper

1. Markov Chain Monte Carlo: Creating a Markov chain to find a transition matrix such that a given probability distribution is stationary with respect to that transition matrix
2. Monte Carlo Integration: Approximating the area of higher dimension functions using Monte Carlo Simulation techniques
3. Science Models: ASEP Model, Ising Model, Ehrenfest urn Model, Kirchoff's Node Laws, and many others

Applications Mentioned in Paper

1. Markov Chain Monte Carlo: Creating a Markov chain to find a transition matrix such that a given probability distribution is stationary with respect to that transition matrix
2. Monte Carlo Integration: Approximating the area of higher dimension functions using Monte Carlo Simulation techniques
3. Science Models: ASEP Model, Ising Model, Ehrenfest urn Model, Kirchoff's Node Laws, and many others
4. Computer Science Models: Google Pagerank, Autocorrect, etc.

Applications Mentioned in Paper

1. Markov Chain Monte Carlo: Creating a Markov chain to find a transition matrix such that a given probability distribution is stationary with respect to that transition matrix
2. Monte Carlo Integration: Approximating the area of higher dimension functions using Monte Carlo Simulation techniques
3. Science Models: ASEP Model, Ising Model, Ehrenfest urn Model, Kirchoff's Node Laws, and many others
4. Computer Science Models: Google Pagerank, Autocorrect, etc.
5. Cryptogrophy: Useful for deciphering a cipher's content by studying the likelihood of letters being english letters based on their positioning/frequency in a corpus of text

Thank You for Listening

Check out my paper for more on this topic!