Math 456: Mathematical Modeling

Tuesday, February 19th 2018

Markov Chains: Basic Examples

Tuesday, February 19th 2018

Last time

1.Gambler's Ruin

A sequence of random variables taking values in $\{0, 1, 2, \ldots, M\}$

$$X_0, X_1, X_2, \ldots$$

If $0 < i < M$, then

$$\begin{aligned} \mathbb{P}(X_{n+1} = i - 1 \mid X_n = i) &= p, \\ \mathbb{P}(X_{n+1} = i + 1 \mid X_n = i) &= q, \\ \mathbb{P}(X_{n+1} = j \mid X_n = i) &= 0, \text{ otherwise,} \end{aligned}$$

Last time

1.Gambler's Ruin

The sequence stops at 0 or M,

$$\mathbb{P}(X_{n+1} = 0 \mid X_n = 0) = \mathbb{P}(X_{n+1} = M \mid X_n = M) \qquad = 1.$$

The sequence also has what we called the Markov property.

Last time

2.The Markov Property

Last time

2.The Markov Property

The transition probability matrix p has the following properties

i. All of its entries are nonnegative and no larger than 1.
ii. The sum of each of its rows is equal to 1.
iii. It encodes all the information about the chain.

Last time

3.The Chapman-Kolmogorov equation

For a homogeneous Markov chain, if we introduced the multi-step probabilities,

$$\mathrm{p}^n(i,j) = \mathrm{P}(X_n = j \mid X_0 = i),$$

where $i, j = 1, \ldots, N$. Then, we have the equation

$$\mathrm{p}^{n+m}(i,j) = \sum_{k=1}^{N} \mathrm{p}^n(i,k)\mathrm{p}^m(k,j)$$

Today

Today we will see:

1. Examples, lots of examples.
2. Discussion regarding what $p^n(i, j)$ looks like for n large.
3. Markov Chains starting from a random state.
4. Stationary distributions.

Now, some examples of Markov chains and transition probability matrices.

Example: The Ehrenfest Chain

We have N marbles distributed in two bowls labeled 1 and 2.

At each step, one of the marbles is chosen randomly –uniformly, and regardless of how they are divided between the bowls. The chosen marble is moved to the **other** bowl.

Then, the state of the system is described by

$$X_n = \#\Big\{ \text{ of marbles in bowl 1 at the end of the } n\text{-th step.} \Big\}$$

Example: The Ehrenfest Chain

It turns out X_n is a homogeneous chain **(why?!)**

At each step, X_n can only change by 1, so

$$\mathrm{p}(i, j) = 0 \text{ if } j \neq i \pm 1.$$

On the other hand, it should be clear that

$$\mathrm{p}(i, i-1) = \frac{i}{N}, \;\; \mathrm{p}(i, i+1) = \frac{N-i}{N}.$$

Example: The Ehrenfest Chain

For instance, if $N = 4$, then the transition probability matrix is

$$\mathrm{p} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Example: The Ehrenfest Chain

Meanwhile (still $N = 4$) the 2-step probability matrix is

$$\mathrm{p}^2 = \begin{pmatrix} 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 5/8 & 0 & 3/8 & 0 \\ 1/8 & 0 & 3/4 & 0 & 1/8 \\ 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \end{pmatrix}$$

Example: The Ehrenfest Chain

Exercise 1: Consider an Ehrenfest chain with N marbles.

Fix $i = 1, \ldots, N$. Determine $\mathrm{p}^n(i, i)$ for $n = 2, 3$. What can be said about it for general n?

Exercise 2: Consider an Ehrenfest chain with 3 marbles.
Find a distribution (well, vector) q such that

$$q\mathrm{p} = q$$

where the entries of q are nonnegative and add up to 1.

Example: The Ehrenfest Chain

The Ehrenfest chain is a very simple stochastic model for particles bouncing all over two chambers connected by a small tunnel.

See for instance, this simulation (with deterministic dynamics), and compare with what happens with the stochastic model:

`https://www.youtube.com/watch?v=pK1NPKm2Dfc`

Example: Random shuffling

Consider a deck of N cards, very creatively labeled,

$$\{1, \ldots, N\}$$

A permutation is simply a one to one function

$$\sigma : \{1, \ldots, N\} \to \{1, \ldots, N\}$$

A special type of permutation is a **transposition**:

$$\begin{aligned} \sigma(i) &= i \ \ \forall\, i \neq j, k \\ \sigma(j) &= k \\ \sigma(k) &= j \end{aligned}$$

Example: Random shuffling

Permutations can be composed with one another, and there are exactly $N!$ of them. The set of permutations is called the symmetric group $\mathcal{S}_N$.

A Markov chain with state space $\mathcal{S}_N$ is given as follows

$$X_0 = \sigma_e \text{ (the identity permutation)}$$

Example: Random shuffling

Given X_n, X_{n+1} is obtained by

$$X_{n+1} = \sigma_n \circ X_n$$

where σ_n is a sequence of i.i.d. random transpositions.

$$\mathbb{P}(\sigma_n = \text{any non-trivial transposition}) = \frac{1}{N^2},$$
$$\mathbb{P}(\sigma_n = \sigma_e \text{ (the identity permutation)}) = \frac{1}{N}.$$

Example: Random shuffling

The Markov chain is a model for how much can one reshuffle a deck of card by randomly shuffling two cards at any time.

Note that for $N = 3$, the state space has $3! = 6$ elements, so the transition matrix for the chain would be a 6×6 matrix.

Examples: The Wright-Fisher Model

In this model, the chain X_n varies over the set of states

$$S = \{0, 1, 2, \ldots, N\}$$

For some $N \in \mathbb{N}$, which usually (as we discuss below), is even.

The transition probabilities are given, for $i, j = 0, 1, \ldots, N$, by

$$\mathrm{p}(i,j) = \binom{N}{j} \left(\frac{i}{N}\right)^{j} \left(1 - \frac{i}{N}\right)^{N-j}$$

Examples: The Wright-Fisher Model

$$p(i,j) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$$

This chain is a simplified model for genetic drift. As follows:

- At each stage, X_n describes the distribution of two genes within a population. Thus, $X_n = \#$ of type A genes, and $N - X_n = \#$ of type B genes.
- The size of the population stays fixed in each generation, and generations do not overlap.
- The population at time $n+1$ is obtained by **drawing with replacement** N times from the population at time n.

Examples: The Wright-Fisher Model

The transition matrix, if say, $N = 4$, looks as follows

$$\begin{pmatrix} (1)^4 & 4(0)^1(1)^3 & 6(0)^2(1)^2 & 4(0)^3(1) & (0)^4 \\ (3/4)^4 & 4(1/4)(3/4)^3 & 6(1/4)^2(3/4)^2 & 4(1/4)^3(3/4) & (1/4)^4 \\ (1/2)^4 & 4(1/2)(1/2)^3 & 6(1/2)^2(1/2)^2 & 4(1/2)^3(1/2) & (1/2)^4 \\ (1/4)^4 & 4(3/4)(1/4)^3 & 6(3/4)^2(1/4)^2 & 4(3/4)^3(1/4) & (3/4)^4 \\ (0)^4 & 4(1)(0)^3 & 6(1)^2(0)^2 & 4(1)^3(0) & (1)^4 \end{pmatrix}$$

Examples: The Wright-Fisher Model

Simplifying a bit, p is given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 81/256 & 27/64 & 54/256 & 3/64 & 1/256 \\ 1/16 & 1/4 & 6/16 & 1/4 & 1/16 \\ 1/256 & 3/64 & 54/256 & 27/64 & 81/256 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We observe that $\mathrm{p}(x, x) = 1$ if $x = 0$ or $x = 1$.
We also note that $0 < p(x, y) < 1$ as soon as $x \neq 0, 1$.

Examples: The Wright-Fisher Model

Exercise 3:
Write a code that takes a $N \times N$ transition probability matrix and a positive number n, and produces the n-th power of a transition probability matrix, presenting the output in visual form (i.e. writing the rows and columns of the matrix).

Use this to calculate $\mathrm{p}^2, \mathrm{p}^5, \mathrm{p}^{10}, \mathrm{p}^{20}$ and p^{40} for: the Gambler's ruin (with $M = 4$), the Ehrenfest chain (with $N = 4$), and the Wright-Fisher model (with $N = 4$).

What pattern do you see as n increases for each matrix?.

Examples: Random walk on a graph

Random walk on a graph
In order to discuss random walks on graphs, let us review the basic definitions surrounding graphs.

A **simple graph** G is a pair of sets $G = (V, E)$, where

• V is a set of nodes or **vertices**
• E is a set of links or **edges** between vertices, that is, the elements of E are unordered pairs of elements of V.

If $x, y \in V$ are such that $\{x, y\} \in E$, we say there is an edge between them, or that they are neighbors.

Examples: Random walk on a graph

We only deal with finite graphs, which, accordingly, also have finitely many edges.

Degree of a vertex
Given $x \in V$, the set of neighboring vertices to x is denoted N_x.

Note that x may, or may not belong to N_x. The number of elements of N_x is called the **degree of** x is denoted d_x.

Examples: Random walk on a graph

Given a graph G, we define a Markov chain X_n through the following transition probabilities:

$$\mathrm{P}(X_{n+1} = y \mid X_n = x) = \begin{cases} 0 & \text{if } y \notin N_x \\ \frac{1}{d_x} & \text{if } y \in N_x \end{cases}$$

That is, at each stage, the process jumps from its current vertex, to one of the neighboring vertices, each of these vertices being equally likely.

Examples: Random walk on a graph

Some examples of graphs:

1. The complete graph in N vertices.
2. Bipartite graphs.
3. Trees
4. Regular graphs

Examples: Random walk on a graph

Graphs are used to describe, among **many** things,

1. Crystals, molecules, atomic structures.
2. Neural networks.
3. Artificial neural networks.
4. Language.
5. Transportation networks and other infraestructure.
6. Social networks.

Examples: Random walk on a graph

The adjacency matrix

If one labels the vertex of a graph from 1 to N(=total number of vertices), this determines a $N \times N$ matrix known as the graph's adjacency matrix. If A denotes the matrix, then

$$A_{ij} = \begin{cases} 1 & \text{if } i, j \text{ are neighbors} \\ 0 & \text{if } i, j \text{ are } \textbf{not} \text{ neighbors} \end{cases}$$

Examples: Random walk on a graph

The adjacency matrix
Observe three important things:

1) The adjacency matrix is a symmetric matrix.

2) If one sums the elements in a the i-th row, that returns the degree corresponding to the vertex i.

3) The adjacency matrix looks a lot like the transition probability matrix for the respective random walk.

Math 456: Mathematical Modeling

Thursday, February 21st 2018

Markov Chains:
More examples and stationary distributions

Thursday, February 21st 2018

Examples

A few words about **weighted graphs**.

Examples

Weighted graph

A **weighted graph** G is a set of vertices V together with a non-negative function

$$w : V \times V \mapsto \mathbb{R}$$

Known as the **weight**. If $w_{ij} > 0$ we say the vertex j is a **neighbor** of i.

The **degree** of i, d_i, is defined by $d_i := \sum_j w_{ij}$

Examples

Weighted graph

An important class of concrete weighted graphs is the following:

You are given vectors $x_1, \ldots, x_N$ in $\mathbb{R}^d$ and $\varepsilon > 0$.

Then, take $V = \{x_1, \ldots, x_N\}$ with weight $w_{x_i x_j}$ given by

$$w_{x_i x_j} := e^{-\frac{|x_i - x_j|^2}{\varepsilon}}$$

This type of weighted graph is ubiquitous in statistical inference, image processing, and machine learning.

Examples

Weighted graph

Note: It should be clear that a graph as defined last time yields a special type of weighted graph:

$$w_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \text{ is an edge,} \\ 0 & \text{otherwise.} \end{cases}$$

Random walk on a weighted graph:
This is a random walk where the transition probabilities are proportional to the respective edge weight.

As such, the transition matrix for this process is given by

$$\mathrm{p}(i, j) := \frac{w_{ij}}{d_i}.$$

The weight matrix
Labeling the vertices of the graph from 1 to N (N=total number of vertices), we have a $N \times N$ matrix $(w_{ij})_{N\times N}$ known as the graph's **weight matrix.**

For a simple graph, the weight matrix has only 1's and 0's, it coincides with the adjacency matrix.

Examples

Random walk on a graph

The weight matrix
Labeling the vertices of the graph from 1 to N (N=total number of vertices), we have a $N \times N$ matrix $(w_{ij})_{N\times N}$ known as the graph's **weight matrix.**

For a simple graph, the weight matrix has only 1's and 0's, it coincides with the adjacency matrix.

It turns out, weighted graphs exhaust all Markov chains: *every Marko Chain is a random walk in some weighted graph*

Markov chains $\Leftrightarrow$ Random walk on weighted graphs

Markov chains with random initial state

If X_0 is random, take the vectors given by

$$q_0(i) := \mathrm{P}(X_0 = i) \text{ and } q_n(i) := \mathrm{P}(X_n = i)$$

where again $i = 1, \ldots, N$. Then, we have

$$q_{n+1}(j) = \sum_{j=1}^{N} \mathrm{p}(i,j) q_n(i)$$

$$q_n(j) = \sum_{j=1}^{N} \mathrm{p}^n(i,j) q_0(i)$$

Markov chains with random initial state

As we saw last time, thinking of q, q_n as **row vectors**, these two equations can be written via matrix multiplication **from the right**, as

$$q_{n+1} = q_n \mathrm{p}, \quad q_n = q_0 \mathrm{p}^n.$$

Writing q, q_n instead as **column vectors**, and using matrix multiplication by the transpose matrix, from the left, the equations become

$$q_{n+1} = \mathrm{p}^t q_n, \quad q_n = (\mathrm{p}^t)^n q_0.$$

Markov chains with random initial state

If there is a distribution π (remember, a vector) such that

$$\pi \mathrm{p} = \pi$$

then π is said to be **stationary** or **invariant** respect to p.
Such a π represents a kind of statistical equilibrium for the chain:
even if the random state might still be changing from step to step, if the initial state is distributed as π, then all subsequent states will be identically distributed!

Question: when does a chain have a stationary distribution? how many does it have?

Irreducible chains and stationary distributions

Definition

A distribution π is said to be **stationary** with respect to a Markov chain with transition probability matrix p, if

$$\mathrm{p}^t \pi = \pi$$

In other words, a stationary distribution is an eigenvector of π, and such that all the coefficients of the vector are nonnegative, and add up to 1 (since it must be a distribution).

Markov chains with random initial state

A complete answer to this question requires two notions known as **closedness** and **irreducibility**.

Types of states and irreducibility

Consider a Markov chain with transition matrix $\mathrm{p}(i, j)$.

A **path** of length m going from the state x to the state y is a sequence of states such that

$$x = x_0, x_1, \ldots, x_{m-1}, x_m = y$$
$$\mathrm{p}(x_k, x_{k+1}) > 0 \text{ for each } k = 0, 1, \ldots, m-1$$

If there exists a path going from x to y, we say that x **communicates with** y, and this is written $x \to y$.

Types of states and irreducibility

One way to think about this: $x \to y$ if there is a positive probability that the system, starting from x, reaches the state y at some later time.

Accordingly, a **path** is nothing but a trajectory for the Markov chain that has positive probability of taking place.

Types of states and irreducibility

Examples

1. Take Gambler's ruin, which states communicate with which states?
2. How about the Ehrenfest chain?
3. For the random walk on a graph, what does it mean for x to communicate with y, in terms of the graph geometry?

Types of states and irreducibility

Closedness and irreducibility

Given A, a set of states for a chain, we say it is **closed** if

$$\mathrm{p}(i,j) = 0 \text{ whenever } i \in A,\ j \notin A$$

That is, if the system's state lies in A, it is impossible for the system to reach a state outside A at a later time.

Likewise, a set A is called **irreducible** if for any pair of states $x, y \in A$, we have that $x \to y$.
A chain with a set of states S is said to be irreducible if S itself is irreducible.

As it turns out, if the Markov chain is irreducible, there is always one, and only one, stationary distribution.

Theorem (Theorem 1.14 in Durrett)

For an irreducible Markov chain, there is one, and only stationary distribution π. Moreover $\pi(x) > 0$ for each state x.

The proof of this theorem will put our linear algebra skills to good use, since it boils down to finding an eigenvector and showing it is unique.

Irreducible chains and stationary distributions

Proof of the theorem

The proof has three broad steps

1. Show that transition matrix p must have at least one eigenvector π with eigenvalue equal to 1
2. Show that any eigenvector π of p must have coordinates which are either all positive, or all negative.
3. Show thatt the space of eigenvectors for 1 is a set of one dimension, and conclude that the unique π we want lies along this one dimensional space.

Irreducibility: what did it mean again?

Irreducibility: what did it mean again?

Let us keep our convention of using S to denote the state space.
A chain is irreducible if $x \mapsto y$ for any pair of states $x, y \in S$.

Irreducibility: what did it mean again?

Let us keep our convention of using S to denote the state space. A chain is irreducible if $x \mapsto y$ for any pair of states $x, y \in S$.

In terms of the transition matrix, irreducibility is equivalent to

$$\forall\, x, y \in S \text{ there is some } k \in \mathbb{N} \text{ such that } \mathrm{p}^k(x, y) > 0$$

Note: it could be that different pairs of x, y require different k (remember the Ehrenfest chain!)

An important fact is the following.

Lemma

If S is irreducible and $\mathrm{p}(x,x) > 0$ for every state x then

$$\mathrm{p}^{N-1}(x,y) > 0 \text{ for all states } x, y \in S.$$

Here, N is the number of states in the chain.

Note: The Ehrenfest chain does not have this property!.

Irreducible chains and stationary distributions

Concrete example

Before the proof, a concrete example

$$\begin{pmatrix} 0.5 & 0.0 & 0.2 & 0.3 \\ 0.0 & 0.3 & 0.6 & 0.1 \\ 0.2 & 0.6 & 0.2 & 0.0 \\ 0.3 & 0.1 & 0.0 & 0.6 \end{pmatrix}$$

Check: $\mathrm{p}^2(i,j) > 0 \ \ \forall\, i,j$. This means the chain is irreducible.

Before the proof, a concrete example

$$\begin{pmatrix} 0.5 & 0.0 & 0.2 & 0.3 \\ 0.0 & 0.3 & 0.6 & 0.1 \\ 0.2 & 0.6 & 0.2 & 0.0 \\ 0.3 & 0.1 & 0.0 & 0.6 \end{pmatrix}$$

Note that the matrix is symmetric.
From here, it is easy to check that $(1, 1, 1, 1)$ is an eigenvector.

Irreducible chains and stationary distributions

Proof of the theorem

The proof has three steps

1. Show that transition matrix p^t must have at least one eigenvector q with eigenvalue equal to 1
2. Show that any eigenvector q of p^t must have coordinates which are either all positive, or all negative.
3. Show that the space of eigenvectors for 1 is a set of one dimension, and conclude that the unique π we want lies along this one dimensional space.

Proof of Step 1.

The entries in each column of the matrix p^t add up to 1.

This means the entries in each column of $\mathrm{p}^t - I$ add up to 0.

Accordingly, the image of $\mathrm{p}^t - \mathrm{I}$ is contained in the ortoghonal complement to $q = (1, 1, \dots, 1)$

As such, $\mathrm{p}^t - \mathrm{I}$ is a $N \times N$ matrix with rank $\leq N - 1$. □

Proof of Step 2

Let q be an eigenvector of p^t with eigenvalue 1

Then q is also an eigenvector with eigenvalue 1 for

$$A = \left(\frac{1}{2}\mathrm{I} + \frac{1}{2}(\mathrm{p}^t)\right)^{N-1}$$

Check: (*this is where irreducibility is used!)

A is a stochastic matrix with **strictly positive entries***:

$$\sum_{j=1}^{N} A_{ij} = 1 \text{ for } i = 1, \dots, N$$

$$A_{ij} > 0 \text{ for } i, j = 1, \dots, N.$$

Proof of Step 2 (continued)

If the entries of q are not all nonnegative nor all nonpositive, then, on account that $A_{ij} > 0$, we have

$$|q_j| = \left|\sum_{i=1}^{N} A_{ij} q_i\right| < \sum_{i=1}^{N} |A_{ij} q_i| = \sum_{i=1}^{N} A_{ij} |q_i|$$

for $j = 1, \ldots, N$.

Proof of Step 2 (continued)

Adding these inequalities for $j = 1, \ldots, N$ we obtain

$$\sum_{j=1}^{N} |q_j| < \sum_{j=1}^{N} \sum_{i=1}^{N} A_{ij} |q_i| = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} A_{ij} \right) |q_i| = \sum_{i=1}^{N} |q_i|.$$

NOTE: We used that the entries of any row of A_{ij} add up to 1.

Proof of Step 2 (continued)

Then, we obtain that

$$\sum_{j=1}^{N} |q_j| < \sum_{i=1}^{N} |q_i|$$

clearly a contradiction.

We conclude all the entries of q are nonnegative or nonpositive.

Proof of Step 2 (finished).

In light of this, using that $A_{ij} > 0$ and the identity

$$q_j = \sum_{i=1}^{N} A_{ij} q_i$$

we conclude that each q_j is non-zero, so the entries of q are either all strictly positive, or strictly negative.

This finishes step 2.

□

Proof of Step 3.

What if the eigenspace of p^t for 1 is of dimension > 1?

There are q and q' in this space which are **orthogonal**.

By Step 2, the entries of q and q' are all strictly positive or strictly negative, in particular

$$q \cdot q' = \sum_{i=1}^{N} q_i q'_i \neq 0$$

A contradiction!, therefore the eigenspace is 1-dimensional. □