Higher Order Markov Chains and Spacey Walks

Nicolas Nytko

December 6, 2020

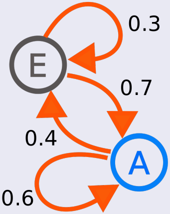
Overview

1. Introduction/review of first-order Markov chains
2. Higher order Markov chains, relation to tensors
3. Stochastic walks on higher order Markov chains
4. Applications of the Spacey walk

Markov Chains

Markov Chains are stochastic models describing a sequence of states with a probability of transitioning between each.

- Defined as matrix: $\mathrm{P} \in \mathbb{R}^{n \times n}$, P_{ij} gives probability of moving to state i from j, for chain with n states.
- Given distribution/state vector $\mathrm{x}^{(n)}$, one step of a random walk is represented by matrix-vector product $\mathrm{x}^{(n+1)} = \mathrm{P}\mathrm{x}^{(n)} \quad \Leftrightarrow \quad x_i^{(n+1)} = \sum_j P_{ij} x_j^{(n)}$.



Transition matrix:

$$\mathrm{P} = \begin{array}{c} \\ A \\ E \end{array} \begin{array}{c} \begin{array}{cc} A & E \end{array} \\ \begin{bmatrix} 0.6 & 0.7 \\ 0.4 & 0.3 \end{bmatrix} \end{array}$$

Figure from `https://commons.wikimedia.org/wiki/File:Markovkate_01.svg`

Definitions

Time Homogeneous

Each transition probability is independent of the time step t.

Irreducible

Every state can be reached from every other state.

Periodic

Periodic markov chain:

$$\exists i \in S \quad s.t. \quad P_{ii} = 0$$

Periodicity implies a state may only return to itself after some multiple d "hops". If every state has self loops the chain is **aperiodic**.

Canonical Examples

Google PageRank

- Markov chain represents the "idealized" web surfer. States are webpages and transition probability is given by number of outgoing links between pages.
- Ranking is determined by probability of being on certain page.

Board Games

- Each spot on game board is a state.
- Probability of moving between spots is given by dice, spinner, etc.

Weather

- States are possible weather conditions.
- Tomorrow's weather is randomly chosen based on today's conditions.

Steady-State Distribution

Steady state vector $\boldsymbol{\pi} \in \mathbb{R}^n$ does not change when multiplied by Markov chain:

$$\boldsymbol{\pi} = \mathrm{P}\boldsymbol{\pi}$$

- Average distribution as random walk tends to length infinity, $\lim_{k\to\infty} \mathrm{P}^k \mathrm{x}$
- Is equivalent to **Perron vector** of matrix, dominant eigenvector with nonnegative components.
- Compute iteratively with power method: $\mathrm{x}^{(k+1)} = \mathrm{P}\mathrm{x}^{(k)}$.

Unique steady states are guaranteed for **irreducible**, **aperiodic** Markov chains.

Markov Chains with Memory

What if we want our chain to **remember** where it has been? Introduce ***m*-order Markov chain**, where transition probability depends on previous *m* states.

- Probability is now stored as $\mathrm{P} \in \mathbb{R}^{n \times n^m}$, where columns are a *m*-tuple of each state permutation.
- Random step becomes slightly more complicated than just a *mat-vec*, will see a better way of doing this.

$$\mathrm{P} = \begin{array}{c} \\ s_0 \\ s_1 \end{array} \begin{array}{c} \begin{array}{cccc} (s_0, s_0) & (s_0, s_1) & (s_1, s_0) & (s_1, s_1) \end{array} \\ \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \end{bmatrix} \end{array}$$

(Example indexing for 2nd order chain with $n = 2$ states.)

Higher Order Markov Chains

We can **fold the last dimension** of the higher-order Markov matrix to obtain a higher dimensional **tensor**.

Order m Markov chain with n states is order $m+1$ tensor

$$\mathcal{P} \in \mathbb{R}^{\overbrace{n \times n \times \cdots \times n}^{m+1 \text{ dimensions}}}$$

Transition probability indexed by $P_{i_{(n+1)}, i_{(n)}, i_{(n-1)}, \ldots, i_{(n-m+1)}}$

- State/distribution is stored as order m tensor.
- Example random step for order 2 Markov chain: $X_{ij}^{(n+1)} = \sum_k P_{ijk} X_{jk}^{(n+1)}$
- Tensor form gives nice analysis properties — will see shortly.

Higher Order Steady-State Distribution

Steady state tensor $\mathrm{X} \in \mathbb{R}^{n \times n}$ does not change when multiplied along first mode.

$$X_{ij} = \sum_k P_{ijk} X_{jk}$$

- Conceptually simple, but requires $\mathcal{O}\left(n^m\right)$ space to store state tensor.

Replace state tensor with **low rank approximation** $\mathrm{x} \in \mathbb{R}^n$:

$$x_i = \sum_{jk} P_{ijk} x_j x_k$$

Actually just the ***z*-eigenvector** of $\mathcal{P}$!

Stochastic Processes

- For first order, steady state/eigenvector can be represented by a **random walk**.
- Our low rank transformation:

$$X_{ij} = \sum_{k} P_{ijk} X_{jk} \quad \Rightarrow \quad x_i = \sum_{jk} P_{ijk} x_j x_k$$

is algebraic, is there a **stochastic** process?

Present the concept of a **spacey walk**: upon reaching a new state we *space out* and forget all previous states. State history is then **randomly made up** from previously seen states.

Let $X(0), X(1), \ldots, X(n)$ be the sequence of visited states and N to be the number of states. Define $Y(n)$ as a random previously seen state, each weighted by number of occurrences:

$$P(Y(n) = k \mid \mathcal{F}_n) = \frac{1}{n+N}\left(1 + \sum_{s=1}^{n} \mathbb{1}(X(s) = k)\right)$$

Transition probability is then given by:

$$P(X(n+1) = i \mid X(n) = j, Y(n) = k) = P_{ijk}$$

Vertex-reinforced random walks

This spacey walk is an instance of a **vertex-reinforced random walk**. I.e., a walk that always picks next state based on transition probabilities, but probabilities evolve over time.

Vertex-reinforced walk

$$w_i(n) = \frac{1}{N+n}\left(1 + \sum_{s=1}^{n} \mathbb{1}\left(X(s) = i\right)\right)$$

$$P\left(X(n+1) = i \mid \mathcal{F}_n\right) = \left[\mathrm{M}\left(\mathrm{w}(n)\right)\right]_{i,X(n)}$$

Where M maps the *occupation vector* w to a stochastic matrix $\mathrm{P} \in \mathbb{R}^{N \times N}$.

For spacey walks, define:

$$M(\mathrm{w}) = \sum_{k=1}^{N} P_{k,:,:} w_k$$

Vertex-reinforced walks and ODEs

Why care about turning the **spacey walk** into a **vertex-reinforced walk**?

Gives us the key relationship:

$$\frac{dx}{dt} = \pi\left(M\left(x\right)\right) - x$$

For some π that maps transition matrices to a stationary distribution.

So, to find the tensor eigenvector x, we can solve the ODE for its **fixed points**.

Solution to Steady State

- Now reduced eigenvector problem to matter of **integrating an ODE**.
- Can offer better convergence properties than traditional power method/fixed point iteration.
- Take following periodic Markov chain as example:

$$\mathcal{F}_{(1)} = \left[\begin{array}{cc|cc} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{array}\right]$$

When using power iteration:

$$x_i^{n+1} = \sum_{jk} P_{ijk} x_j x_K$$

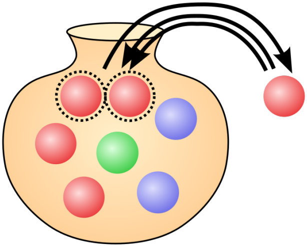
Oscillates between $\begin{bmatrix} 0 & 1 \end{bmatrix}^T$ and $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$.

Forward-stepping ODE

$$\frac{dx}{dt} = \pi \left(M \left(x \right) \right) - x$$

with Euler converges to analytical solution $\left[\left(\sqrt{2} - 1 \right) / 2 \quad \left(3 - \sqrt{5} \right) / 2 \right]^T$.

Pólya Urn



For simple example of the Spacey walk, consider **Pólya Urn** process. Have urn with red and green balls inside, and repeat the following process:

1. Select a random ball from the urn
2. Put the random ball back in the urn
3. Put another ball of same selected color in the urn

$$\mathcal{P}_{(1)} = \left[\begin{array}{cc|cc} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

Population Genetics

References

Austin R. Benson, David F. Gleich, and Lek-Heng Lim.
The spacey random walk: A stochastic process for higher-order data.
SIAM Review, 59(2):321–345, January 2017.

Hassan Bozorgmanesh and Masoud Hajarian.
Convergence of a transition probability tensor of a higher-order markov chain to the stationary probability vector.
Numerical Linear Algebra with Applications, 23(6):972–988, August 2016.