

Higher Order Markov Chains and Spacey Walks

Nicolas Nytko

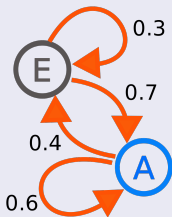
December 9, 2020

- ➊ Introduction/review of first-order Markov chains
- ➋ Higher order Markov chains, relation to tensors
- ➌ Stochastic walks on higher order Markov chains
- ➍ 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: $P \in \mathbb{R}^{n \times n}$, P_{ij} gives probability of moving to state i from j , for chain with n states.
- Given distribution/state vector $x^{(n)}$, one step of a random walk is represented by matrix-vector product
$$x^{(n+1)} = Px^{(n)} \Leftrightarrow x_i^{(n+1)} = \sum_j P_{ij}x_j^{(n)}.$$



Transition matrix:

$$P = \begin{matrix} & \begin{matrix} A & E \end{matrix} \\ \begin{matrix} A \\ E \end{matrix} & \begin{bmatrix} 0.6 & 0.7 \\ 0.4 & 0.3 \end{bmatrix} \end{matrix}$$

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:

$$\exists i \in S \quad s.t. \quad \left(P^d\right)_{ii} \neq 0; \quad d \in \mathbb{Z}, d > 1$$

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

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

$$\pi = P\pi$$

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

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 $P \in \mathbb{R}^{n \times n^m}$, 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.

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

(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 $P_{i_{(n+1)}, i_{(n)}, i_{(n-1)}, \dots, i_{(n-m+1)}}$

- State/distribution is stored as order m tensor.
- Example random step for order 2 Markov chain:

$$X_{ij}^{(n+1)} = \sum_k P_{ijk} X_{jk}^{(n+1)}$$

- Will write tensors as flattening along first mode, i.e. for $\mathcal{P} \in \mathbb{R}^{2 \times 2 \times 2}$:

$$\mathcal{P} = \left[\begin{array}{cc|cc} p_{111} & p_{121} & p_{112} & p_{122} \\ p_{211} & p_{221} & p_{212} & p_{222} \end{array} \right]$$

- Each slice is referred to as a **panel**, enumerates transition probabilities given some previous state.
- Note each panel is its own **stochastic matrix**.

Higher Order Steady-State Distribution

Assume 2nd order chains in following slides for notational simplicity – everything generalizes to higher orders though.

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

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

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

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

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

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

- 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.

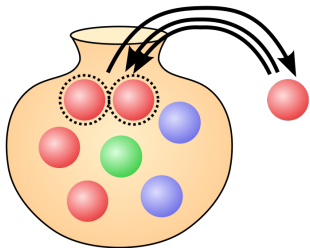
Let $X(0), X(1), \dots, X(n)$ be the sequence of visited states and N to be the number of states. Define $Y(n)$ as a random previously seen state, each weighted by number of occurrences:

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

Transition probability is then given by:

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

Pólya Urn



[https://commons.wikimedia.org/wiki/
File:Urn_problem_qtl5.svg](https://commons.wikimedia.org/wiki/File:Urn_problem_qtl5.svg)

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

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

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

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

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

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

For spacey walks, define:

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

Vertex-reinforced walks and ODEs

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

Lets us model chain as a **dynamical system**:

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

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**:

$$\frac{dx}{dt} = 0 \quad \Longleftrightarrow \quad x_i = \sum_{jk} P_{ijk} x_j x_k$$

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

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

with Euler converges to analytical solution

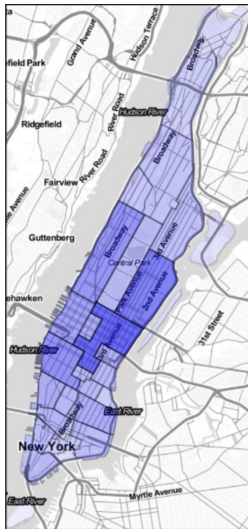
$$\left[(\sqrt{2} - 1) / 2 \quad (3 - \sqrt{5}) / 2 \right]^T.$$

Population genetics study dynamics of allele distribution, or type of each organism. Suppose we have n types, the possibility of passing each type to child is given by:

$$P(\text{child is type } i \mid \text{parents are types } j, k) = P_{ijk}$$

- Parent ordering doesn't matter, have natural symmetry $P_{ijk} = P_{ikj}$.
- Spacey walk models how population dynamics evolve over time.
- Stationary type distribution satisfies $x_i = \sum_{jk} P_{ijk} x_j x_k$.

Taxi Trajectory Modeling



(Benson, Gleich, Lim)

- Model taxicab as **spacey walk** P_{ijk} , with N states being possible locations.
- Passenger at location k gets picked up at j and goes to i with probability P_{ijk} .
- Assume passengers are likely to be picked up nearby where they are dropped off, i.e. at airport.
- Difficult to solve, higher order gives better performance (RMSE) than first order:

First Order	Second Order	Spacey Walk
0.846	0.835	0.835

Conclusion

- Introduce high order Markov chain as having **memory of previous states**.
- High order Markov chains can be represented as **tensor** of transition probabilities.
- Tensor form naturally emits eigenvector steady-state.
- Z-eigenvector is the limiting distribution of the **spacey random walk**.
- Random walk can be turned into equivalent ODE construction.
- We can solve an ODE to find the tensor eigenvector!
- The Spacey walk is a generalization of the **Pólya process**, has interesting applications.



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.