Sampling over tilings of the plane: A computational approach against political gerrymandering

Janabel Xia

Mentor Younhun Kim

May 18–19, 2019

MIT PRIMES Conference

Motivation

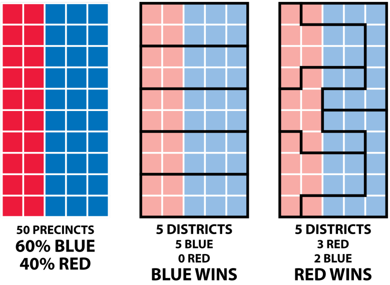
What is political gerrymandering?

https://www.fairvote.org/new_poll_everybody_hates_gerrymandering

Motivation

What is political gerrymandering?

Gerrymandering is the practice of drawing boundaries of electoral districts in a way that gives one party an unfair advantage over the others.



https://www.fairvote.org/new_poll_everybody_hates_gerrymandering

Motivation

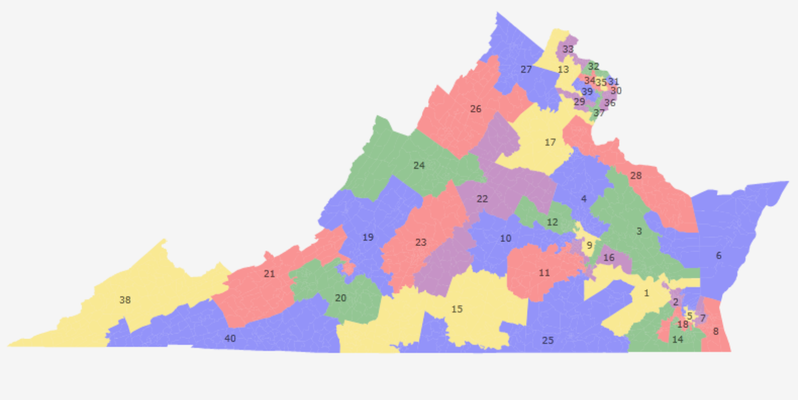


Figure 1: Map of Virginia's electoral districts, 2011

https://rrhelections.com/index.php/2013/01/29/the-new-virginia-senate-map/

Uniform Sampling

We can think of electoral redistricting as tiling a grid of cells.

How can we determine if a tiling is fair?

- Sampling repeatedly for statistics — Law of Large Numbers

How do we sample tilings of the plane uniformly at random?

- Distribution is very complicated
- Potential solution: use Markov chains

Markov chains

Definition

A sequence of random variables $(X_0, X_1, \ldots)$ is a *Markov chain* with state space Ω and transition matrix P if for all states $x, y \in \Omega$, we have

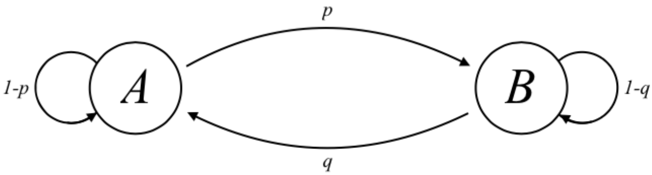
$$Pr\{X_{t+1} = y \mid X_t = x\} = P(x, y).$$

Markov chains

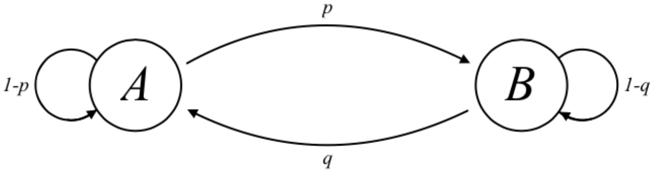
Definition

A sequence of random variables $(X_0, X_1, \dots)$ is a *Markov chain* with state space Ω and transition matrix P if for all states $x, y \in \Omega$, we have

$$Pr\{X_{t+1} = y \mid X_t = x\} = P(x, y).$$



Example (continued)



$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

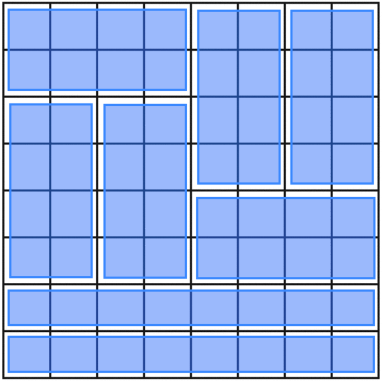
We can repeatedly multiply by P to get

$$\mu_t = \mu_0 P^t = \begin{bmatrix} \pi_A & \pi_B \end{bmatrix} \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}^t$$

Our Proposed Markov chain

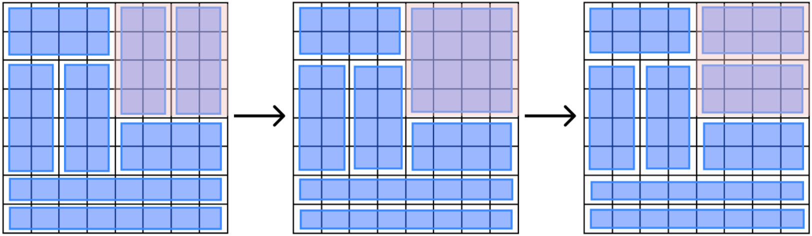
State space: tiling $n \times n$ with tiles of area n

- Simple case: rectangular tilings



Our Proposed Markov chain

Transitions: merge and re-split



What is this Markov chain doing?

Stationary Distributions

Definition

A distribution π is called a *stationary distribution* of the Markov chain if it satisfies the following property:

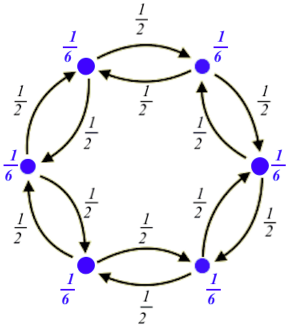
$$\pi = \pi P$$

Stationary Distributions

Definition

A distribution π is called a *stationary distribution* of the Markov chain if it satisfies the following property:

$$\pi = \pi P$$



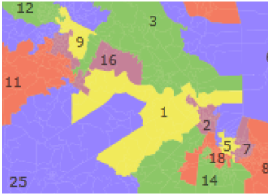
Stationary Distributions

Definition

A distribution π is called a *stationary distribution* of the Markov chain if it satisfies the following property:

$$\pi = \pi P.$$

- Fair tilings should be natural — "compact"
- Ideally, stationary distribution of our chain favors "compact" tiling



https://rrhelections.com/index.php/2013/01/29/the-new-virginia-senate-map/

Irreducibility

Definition

A chain P is called *irreducible* if for any two states $x, y \in \Omega$, there exists an integer t (possibly depending on x and y) such that

$$P^t(x, y) > 0.$$

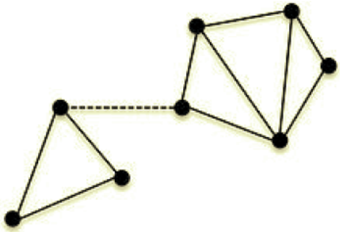
https://en.wikipedia.org/wiki/Connectivity_(graph_theory)

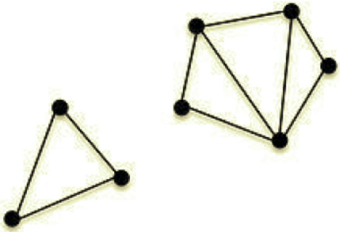
Irreducibility

Definition

A chain P is called *irreducible* if for any two states $x, y \in \Omega$, there exists an integer t (possibly depending on x and y) such that

$$P^t(x, y) > 0.$$



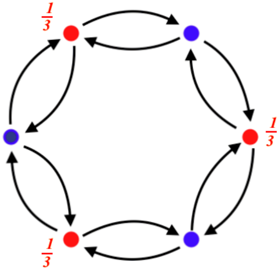


https://en.wikipedia.org/wiki/Connectivity_(graph_theory)

Aperiodicity

Definition

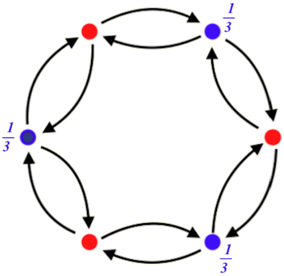
Let $T(x) := \{t \geq 1 \colon P^t(x,x) > 0\}$. The chain P is *aperiodic* if for all states $x \in \Omega$, the $gcd(T(x))$ is 1.



Aperiodicity

Definition

Let $T(x) := \{t \geq 1 \colon P^t(x,x) > 0\}$. The chain P is *aperiodic* if for all states $x \in \Omega$, the $gcd(T(x))$ is 1.



Convergence Theorem

Theorem (Convergence Theorem)

Suppose that P is irreducible and aperiodic, with stationary distribution π. Then $P^t \pi_0 \rightarrow_{t\rightarrow\infty} \pi$ for any π_0.

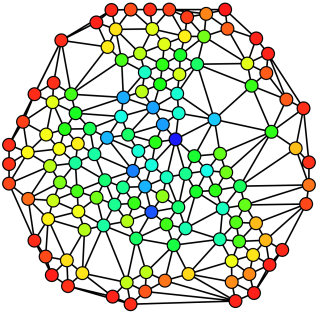
Why is this important?

- Want to use Markov Chain for sampling
- Running Markov chain algorithmically — want to draw from stationary distribution
  - If we can show irreducible and aperiodic, then this is possible
- Allows us to compute statistics like averages

Mixing Time of a Markov chain

How long should we run this Markov chain?

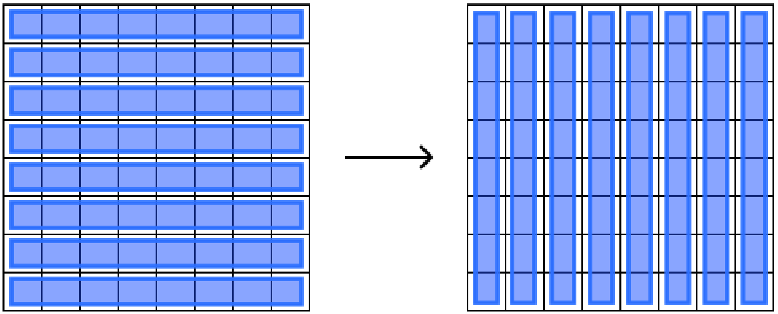
- e.g. the larger the diameter, the longer this takes
- Vertices = states, edges = nonzero probabilities



https://en.wikipedia.org/wiki/Betweenness_centrality

Diameter of our Markov chain

Conjecture: diameter $= O(n \log n)$



- Diameter over rectangular tilings explains mixing time
- Count tilings to understand how large graph on state space is

Future Work

- Prove irreducibility
- Calculate diameter/bottleneck/mixing time
- Understand stationary distribution of current Markov chain — need to run experiments
- Does our Markov chain favor "compact" tilings?
  - If not, can weigh transitions in Markov chain

Acknowledgements

Thank you to:

- My mentor, Younhun Kim
- MIT PRIMES
- MIT Mathematics Department
- Dr. Tanya Khovanova, Dr. Slava Gerovitch, and Dr. Claude Eicher
- My parents