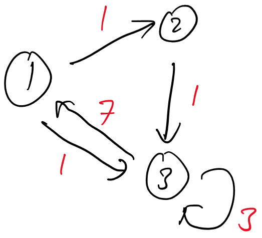
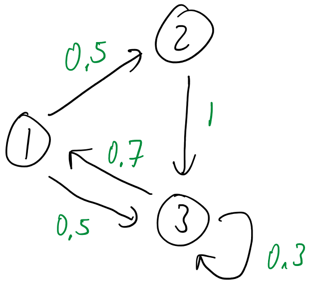
11-Markov-chains

Friday, April 3, 2026 2:32 PM

Markov chains + random walks on graphs

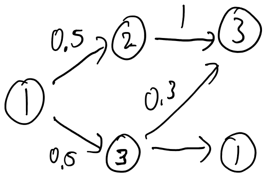


- Consider a directed graph $G=(V,E)$, $E \subseteq V^2$
- A graph G is strongly connected if $\forall x, y \in V$, $\exists$ a path from x to y.
- A weighted graph additionally assigns a positive value to each edge.
- A random walk on a graph is a path generated by starting from a vertex, and then iteratively choosing the next vertex by travelling along edges.
  - ↳ The standard random walk fixes the transition probabilities as proportional to edge weight.



Ex time: 1 2 3 ......

If we start at node 1.



......

prob dist. at time t

Let the matrix P have $p_{ij} = \text{Prob}(\text{transition from } i \text{ to } j)$. Then $\vec{p}(t+1) = \vec{p}(t) P$

Ex.

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \\ 0.7 & 0 & 0.3 \end{bmatrix}$$

$$[1 \quad 0 \quad 0] P = [0 \quad 0.5 \quad 0.5]$$

$$[1 \quad 0 \quad 0] P^2 = [0.35 \quad 0 \quad 0.65]$$

Random walk on graph $\iff$ Markov chain

vertices $\iff$ states

strongly connected graph $\iff$ connected Markov states

We will study Markov chains as both a method for sampling probability distributions + as a basis for modelling phenomena.

Let $\vec{p}(t)$ be the probability distribution after t steps of a random walk.

Let $\vec{p}(t)$ be the probability distribution after t steps of a random walk.

Def. The long-term average prob. dist. $\vec{a}(t)$ is

$$\vec{a}(t) = \frac{1}{t}\sum_{s=0}^{t-1}\vec{p}(0).$$

Goal: $\lim_{t\to\infty}\vec{a}(t) = \vec{x}$ s.t. $\vec{x}P = \vec{x}$ for a connected Markov matrix.

(technical)

Lemma: Let P be a transition matrix for a connected Markov chain. The $n\times(n+1)$ matrix $A = [P - I, \vec{1}_n^T]$ has rank n. $\left(\vec{1}_n^T = \begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}\right)$

proof. Suppose $\operatorname{rank}(A)\neq n$. Then $\operatorname{rank}(A) < n \Rightarrow \dim(\operatorname{Null}(A)) \geq 2$.

$P\vec{1}_n^T = \vec{1}_n^T$ because each row in P sums to 1 as a prob. dist.

(we use connectedness here)

Then $A\begin{bmatrix}\vec{1}_n^T\\ 0\end{bmatrix} = (P-I)\vec{1}_n^T = 0$

Assume $\exists\,[\vec{x}\ \ \alpha] \perp [\vec{1}_n\ \ 0]$ s.t. $A[\vec{x}, \alpha]^T = 0$ (second solution)

$$\Rightarrow [P - I, \vec{1}_n^T]\begin{bmatrix}\vec{x}^T\\ \alpha\end{bmatrix} = (P-I)\vec{x}^T + \alpha\vec{1}_n^T = 0$$

Thus, $\forall i$, $\sum_{j=1}^{n} p_{ij}x_j - x_i + \alpha = 0 \Rightarrow x_i = \sum_{j=1}^{n} p_{ij}x_j + \alpha$ (each x_i is a convex combination of x_j's + α)

Since $\vec{x}\perp\vec{1}_n$, if $\vec{x}\neq 0$, then some $x_i > 0$ and some $x_j < 0$.

Let $x_i \geq x_j$ for all j. Then $\sum_{j=1}^{n} p_{ij}x_j < \sum_{j=1}^{n} p_{ij}x_i = x_i\sum_{j=1}^{n} p_{ij} = x_i$

But $x_i = \sum_{j=1}^{n} p_{ij}x_j + \alpha$, so $\alpha > 0$.

However, repeat above logic letting $x_i \leq x_j$ for all j. Then $\alpha < 0$.

Contradiction, so $\operatorname{rank}(A) = n$. $\rightarrow|\leftarrow$

Fundamental Thm of Markov Chains For a connected Markov chain,

Fundamental Thm of Markov Chains For a connected Markov chain, there is a unique prob. vector $\vec{\pi}$ satisfying $\vec{\pi} P = \vec{\pi}$.

Moreover, for any starting dist. $\vec{p}(0)$, $\lim_{t\to\infty} \vec{a}(t) = \vec{\pi}$.

proof. $\vec{a}(t) = \frac{1}{t}\left(\vec{p}(0) + \cdots + \vec{p}(t-1)\right)$, so $\vec{a}(t)$ is also a prob. dist.

Let $\vec{b}(t) = \vec{a}(t)P - \vec{a}(t) = \frac{1}{t}\left(\vec{p}(t) - \vec{p}(0)\right)$.

Then $|\vec{a}(t)P - \vec{a}(t)| = \frac{1}{t}|\vec{p}(t) - \vec{p}(0)| \leq \frac{1}{t}|\vec{p}(t)| + \frac{1}{t}|\vec{p}(0)| = \frac{2}{t} \to 0$ as $t \to \infty$

Aside: we use the 1-norm: $|\cdot| = \|\cdot\|_1$

So the limit exists.

By Lemma, let $A = [P - I, \mathbb{1}_n^T]$. $\operatorname{rank}(A) = n$

$A = [C_1 \; C_2 \; \cdots \; C_{n+1}]$, $C_{n+1} = \mathbb{1}_n^T$.

Let $B = [C_2 \; C_3 \; \cdots \; C_{n+1}]$, an $n \times n$ submatrix.

$\operatorname{rank}(B) = n$ because $C_1 + C_2 + \cdots + C_n = 0$, so that set is linearly dep.

$\Rightarrow B$ is invertible.

Let $\vec{c}(t)$ be $\vec{b}(t)$ with 1st col removed.

Note $\vec{b}(t) = \vec{a}(t)[P - I]$

So $\vec{a}(t) A = [\vec{b}(t), \; \vec{a}(t)\vec{\mathbb{1}}_n^T] = [\vec{b}(t), 1]$

$\vec{a}(t) B = [\vec{c}(t), 1]$

$\vec{a}(t) = [\vec{c}(t), 1] B^{-1} \to [0, 1] B^{-1}$ as $t \to \infty$.

Thus, $\vec{\pi} = [0, 1] B^{-1}$ satisfies the thm.

$\blacksquare$

Exercise: Why didn't we just prove $\vec{\pi} = \lim_{p\to\infty} \vec{p}(t)$?

Exercise: Consider a random walk on an undirected, unweighted strongly connected graph. What is the limiting distribution?

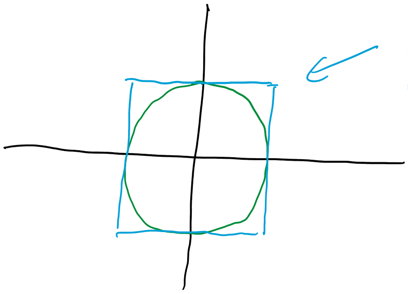
Monte Carlo Methods

In earlier lectures, we saw randomness used in sketching algorithms. Those methods can be thought of as examples more generally of repeated sampling for computational experiments, collectively known as Monte Carlo methods in numerics/optimization.

Steps:
(1) Define domain of possible inputs
(2) Generate inputs randomly from appropriate probability dist.
(3) Perform deterministic computation on outputs

(2) Generate inputs randomly from appropriate probability dist.
(3) Perform deterministic computation on outputs
(4) Aggregate results

Classic example: Compute π by inscribing circle in square + generating random points.



easy because square is easy to sample from

Notice: What else is easy to sample from? Markov chains!

Markov Chain Monte Carlo (MCMC)

Given a prob. dist $p(\vec{x})$, want to estimate $\mathbb{E}f = \sum_{\vec{x}} p(\vec{x}) f(\vec{x})$

If each x_i has at least 2 possibilities, then exponentially many possible $\vec{x}$.

Rather want to sample points $\vec{x}$ according to p, so we don't need to evaluate everywhere.

e.g. equivalent to finding mean by drawing random samples.

MCMC draws a sample $\vec{x}$ from $p(\vec{x})$ by designing a Markov chain whose stationary dist. is $p(\vec{x})$

↳ common variants include Metropolis-Hasting + Gibbs sampling.

$$\mathbb{E}_{\vec{x} \sim p} f(\vec{x}) = \sum_{\vec{x}} p(\vec{x}) f(\vec{x}).$$

Notation: $\mathbb{E}f = \sum_i p_i f_i$, where i is our state.

Consider a random walk on our Markov chain.

Let γ be the average of f along nodes in a t-step walk $(w_0, \dots, w_t)$

Then γ is an estimator for $\mathbb{E}f$ as $t \to \infty$.

$$\underset{\nearrow}{\mathbb{E}}\gamma = \sum_i f_i \cdot \left(\frac{1}{t} \sum_{j=1}^{t} \text{Prob}(w_j = i) \right) = \sum_i f_i \cdot a_i(t).$$

over t-step walks

Let $f_{max} = \max_i |f_i|$.

Then $$\left| \sum_i f_i p_i - \mathbb{E}\gamma \right| \leq f_{max} \sum_i |p_i - a_i(t)| = f_{max} |\vec{p} - \vec{a}(t)|$$

So we can bound the performance of MCMC estimate by the f_{max}

So we can bound the performance of MCMC estimate by the f_{max} and the 1-norm of distance between $\vec{p}$ + $\vec{a}(t)$ prob. dist.

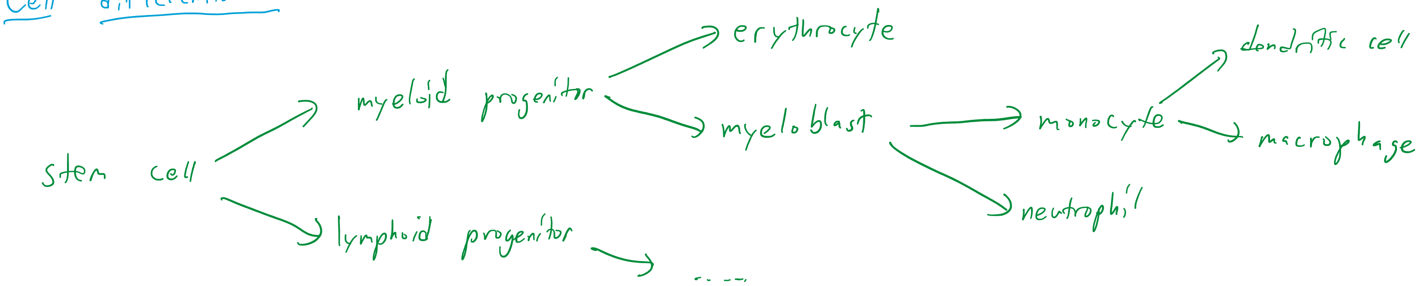
The rate of convergence depends on how quickly $\vec{a}(t) \to \vec{p}$, so we want to define Markov chains that rapidly mix.

Exercise: Can you derive a relationship between the Fiedler eigenvalue of the Laplacian of a graph and how quickly the Markov chain defined by the standard random walk on that graph mixes?

More generically, while MCMC is used in statistics to refer to sampling from a complex distribution, in bioinformatics it is sometimes used to refer to any repeated sampling on a Markov chain.

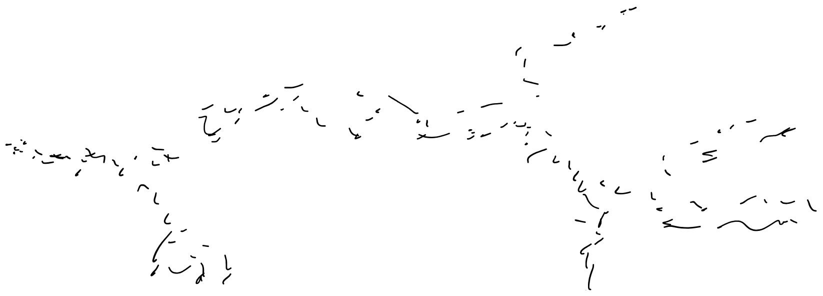
Exercise: I would like to determine the importance of edges in an undirected graph. Suppose I compute the stationary probabilities p_i for the standard random walk on the graph, for being at vertex i. Let d_i be the degree, so the frequency that edge (i,j) is traversed from $i \to j$ would be $\frac{1}{d_i} p_i$ and $j \to i$ would be $\frac{1}{d_j} p_j$. I define an importance of $\left| \frac{1}{d_i} p_i - \frac{1}{d_j} p_j \right|$ to the edge. What goes wrong with this definition?

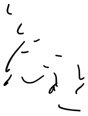
Cell differentiation



Cells differentiate. We can trace this progression with single-cell RNA-seq.

Last time: low-dimensional embedding of expression vector





But single-cell RNA-seq is destructive, so we cannot actually trace a single cell as it differentiates, and even if we could, it would only take one path.

Can we reconstruct trajectories from expression data?

Intuition: This problem is similar to phylogenetic tree reconstruction.

Problem: possible for expression patterns to cycle.

So instead of trying to construct a tree, maybe we can construct a graph instead.

Idea: Build undirected k-nearest neighbors graph, + treat every measured cell as an vertex / state in a Markov chain.

Pseudotime: We can then assign a "pseudotime" to each cell depending on how far away it is from a known root.

RNA velocity: Some methods also incorporate the ratio of spliced to unspliced mRNA to determine if a gene is being induced or repressed. That can add directionality to edges.

Idea: Maybe not every cell corresponds to a state.

Exercise: Can we aggregate many cells into a single macrostate?

Idea: What if there are hidden states that have multiple possible expressions

↳ idea leads to hidden Markov Models for next week.