$\boxed{\text{1 Dec 2025}}$ Markov Chain intro

Announcements:

- Problem Set 2 graded
- Prelim 2 getting close

Problem. Draw a sample from the following distribution:

Output	Probability	Partial Sums
A	0.2	0.2
B	0.3	0.5
C	0.1	0.6
D	0.4	1.0

For any sampling problem where the probabilities are explicitly given as a table, the above alg. solves random sampling in $O(n)$ time.

Random sampling becomes tough when you're drawing samples from an implicitly normalized distribution on an exponentially large set.

Def. If X is a set (at most countable) an implicitly normalized distrib. on X

is a function $w: X \to \mathbb{R}_{\geq 0}$

s.t. $Z = \sum_{x \in X} w(x) > 0$.

The probability distrib defined by w

is $p(x) = \frac{1}{Z} \cdot w(x)$.

Typically you have a subroutine for evaluating w fast, but you don't know Z and the set X is too large to compute $\sum w(x)$ directly.

Ex. 1 Random q-coloring of a graph.

Say $G = (V,E)$ is given, and every vertex has $< q$ neighbors.

Uniformly random q-coloring corresponds to

X = all functions $V \to [q] = [q]^V$

$$w(x) = \begin{cases} 1 & \text{if } x \text{ is a proper coloring} \\ \emptyset & \text{if not.} \end{cases}$$

Ex. 2. q-state Potts model $X = [q]^V$

$$w(x) = \prod_{\{u,v\} \in E} \begin{cases} \lambda & \text{if } x(u) = x(v) \\ 1 & \text{if } x(u) \neq x(v) \end{cases}$$

Generalizes previous example which corresponds to $\lambda = 0$.

Ex. 3 w = score function given by a neural network

Draw random outputs with probability proportional to score.

Ex. 4. Suppose given $p : X \to [0,1]$ s.t. $\sum_{x} p(x) = 1$ and a conditioning event $\mathcal{E} \subset X$ s.t.

$\sum_{x \in \mathcal{E}} p(x) > 0$. Drawing samples from the cond. distrib of p given $\mathcal{E}$ is equiv't to the

impl. normalized

$$w(x) = p(x) \cdot \mathbb{1}[x \in \mathcal{E}].$$

MARKOV CHAINS

An iterative method for sampling from a distribution.

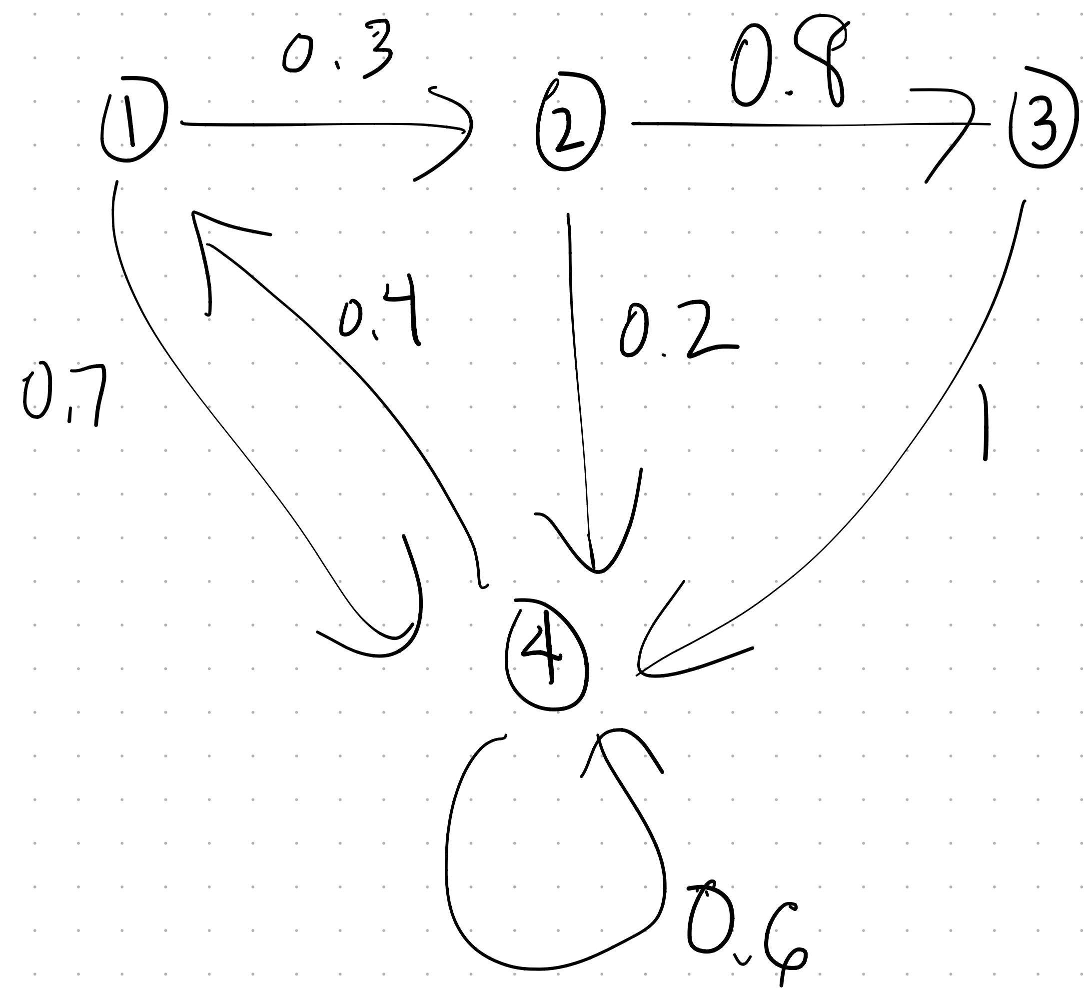
A Markov Chain with state space $\mathcal{X}$ and transition matrix P is a sequence of rand vars

$$x_0, x_1, \ldots$$

such that $\forall t > 0$, $\forall x, y \in \mathcal{X}$

$$\Pr\left(x_{t+1} = y \mid x_t = x\right) = P_{xy}$$

Visualize using transition graph



$$\begin{bmatrix} 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 1 \\ 0.4 & 0 & 0 & 0.6 \end{bmatrix}$$

A Markov transition matrix

<u>always</u> satisfies $\forall_{x,y}\ P_{xy} \geq 0$

and $P \vec{\underline{1}} = \vec{\underline{1}}$.

<u>Definition</u>: A probab. distrib.

$\pi : X \to [0,1]$ is stationary

for P if

$\forall y \in X \quad \pi^T \cdot P = \pi^T$

$$\underbrace{\sum_{x \in X} \pi(x) \cdot P_{xy}}_{\text{Draw } x \text{ from } \pi \text{ and take one Markov transition to land on } y} = \underbrace{\pi(y)}_{\text{Draw } y \text{ from } \pi}$$

Draw x from π and take one Markov transition to land on y == Draw y from π

<u>Lem.</u> Every Markov chain[1] has (at least one) stationary distribution.

[1] with finite state space

<u>Proof.</u> Let $\Delta^X = \{$ row vectors representing prob distrib on $X \}$

$$= \left\{ \xi \;\middle|\; \xi_x \geq 0 \;\forall x, \; \sum \xi_x = 1 \right\}$$

The function $\xi \mapsto \xi \cdot P$

maps Δ^X to itself.

If $\zeta = \xi \cdot P$ then

$$\zeta_y = \sum_x \xi_x P_{xy} \geq 0$$

$$\sum_y \zeta_y = \zeta \cdot \vec{1} = \xi P \vec{1}$$
$$= \xi \vec{1} = 1.$$

Brouwer FP Theorem $\Rightarrow$

$\exists\, \xi \in \Delta^X$ st. $\xi \cdot P = \xi$.

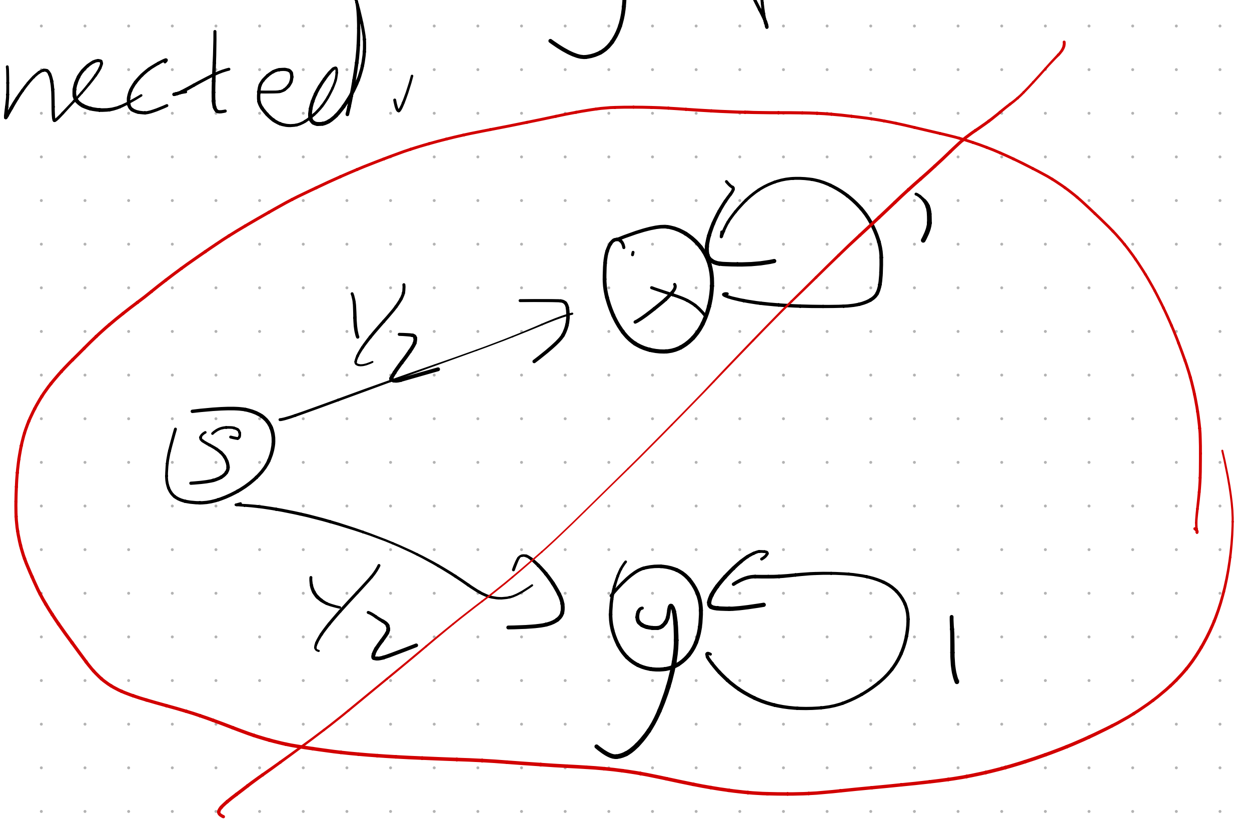
$\pi = \xi^T$ is a stationary distrib.

Which MCs have a unique stat distrib?

If $X_0, X_1, \ldots$ has trans matrix P is the marginal distrib of X_t guaranteed to converge to π as $t \to \infty$?

Answers. 1 For uniqueness, it suffices that the MC is irreducible ...

state trans. graph strongly connected.



2) For convergence, it suffices that the MC is irreducible and aperiodic

"ergodic"

meaning $\exists\, t < \infty$ s.t. every component of P^t is > 0.

For some $t > 0$ we can find a path of length t in state transition graph from any source to any dest.