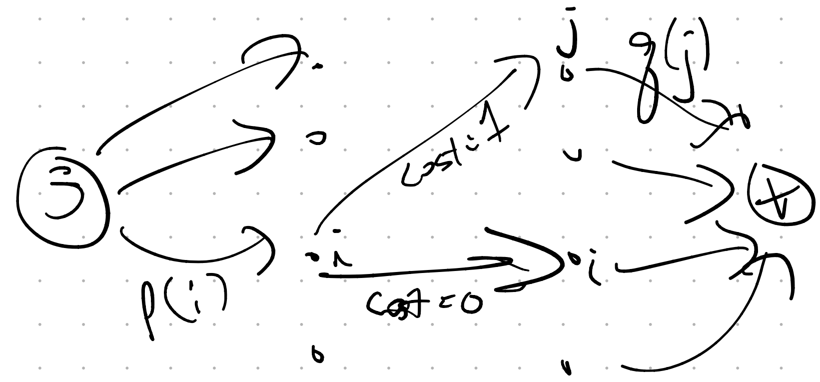
$\boxed{\text{4 Dec 2024}}$ Markov Chains & Metropolis – Hastings

Def. If p, q are two prob. distributions on the same finite ~~set~~ Ω their total variation (TV) distance is

$$\|p-q\|_{TV} = \max_{S \subset \Omega} \left| p(S) - q(S) \right| = \frac{1}{2} \|p-q\|_1$$

Lemma. (Coupling Lemma) $\|p-q\|_{TV}$ equals

$$\inf \left\{ \Pr(x \neq y) \;\middle|\; \begin{array}{l} \text{Joint distribs on pairs } (x,y) \\ \text{s.t. marginals of } x, y \text{ are } p, q \end{array} \right\}$$



Given unnormalized distribution

$w : \Omega \to \mathbb{R}_{\geq 0}$ $\quad p = \frac{w}{Z}$ normalization of w

to solve the w-sampling problem with approximation ε means:

to design an algorithm that draws samples from a distribution q s.t. $\| p - q \|_{TV} \leq \varepsilon$.

<u>Typical goal:</u> Run in time $\mathrm{poly}\left(\log|\Omega|, \log\left(\frac{1}{\varepsilon}\right)\right)$.

<u>Typical method:</u> Design a <u>Markov chain</u> whose <u>stationary distribution</u> is p.

Run the chain for T steps where $T \geq \tau_{mix}(\varepsilon)$ "<u>ε-mixing time</u>" and output the final state.

"Markov Chain Monte Carlo."

Def. A Markov chain with state set S and transition matrix P is a sequence of random variables

$$X_0, X_1, \ldots,$$

taking values in S s.t.

$$\forall t > 0 \quad \forall s_0, \ldots, s_{t-1} \in S$$

$$\Pr(X_t = s_t \mid X_0 = s_0, X_1 = s_1, \ldots, X_{t-1} = s_{t-1})$$

$$= P_{s_{t-1}, s_t}$$ ← "probability of transitioning from s_{t-1} to s_t"

Note $P_{ij} \geq 0$ and $\sum_j P_{ij} = 1$,

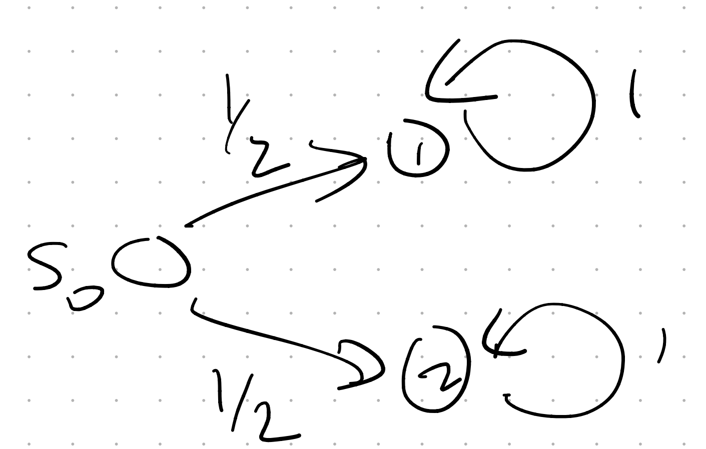
$$P \cdot \vec{1} = \vec{1}.$$

$\exists$ a vector π s.t. $\pi^T P = \pi^T$

Thm. (Perron–Frobenius) This π is unique up to scaling if P is irreducible (the graph of pairs (i,j) s.t. $P_{ij} > 0$ is strongly connected) and π can be chosen s.t.

$$\pi \geq 0, \quad \sum_{i \in S} \pi_i = 1.$$

This π is called the stationary distribution of P.



$$\begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Furthermore if P is aperiodic (i.e. $\exists\, k < \infty$ s.t. all entries of P^k are > 0) then for all initial state distributions π_0

the sequence of marginals $\pi_0^T P^t$ converges to π^T as $t \to \infty$.

Suppose $X_0, X_1, \ldots$ is a Markov chain with trans mtx P. Let π_t denote marginal distrib of X_t. Then

$$\underbrace{\pi_{t+1}^T}_{j^{th} \text{ coord equals}} = \pi_t^T P.$$

$$\Pr(X_{t+1} = j) = \sum_i \Pr(X_t = i) \cdot P_{ij}$$

By induction $\pi_t^T = \pi_0^T P^t$.

Designing P with a specified stationary distribution, $\frac{w}{Z}$ ...

<u>Def.</u> P is reversible w.r.t. π if

$$\forall i,j \quad \pi_i P_{ij} = \pi_j P_{ji}$$

<u>Lem.</u> If P is a Markov trans mtx reversible w.r.t. π then π is stationary for P.

<u>Proof.</u> Need to show

$$\sum_{i \in S} \pi_i P_{ij} = \pi_j$$

$$\| \qquad\qquad \|$$

$$\sum_{i \in S} \pi_j P_{ji} = \pi_j \left(\sum_{i \in S} P_{ji} \right)$$

<u>The Metropolis algorithm.</u>

Given Ω and $w : \Omega \to \mathbb{R}_{\geq 0}$

and a <u>symmetric</u> $\Omega \times \Omega$ matrix, K,

"proposal distribution"

with row sums $= 1$.

The Markov chain (S, P) has

$$S = \Omega$$

$$P_{ij} = K_{ij} \cdot \min\left\{1, \frac{w(j)}{w(i)}\right\}$$