This course is about Markov chain Monte Carlo (MCMC) algorithms.
These are often easy algorithms to design but their analysis is difficult/impossible.
This is a theory course — the focus is on settings where we can analyze (formally) the convergence rate & techniques to do so.
We are not studying ~~too~~ heuristics for MCMC implementations.
We will see some advanced tools such as simulated annealing & MC^3.

Motivating example:

For a graph $G=(V,E)$,

a matching is a subset $M \subset E$ where each vertex is incident ≤ 1 edge in M

(perfect if every vtx incident $=1$ edge in M)

let $\mathcal{M} = \mathcal{M}(G)$ = collection of all matchings in G

$$= \{ S \subset E : \forall v \in V, \underbrace{\deg_S(v)}_{\text{degree of } v \text{ in } S} \le 1 \}$$

Sampling problem: Can we sample uniformly at random from $\mathcal{M}$?

Counting: Can we compute $|\mathcal{M}|$?

Note, $|\mathcal{M}|$ is typically $\exp(\Omega(n))$.

Can we do the sampling/counting in time $\mathrm{poly}(n)$?

More generally, can have weights on matchings

$$w : \mathcal{M} \to \mathbb{R}_+$$

- Simple example: parameter $\lambda > 0$

$$w(M) = \lambda^{|M|}$$

(called hard-core distribution)

Gibbs distribution

Sampling: Sample from $\mu(M) = \frac{w(M)}{Z}$

where $Z = \sum_{M' \in \mathcal{M}} w(M')$

called partition function

Counting: Estimate Z

Note, if $w(M) = 1$ forall $M \in \mathcal{M}$,

then $\mu = \text{uniform}(\mathcal{M})$

& $Z = |\mathcal{M}| = \#$ of matchings in G.

Exact computation of $|M|$ is #P-complete

counting analog of NP.

So no poly-time alg. unless P = NP (& P = #P).
(but can do on planar graphs — we'll see)

Can we sample from μ?
& approx. $|M|$?

Typically:

$$\begin{tikzcd}
\text{Exact counting} \arrow[rr] & & \text{~~Ap~~ Exact sampling } \mu \arrow[dll] \\
\text{Approx. counting (FPRAS for } Z\text{)} \arrow[rr, leftrightarrow] & & \text{Approx. sampling } \mu
\end{tikzcd}$$

We'll see later: a) Exact computation of $|M|$ is #P-complete
b) Approx. counting $\Longleftrightarrow$ Approx. sampler.

& we'll use (c) MCMC techniques
to get approx. sampler for M
(for any G).
(open: exact sampler for M).

Permanent: $|P|$ = # of perfect matchings of bipartite G

we'll see (d) FPRAS for $|P|$
& approx. sampler

FPRAS for Z:

fully-polynomial time randomized approximation scheme.

Input: $G=(V,E)$, $\epsilon>0$, $\delta>0$,

output is a number OUT where:

$$\Pr\left(\text{OUT}(1-\epsilon) \le Z \le \text{OUT}(1+\epsilon)\right) \ge 1-\delta$$

in time $\text{poly}(n, \frac{1}{\epsilon}, \log(1/\delta))$

$n=|V|$.

Note, if $\Pr(\quad) \ge 3/4$

then can boost by taking median of $O(\log(1/\delta))$ trials.

FPTAS: achieves $\delta=0$ (Deterministic analog).

Approx. sampler:

Input: $G=(V,E)$, $\delta>0$

Output: matching M from distribution π where:

$$d_{TV}(\mu, \pi) \le \delta$$

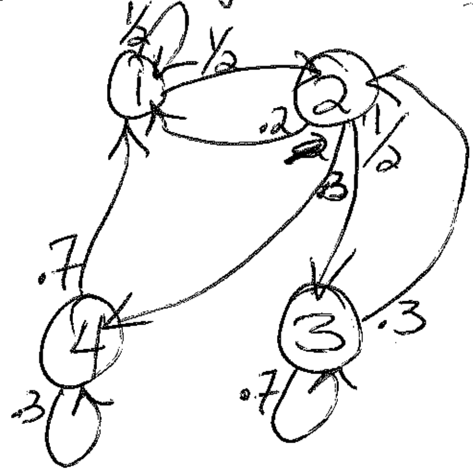
in time $\text{poly}(n, \log(1/\delta))$.

for μ, π on Ω: $d_{TV}(\mu,\pi) = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \pi(x)| = \sup_{S\subset\Omega} \mu(S)-\pi(S)$.

What's a Markov chain?

Example: States $\{1,2,3,4\}$

Graphical representation:



Algebraic representation:

$$P = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .2 & 0 & .5 & .3 \\ 0 & .3 & .7 & 0 \\ .7 & 0 & 0 & .3 \end{bmatrix}$$

P is stochastic (rows sum to 1)
b/c it's the distribution for next state.

In general, state space Ω
(in this class, $|\Omega|$ is always finite)

random variable $X_t \in \Omega$
X_t = state at time t
discrete time $t = 0, 1, 2, \ldots$

transition matrix P is a $N \times N$ matrix, $N = |\Omega|$.
P is non-negative stochastic.

For $i, j \in \Omega$, $\Pr(X_{t+1} = j \mid X_t = i) = P(i,j)$.

Markovian property:

for states $k_0, k_1, \ldots, k_{t-1}, i, j \in \Omega$,

$$\Pr(X_{t+1}=j \mid X_0=k_0, X_1=k_1, \ldots, X_{t-1}=k_{t-1}, X_t=i)$$
$$= \Pr(X_{t+1}=j \mid X_t=i)$$
$$= P(i,j).$$

2-step transition probabilities:

$$\Pr(X_{t+2}=j \mid X_t=i)$$
$$= \sum_{k\in\Omega} \Pr(X_{t+2}=j, X_{t+1}=k \mid X_t=i)$$
$$= \sum_k \Pr(X_{t+2}=j \mid X_{t+1}=k)\Pr(X_{t+1}=k \mid X_t=i)$$
$$= \sum_k P(k,j)P(i,k)$$
$$= \sum_k P(i,k)P(k,j)$$
$$= P^2(i,j)$$

In general for $\ell \geq 1$,

$$\Pr(X_{t+\ell}=j \mid X_t=i) = P^\ell(i,j)$$

If $X_0 \sim \mu_0$ (so X_0 is sampled from distribution μ_0)

then $X_t \sim \mu_t$ where $\mu_t = \mu_0 P^t$

Earlier example:

$$P^{20} = \begin{bmatrix} .244190 & .244187 & .406971 & .104652 \\ .244187 & .244186 & .406975 & .104651 \\ .244181 & .244185 & .406984 & .104650 \\ .244195 & .244188 & .406966 & .104652 \end{bmatrix}$$

$\exists$ distribution $\pi \approx [.2442, .2442, .4070, .10465]$

where $\lim_{t\to\infty} P^t = \begin{bmatrix} \pi \\ \pi \\ \pi \\ \pi \end{bmatrix}$

thus, for any X_0, any j, $\lim_{t\to\infty} \Pr^*(X_t = j) = \pi(j)$

& $\pi P = \pi$

So we eventually reach π (no matter where we start)

& once we reach it we stay there.

(i.e., π is an eigenvector with eigenvalue 1)

π is a stationary distribution if

$\pi P = \pi$ (eigenvector with eigenvalue 1)

Every stochastic matrix has ≥ 1 such

& all eigenvalues are between ~~± 1~~ $[-1, 1]$

When is there a unique stationary π

& its limiting (converge to it from any X_0)?

<u>Ergodic</u>: $\exists t$ s.t. $\forall i, j \in \Omega$, $P^t(i,j) > 0$

(i.e., graph (Ω, P^t) is the complete graph)

<u>Irreducible</u>: $\forall i, j \in \Omega$, $\exists t$, $P^t(i,j) > 0$

(graph (Ω, P) is 1 SCC)

i.e. strongly connected

<u>Aperiodic</u>: For $i \in \Omega$, period of ~~#~~ i is

$\gcd(T_i)$ where $T_i = \{t : P^t(i,i) > 0\}$.

<u>aperiodic</u> if $\forall i \in \Omega$, $\gcd(T_i) = 1$.

(example of <u>periodic</u> is bipartite)

Ergodic $\iff$ Irreducible & Aperiodic
(see [LPW, Section 1.3]

Fundamental Theorem of Markov Chains:

For a finite, ergodic MC, there is a unique stationary distribution π.
Moreover, for all $i,j \in \Omega$,

$$\lim_{t\to\infty} P^t(i,j) = \pi(j)$$

We'll prove this in a few lectures.

Easy to verify if a MC is ergodic, but how to find π?
$N = |\Omega|$ is huge so can't do Gaussian elimination.

If P is symmetric then $\pi = \text{uniform}(\Omega)$.
$P(i,j) = P(j,i)$

<u>Proof:</u> Let $\pi = \text{uniform}(\Omega)$
Need to verify $\pi P = \pi$.
For $i \in \Omega$, need to check that $(\pi P)(i) = \pi(i)$.

$$(\pi P)(i) = \sum_{k \in \Omega} \pi(k) P(k,i)$$

$$= \sum_k \pi(k) P(i,k) \quad \text{since } P \text{ is symmetric}$$

$$= \frac{1}{N} \sum_k P(i,k) \quad \text{since } \pi = \text{uniform}(\Omega)$$

$$= \frac{1}{N} \quad \text{since } P \text{ is stochastic}$$

$$= \pi(i) \quad \blacksquare$$

Weighted symmetric?
P is <u>reversible</u> wrt π if:
for all $i,j \in \Omega$, $\pi(i)P(i,j) = \pi(j)P(j,i)$.
if this is the case then
π is a stationary distribution.

Proof:

$$(\pi P)(i) = \sum_k \pi(k) P(k,i)$$
$$= \sum_k \pi(i) P(i,k) \quad \text{since reversible}$$
$$= \pi(i) \sum_k P(i,k)$$
$$= \pi(i) \quad \blacksquare$$

So if P is reversible it's easy to determine π but in general it's very difficult.

Random walk on d-regular connected undirected $G=(V,E)$

for $(i,j) \in E$, $P(i,j) = 1/d$.

this is symmetric so $\pi = \text{uniform}(V)$.

What if non-regular (but connected)?

then $\pi(i) = \frac{d(i)}{Z}$ where $d(i)$ = degree of i & $Z = \sum_j d(j) = 2m$.

check: $\pi(i) \cancel{P}(i,j) = \frac{d(i)}{Z}\frac{1}{d(i)} = \frac{1}{Z} = \pi(j) P(j,i)$.

What if G is directed?

then $P(i,j)$ might be >0

& $P(j,i)=0$

So it's not reversible &

no idea of the stationary distribution.

(PageRank!)