Logistics

- Optional Problem Set 5 due 5/13 at 11:59pm.
- Final exam will be Tuesday 5/14, 10:30-12:30pm in the classroom. Study materials to be posted soon.
- Final project due the last day of finals: Friday 5/17.
Summary

Last Time:

- Finish up coupling. Example applications to shuffling, random walks on hypercubes, and exponential convergence of TV distance.
- Markov Chain Monte Carlo – example of sampling random independent sets.
- Start on Metropolis Hastings algorithms and application to sampling from the hardcore model.

Today:

- Finish the Metropolis Hastings algorithm.
- Sampling to counting reduction for independent sets.
A Markov chain is **reversible** if $\pi(i)P_{ij} = \pi(j)P_{ji}$ for all $i,j$. I.e., if the probability of transitioning from state $i$ to state $j$ is equal to the probability of transitioning from state $j$ to state $i$ in the steady state distribution. *’Detailed balance’ condition.*

- If the chain is irreducible and reversible, $P$ has all real eigenvalues, $1 = \lambda_1 > \lambda_2 \ldots > \lambda_n$.
- The eigenvalue gap is $\gamma = \lambda_1 - \max\{|\lambda_2|,|\lambda_n|\}$.
- The mixing time is equal to $\tau(\epsilon) = \tilde{O}(\frac{1}{\gamma})$. 

**Mixing Time and Eigenvalues**
Claim: If a Markov chain is reversible (i.e., \(\pi(i)P_{ij} = \pi(j)P_{ji}\) for all \(i,j\)), then \(P\) has all real eigenvalues.

Proof:

- Let \(D = diag(\pi)\). Then \(D^{-1/2}PD^{1/2}\) is symmetric (and thus has real eigenvalues).

- The above is a similarity transform. The eigenvalues of \(P\) are identical to the eigenvalues of \(D^{-1/2}PD^{1/2}\) and are thus real.
MCMC Methods Continued
Achieving a Non-Uniform Stationary Distribution

Suppose we want to sample an independent set $X$ from our graph with probability:

$$
\pi(X) = \frac{\lambda^{|X|}}{\sum_{Y \text{ independent}} \lambda^{|Y|}},
$$

for some ‘fugacity’ parameter $\lambda > 0$.

Known as the ‘hard-core model’ in statistical physics.
Metropolis-Hastings Algorithm

A very generic way of designing a Markov chain over state space $[m]$ with stationary distribution $\pi \in [0, 1]^m$.

- Assume the ability to efficiently compute a density $p(X) \propto \pi(X)$.
- Assume access to some symmetric transition function with transition probability matrix $Q \in [0, 1]^{m \times m}$.
- At step $t$, generate a ‘candidate’ state $X_{t+1}$ from $X_t$ according to $Q$.
- With probability $\min \left( 1, \frac{p(X_{t+1})}{p(X_t)} \right)$, ‘accept’ the candidate. Else ‘reject’ the candidate, setting $X_{t+1} = X_t$. 
Metropolis-Hastings Intuition
Need to check that for the Metropolis-Hastings algorithm, $\pi P = \pi$.

Suffices to show that $pP = p$ where $p(i) \propto \pi(i)$ is our efficiently computable density.

$$\begin{align*}
[pP](i) &= \sum_j p(j) \cdot Q_{j,i} \cdot \min \left(1, \frac{p(i)}{p(j)}\right) + p(i) \cdot \sum_j Q_{i,j} \left(1 - \min \left(1, \frac{p(j)}{p(i)}\right)\right) \\
&= \left\{\begin{array}{ll}
\text{acceptances} & \\
\text{rejections} & \\
\end{array}\right.
\end{align*}$$

$$\begin{align*}
&= \sum_j Q_{i,j} \cdot \min (p(j), p(i)) + p(i) \cdot \sum_j Q_{i,j} - \sum_j Q_{i,j} \cdot \min(p(i), p(j)) \\
&= p(i) \cdot \sum_j Q_{i,j} = p(i).
\end{align*}$$
Metropolis-Hastings for the Hard-Core Model

Want to sample an independent set $X$ with probability

$$\pi(X) = \frac{\lambda^{|X|}}{\sum_{Y \text{ independent}} \lambda^{|Y|}}.$$  

- Let $p(X) = \lambda^{|X|}$ and let the transition function $Q$ be given by:
  - Pick a random vertex $v$.
  - If $v \in X_t$, set $X_{t+1} = X_t \setminus \{v\}$ with probability $\min(1, 1/\lambda)$.
  - If $v \notin X_t$ and $X_t \cup \{v\}$ is independent, set $X_{t+1} = X_t \cup \{v\}$.
  - Else set $X_{t+1} = X_t$ with probability $\min(1, \lambda)$.

- Need to accept the transition with probability $\min \left( 1, \frac{p(X_{t+1})}{p(X_t)} \right)$.

The key challenge then becomes to analyze the mixing time.

For the related Glauber dynamics, Luby and Vigoda showed that for graphs with maximum degree $\Delta$, when $\lambda < \frac{2}{\Delta - 2}$, the mixing time is $O(n \log n)$. But when $\lambda > \frac{c}{\Delta}$ for large enough constant $c$, it is NP-hard to approximately sample from the hard-core model.
MCMC for Approximate Counting
Counting to Sampling Reductions

Often if one can efficiently sample from the distribution 
\[ \pi(X) = \frac{p(X)}{\sum_Y p(Y)} \], one can efficiently approximate the normalizing constant \( Z = \sum_Y p(Y) \) (often called the partition function).

- If \( Z \) is hard to approximate, then this can give a proof that sampling is hard, and thus it is unlikely that any simple MCMC method for sampling from \( \pi \) mixes rapidly.

- This is e.g., how one can show that sampling from the hard-core model is hard when \( \lambda = \Omega(1/\Delta) \).

- Let’s consider the simple case of \( \lambda = 1 \). I.e., we want to sample a uniformly random independent set.

- In this case, \( Z = |S(G)| \), the number of independent sets in \( G \). It is known that approximating \( |S(G)| \) even up to a \( \text{poly}(n) \) factor is NP-Hard.
Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

Let $G_0, G_1, \ldots, G_m$ be a sequence of graphs with $G_m = G$ and $G_i$ obtained by removing an arbitrary edge from $G_{i+1}$.

We can write:

$$|S(G)| = \frac{|S(G_m)|}{|S(G_{m-1})|} \cdot \frac{|S(G_{m-1})|}{|S(G_{m-2})|} \cdots \frac{|S(G_1)|}{|S(G_0)|} \cdot |S(G_0)|.$$
Counting Independent Sets

\[ |S(G)| = \frac{|S(G_m)|}{|S(G_{m-1})|} \cdot \frac{|S(G_{m-1})|}{|S(G_{m-2})|} \cdot \ldots \cdot \frac{|S(G_1)|}{|S(G_0)|} \cdot |S(G_0)| 2^n = 2^n \cdot \prod_{i=1}^{m} r_i, \]

where \( r_i = \frac{|S(G_m)|}{|S(G_{m-i})|} \). If we can estimate each \( r_i \) with \( \tilde{r}_i \) satisfying

\[
\left(1 - \frac{\epsilon}{2m}\right) \cdot r_i \leq \tilde{r}_i \leq \left(1 + \frac{\epsilon}{2m}\right) \cdot r_i,
\]

then:

\[
(1 - \epsilon) \cdot |S(G)| \leq 2^n \cdot \prod_{i=1}^{m} \tilde{r}_i \leq (1 + \epsilon) \cdot |S(G)|
\]

since \( (1 + \frac{\epsilon}{2m})^m \leq 1 + \epsilon \) and \( (1 - \frac{\epsilon}{2m})^m \geq 1 - \epsilon. \)
Consider the ratio \( r_i = \frac{|S(G_i)|}{|S(G_{i-1})|} \). Observe that \( r_i \leq 1 \).

Further, \( r_i \geq 1/2 \). Let \((u, v)\) be the edge removed from \( G_i \) to obtain \( G_{i-1} \). Then each independent set in \( S(G_{i-1}) \setminus S(G_i) \), must contain both \( u \) and \( v \).

So, we can map each set in \( S(G_{i-1}) \setminus S(G_i) \) to a unique set in \( S(G_i) \) by simply removing \( v \).

\[
\begin{align*}
\frac{|S(G_i)|}{|S(G_{i-1})|} &= \frac{|S(G_i)|}{|S(G_i)| + |S(G_{i-1}) \setminus S(G_i)|} \\
&\geq \frac{1}{2}.
\end{align*}
\]
So Far: We have written $|S(G)| = 2^n \cdot \prod_{i=1}^{m} r_i$ where $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Need to get a $1 \pm \epsilon/m$ estimate to each $r_i$ to get a $1 \pm \epsilon$ estimate to $|S(G)|$.

Let $X$ be a random variable generated as follows: pick a random independent set from $G_{i-1}$ and let $X = 1$ if the set is also independent in $G_i$. Otherwise let $X = 0$.

What is $\mathbb{E}[X]$?

How many samples of $X$ do we need to take to obtain a $1 \pm \epsilon/m$ approximation to $r_i$ with high probability?
Counting Independent Sets

Upshot: For a graph $G$ with $m$ edges, making $\tilde{O}(m^2/\epsilon^2)$ calls to a uniform random independent set sampler on $G$ or its subgraphs suffices to approximate the number of independent sets in $G$ up to $1 \pm \epsilon$ relative error.

- So a polynomial time algorithm for uniform random independent set sampling, would lead to a polynomial time algorithm for counting independent sets, and hence the collapse of $NP$ to $P$.
- Observe that near-uniform sampling (as would be obtained e.g., with an MCMC method) would also suffice.