Matroids:

Matroid $M = (E, I)$ consists of:

- ground set E
- set of subsets $I \subseteq 2^E$ called independent sets

where:

P1. Downward closure:

every subset of an independent set is also an independent set, i.e.,
if $S \in I$ & $T \subseteq S$ then $T \in I$.

P2. Augmentation: (exchange property)

if $S \in I$ & $T \in I$ & $|S| < |T|$
then $\exists e \in S \setminus T$ where $T \cup e \in I$

Note, $\emptyset$ is always an independent set.
A set $R \subset E$ which is not an independent set is called dependent.

A <u>basis</u> is a <u>maximal</u> independent set,

i.e., an independent set S where for all $e \notin S$, $S \cup e \notin I$.

All bases have the same size called <u>rank</u> of M.

<u>Example:</u>

1. <u>Graphic</u> matroid:

Let $G=(V,E)$ be a connected graph.

Then $M=(E,I)$ is a matroid where:

E = edges of G,

I = forests of G (acyclic subgraphs)

The bases are maximal forests = spanning trees.

2. Transversal matroids:

Let $G=(L \cup R, E)$ be a bipartite graph.

then $M=(E, I)$ where:

$E = L$,

I = subsets of L that are endpoints of some matching of G.

bases are subsets of L covered by a maximum matching (note, $S \in I$ can correspond to multiple max matchings)

Basis exchange process:

Let Ω = bases of matroid $M = \{S \subseteq E : S$ is a basis for $M\}$

(think spanning trees of G)

From $B_t \in \Omega$,

1. Pick $e \in B_t$ uar
2. Let $F = \{f \in E : B_t \setminus e \cup f \in I\}$ denote edges of E that we can add to $B \setminus e$ while remaining indpt.
3. Pick $f \in F$ uar & let

$$B_{t+1} = B_t \setminus e \cup f.$$

Basis-exchange walk is ergodic.

Aperiodic: $e \in F$, hence $P(B_t, B_t) > 0$.

Irreducible: Consider $B, B' \in \Omega$ so $|B| = |B'| = r$.

Let $B \oplus B' = \{e_1, \dots, e_k\} \cup \{f_1, \dots, f_k\}$

where $\{e_1, \dots, e_k\} \subseteq B$ & $\{f_1, \dots, f_k\} \subseteq B'$.

Induct on k:

If $k=1$, then let $\hat{B} = B \setminus e_1$.

By P2, $\exists e' \in B' \setminus B$ so that

$\hat{B} \cup e' \in \Omega$ & it must be that

$e' = f_1$ since $|B \oplus B'| = 2$.

For $k > 1$, follow the same approach & can reduce the symmetric difference by 2.

Note, the basis exchange walk is symmetric so

$\pi = \text{uniform}(\Omega)$.

[Anari, Liu, Oveis-Gharan, Vinzant '18] showed $T_{mix} = O(n^2 \log n)$

& [Cryan, Guo, Mousa '19] showed $T_{mix} = O(n \log n)$

Will analyze using functional analysis introduced last class.

For a MC defined by (P, Ω, π):

for a function $f: \Omega \to \mathbb{R}$,

$$\text{let } E_\pi f = \mu(f) = \sum_{x \in \Omega} \pi(x) f(x)$$

$$\& \ \mathrm{Var}_\pi f = \sum_{x \in \Omega} \pi(x)\left(f(x) - E_\pi f\right)^2$$

We saw that:

$$\mathrm{Var}_\pi f = \frac{1}{2} \sum_{x,y \in \Omega} \pi(x)\pi(y)\left(f(x) - f(y)\right)^2$$

(this measures global variation)

local variation wrt P is measured by the

Dirichlet form: $\mathcal{E}_P(f,f) = \frac{1}{2} \sum_{x,y \in \Omega} \pi(x) P(x,y) \left(f(x) - f(y)\right)^2$

(measures how f varies over transitions)

Poincare constant: $\alpha = \inf_{f:\Omega \to \mathbb{R},} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi f}$

only consider f not constant

So $\mathrm{Var}_\pi f \neq 0$.

<u>Theorem:</u> $T_{mix} = O\left(\frac{1}{\alpha} \log\left(\frac{1}{\pi_{min}}\right)\right)$

<u>Key fact:</u> $\mathrm{Var}_\pi(Pf) \leq \mathrm{Var}_\pi f - \frac{1}{2}\mathcal{E}_P(f,f)$

then $\mathrm{Var}_\pi(P^t f) \leq \left(1 - \frac{\alpha}{2}\right)^t \mathrm{Var}_\pi f$

Letting $f = 1(A)$ for $A \subset \Omega$

then $\mathrm{Var}_\pi(P^t f)$ measures ~~how far~~

$|P^t(X_0, A) - \pi(A)|$

So setting $t =$ we get that $\leq \epsilon$,

for all $A \subset \Omega$ so the total variation distance of X_t & π is $\leq \epsilon$

See Jerrum P.51 in chapt 5 for Details

Statistical inference: (Bayesian inference)

Goal: underlying model with parameters Θ

for example, Θ is a graph $G=(V,E)$ with edge parameters β for an Ising model

Given: observed data X (generated from Θ)

for example, vertices are assigned labels $\{+1,-1\}$ (maybe dark/light pixels)

Maximum likelihood: find Θ which $\max_{\Theta} \Pr(\Theta \mid X)$

More robust, sample from the posterior distribution:

$$p(\Theta) := \Pr(\Theta \mid X) = \frac{\Pr(X \mid \Theta)\Pr(\Theta)}{\Pr(X)}$$

Bayes rule

Three terms there: $Pr(X|\theta)$, $Pr(\theta)$, $Pr(X)$

- $Pr(\theta)$ = <u>Prior</u>: we need to specify the prior.
  Can set it to be the same for all θ (uniform prior)
  & then $Pr(\theta|X) \propto Pr(X|\theta)$

  this is the likelihood

- $Pr(X|\theta)$ = likelihood of data X in model θ
  e.g., given Ising model $\theta = (G, \beta)$
  what's the prob. of seeing configuration(s) X?
  easy to compute $w(X)$ but not prob., need the normalizing factor.

- $Pr(X)$ = Prob. X is observed (without knowledge of θ)

For model Θ, let $\omega(\Theta) = Pr(X|\Theta)Pr(\Theta)$

then the posterior distribution:

$$\pi(\Theta) = Pr(\Theta|X) = \frac{\omega(\Theta)}{Z}$$

where $Z = \sum_{\Theta'} \omega(\Theta')$

Can design a MC whose stationary distribution is $\pi(\Theta)$, but this is on the space of $\Theta = (G, \beta)$ so we vary G &/or β.

For a specific $\Theta = (G, \beta)$ need to compute $\omega(\Theta)$.

for uniform priors, need to compute $Pr(X|\Theta)$

& for this we need $\cancel{Z_\Theta = \sum_{X'} Pr(X'|\Theta)}$

$$Z(\Theta) = \sum_{X'} \omega(X'|\Theta)$$

= Partition function for Ising model

We saw examples of MC's for sampling.

For example, given $G=(V,E)$, generate a random matching of G.

$\Rightarrow$ Can use this to approximate # of matchings of $G = |M(G)|$.

How? fix edge e & generate samples to determine $Pr(e \in M)$ vs. $Pr(e \notin M)$, then recurse.

How to use sampler to approximately count in general?

Consider Ising model on (G, β).

Let $B = e^{\beta}$ so that for $\sigma \in \{+1,-1\}^V$,

$$\omega(\sigma) = B^{|M(\sigma)|}$$

where $M(\sigma)$ = monochromatic edges of G in σ.

$B > 1$: ferromagnetic/attractive; $B < 1$: antiferro/repulsive.

Say $B>1$, goal is to compute Z_{B^*} for $B^*=B$.

Define sequence:

$$B_0 = 1 < B_1 < \cdots < B_{\ell-1} < B_\ell = B^*$$

Note,

$$Z(B^*) = \frac{Z(B_\ell)}{Z(B_{\ell-1})} \times \frac{Z(B_{\ell-1})}{Z(B_{\ell-2})} \times \cdots \times \frac{Z(B_1)}{Z(B_0)} \times Z(B_0)$$

$$\& \; Z(B_0) = 2^n$$

Let $\rho_i = \dfrac{Z(B_i)}{Z(B_{i-1})}$ so $Z(B^*) = 2^n \prod_{i=1}^{\ell} \rho_i$

Generate $\sigma \sim \mu_{B_{i-1}}$ & look at $W_i = \dfrac{w_{B_i}(\sigma)}{w_{B_{i-1}}(\sigma)} = \left(\dfrac{B_i}{B_{i-1}}\right)^{M(\sigma)}$

Note,

$$E_{\mu_{B_{i-1}}}[W_i] = \sum_{\sigma \in \{+1,-1\}^V} \frac{\cancel{w_{B_{i-1}}(\sigma)}}{Z(B_{i-1})} \frac{w_{B_i}(\sigma)}{\cancel{w_{B_{i-1}}(\sigma)}} = \frac{1}{Z(B_{i-1})} \sum_{\sigma} w_{B_i}(\sigma)$$

$$= Z_{B_i}/\; = \rho_i$$

Thus, if we can generate samples from $\mu_{B_{i-1}}$ using our MC, then the estimator

$$W_i = \frac{w_{B_i}(\sigma)}{w_{B_{i-1}}(\sigma)} = \left(\frac{B_i}{B_{i-1}}\right)^{|M(\sigma)|}$$

is an unbiased estimator for ρ_i.

How many samples from $\mu_{B_{i-1}}$ are needed to get a good estimate of ρ_i?

If $B_i = \left(1+\frac{1}{2m}\right)B_{i-1}$ then

$$W_i \le \left(\frac{B_i}{B_{i-1}}\right)^{|M(\sigma)|} \le \left(1+\frac{1}{2m}\right)^m \le e^{1/2}$$

for $m = |E|$.

Take $s = O\left(\frac{\ell^3}{\epsilon^2}\right)$ samples from $\mu_{B_{i-1}}$ & look at sample mean as estimate of ρ_i.

Then by Chebyshev's ineq., $\Pr\left(\overline{W_i} \ge \rho_i\left(1+\frac{\epsilon}{2\ell}\right)\right) \le \frac{1}{8\ell}$

So with prob. $\ge 1 - \frac{\ell}{8\ell} = \frac{7}{8}$, $2^n \prod_i \overline{W_i} \le 2^n \prod_i \rho_i \left(1+\frac{\epsilon}{2\ell}\right)^\ell$

Hence, we need $O\left(\frac{\ell^3}{\epsilon^2}\right)$ samples per B_i

& $O\left(\frac{\ell^{4}}{\epsilon^2}\right)$ samples in total.

What's ℓ? $\quad \ell = O(m \ln \beta^*)$

$O^*(\ )$ hides log factors & $\frac{1}{\epsilon}$ dependence.

Possible to do with $\ell = O^*(\sqrt{n})$,

and by looking at $Var(W_1 W_2 \cdots W_\ell)$

we can get it down to $O\left(\frac{\ell}{\epsilon^2}\right)$ sampes per i,

& $O\left(\frac{\ell^2}{\epsilon^2}\right)$ samples in total.

For $\ell = O^*(\sqrt{n})$ we need $O^*(n)$ total

samples from our MC.

Chebyshev's ineq.: $\Pr(|X-\mu| \geq c) \leq \frac{Var(X)}{c^2}$

Want $X \leq (1+\frac{\epsilon}{2})\mu$

So set $c = \frac{\epsilon}{2}\mu$

then: $\Pr(|X-\mu| \geq \frac{\epsilon}{2}\mu) \leq \frac{4}{\epsilon^2}\frac{Var(X)}{\mu^2}$

So need to bound $\frac{Var(X)}{\mu^2}$.

Note, $W_i = O(1)$ so $Var(W_i) \leq O(1)$ $\left(\begin{array}{l}\text{if } X \leq c \\ \text{then } Var(X) \leq \frac{c^2}{4}\end{array}\right)$

Let $\overline{W_i}$ be the sample mean of s samples, then

$$Var(\overline{W_i}) = \frac{Var(W_i)}{s}$$

taking $s = O\left(\frac{l}{\epsilon^2}\right)$ then $\frac{Var(\overline{W_i})}{\mu_i^2} = \frac{Var(\overline{W_i})}{\rho_i^2} \leq O\left(\frac{\epsilon^2}{l}\right)$

Now for $X = \overline{W_1}\,\overline{W_2}\cdots\overline{W_l}$ we have,

$$\frac{Var(\overline{W_1}\cdots\overline{W_l})}{(\mu_1\cdots\mu_l)^2} = \frac{E[\overline{W_1}^2\,\overline{W_2}^2\cdots\overline{W_l}^2]}{\mu_1^2\cdots\mu_l^2} - 1$$

$$= \left(\prod_i \frac{E[\overline{W_i}^2]}{\mu_i^2}\right) - 1 \quad \text{Since } \overline{W_i}\text{'s are indpt.}$$

$$= \prod_i \left(1 + \frac{Var(\overline{Z_i})}{\mu_i^2}\right) - 1 \leq \left(1 + \frac{c\epsilon^2}{l}\right)^l - 1 \leq e^{\epsilon^2 c\,\cancel{/4}} - 1$$

$$\leq \epsilon^2 .$$

Thus by taking $s = O\left(\frac{\ell}{\epsilon^2}\right)$ samples per i then we get an estimate of $\mu_1 \cdots \mu_\ell = \rho_1 \cdots \rho_\ell$ which is within $\leq \frac{\epsilon \mu_1 \cdots \mu_\ell}{2}$ with prob. $\geq \frac{3}{4}$.