# Probability Seminar Notes

Jared Marx-Kuo

Oct. 4th - Dec. 6th, 2019

## 10/4: Yuval's talk on "Expander Graphs & Their Applications"

1. Simple Random walk on a graph, $G = (V, E)$ with $|V| = n$. Transition matrix $M \in M_{n,n}$ s.t.

$$M = (m_{ij}) \quad m_{ij} = \begin{cases} \frac{1}{d_j} & \text{vertex i adjacent to vertex j} \\ 0 & \text{else} \end{cases}$$

2. If $p$ a prob vector, $p \in \mathbb{R}^n$ s.t. $p_i \geq 0$ and $\sum_i p_i = 1$, then $Mp$ is the prob vector after one step of the random walk, if we started with $p$. This is because

$$(Mp)_i = \sum_j m_{ij} p_g = \sum_{j \sim i} \frac{1}{d_j} p_j$$

where $j \sim i$ denotes vertex $j$ adjacent to vertex $i$.

3. Let $\vec{d} = (d_i)$ be the vector of degrees, i.e. vector whose ith coordinate is the degree of vertex $i$. Then

$$(Md)_i = \sum_{j \sim i} \frac{1}{d_j} d_j = d_i$$

thus $d$ is an eigenvector of $M$ with eigenvalue equal to 1

4. $\vec{d}$ normalized will be the stationary distribution of this walk, i.e. $\frac{1}{\sum_i d_i} \vec{d}$

5. From now on, $G$ is d-regular (i.e. some unknown number of vertices $n$, but each vertex has degree $d$), so:
   (a) the stationary distribution is uniform, i.e. $u = (1/n, \dots, 1/n)$
   (b) $M$ is symmetric, $M = \frac{1}{d} A$ where $A$ is the adjacency matrix
   (c) Let eigenvalues of $A$ be $d = \lambda_1, \dots, \lambda_n$

6. Facts:
   (a) $\lambda_1 = \lambda_2 \iff G$ disconnected
   (b) $\lambda_n = -\lambda_1 \iff G$ bipartite
   (c) Note that in order for our SRW (simple random walk) to converge to a stationary distribution, we should have that $G$ is both connected and not bipartite

7. **Question**: Starting with some distribution $p$, how does $M^t p \to u$? Here, $u$ is the stationary distribution and $t \in \mathbb{Z}^+$.

8. Let $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$ and $\alpha = \lambda / d$

9. **Theorem:** : For any distribution $p$, $||M^t p - u|| \leq \alpha^t$. Note: in the disconnected and bipartite case, this bound is 1
   **Proof:** : Note that

$$\langle p - u, u \rangle = \frac{1}{n} \sum_i (p_i - u) = \frac{1}{n} \left( \sum_i p_i - 1 \right) = 0$$

Note too that $Mu = u$, second largest e-value of $M$ is $\alpha$ by correspondence between $M$ and $A$, the adjacency matrix. Now

$$||M^t p - u||_2 = ||M^t(p-u)||_2 \leq \alpha^t ||p-u||_2 \leq \alpha^t$$
$$||p-u||_2^2 = \langle p, p-u\rangle - \langle u, p-u\rangle$$
$$= \sum p_i(p_i - 1/n) \leq \sum_i p_i = 1$$

Note that we showed that $p - u \perp u$, which is the eigenvector with eigenvalue $\lambda_1$, so we can bound by the second largest eigenvalue. Also in the above we use $|p_i - 1/n| \leq 1$. □

10. **Corollary:** $||M^t p - u||_1 \leq \sqrt{n}\alpha^t$
    **Proof:** Cauchy Schwarz: $\forall v,\ ||v||_1 \leq \sqrt{n}||v||_2$. □

11. Now note
$$t \geq \frac{\frac{1}{2}\log n}{\log(1/\alpha)} + C \implies ||M^t p - u||_1 \leq e^{-C}$$

12. We want $\alpha \leq 1-\epsilon$, this is called the spectral gap. A graph is called an $(n, d, \epsilon)$-spectral **EXPANDER** if $n$ vertices, $d$-regular, and $\alpha \leq 1 - \epsilon \iff \lambda \leq (1-\epsilon)d$ for $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$ as before

13. **Question**: How small can $\lambda$ be in terms of $d$. Answer: $G = K_n$, which is $(n-1)$-regular, implies that $d = n-1$ all other eigenvalues are $-1$, so $\lambda = 1$

14. In light of the above, we want to restrict to the case when $d = o(n)$ or $d = O(1)$, i.e. degree is growing slowly w.r.t. the number of vertices

15. **Theorem:** If $d = o(n)$, then $\lambda > \sqrt{d}(1 - o(1))$
    **Proof:** : Consider $\text{tr}(A^2) = \sum_i \lambda_i^2 \leq d^2 + \lambda^2(n-1)$. But also $\text{tr}(A^2) = dn$ because the $(i,i)$th entry is $d$ which can be seen by doing the matrix multiplication out. Thus
$$dn \leq d^2 + \lambda^2(n-1) \implies \lambda^2 \geq d\frac{n-d}{n-1} \implies \lambda \geq \sqrt{d}(1-o(1))$$

16. **Theorem:** (Alon - Boppana) $\lambda \geq 2\sqrt{d-1}(1-o(1))$ for $d$ constant

17. **Question**: Do Expanders even exist??? yeS

18. How people came up with expanders: Random constructions, i.e. if $G$ is a random $d$-regular graph on $n$ vertices, want to understand $\lambda(G)$.

19. **Theorem:** (Broder-Shamir) $\lambda(G) = O(d^{3/4})$ almost surely! (a.a.s) - I don't know what this means. Could apply in the context of random matrix theory/graphs

20. **Theorem:** (Friedman) $\forall \epsilon > 0$, $\Pr[\lambda(G) > 2\sqrt{d-1} + \epsilon] = o(1)$ when $d \geq 3$

21. Explicit constructions: a graph with $\lambda = 2\sqrt{d-1}$ is called a Ramanujan graph. See Lubotzky-Philips-Sarnak: Ramanujan graphs exist for all $d = p+1$ where $p \equiv 1 \mod 4$ and $p$ is a prime. The graph is a Cayley graph of $PSL_2(\mathbb{F}_q)$, other $q$, infinitely many $n$

22. **Conjecture:** 275 of graphs are Ramanujan for all $d$

23. Let $V(G) = \mathbb{F}_p$ and $x \sim x+1$ and $x \sim x-1$ and $x \sim x^{-1}$ when $x \neq 0$. Turns out this is an expander, but this requires hard number theory

24. **Lemma:** (Expander Mixing Lemma): $\forall S, T \subseteq V$, we have
$$\left|E(S,T) - \frac{d}{n}|S||T|\right| \leq \lambda\sqrt{|S|\ |T|}$$
    where $E(S,T)$ is the number of edges between $S$ and $T$, when $S \cap T = \emptyset$ and is some modification when they are not disjoint. This is proved doing similar tricks as before with
$$E(S,T) = \mathbb{1}_S A \mathbb{1}_T^T$$

25. A calculation
$$\Big|\frac{E(S,T)}{dn} - \frac{|S||T|}{n^2}\Big| \le \frac{\lambda}{dn}\sqrt{|S||T|} \le \frac{\lambda}{d} = \alpha$$
This of this as picking $x, y$ random vertices of $V$, then $Pr[x \in S, y \in T]$ relates to the above. Pick $x$ at random, $y \sim x$ another random choice

26. Fix a "bad" set, $B \subseteq V$, $|B| = \beta n$. Pick $t+1$ vertices, we want one of them to be outside of $B$. If I choose independently, $Pr[\text{failure}] = \beta^{t+1}$

27. **Theorem:** (Ajtai - Komlos - Szemeredi) If instead I pick $X_0$ random, $X_1, \ldots, X_t$ to be a SRW, then
$$Pr[X_i \in B\ \forall i] \le (\beta + \alpha)^t$$

28. **Theorem:** (Expander Chernoff bound) (Gillman) Let $f : V \to [0,1]$, let $\mu = \mathbb{E}f$. SRW $X_0, \ldots, X_t$ starting at random vertex. Then
$$\forall \epsilon > 0 \qquad Pr[|\frac{1}{t}\sum_i f(X_i) - \mu| > \epsilon] \le e^{-c\epsilon^2 t(1-\alpha)}$$

# 10/11: Mark Selke's talk on Cheeger Inequalities

1. Mark Selke is giving a talk about Cheeger stuff. We'll go over
   (a) Cheeger's inequality
   (b) Multiway cheeger
   (c) **Theorem:** Let $T$ be a self-adjoint ergodic Markov Operator with
   $$||T||_{L^1 \to L^p} < \infty\ \forall p > 2, \qquad \text{then } spec(T) \subseteq \{1\} \sqcup [-1, 1-\epsilon]$$
2. Let $G$ be a $d$-reguler graph, $n$ vertices $V$
3. Normalized Laplacian:
$$L_G = I - \frac{A}{d}, \quad L_G(f)(v) = f(v) - \frac{1}{d}\sum_{w \sim v} f(w)$$
where $w \sim v$ means the two vertices are adjacent
4. $L_G$ has eigenvalues $0 = \lambda_1 \le \lambda_2 \le \cdots \le 2$
5. Spectral gap aims to measure $\lambda_2 - \lambda_1$
6. If the spectral gap is big (i.e. $\lambda_2 > 0$ and large, maybe $\Omega(1)$, i.e. is greater than some constant $c \ge 0$), then this corresponds to fast mixing of random walks
7. This connects with things called "sparse bottleneck" which is when we only have a few edges between two components of the graph.
8. The idea is that no sparse bottlenecks is equivalent to a large spectral gap
9. Bottleneck: $S \subseteq V$ has a bottleneck ratio of
$$\phi(S) = \frac{E(S, S^c)}{d \cdot \min(|S|, |S|^c)}$$
Note that
$$\phi(G) = \min_{S \subseteq V} \phi(S)$$
10. **Theorem:** (Cheeger)
$$\frac{\lambda_2}{2} \le \phi(G) \le \sqrt{2\lambda_2}$$
11. When is either side sharp? Consider $K_n$, the complete graph of size $n$, for which $\lambda_2 = 1$ and $\phi(G) = 1/2$

12. For a cycle graph $C_n$, we have $\lambda_2 \sim \frac{1}{n^2}$ and $\phi \sim \frac{1}{n}$ so this shows us Cheeger is tight (up to a constant, which I guess doesn't really matter?)

13. For the hypercube $\{0,1\}^n$ we have $\lambda_2 \sim \frac{1}{n}$ and $\phi = \frac{1}{n}$

14. Rayleigh Quotient

$$f : V \to \mathbb{R}, \qquad R(f) = \frac{\sum_{v \sim w}(f(v) - f(w))^2}{d \cdot \sum_v f(v)^2}$$

think of the numerator as $\langle Lf, f \rangle$ for the laplacian (its actually $L_G$ but there's some tricky grouping to get this to work out). If $f$ has an eigendecomposition, then

$$f = \sum_i \alpha_i f_i \qquad R(f) = \frac{\sum_i \lambda_i \alpha_i^2}{\sum_i \alpha_i^2}$$

for $\{f_i\}$ the eigenfunctions, which are pairwise orthogonal. Another way to think about this is

$$\langle Lf, f \rangle = \sum_{v \sim w} (f(v) - f(w))^2 = \langle \nabla f, \nabla f \rangle$$

15. Suppose that $\alpha_1 = 0$, then $R(f) \geq \lambda_2$

16. Part 1: $\lambda_2/2 \leq \phi(G)$

    (a) Say $S$, $\phi(S) = \frac{E(S,S^c)}{d \cdot \min(|S|,|S^c|)} = \phi(G)$

    (b) Then define

$$g_S(V) = \begin{cases} \frac{|S|}{n} & v \notin S \\ \frac{|S|-n}{n} & v \in S \end{cases}$$

Note that $g_s(V)$ has mean zero. Moreover

$$\lambda_2 \leq R(g_s) = \frac{E(S,S^c)}{d\left(\frac{|S|}{n}\right)^2 (n - |S|) + \left(\frac{n-|S|}{n}\right)^2 |S|} = \frac{E(S,S^c)}{d\frac{|S||S|^c}{n}} \cong \phi(G)$$

Note that up to a factor of two, we have

$$d\frac{|S||S|^c}{n} \sim d\min(|S|, |S|^c)$$

17. Part 2: $\phi(G) \leq \sqrt{2\lambda_2}$

    (a) Start with $f_2$ such that $R(f_2) = \lambda_2$ because $f_2$ is the second eigenfunction of $f$

    (b) Want to get a set $S$ with $\phi(S)$ small

    (c) Natural approach: take $\alpha \sim xdx$, i.e. it is distributed according to $xdx$, let

$$S_\alpha = \{v \mid f_i(v) > \alpha\}$$

Then for $v, w \in V$, we have that

$$\mathbb{P}(\text{v,w are on opposite sides }) = |f(v)^2 - f(w)^2|$$

where "opposite sides" means that one lies in $S_\alpha$ and the other in $S_\alpha^c$

    (d) Now replace $f_2$ with either the positive or negative part, i.e.

$$g = (f_2)_+ \text{ or } g = (f_2)_-$$

then $R(g) \leq 2\lambda_2$.

(e) Now we ask what is the expected value of the number of edges between $S$ and $S^c$, i.e.

$$\mathbb{E}[E(S_\alpha, S_\alpha^c)] = \sum_{v\sim w} |g(v)^2 - g(w)^2|$$

We also have

$$\mathbb{E}[|S_\alpha|] \cong \sum_v g(v)^2$$

(f) We have the following lemma: If $\mathbb{E}(A)/\mathbb{E}(B) \leq c$, then $\mathbb{P}[A/B \leq c] > 0$, by looking at $\mathbb{E}[A - cB]$

(g) Now going back, we have

$$\mathbb{E}[E(S_\alpha, S_\alpha^c)] = \sum_{v\sim w} |g(v)^2 - g(w)^2| \leq \sqrt{\sum_{v\sim w}(g(v) - g(w))^2} \times \sqrt{\sum_{v\sim w}(g(v) + g(w))^2}$$

via cauchy schwartz. Further note that

$$\sum_{v\sim w}(g(v) + g(w))^2 \cong d\sum_v g(v)^2$$

(h) The above yields

$$\frac{\mathbb{E}[E(S, S^c)]}{\mathbb{E}[|S_\alpha|]} \lesssim \sqrt{\frac{\sum_{v\sim w}(g(v) - g(w))^2}{\sum_v g(v)^2}} \lesssim \sqrt{\lambda_2}$$

18. Higher Order Cheeger

(a) We have the following bound

$$\frac{\lambda_k}{2} \leq p_k(G) \leq k^{O(1)}\sqrt{\lambda_k}$$

we also have

$$p_k(G) \leq \sqrt{\lambda_{2k}\log k}, \qquad p_k(G) \leq \ell^{O(1)}\frac{\lambda_k}{\sqrt{\lambda_\ell}} \quad \forall \ell \geq k$$

We have that

$$p_k(G) = \min_{\substack{S_1,\dots,S_k \\ \text{partition}}} \max_{j\leq k} \phi(S_j)$$

19. Application: Clustering Data

(a) Form weighted similarity graph

(b) Map to $\mathbb{R}^k$, i.e. $v \mapsto (f_1(v), \dots, f_k(v))$

(c) Do some clustering

## 10/18: Jared Marx-Kuo's talk on Hypercontractivity and Log-Sobolev Inequalities

## 10/25: Mark Perlman's talk entitled "Trailing 'Trailing the Dovetail shuffle to its lair' to its lair"

1. **Definition:** TFAE definitions of a riffle shuffle

(a) Sequential:

i. Step 1: cut deck into 2 piles, choice of pile is distributed $\sim Bin(n, 1/2)$. So cutting the deck in half in a binomial way

ii. Interleave them sequentially, #(pile 1) $= A$, #pile 2 $= B$, and $\mathbb{P}$(next card falls from pile 1) $= A/(A+B)$

(b) Entropy: Each valid tuple (pile size, interleaving) (here, pile size refers to the size of pile 1, and there's only two piles in play) is equally likely

(c) Geometric: have $n$ i.i.d. variables distributed uniformily on $[0,1]$. Then we map $X \mapsto 2X \mod 1$

2. **Remark** The riffle shuffle procedure defines a markov chain on $S_n$

3. Our goal: study $\lim_{m\to\infty} ||Q_m - U||_{TV}$ for $n$ fixed, where $U$ is the uniform measure on $S_n$. Here $Q_m = Q^m v_0$ where $v_0$ is our initial vector and $Q^m$ is our markov chain raised to the $m$th power

4. We now prove equivalence of these definitions of shuffles:
**Proof:** (Sequential = Entropic)
1. Need to show same distribution on pile sizes, so the plan: show for entropy, pile size $\sim Bin(n, 1/2)$.

For a given pile size $k$, how many valid interleavings are there? This is a stars and bars argument, and it's just $\binom{n}{k}$ which matches up with the binomial distribution.
2. Given a pile size, want same distribution on interleavings. The plan: in sequential, show interleavings are equally likely

$$\mathbb{P}(\text{some interleaving} \mid k) = \frac{k!(n-k)!}{n!}$$

(maybe some other denominator, but there are definitely $k! \cdot (n-k)!$ choices to be made).

(Geometric = Others proof):
1. Pile size distribution: how many points will be $> 1/2$? This is $Bin(n, \frac{1}{2})$ which is the same as the sequential model so we're good.

2. Interleavings have the same distribution given a pile size: We have $[0,1]$ and uniform rand variables in that space. Any interleaving should be equally likely once we bring the variables who's values are $> 1/2$ back to $[0,1]$ (by doing $2x \mod 1$ operation). □

5. **Definition:** /**Theorem:** TFAE definitions of an a-shuffle

(a) Sequential:

i. Step 1: Cut into $a$-piles. Distribute the shuffle multinomially, i.e. $\sim Multi(n, (\frac{1}{a}, \frac{1}{a}, \ldots, \frac{1}{a}))$

ii. Step 2: $\mathbb{P}(\text{next card falls from pile i}) = \frac{k_i}{\sum_j k_j}$

(b) Entropy: every valid tuple $((k_1, k_2, \ldots, k_a), \text{interleaving})$ is equally likely

(c) Geometric: $X \mapsto aX \mod 1$

6. Rule: applying an $a$-shuffle, then a $b$-shuffle is the same as applying an $ab$-shuffle (this follows by the definition of a geometric shuffle, which is equivalent to all of the other ones)

7. So applying $m$ 2-shuffles is equivalent to applying $2^m$ shuffles! Neat!

8. Rising Sequences:

(a) How many times did I shuffle this deck?

9 1 5 6 10 2 11 7 3 13 12 14 4 15 8 16

twice! This is because we can find

$$[9]\ (1)\ \{5\}\ \{6\}\ [10]\ (2)\ [11]\ \{7\}\ (3)\ \boxed{13}\ [12]\ \boxed{14}\ (4)\ \boxed{15}\ \{8\}\ \boxed{16}$$

4 rising sequences, and each shuffle yields two new rising sequences assuming the shuffle is non-trivial

(b) **Definition:** A rising sequence is a maximal increasing, sequential progression

(c) **Remark** Any $\pi \in S_n$ can be decomposed (partitioned) into rising sequences

(d) **Theorem:**

$$\mathbb{P}(\text{a-shuffle gives } \pi \text{ with r rising sequences}) = \frac{\binom{a+n-r}{n}}{a^n}$$

**Proof:** Given pile sizes, then there is at most 1 way to make $\pi$. So, we need a way to calculate

$$\#\{\text{piles sizes: they can lead to } \pi\}$$

We require $a \geq r$ from the onset. If $a = r$, then there is only 1 such permutation. If $a > r$, then we need to count the number of ways to refine $r$ rising sequences into $a$ piles

(e) FACT: $r(\pi) =$ (number of rising sequences in $\pi$) is also a markov chain under shuffling.

9. Asymptotics

(a) **Theorem:** If $n$ cards are shuffled $m$ times, then if $m = \frac{3}{2}\log_2 n + \theta$, then

$$||Q_m - U||_{TV} = 1 - 2\Phi\left(\frac{-2^{-\theta}}{4\sqrt{3}}\right) +)(n^{-1/4})$$

# 11/1: Kevin Yang's talk on Random Matrix Theory

1. Wigner's Matrix Ensemble

(a) We have $H = (h_{ij})_{i,j=1}^N$, real symmetric

(b) $\{h_{ij}\}$ is independent up to $h_{ij} = h_{ji}$

(c) $E(h_{ij}) = 0$ with $Eh_{ij}^2 \sim \frac{1}{N}$ and

$$\sup_{i,j \leq N} E|h_{ij}|^{4+\epsilon} < \infty$$

(d) We also have

$$\sum_j E|h_{ij}|^2 = 1$$

(e) Ex: (GOE) for $h_{ij} \overset{iid}{\sim} N(0, 1/N)$ with $i \neq j$ and $h_{ii} \overset{iid}{\sim} N(0, 2/N)$ for $i = 1, \ldots, N$. Then FACT: GOE is invariant under orthogonal conjugation

2. Wigner semicircle law

(a) We want to look at

$$\rho_{sc}(x)dx = \frac{1}{2\pi}\sqrt{(4-x^2)_+}dx$$

(b) We want to look at the following measure

$$\mu_N = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j}$$

for $\lambda_1 \leq \cdots \leq \lambda_N$ eigenvalues of $H$ and $\delta_x =$ dirac mess at $x$

(c) Note: $\mu_N$ is a random probability measure on $\mathbb{R}$

(d) **Theorem:**

$$\int_{-\infty}^{\infty} f d\mu_N \overset{N\to\infty}{\to} \int_{-\infty}^{\infty} f(x)\rho_{sc}(x)dx$$

**Proof:** Compute moments of LHS and they show that they converge to the moments of $\rho_{sc}(x)dx$, i.e. take $f(x) = x^k$ for $k \geq 0$ and show convergence. Via Weierstrass approximation, this will give us the desired weak convergence, but we need some **tightness**, i.e. need to ensure that the eigenvalues are not too big with high probability

3. Work by Dyson, Gaudin, Mehta

(a) They explcitly computed the gap statistics, i.e. looking at the random variables $N(\lambda_i - \lambda_j)$, for GOE (Gaussian Orthogonal Ensembles)

(b) Others did the same for invariant ensembles

(c) Johansson in 2001: For $H = H_0 + aW$ for $H_0$ random, $a > 0$, and $W$ is a GOE, then the gap statistics are equal to those of the GOE. Here all of the above matrices are complex hermitian

4. Dyson Brownian Motion

(a) Dyson Brownian Motion is a modified Brownian motion

(b) First realize that the real symmetric space is a real dimensional space, so we can do a brownian motion on it

(c) Then we have

$$dH(T) = \frac{1}{\sqrt{N}} dB(t) - \frac{1}{2} H(T) dT$$

(d) We define $B(T) = (B_{ij}(T))$ where $B_{ij}(T) \overset{iid}{\sim}$ Brownian motions for $i \neq j$ (up to symmetric) and $B_{ii}(T) \overset{iid}{\sim} \frac{1}{\sqrt{2}}$Brownian Motions

(e) $B(T)$ is an evolving path in the space of real symmetric matrices, so the eigenvalues are evolving as well

(f) **Proposition:**

$$d\lambda_j(T) = \sqrt{\frac{2}{N}} dB_j(T) + \left( -\frac{\lambda_j}{2} + \frac{1}{N} \sum_{\ell \neq j} \frac{1}{\lambda_j - \lambda_\ell} \right) dT$$

**Proof:** Perturbation theory from QM and Ito's formula and some stochastic analysis. □

(g) Dyson's conjecture:

i. globally, the above relaxes on an $O(1)$ time-scale

ii. The gap statistics converge on an $O(N^{-1})$ time-scale

5. Bakery-Emory Theory

(a) For our set up, the generator is

$$L = \frac{1}{N} \Delta - v \cdot \nabla, \qquad v_j = \frac{x_j}{2} - \frac{1}{N} \sum_{\ell \neq j} \frac{1}{x_j - x_\ell}$$

(b) **Lemma:** $v_j = \partial_j H$ and $H = \frac{1}{2} \sum_{j=1}^N x_j^2 - \sum_{i \neq j} \log |x_i - x_j|$

(c) What is the generator?

$$\mu_\infty(dx) = e^{-NH}/Z \; dx$$

(d) Fact: if $f_0 \geq 0$ and $\int_{f_0} d\mu_\infty = 1$, then $\partial_t f_t = L f_t$ is the probability density w.r.t. $\mu_\infty$ under the Dyson Brownian Motion

(e) **Definition:**

$$H(T) = \int f(T) \log f(T) d\mu_\infty, \qquad D(T) = \frac{1}{N} \int |\nabla f(T)|^2 d\mu_\infty$$

(f) **Theorem:** $N \nabla^2 H \geq Nk > 0$ for $k > 0$

(g) Consequence: $H(T) \leq N D(T)$ and $H(T) \leq e^{-k'T} H(0)$ and $D(T) \leq e^{-k'} T D(0)$

(h) This tells us that the eigenvalues can't get too close to each other, so in the convergence of the (normalized) eigenvalues to the semicircle, each eigenvalue stays in some small subsection of the semicircle

## 11/8: Huy Pham's talk on Clumpy and Dealer Shuffles

1. Given $p \in (0,1)$, generate a binary sequence $B$ of length $n$ following the markov chain

$$\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

2. We split cards to 2 decks, the 0-deck has size equal to the number of 0's in our binary string, and the 1-deck has size equal to the number of 1's in our string, $B$.

3. We go through $B$, if we see a 1, we drop the bottom card from the 1-deck, and if we see a 0 we drop the bottom card from a 0-deck

4. If $p = 1/2$, then this is the standard riffle shuffle which Mark told us about

5. If $p > 1/2$, then we call this a **clumpy** shuffle

6. If $p < 1/2$, this is called the dealer's shuffle because we're more likely to alternate, which is not standard in the riffle shuffle

7. Example: Take cards $\{1, 2, 3, 4, 5, 6\}$ and $B = 000110$, then we have two decks

$$\{1, 2, 3, 4\},\ \{5, 6\}$$

and our shuffle will be $\{1, 2, 3, 5, 6, 4\}$

8. Now let $p \in (0,1)$ be fixed

9. Call

$$\pi_t = \sigma_t \sigma_{t-1} \cdots \sigma_1 \pi_1$$

such that $\sigma_i$ is iid according to the shuffle alogirthm

10. define $\tau_{mix} = \inf\{t\ :\ ||\pi_t - U||_{TV} \leq \frac{1}{4}\}$ where $U$ is the uniform measure

$$||P||_{TV} = \frac{1}{2} \sum_{\pi \in S_n} |P(\pi)|$$

is the total variation

11. **Theorem:**

$$\tau_{mix}(p) \leq C_p (\log n)^4$$
$$\tau_{mix}(p) \geq \frac{\log n - O(1)}{|\log p|}$$

12. Time reversal: If the shuffle $\pi_{t+1} = \sigma_{t+1}\pi_t$, then the backward shuffle is defined by $\tilde{\pi}_{t+1} = \tilde{\sigma}_{t+1}^{-1}\tilde{\pi}_t$. Note that $\tilde{\sigma}_{t+1}$ has the same distribution as $\sigma_{t+1}$

13. **Proposition:** $||\pi_{t+1} - U||_{TV} = ||\tilde{\pi}_{t+1} - U||_{TV}$

14. Time reversal of $p$-riffle shuffle: generate a binary sequence $B$ according to Markov Chain, $M_p = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$, then draw all cards with a 0 in t he sequence our and place them at the bottom (in order). Draw all cards with 1 and place on top

15. Proof of lower bound on mixing time: Call $(i, i+1)$ an adjacent pair, if $\pi(i+1) = \pi(i) + 1$. For each pairs of cards $(i, i+1)$, they will be adjacent if $MC_p$ did not change state, which happens with probability $p$. After $t$ steps, $(i, i+1)$ remains an adjacent pair with probability $p^t$.

16. The number of adjacent pairs in $\pi_t$ is distribution $Bin(n-1, p^t)$. Choose

$$t = \frac{\log n - C}{|\log p|}$$

Hence

$$P(\# \text{ of adjacent pairs } \leq \frac{e^C}{4C}) \leq \frac{1}{8}$$

We also know that for sufficiently large $C$, we get

$$P_U(\# \text{ of adjacent pairs } \leq \frac{e^C}{4C}) > \frac{7}{8}$$

along with the facts that $E_U(\# \text{ of adjacent pairs }) = (n-1)/n$ and $Var_U(\# \text{ of adjacent pairs }) < 2$, then we get

$$||\pi_t - U||_{TV} \geq \frac{3}{4}$$

17. Entropy, Collision and the upper bound

18. **Definition:** $H(X) = -\sum_{x \in supp(X)} p(x) \log p(x)$

19. **Definition:** $H(X \mid Y)(y) = H(X \mid Y = y)$

20. We have $H(X) \leq \log |supp(X)|$

21. **Proposition:** The pinsker inequality says

$$||X - U||_{TV} \leq \sqrt{\frac{1}{2}(H(U) - H(X))}$$

22. **Definition:** (Jensen-Shannon Divergence): For $p, q \geq 0$,

$$d(p,q) = \frac{1}{2} p \log p + \frac{1}{2} q \log q - \frac{p+q}{2} \log \frac{p+q}{2}$$
$$d(X,Y) = \sum_s d(P(x=s), P(y=s))$$

where $(X+Y)/2$ is a distribution where the probability that $s$ appears is

$$\frac{P(x=s)}{2} + \frac{P(Y=s)}{2}$$

23. Note that $d(X,Y) \geq 0$ and $d(X,Y) \leq \log 2$

24. This $d$ function is also convex in the sense that

$$d(pX_1 + (1-p)X_2, Y) \leq pd(X_1, Y) + (1-p)d(X_2, Y)$$

25. **Lemma:** $X, Y$ RVs over $S$, $g : S \to T$, then

$$d(g(X), g(Y)) \leq d(X, Y)$$

26. **Lemma:** We have

$$d(X, U_s) \geq \frac{c}{\log |S|}(\log |S| - H(x))$$

for some universal constant $c$

27. Collisions and the Monte Shuffle

28. **Definition:**

$$c(a,b) = \begin{cases} Id & \text{with probability } \frac{1}{2} \\ (a,b) & \text{with probability } \frac{1}{2} \end{cases}$$

where $(a,b)$ is the permutation switching $a$ and $b$

29. **Definition:** $\pi_t$ is a Monte shuffle if

$$\pi_{t+1} = \nu c(a_1, b_1) \cdots c(a_k, b_k)\pi_t$$

where $\nu$ is a random variable, the $\{(a_i, b_i)\}$ are all disjoint, each $(a_i, b_i)$ can depend on $\pi_t$, but conditional on $\pi_t$ and $\nu$, and the randomness of the collisions are independent.

30. **Theorem:** Let $\pi$ be of the form of a Monte shuffle. Let $T$ be a RV with values in $[1, t]$.

31. For each card $x$, we let $b(x)$ be the first card after time $T$ for which there exists a $c(i, j)$ such that $\pi_{t_1}(x) = i$ and $\pi_{t_1}(y) = j$ and in the definition of $\pi_{t+1}$, we have a $c(i, j)$.

32. Let $m(x) = b(x)$ if $b(b(x)) = x$. Otherwise let $m(x) = x$. Let $A_i$ be such that $P(m(i) = j) \geq \frac{A_i}{i}$ for all $j \leq i$. Then let $\mu$ be a random permutation, independent of $\pi$, with

$$H(\pi_t\mu) - H(\mu) \geq \frac{C}{\log n}\sum_{k=1}^{n} A_k E_k$$

where

$$E_k = \log k - E\left[H(\mu^{-1}(k))(\mu^{-1}(k+1))\mu^{-1}(n)\right]$$

33. We apply this to the $p$-riffle shuffle. Start with a binary sequence $B$, at this point I stopped paying attention.

## 11/15: Andy Tsao's talk on "Bounding Mixing Times via Coupling"

1. **Definition:** A coupling of 2 distributions $(\mu, \nu)$ is a pair $(X, Y)$ defined on the same space such that $X \sim \mu$ and $Y \sim \nu$.

2. **Proposition:** We have

$$||\mu - \nu||_{TV} = \inf\{\mathbb{P}(X \neq Y) \ : \ (X, Y) \text{ is a coupling of } \mu, \nu\}$$

**Proof:** For $A$ some event, then

$$\mu(A) - \nu(A) = P(X \in A) - P(Y \in A) \leq P(X \in A, \ Y \notin A) \leq P(X \neq Y)$$

this proves the left hand side is less than or equal to the right hand side. To get equality, we construct coupled random variables: Flip a coin with probability heads of $1 - ||\mu - \nu||_{TV}$ and then we get two variables from this somehow. □

3. How does this extend to markov chains?

4. A coupling of Markov chains with transition matrix $P$ is a process $(X_t, Y_t)_{t=0}^{\infty}$ such that $(X_t)_{t=0}^{\infty}$, $(Y_t)_{t=0}^{\infty}$ are both markov chains with transition matrix $P$.

5. From our proposition:

$$||P^t(x, \cdot), P^t(y, \cdot)||_{TV} \leq P_{x,y}(\tau_{couple} > t)$$

where $\tau_{couple}$ is the first time that our independent markov chains starting at $x, y$ agree.

6. Now note

$$d(t) = \sup_x ||P^t(x, \cdot) - \pi||_{TV} \leq ||P^t(x, \cdot) - P^\tau(y, \cdot)||_{TV}$$

and then we define

$$t_{mix}(\epsilon) = \min(d(t) < \epsilon) \leq \max_{x,y} \frac{\mathbb{E}\tau_{couple}}{\epsilon}$$

7. Path coupling: why we care
We have a state space, $\mathcal{X}$, a graph, $(\mathcal{X}, E)$ and

$$\Phi(x, y) \geq \delta(x, y)$$

is a potential function which is greater than

$$\delta(x,y) = \begin{cases} 1 & x, y \text{ are adjacent} \\ 0 & \text{not adjacent} \end{cases}$$

where a path $\gamma = (x_0, x_1, \ldots, x_r)$ where $(x_{i-1}, x_i)$ adjacent and

$$\rho(x,y) = \min\{\Sigma_{i=1}^{r} \Phi(x_{i-1}, x_i) \ : \ (x_0, \ldots, x_r) \text{is a path from } x \to y\}$$

8. For any pair of Random variables, $X$ and $Y$, then $\mathbb{P}(X \neq Y) = \mathbb{E}\mathbb{1}_{X \neq Y} \leq \mathbb{E}\rho(X,Y)$. If $(X,Y)$ is a coupling of $\mu, \nu$ then

$$||\mu - \nu||_{TV} \leq \mathbb{P}(X \neq Y) \leq \mathbb{E}\rho(x,y)$$

and

$$\rho_k(\mu, \nu) = \inf\{\mathbb{E}\rho(X,Y) \ : \ (X,Y) \text{ is a coupling of } \mu, \nu)\}$$

9. Path coupling: Suppose that, for each edge $(x,y) \in E$, there exists a coupling $(X_1, Y_1)$ of the distributions $P(x, \cdot)$ and $P(y, \cdot)$ such that

$$\mathbb{E}_{x,y}\rho_K(X_1, Y_1) \leq e^{-\alpha}\rho(x,y)$$

Then for any measures, $\mu$ and $\nu$, we have

$$\rho_k(\mu P, \nu P) \leq e^{-\alpha}\rho_k(\mu, \nu)$$

Consequently

$$t_{mix}(\epsilon) \leq \lceil \frac{\log(D\epsilon^{-1})}{\alpha} \rceil$$

where $D$ is the diameter of $G$, defined to be

$$D = \max_{x,y} \rho(x,y)$$

10. Sampling independent sets via Glauber

11. define the **Hard-Core** measure

$$\mu_G(\sigma) = \frac{\lambda^{|\sigma|}}{Z_G}$$

The dynamics are as follows: Starting from an independent set $\sigma$,

   (a) Choose a vertex, $v$, of $G$ uniformily at random
   (b) Flip a coin with probability of heads equal to $\lambda/(1+\lambda)$, if heads then add $v$, if tails, remove $v$
   (c) Then make the transition $\sigma \mapsto \sigma'$ if $\sigma' \in Ind(G)$

12. We check $\mu_G(\sigma)$ satisfies detailed balance: If $\sigma' = \sigma \sqcup \{v\}$

$$\frac{\lambda^{|\sigma'|}}{Z_G}P(\sigma, \sigma') = \frac{\lambda^{|\sigma|+1}}{Z_G}\frac{1}{|v|}\frac{1}{\lambda+1} = \frac{\lambda^{|\sigma|}}{Z_G}\frac{1}{|v|}\frac{\lambda}{\lambda+1} = \frac{\lambda^{|\sigma|}}{Z_G}P(\sigma, \sigma')$$

13. Our state space $\mathcal{X}$ = collection of independent sets, then

$$(\sigma_1, \sigma_2) \in E \quad \text{ if } \sigma_1 \neq \sigma_2, \quad \sigma_2 = \sigma_1 \cup \{v\}, \quad \text{or} \quad \sigma_2 = \sigma_1 \backslash \{v\}$$

14. We have additional assumptions

   (a) Maximum degree is $\delta$ for some $\delta$
   (b) $\lambda < \frac{2}{\delta - 2}$ with $(\lambda = (1-\alpha)\frac{2}{\delta-2})$
   (c) $G$ is triangle-free

15. **Theorem:** (Main theorem) Under assumptions

$$t_{mix}(\epsilon) \leq O(n \log(n/t))$$

where the actual constants are $\frac{4\delta n}{\alpha} \log \frac{3n\delta}{t}$

16. The coupling is: starting from states $\sigma_1$, $\sigma_2$ try to apply some update

$$\Phi(\sigma, \sigma_v) = \delta_v - c|B(\sigma, v)|$$

where $\delta_v$ means the degree of $v$ where $B(\sigma, v)$ is the set of blocked vertices from $\sigma$, i.e.

$$w \in B(\sigma, v) \iff w \in \Gamma(v), \qquad \sigma \cup \{w\} \text{ is not an independent set}$$

where $\Gamma(v)$ denotes the set of all neighbors of $v$

17. Now we look at

$$\mathbb{E}\Delta^{+x} = \mathbb{E}(\Delta\Phi \mid \text{MC attempts to add x})$$
$$\mathbb{E}\Delta^{+x} = \mathbb{E}(\Delta\Phi \mid \text{MC attempts to remove x})$$
$$\mathbb{E}[\Delta^{+x}\Phi] = \frac{\lambda}{1+\lambda}\mathbb{E}\Delta^{+x}\Phi + \frac{1}{1+\lambda}\mathbb{E}\Delta^{-x}\Phi$$

The cases are as follows:

(a) MC transitions at $v$
(b) MC transitions at $w \in \Gamma(V)$
(c) MC transitions at $x$, a neighbor of a neighbor of $v$

The point is that the no triangles in the graph condition means that the last two cases are distinct

18. We have

(a) In this case, $E\Delta^V\Phi = -\delta_v + c|B(\sigma, v)|$
(b) If $w \in \Gamma(v)$, then $w \notin \sigma$, $w \notin \sigma_v$, then

$$\mathbb{E}\Delta^{+w}\Phi = ???$$

If $w$ is blocked, then cannot add $w$ either $\sigma$ or $\sigma_v$

$$\implies \mathbb{E}\Delta^{+w}\Phi = 0$$

If $w$ is not blocked, then $\sigma \cup \{w\} = \sigma_w$.

$$\Delta^{+w}\Phi = \Phi(\sigma_w, \sigma_v) - \Phi(\sigma, \sigma_v) \leq \Phi(\sigma_w, \sigma) + \Phi(\sigma, \sigma_v) - \Phi(\sigma, \sigma_v) = \Phi(\sigma, \sigma_w)$$

Then

$$E\Delta^w\Phi = \frac{\lambda}{1+\lambda}(\delta_w - c|B(\sigma, w)|) \quad \text{if w is not blocked}$$

(c) More cases if: a) $x$ is in both independent sets, a change of $\Phi$ from removing $x$ comes from unblocking $w \in \Gamma(v)$ and b) $x$ is in neither independent set

## 11/22: Max Xu's talk on "Mixing time of Markov Chain $AX + B$ in $\mathbb{F}_p$ and its connection to number theory"

1. Let $p$ be a prime (integers). Consider $X_{n+1} = aX_n + b_n$ in $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for $a \neq 0$

2. $b_n$'s are i.i.d. with law (i.e. distributed by) $|mu|$ with $|supp(\mu)| \geq 2$, e.g. $\mu(1) = \mu(0) = \mu(-1) = 1/3$

3. Question: how fast does $X_n \to U$, the uniform distribution

4. Mixing time: $T(\delta) = \inf(\{n \ : \ ||\mu_a^n - U||_{TV} \leq \delta\})$

5. Here, we have

$$||\pi_1 - \pi_2||_{TV} = \sup_{A \subseteq \mathbb{F}_P} |\pi_1(A) - \pi_2(A)| = \frac{1}{2}||\pi_1 - \pi_2||_1$$

6. We want to bound $T(S)$ in terms of $p$

7. Specific SettingL (Chung, Diaconis, Graham paper setting): If we fix $a$, then what conditions on it tell us that the mixing time is small or large. If $p$ important? Note that $p$ is not necessarily prime

8. Global setting: What happens if $a$ is arbitrary in $\mathbb{F}_p$

9. **Theorem:** : (Konyagin; Bukh; Harper; Helfgott) - suppose $a$ has multiplicative order $a$ at least $c \log p(\log\log p)$, then for every $\delta \in (0, 1/2)$, we have

$$T(\delta) \leq \frac{c}{1 - ||\mu||_{L^2}^2}\left(\log(1/\delta) + (\log(p))^2(\log\log p)^5\right) = O((\log p)^{2+o(1)})$$

10. **Theorem:** 2 (observation) If one can prove the following statement (which is open), then the upper bound will be $O(\log p)$: For
    (a) $a$ with sufficient large multiplicative order and
    (b) most primes $p$

11. If the above statement is true, then this will prove Lehmer's conjecture

12. Lehmer's conjecture: For all $p(x) \in \mathbb{Z}[x]$, there exists a constant $c > 1$ such that one of the following is true
    (a) $\mu(p(x)) \geq c > 1$
    (b) $\mu(p(x)) = 1$

    here $\mu(p(x)) = |a_0| \prod_{i=1}^{n} \max\{1, |\alpha_i|\}$, where

$$p(x) = a_0(x - \alpha_1) \cdots (x - \alpha_n)$$

13. **Theorem:** For almost all $a$, we have
    (a) The $O(\log p \log\log p)$ bound holds for all primes
    (b) $O(\log p)$ holds for "most" primes
    (c) The Markov chain has a cut-off phenomena at $n = \frac{\log p}{H(\mu)}$, if the general riemann hypothesis is true. This means that the markov chain is close to uniform after this number of steps

$$|T(\delta) - \frac{\log p}{H(\mu)}| \leq C_{\epsilon,\mu}\sqrt{\log p}$$

    In the second point, the notion of "most" primes is somehow different than a density in the limit as $n \to \infty$ notion of "most"

14. **Theorem:** : For $X_{n+1} = X_n + b_n$ (i.e. $a = 1$) and $\mu(1) = \mu(-1) = \mu(0) = \frac{1}{3}$, then there exists $\alpha, \beta$ such that $e^{-\alpha N/p^2} \leq ||\mu^{(N)} - U|| \leq e^{-\beta N/p^2}$.
    **Proof:** We write

$$\hat{\mu}(j) = \sum_{k \in \mathbb{F}_p} e(jk/p)\mu(k) \qquad e(x) = e^{2\pi i x}$$

$$\mu(k) = \frac{1}{p}\sum_{j} e(-jk/p)\hat{\mu}(j)$$

$$\frac{1}{p}\sum_{j \in \mathbb{F}_p} |\hat{\mu}(j)|^2 = \sum |\mu(k)|^2$$

Now

$$||\mu^{(N)} - U||^2 = \frac{1}{4}\left(\sum_k |\mu(k) - U(k)|\right)^2 \le \frac{1}{4}p\sum_k |\mu(k) - U(k)|^2 = \frac{1}{4}\sum_{j\neq 0} |\hat{\mu}(j)|^2$$

$$= \frac{1}{4}\sum_{j\neq 0} \hat{\mu}^{2N}(j) \qquad \text{using that } \mu \hat{*} \nu = \hat{\mu}\cdot\hat{\nu}$$

$$= \frac{1}{4}\sum_{j\neq 0}\left(\frac{1}{3} + \frac{2}{3}\cos\left(\frac{2\pi i}{p}\right)\right)^{2N}$$

with some more work, we can get the desired bound. □

15. **Theorem:** We have the upper bound of $C\log p \log\log p$ for all $p$, where the markov chain is

$$X_N = \sum_{i=0}^{N-1} 2^{N-1-j} b_j = \sum_{j=0}^{N-1} 2^j b_j$$

**Proof:** (sketch) mess with convolutions, get another product of cosines, and bound this product using step functions and a binary expansion trick. □

16. Remark: when $p = 2^t - 1$, then the bound of $\log p \log\log p$ is sharp!

17. **Theorem:** When $a = 2$, we have that for almost all odd $p$, if

$$N \ge \frac{\log p}{\log(9/5)} + c$$

then $||\mu^{(N)} - U|| = O((5/9)^{C'/2})$

18. Consider

$$P(X) = b_0 X^{n-1} + \cdots + b_{n-2}X + b_1$$

where $b_2 \in Z$ and i.i.d. with law $\mu$, then $P(a)$ has law $\mu_a^( n)$

## 12/6: Andrea Ottolini's Talk on Random Walks and Cut Offs for Mixing Times

1. Set up is: $G_N = (V_N, E_N)$ for $|V_N| = N$

2. For each $e \in E_N$, put an exponential clock with norm 1

3. When a clock rings, exchange the cards (i.e. the vertices that the edge connects is represented by a card)

4. $\nu$ is the initial distribution on $S_N$

5. $P_t^\nu$ is the distribution after time $t$ starting from $\nu$

6. In this talk, we consider $G_N$ to be the graph $0 \to 1 \to 2 \to \cdots \to N$

7. $\mu =$ uniform distribution on $N$

8. $d^N(t) = \max_\nu ||P_t^\nu - \mu||$

9. $T_{mix}^N(\epsilon) = \inf\{t > 0 \ : \ d^N(t) \le \epsilon\}$

10. **Theorem:**

$$\lim_{N\to\infty} \frac{T_{mix}^N(\epsilon)}{\frac{N^2\log N}{2\pi^2}} = 1$$

11.

$$||P_{\frac{N^2\log N}{2\pi^2}(1+\delta)} - \mu|| \to \begin{cases} 1 & \delta < 0 \\ 0 & \delta \ge 0 \end{cases}$$

12. Given $\sigma \in S_N$, define $\tilde{\sigma} : \{0, \ldots, N\} \to \mathbb{R}$ by

$$\tilde{\sigma}(x, y) = \sum_{z=1}^{x} \mathbb{1}_{\sigma(z) \leq y} - \frac{xy}{N}$$

this is like counting minus a normalized area

13. As an example, check the counts when $\sigma = (2\ 3)$ on $\{1, 2, 3\}$ and look at the map

$$x \mapsto \sum_{z=1}^{x} \mathbb{1}_{\sigma(z) \leq y}$$

14. Note we can define a partial order on permutations by

$$\sigma \leq \sigma' \iff \tilde{\sigma}(x, y) \leq \tilde{\sigma}'(x, y)$$

15. Given an update $\sigma \to \sigma(x\ x+1)$, then $\tilde{\sigma}(x, y)$ is affected if and only if $\tilde{\sigma}(\cdot, y)$ has a max/min at $x$

16. We want to show that

$$f(x, t;\ y) = \mathbb{E}[\tilde{\sigma}_t(x, y)]$$

satisfies some sort of discrete heat equation

17. We do this is by computing

$$\frac{\partial}{\partial t} f(x, t;\ y) = \mathbb{E}[\tilde{\sigma}_t(x+1), y] + \tilde{\sigma}(x-1, y) - 2\tilde{\sigma}_t(x, y) =: \Delta_x f$$

the idea is that we compute that variation of $f$, and this will be 0 unless we have a maximum or minimum, in which case the paranthesized expression captures the variation w.r.t. to $t$. We also have the initial conditions

$$f(0, t;\ y) = 0 = f(N, t;\ y), \qquad f(x, 0;\ y) = \mathbb{E}[\tilde{\sigma}_0(x, y)]$$

18. **Lemma:** We have

$$\max_{x \in \{0, \ldots, N\}} f(x, t, y) \leq 2 \min(y, N - y) e^{-\lambda_N t}$$

where $\lambda_N$ is the last eigenvalue of the laplacian and

$$\lambda_N = 2(1 - \cos(\pi/N)) = \frac{\pi^2}{N^2}(1 + o(1))$$

Also if $\tilde{\sigma}_0 = \delta_{\tilde{Id}}$ then

$$f(x, t;\ y) \geq \frac{\min(y, N-y)}{\pi} \sin\left(\frac{\pi x}{N}\right) e^{-\lambda_N T}$$

19. Now start from the identity, and look at the number of red cards in the first half of the deck - take an average

20. Under the uniform distribution, $f_\mu(\frac{N}{2}, t;\ \frac{N}{2}) = 0$ and $\tilde{\sigma}_t^\mu(\frac{N}{2}, \frac{N}{2}) = \sqrt{N} Z(0, 1)$

21. Then the idea is we can bound

$$f\left(\frac{N}{2}, t, \frac{N}{2}\right) \geq c N^{1/2 + \delta}$$

for $t = \frac{N^2 \log N}{2\pi^2}(1 - \delta)$

22. **Definition:** A censoring scheme is a cadlog function $C : \mathbb{R}^+ \to P(\{1, \ldots, N+1\})$ (where $P$ denotes the power set). If $x$ is selected, then do the update iff $x \in C(t)$

23. **Theorem:** $||P_t^{\nu, C} - \mu||_{TV} \geq ||P_t^\nu - \mu||_{TV}$ if $\nu$ is increasing (here, we're using the order on the permutations)

24. Fix $K$, define the semi-skeleton, $\hat{\sigma} : \{0, \ldots, N\} \times \{0, K\} \to \mathbb{R}$ such that $(x, i) \mapsto \tilde{\sigma}(x, x_i)$ and $x_i = \lceil \frac{iN}{K} \rceil$ and the skeleton $\overline{\sigma} : \{0, \ldots, K\} \times \{0, \ldots, K\} \to \mathbb{R}$ such that $(i, j) \mapsto \tilde{\sigma}(x_i, x_j)$

25. As an example: $K = 2$, then $\hat{\sigma}$ encodes the position of the red cards and $\overline{\sigma}(1,1) =$ the number of red cards in the first half

26. Given $\nu$ a measure on $\wedge\sigma$, define

$$\hat{\nu} = \text{push forward of } \nu \text{ under the map } \tilde{\sigma} \to \hat{\sigma}$$
$$\overline{\nu} = \text{push forward of } \nu \text{ under the map } \tilde{\sigma} \to \overline{\sigma}$$
$$\tilde{\nu} = \text{measure on } S_N\text{; given } \overline{\nu}\text{, all ``something'' are equally likely}$$

27. Fix $\epsilon > 0$, $\delta > 0$ and $N$ large. Assume

$$||P_{t_3}^{\nu} - \mu||_{TV} \leq \epsilon$$

and

$$t_1 = \frac{\delta}{3}\frac{N^2}{2\pi^2}\log N$$
$$t_2 = (1 + 2/3\delta)\frac{N^2}{2\pi^2}\frac{N^2}{2\pi^2}\log N$$
$$t_3 = (1 + \delta)\frac{N^2}{2\pi^2}\frac{N^2}{2\pi^2}\log N$$

28. Fix $K = \lfloor \frac{1}{\delta} \rfloor$, run the censoring scheme where in $[0, t_1] \cup [t_2, t_3]$ all moves at $x_i$ for $i = 1, \ldots, k$ are censored. Then

$$||P_{t_3}^{\nu} - \mu||_{TV} \leq ||P_{t_3}^{\nu,C} - \mu|| \leq ||P_{t_1}^{\nu,C} - \tilde{P}_{t_1}^{\nu,C}|| + ||\tilde{P}_{t_3}^{\nu,C} - \mu||_{TV} \leq \sum_{i=1}^{K} ||\nu_t^i - \mu_t^i||_{TV}$$

where $||\tilde{P}_{t_3}^{\nu,C} - \mu||_{TV} = ||\tilde{P}_{t_3}^{\nu,C} - \hat{\mu}||_{TV}$