Algorithms, 2023 Fall, Homework 13
(Due: Dec 27)

December 18, 2023

Reading:

Problem 1 : Suppose $n \geq 2$ and $H = (V, E)$ be an n-uniform hypergraph (i.e., each hyperedge contains exactly n-vertices) with $|E| = 4^{n-1}$. Show there is a coloring of V by four colors so that no edge is monochromatic. (Hint: Consider a random coloring.)

Problem 2 : KT book pp 703, problem 2.

Problem 3 : This problem relates to the algorithm for solving the distinct element problem. Suppose we have N points uniformly distributed in the unit interval $[0, 1]$. Suppose they are $0 < x_1 < x_2 < \ldots < x_n < 1$ (with probability 1, no two points have the same position). Let $t = 0.2N$ be an integer and ϵ a sufficiently small positive constant (say 0.001). Show that with sufficiently large N (as a function of ϵ) and probability 0.99,

$$x_t \in (1 \pm \epsilon)0.2.$$

Problem 4 : Suppose P is the transition matrix of a Markov Chain with N states. We further assume P is symmetric $P_{ij} = P_{ji}$. Show that 1 is the largest eigenvalue of P, $\frac{1}{\sqrt{N}}(1, 1, \ldots, 1)$ is the corresponding eigenvector, and $\frac{1}{N}(1, 1, \ldots, 1)$ is the stationary distribution.

(Hint: there are multiple ways to prove this. One way to do this is as follows: first show that such a transition matrix P can be decomposed into a convex combination of permutation matrix. A permutation matrix is a matrix where each row and each column has exactly one being '1' and the rest '0'. Also note that the maximum eigenvalue of P is $\sup_{\|x\|_2=1} \|Px\|_2$ (this is the variational definition of the maximum singular value of P, which coincides with the maximum eigenvalue of P since P is symmetric). Hence, you need to show that $\|Px\|_2 \leq \|x\|_2$ for any vector x.)

Problem 5 : Total variational distance is an important distance measuring the difference of two probability distributions. We are given two discrete distributions p and q (represented as d-dim vectors). It is defined as follows:

$$d_{tv}(p, q) = \frac{1}{2}\|p - q\|_1.$$

Prove that the following two other definitions are equivalent to the above one: (1)

$$d_{tv}(p,q) = \sup_{S \subseteq [d]} |\sum_{i \in S} p_i - \sum_{i \in S} q_i|.$$

(2)

$$d_{tv}(p,q) = \inf_J \mathbb{E}_{(x,y)\sim J}[1(x,y)].$$

where $1(x,y) = 1$ if $x \neq y$ and 0 otherwise. Here inf is over all joint distributions of (x,y) such that the marginal distribution of x is the same as p and the the marginal distribution of y is the same as q. (for discrete case, such joint distribution J can be represented as a matrix such that the vector of row sums is p and the vector of col sums is q.)

sidenote: if we replace the distance $1(x,y)$ with other distance such as $|x-y|$, we will get another popular probability distance measure, called *earth mover distance*, or *transportation distance*, or *Wassenstein distance*.

Problem 6 : Recall in linear algebra, a symmetric matrix P has an eigendecomposition as follows:

$$P = \sum_i \lambda_i u_i u_i^T = U \Sigma U^T.$$

where Σ is the diagonal matrix $diag(\lambda_1, \lambda_2, \dots, \lambda_N)$ and $U = (u_1, \dots, u_N)$.

Now, we prove the statement I made in the class about the relation between mixing time and the eigenvalue gap of the transition matrix:

$$\pi_\epsilon(P) \leq O\left(\frac{\log N + \log 1/\epsilon}{1 - \lambda_{\max}}\right)$$

where $\lambda_{\max} = \max\{|\lambda_2|, |\lambda_N|\}$ and $\pi_\epsilon(P)$ is the mixing time defined as

$$\pi_\epsilon(P) = \max_a \min\{t \mid d_{tv}(p_t^a, \bar{p}) \leq \epsilon\}.$$

where p_t^a is the state distribution (a vector) at time t (suppose we start from state a) and $\bar{p}$ is the stationary distribution $(1/N, \dots, 1/N)$ (we assume $(1/N, \dots, 1/N)$ is the only stationary distribution for this Markov chain. Please check the reference I provided for more information about when this assumption is true).

Hint: Suppose p_0^a is the initial state distribution (it is an indicator vector). Decompose p_0^a as follows: (Suppose P is of full rank and hence $\{u_i\}_i$ is an orthonormal basis)

$$p_0^a = \sum_i \alpha_i u_i$$

where $\alpha_i = u_i^T p_0^a$. Derive a nice formula for p_t^a and analyze its distance from the stationary distribution $\bar{p}$.

sidenote: The above theorem shows that the Markov Chain mixes fast if the gap between the largest and second largest eigenvalues is non-negligible. Try to make some sense of it by deriving the 2nd eigenvalue of P when the Markov chain contains two disjoint connected components.

In fact, the theory of Markov chain has a continuous analogue. We can define Markov process in continuous space such as $\mathbb{R}^d$. There, the state probability vector will be replaced by probability density functions, and the transition matrix will be replaced by a differential operater. Analogous to discrete random walk, you may need to learn Brownian motions and related concepts to study random walks in continuous spaces. The above result relating the mixing time and eigengap of transition matrix also has continuous analogue: the eigengap will be captured by the famous *Poincare inequality*. These tools are very useful in various optimization and machine learning contexts, such as Bayesian inference, MCMC, SGD, diffusion process etc. I will teach some of these in my graduate course "Advanced Theoretical Computer Science".

Bonus problems: Bonus problems are problems that may be (may be not) more challenging than other problems. You do not lose points if you do not answer bonus problems. But you will get extra points if you get correct answers for such problems. **To make your life easier, we give you two weeks for the bonus problems**).

Problem 1 : Suppose we sample n points independently and uniformly from the unit interval $[0, 1]$. These n points partition the unit interval into $n + 1$ pieces. What is the expected length of the shortest piece? What is the expected length of the longest piece? Provide your answer and the proof. (asymptotic answers, i.e., $O()$, would suffice).