CS294-179, Spring 2025 Homework #5, Due 4/11

1 Cut-Norm and Approximation by Step-functions [6 Points]

Recall that given a function $W : [0,1]^2 \to \Re$ and a partition $P = (Y_1, \ldots, Y_k)$ of [0,1] into disjoint sets, we define W_P as the step-function which is equal to $\beta_{ij} = \frac{1}{|Y_i||Y_j|} \int_{Y_i \times Y_j} W$ on $Y_i \times Y_j$. Also recall the definition of the cut-norm, $\|W\|_{\Box} = \sup_{S,T \subset [0,1]} \left| \int_{S \times T} W \right|$. We say that U is a step function on $P \times P$ if U is constant on all sets of the form $Y \times Y'$ with $Y, Y' \in P$.

a) Show that the map $W \mapsto W_P$ is a contraction with respect to the L_2 norm, i.e., show that

$$\|W_P\|_2 \le \|W\|_2.$$

Hint: Consider $||W - W_P||_2^2$ and then use that $\int W_P W = \int W_P^2$.

b) Show that in the L_2 -norm, the step function W_P is the best possible step-function approximation to W on $P \times P$, i.e., show that

 $||W - W_P||_2 \le ||W - U||_2$ for all step functions U on $P \times P$.

c) Show that when calculating the cut-norm of $||W_P||_{\Box}$, you can restrict yourself to sets S and T which are unions of classes in P, i.e., show that

$$||W_P||_{\Box} = \max_{S,T \in \sigma(P)} \left| \int_{S \times T} W \right|,$$

where $\sigma(P)$ is the set of subsets $S \subset [0,1]$ of the form $S = \bigcup_{i \in I} Y_i$ for some $I \subset [k]$. Hint: Consider a fixed set T, and the function $f_T(x) = \int_T dy W_P(x,y)$ and let S_{\pm} be the sets where f < 0 and f > 0, respectively. Show that for all $S \subset [0,1]$,

$$\left| \int_{S} f_T(x) \right| \le \max\left\{ \int_{S_+} f_T, -\int_{S_-} f_T \right\}$$

to prove that S can be chosen in $\sigma(P)$. After that, fix $S \in \sigma(P)$ and repeat the argument for T.

d) Use (c)) to show that the map $W \mapsto W_P$ is a contraction with respect to the cut-norm, i.e., show that

$$||W_P||_{\square} \le ||W||_{\square}.$$

e) Use (d) and the triangle inequality to prove that up to a factor of two, the step function W_P is the best possible step-function approximation to W on $P \times P$ with respect to the cut-norm:

$$||W - W_P||_{\Box} \le 2||W - U||_{\Box}$$
 for all step functions U on $P \times P$

Hint: Use that $U_P = U$.

2 Concentration once more [4 Points]

We will need the following generalization of the Chernoff bound for i.i.d. Bernoulli random variables: Let X_1, \ldots, X_N independent r.v. with $X_i \sim Be(p_i)$ and let $n_i \in \{0, 1, 2\}$ be non-random.

a) Prove that

$$\Pr\left(\sum_{i} n_i (X_i - p_i) \ge 2\delta \sum_{i} p_i\right) \le e^{-\frac{\delta^2}{2\delta + 2}\sum_{i} p_i}.$$

Hint: Bound $E[e^{tn_i(X_i-p_i)}]$ by $\exp((e^{n_it}-1)p_i-tn_ip_i) \le \exp((e^{2t}-1-2t)p_i)$.

b) In a similar way, show that

$$\Pr\left(\sum_{i} n_i (X_i - p_i) \le -2\delta \sum_{i} p_i\right) \le e^{-\frac{\delta^2}{2}\sum_{i} p_i}.$$

c) Use these two bounds to show that

$$\Pr\left(\left|\sum_{i} n_i (X_i - p_i)\right| \ge 2\epsilon N\right) \le 2e^{-\frac{\epsilon^2}{4}N}.$$

Hint: For $\epsilon > 1$, the bound is trivial (why?), so you may assume without loss of generality that $\epsilon \leq 1$.

3 G(n,p) converges to $W \equiv p$ [**3** Points + 1 Bonus Point]

Recall that the empirical graphon of a graph with $n \times n$ adjacency matrix A (or more general, any $n \times n$ matrix A) is defined as

$$W_A(x,y) = \sum_{i,j=1}^n A_{ij} \mathbb{I}_{x \in I_i, y \in I_j}$$

where I_1, \ldots, I_n is a partition of [0, 1] into adjacent intervals of width 1/n, and \mathbb{I}_B is the indicator function that the event B happening, i.e., it is one if B holds, and 0 otherwise.

(a) [1 Point] Use the results from Problem 1), part (c) to conclude that for any matrix with real valued entries, the cut-norm of W_A is a maximum over a finite number of sets:

$$||W_A||_{\Box} := \sup_{U, V \subset [0,1]} \left| \int_{U \times V} W_A \right| = \frac{1}{n^2} \max_{S, T \subset [n]} \left| \sum_{\substack{i \in U \\ j \in V}} A_{ij} \right|.$$

T

Apply this to the case where A is the adjacency matrix of G(n, p), and use the above expression to express $||W_A - P||_{\Box}$ (where P is the constant function that assigns p to $(x, y) \in [0, 1]^2$) as

$$||W_A - P||_{\Box} = \frac{1}{n^2} \max_{S, T \subset [n]} \left| \sum_{\substack{i \in S \\ j \in T}} (A_{ij} - p) \right| = \frac{1}{n^2} \max_{S, T \subset [n]} \left| -p|S \cap T| + \sum_{\substack{i \in S, j \in T \\ i \neq j}} (A_{ij} - p) \right|$$

- (b) [2 Points]: Calculate the expectation of the sums for fixed S and T, and then use the Chernoff bound from Problem 2 part(c) (with $N = \binom{n}{2}$ and $p_i = p$) to bound the probability that this sum is larger than ϵn^2 in absolute value. Use a union bound to prove that G(n, p) converges to the constant graphon P in probability.
- (c) [1 Bonus Point]: Consider now the case of sparse random graphs, i.e., the case that $p = p_n \to 0$. Show that if it does not go to zero too fast, you get that $\|\frac{1}{p_n}W_A 1\|_{\Box} \to 0$ in probability. How sparse can your graphs be? Can you do $p_n = 1/\log n$, or $p_n = 1/\sqrt{n}$. What prevents you from choosing p = c/n?

4 Graph Limit of Chung-Lu Model [5 Points]

In this exercise, we consider a dense version of the Chung-Lu Model, defined in terms of a function $w : [0,1] \to [0,1]$. In this model, we first draw x_1, \ldots, x_n iid uniformly at random in [0,1], and then connect i < j with probability

$$P_{ij} = w(x_i)w(x_j),$$

giving a graph G_n with adjacency matrix $A_{ij} = A_{ji} \sim Be(P_{ij})$, independently for all i < j. In this exercise, you will show that the sequence G_n converges to the graphon

$$W(x,y) = w(x)w(y)$$

in the cut-metric δ_{\Box} .

(a) [2 Points]: Proceed as in the previous problem to show that

$$||W_A - W_P||_{\Box} \to 0$$
 in probability,

where P is the matrix with entries P_{ij} . Hint: the proof is identical, except that when using the bound (c) from Problem 2, the probabilities p_i now become the probabilities P_{ij} rather than the constant p.

(b) Specialize now to the case where w is a step function,

$$w(x) = \sum_{k=1}^{K} \alpha_k \mathbb{I}_{J_k}$$
 where $J_k = [(k-1)/K, k/K),$

choose $x_1, \ldots, x_n \in [0, 1]$ i.i.d. uniformly at random, and order them such that $x_1 < x_2 < \cdots < x_n$. Defining \tilde{P} to be the matrix P after the reordering is applied, we note that

$$W_{\tilde{P}}(x,y) = \tilde{w}(x)\tilde{w}(y)$$
 where $\tilde{w}(x) = \sum_{k=1}^{K} \alpha_k \mathbb{I}_{\tilde{J}_k},$

with J_k being the union of the intervals $I_i = [(i-1)/n, i/n)$ for all i such that $x_i \in J_k$.

(i) [1 Point] Let N_k be the number of points x_i such that $x_i \in J_k$. Calculate the expectation and variance of N_k , and use this to prove that as $n \to \infty$,

$$\epsilon_k = \max_k \left| |\tilde{J}_k| - |J_k| \right| \to 0$$
 in probability

(ii) [1 Point] Bound the L_1 distance of w and \tilde{w} in terms of ϵ_k and use this to prove that

$$\|W_{\tilde{P}} - W\|_{\Box} \le \|W_{\tilde{P}} - W\|_1 \to 0 \quad \text{ in probability.}$$

- (iii) [1 Point] Show that there is an interval permutation $\phi : [0,1] \to [0,1]$ such that $W_P^{\phi} = W_{\tilde{P}}$ and use what you have shown so far to prove that for the case of step-functions, G_n converges to W in the cut-metric.
- (c) The general case can be reduced to the case of step-functions by noting that any bounded, measurable function can be approximated by step functions. We won't do this here see, however, the bonus problem below, where we address this problem in a more general context.

5 Inhomogneous random graphs converge to the generating graphon [6 Bonus Points]

Recall that a graphon is a symmetric function from $[0,1]^2$ into [0,1], i.e., a function $W : [0,1]^2 \to [0,1]$ such that W(x,y) = W(y,x). We define the sequence $G_n(W)$ of inhomogeneous random graphs generated by W as follows:

- 1. Choose x_1, \ldots, x_n i.i.d. uniformly at random from [0, 1], and define a matrix $P^{(n)} = P^{(n)}(W) \in [0, 1]^{n \times n}$ by setting $P_{ij}^{(n)} = W(x_i, x_j)$.
- 2. The graph $G_n(W)$ on [n] is then defined by choosing, independently for all i < j, an edge ij with probability $P_{ij}^{(n)}$.

In this exercise, we will prove the following theorem, which implies that $\delta_{\Box}(G_n(W), W) \to 0$ in probability.

Theorem 1.1 If W is a graphon, then

$$\mathbb{E}\left[\delta_{\Box}\left(G_n(W),W\right)\right] \to 0.$$

The theorem relies on two lemmas, which we will prove separately.

Lemma 1.1 There exists a constant $D < \infty$, such that if $P \in [0,1]^{n \times n}$ is a symmetric matrix with empty diagonal, and $A \in \{0,1\}^{n \times n}$ is the random, symmetric matrix with empty diagonal obtained from P by setting $A_{ij} = A_{ji} = 1$ with probability P_{ij} , independently for all i < j, then

$$\Pr(\|W_A - W_P\|_{\square} \ge D/n) \le 2e^{-n}.$$

As a consequence

$$E[||W_A - W_P||_{\Box}] \le D/n + 2e^{-n}.$$

To state the second lemma, for an $n \times n$ matrices A and a graphon W, we define

$$\hat{\delta}_1(A, W) = \min \|W - W_{A^{\sigma}}\|_1,$$

where the min is taken over all permutations of [n], and $A_{ij}^{\sigma} = A_{\sigma(i),\sigma(j)}$. Since such a permutation is an interval permutation, and since the cut-norm is bounded by the L_1 norm, we clearly have that $\delta_{\Box}(W_A, W) \leq \hat{\delta}_1(W_A, W)$.

Lemma 1.2 For all symmetric $W : [0,1]^2 \to [0,1]$, define $P^{(n)} = P^{(n)}(W)$ to be the $n \times n$ random matrix with entries $P_{ii}^{(n)} = W(x_i, x_j)$, where x_1, \ldots, x_n are chosen iid uniformly at random in [0,1]. Then

$$\mathbb{E}\left[\hat{\delta}_1\left(P^{(n)},W\right)\right] \to 0.$$

- (a) Use the Chernoff bound from Exercise 2 to prove Lemma 1.1. Hint: at this point the proof is a one-line argument, since you have already done the necessary estimate when solving Part (a) of Problem 4, since the precise form for P_{ij} never entered your proof. [1 Bonus Point]
- (b) Prove Lemma 1.2 for step functions. More precisely, let Q_k be the partition of [0, 1] into k intervals of length 1/k, and prove Lemma 1.2 for functions W which are constant on sets of the form $Y \times Y'$, $Y, Y' \in Q_k$. [1 Bonus Point]
 - Hint: Reorder x_1, \ldots, x_n in such a way that $x_1 < x_2 < \cdots < x_n$, and use that for $n \gg k$, the fraction of variables x_i that fall into the i^{th} interval of the partition Q_k is concentrated around 1/k. Determine how large n has to be (as a function of k), to get enough concentration to imply that $E\left[\hat{\delta}_1(P^{(n)}(W), W)\right] \to 0.$
- (c) Reduce Lemma 1.2 to the case where W is a step function. To this end, use the following approximation to the graphon W: the function W_{Q_k} obtained by averaging W over the blocks $Y \times Y'$, where Y and Y' are classes in Q_k .
 - Prove that for any two graphons U, W,

$$\frac{1}{n^2} E[\sum_{i \neq j} |P_{ij}^{(n)}(U) - P_{ij}^{(n)}(W)| = \frac{n-1}{n} ||U - W||_1.$$

- Use the fact that $||W_{Q_k} W||_1 \to 0$ to reduce the proof of Lemma 1.2 to the case analyzed under b. (For people familiar with measure theory: you can use the Lebesgue differentiation theorem to prove almost sure convergence, which implies L_1 convergence. You are not asked to prove this fact - you can just use it). [3 Bonus Points]
- (d) Prove Theorem 1.1 from Lemmas 1.1 and 1.2.

[1 Bonus Point]