Homework 10

Math 561: Theory of Probability I

Due date: April 13, 2023

Each problem is worth 25 points and solve any TWO problems. Please indicate whom you worked with, it will not affect your grade in any way.

1. Recall the definition of Wasserstein distance between two probability distributions μ, ν on $(\mathbb{R}, \mathcal{B})$ given by

$$\mathrm{d_W}(\mu,\nu) := \sup_{f\,:\,1-\text{Lipschitz}} |\mu(f) - \nu(f)|\,.$$

a) Prove that,

$$\mathrm{d_W}(\mu,\nu) = \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)|\,dx,$$

where F_μ, F_ν are the distribution functions of μ, ν, respectively. Use homework problem 5.4.

b) Prove that,

$$\mathrm{d_W}(\mu,\nu) = \inf_{\text{all couplings } (X,Y) \text{ with marginals } \mu,\nu} \mathbb{E}\,|X - Y|\,.$$

c) Find the Wasserstein distance between $\mathrm{N}(0,1)$ and $\mathrm{N}(0,\sigma^2)$ distributions (or upper and lower bounds differing by a constant multiple) where $\sigma > 1$.

2. Consider three urns numbered $1, 2, 3$; and n distinct balls. The i^{th} ball is placed into the urn with number U_i, where U_i's are i.i.d. random variables with $\mathbb{P}(U_1 = k) = p_k$ for $k = 1, 2, 3$ and $p_1 + p_1 + p_3 = 1$. Let W be the scaled difference between the number of balls in the first and second urn, *i.e.,*

$$W := \frac{\text{number of balls in urn 1}}{p_1} - \frac{\text{number of balls in urn 2}}{p_2}.$$

Prove CLT for appropriately scaled W and bound the Wasserstein distance from $\mathrm{N}(0,1)$.

3. Given n distinct real numbers $x_1, x_2, \ldots, x_n$ with $\sum_{i=1}^n x_i = 0$, consider a random sample of size k **without replacement**, *i.e.,* choose a random subset $\Pi \subseteq \{1, 2, \ldots, n\}$ of size k uniformly from all $\binom{n}{k}$ choices. Assuming that, $k/n \approx p \in (0,1)$, prove CLT for the for appropriately scaled sample average $W := \frac{1}{k}\sum_{i\in\Pi} x_i$ and bound the Wasserstein distance from $\mathrm{N}(0,1)$.

4. Fix $p \in (0,1)$. Let $G_{n,p}$ be an Erdös-Rényi random graph on vertex set $\{1, 2, \ldots, n\}$ with edges present with probability p, independently of each other. Let W be the number of wedges, *i.e.,* paths of length 2, present in $G_{n,p}$. Prove a CLT for W after appropriate centering and scaling.

5. Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}\,X_1 = \mu, \mathrm{Var}(X_1) = 1$ and $\mathbb{E}\,|X_1|^3 = \rho$. For some k with $1 \leqslant k \leqslant n$, define

$$Y_j = \prod_{i=j}^{j+k-1} X_i, \qquad j = 1, \ldots, n,$$

where the bounds of the product are defined modularly (*e.g.*, $Y_n = X_n X_1 \cdots X_{k-1}$). Let

$$W = \sum_{j=1}^{n} (Y_j - \mu^k).$$

Find $\sigma^2 = \mathrm{Var}(W)$ and an upper bound on $\mathrm{d_W}(W/\sigma, Z)$. Identify a regime when the rate goes to zero.

6. * Let π be a uniformly chosen random permutation of $\{1, 2, \ldots, n\}$. Prove CLT for the average gap between neighboring points

$$\frac{1}{n} \sum_{i=1}^{n} |\pi_{i+1} - \pi_i|$$

and bound the Wasserstein distance from $\mathrm{N}(0, 1)$.

7. * Fix $p \geqslant 2$. Consider the n-dimensional unit ℓ^p-sphere $S_{n,p} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_p^p = \sum_{i=1}^n |x_i|^p = 1\}$. For a uniform random point $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ in $S_{n,p}$, prove CLT for $\sum_{i=1}^n X_i$ after appropriate scaling and bound the Wasserstein distance from $\mathrm{N}(0, 1)$.