MathTools HW 8

1. **Total variation distance.** Recall that, by definition, $d_{TV}(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2}\|\boldsymbol{p}-\boldsymbol{q}\|_1$, where $\boldsymbol{p},\boldsymbol{q} \in [0,1]^n$ are probability distributions.

   (a) Prove the following statement from TA8:

   $$d_{TV}(\boldsymbol{p},\boldsymbol{q}) = \max_{S\subset[n]} \left[\boldsymbol{p}(S) - \boldsymbol{q}(S)\right] .$$

   Hint: Find explicitly the maximizing set S. Also, remember that $\sum_{i\in S}(p_i - q_i) + \sum_{i\notin S}(p_i - q_i) = 0$ - explain why.

   (b) Consider the following *hypothesis testing*. One observed a random variable $X \in \{1,\ldots,n\}$ sampled from one of the distributions $\boldsymbol{p}$ or $\boldsymbol{q}$ - we don't know which, and want to guess with the smallest possible error. A (deterministic) test is a set of outcomes $S \subset [n]$ such that: (i) if $X \in S$, we declare "$\boldsymbol{p}$"; (ii) if $X \notin S$, we declare "$\boldsymbol{q}$".

   Denote by $\varepsilon_p(S)$ the error probability when the group truth is $\boldsymbol{p}$: namely, X was sampled from $\boldsymbol{p}$, but we erroneously declared "$\boldsymbol{q}$"; $\varepsilon_q(S)$ is defined likewise. Prove that

   $$\varepsilon_p(S) + \varepsilon_q(S) \geq 1 - d_{TV}(\boldsymbol{p},\boldsymbol{q}) ,$$

   and show that there is a test for which there is equality above.

2. **Easy leftovers from TA.** Let P the transition matrix on an **ergodic** Markov chain.

   (a) Suppose that $P = P^\top$. Prove that its stationary distribution is the uniform distribution.

   (b) Denote by $\mathbf{x}_t$ the marginal distribution of X_t at time t. Prove that the sequence $d_{TV}(\mathbf{x}_t,\boldsymbol{\pi})$ is non-increasingly, namely, $d_{TV}(\mathbf{x}_{t+1},\boldsymbol{\pi}) \leq d_{TV}(\mathbf{x}_t,\boldsymbol{\pi})$.

   (c) Using the notations from TA8, so that

   $$\tau(\varepsilon) = \max_{i=1,\ldots,n} \tau(\varepsilon|\mathbf{e}_i) .$$

   In other words, the "worst" starting distribution in terms of mixing time is one which starts deterministically at some state.

3. Let G be a d-regular, connected non-bipartite graph. As in the TA, let $\tau(\varepsilon)$ be the mixing time of the SRW on G to within ε TV-distance from the stationary (uniform) distribution. Recall: we proved that

$$\tau(\varepsilon) = O\left(\frac{\log n + \log\frac{1}{\varepsilon}}{\gamma}\right)$$

where γ is the spectral gap (throughout, let's assume it is small, so that $\log \frac{1}{1-\gamma} = \Theta(\gamma)$). In particular, when $\varepsilon = 1/\text{poly}(n)$, for example, $\varepsilon = n^{-4}$, we get $\tau(\varepsilon) = O\left(\frac{\log n}{\gamma}\right)$. Prove a *matching lower bound* on the mixing time, for *very small target precision* ε: namely, show that for $\varepsilon = n^{-4}$

$$\tau(n^{-4}) = \Omega\left(\frac{\log n}{\gamma}\right) .$$

Hint: You need to come up with a starting distribution $\boldsymbol{\pi}_0$, such that $d_{TV}(\boldsymbol{\pi}_0 P^t, \boldsymbol{\pi})$ is large for all small t. Explain why the following is true: if $\|u\|_2 = 1$ and $u \perp \mathbf{1}$, then $\pi + n^{-1}u$ is a kosher probability distribution (recall: $\boldsymbol{\pi} = (1/n, \ldots, 1/n)$). Now, basically follow the same proof we did in class.

4. **Lazy random walk.** Let G be a connected graph. Consider the following random walk on G: suppose that in time t, you are in vertex X_t; you flip an even coin (head w.p. $1/2$) - if it is heads, you stay in X_t (meaning $X_{t+1} = X_t$), otherwise X_{t+1} is just a random neighbor of X_t, chosen uniformly (as in a SRW).

   (a) Let P be the transition matrix for the SRW (the "usual" random walk) on G. What is the transition matrix for the LRW (lazy random walk)?

   (b) Show that the stationary distribution of the LRW is the same as that of the SRW.

   (c) Prove that the LRW is always ergodic (even when G is bipartite!).

   (d) Suppose that G is not bipartite. Denote by $\tau_{SRW}(\varepsilon)$ the mixing time for the SRW and by $\tau_{LRW}(\varepsilon)$ the mixing time for the LRW. Suppose that the SRW on G is rapidly mixing, in the sense that for every constant $\varepsilon > 0$, $\tau_{SRW}(\varepsilon) = \text{polylog}(n)$, where ε is thought of as a *constant*. Prove that the LRW is also rapidly mixing, that is, that $\tau_{LRW}(\varepsilon) = \text{polylog}(n)$.[1]

   Hint: I am sure there are many ways to show this. Here is the one I had in mind (but you can prove this any way you like): Let $\mathbf{x}_0$ be the initial dist. and $\mathbf{x}_t$ be the dist. at step t. Show that $d_{TV}(\mathbf{x}_t, \boldsymbol{\pi}) \leq 2^{-t} \sum_{i=0}^{t} \binom{t}{i} d_{TV}(\mathbf{x}_0^\top P^i, \boldsymbol{\pi})$. Take, e.g, $t = [\tau_{SRW}(\varepsilon/2)]^2$; control the first $\sqrt{t}$ terms and the remaining $(t + 1 - \sqrt{t})$ terms separately.

[1] A nitpicky point: since we're talking about rapid mixing (asymptotics in n), in truth we have here a *family* of graph on n vertices, and $n \to \infty$...