SUN Yilin
520030910361

Homework 2

1 Optimal Coupling

1.1 Basic Ideas

To reach the lower bound $D_{TV}(\mu,\nu)$, the intuition is that we maximize the terms where $X=Y$. Denote our optimal coupling by ω^*, then $\forall(x,y)\in\Omega^2$ such that $x=y$, the maximum of $\omega^*(x,y)$ can only be $\min\{\mu(x),\nu(y)\}$. As we want to maximize these cases, we can directly set it as $\min\{\mu(x),\nu(y)\}$.
Then we need a formula to define $\omega^*(x,y)$ where $x\neq y$. Here the intuition is that $\omega^*(x,y)$ should be some constant times $\max\{\mu(x)-\nu(x),0\}\cdot\max\{\nu(y)-\mu(y),0\}$. So up to here, I construct a coupling as follows:

$$\omega^*(x,y)=\begin{cases}\min\{\mu(x),\nu(y)\}, & x=y\\ C\cdot\max\{\mu(x)-\nu(x),0\}\cdot\max\{\nu(y)-\mu(y),0\}, & o.w.\end{cases}$$

Now I will determine C so that ω^* is indeed a valid coupling, and then show $Pr_{(X,Y)\sim\omega^*}(X\neq Y)$ is actually $D_{TV}(\mu,\nu)$.

1.2 Determine C and proof of valid coupling

Define $A=\{x\in\Omega|\mu(x)\geq\nu(x)\}$. It follows that $\bar{A}=\{x\in\Omega|\mu(x)<\nu(x)\}$. We now calculate the marginal distribution of X under ω^*.

$$\begin{aligned}\forall x\in A,\quad \sum_{y\in\Omega}\omega^*(x,y)&=\sum_{y\in\Omega\wedge y=x}\omega^*(x,y)+\sum_{y\in\Omega\wedge y\neq x}\omega^*(x,y) &(1)\\ &=\min\{\mu(x),\nu(x)\}+\sum_{y\in\Omega\wedge y\neq x}\omega^*(x,y) &(2)\\ &=\nu(x)+C\max\{\mu(x)-\nu(y),0\}\sum_{y\in\Omega\wedge y\neq x}\max\{\nu(y)-\mu(y),0\} &(3)\\ &=\nu(x)+C(\mu(x)-\nu(x))\sum_{y\in\bar{A}}(\nu(y)-\mu(y)) &(4)\end{aligned}$$

Note that $\sum_{y\in\bar{A}}(\nu(y)-\mu(y))=D_{TV}(\mu,\nu)$ by our definition of $\bar{A}$, so by setting $C=1/D_{TV}(\mu,\nu)$ we can get

$$\begin{aligned}\forall x\in A,\quad \sum_{y\in\Omega}\omega^*(x,y)&=\nu(x)+\mu(x)-\nu(x) &(5)\\ &=\mu(x) &(6)\end{aligned}$$

We use the same C and consider the cases where $x \in \bar{A}$. Similarly

$$\begin{aligned}
\forall x \in \bar{A}, \quad \sum_{y\in\Omega} \omega^*(x,y) &= \min\{\mu(x),\nu(x)\} + \sum_{y\in\Omega \wedge y\neq x} \omega^*(x,y) &(7)\\
&= \mu(x) + D_{TV}(\mu,\nu)\max\{\mu(x)-\nu(x),0\}\sum_y \max\{\nu(y)-\mu(y),0\} &(8)\\
&= \mu(x) + D_{TV}(\mu,\nu)\cdot 0 \cdot \sum_y \max\{\nu(y)-\mu(y),0\} &(9)\\
&= \mu(x) &(10)
\end{aligned}$$

Now we have proved such a coupling satisfies that the marginal distribution of X is indeed $\mu(x)$. The same argument works for Y as well. So by setting $C = 1/D_{TV}(\mu,\nu)$ we actually constructed

$$\omega^*(x,y) = \begin{cases} \min\{\mu(x),\nu(y)\}, & x = y \\ \frac{1}{D_{TV}(\mu,\nu)} \cdot \max\{\mu(x)-\nu(x),0\} \cdot \max\{\nu(y)-\mu(y),0\}, & o.w. \end{cases}$$

Now we only need to show $Pr_{(X,Y)\sim\omega^*}(X \neq Y)$ is actually $D_{TV}(\mu,\nu)$.

1.3 Proof of optimal coupling

We show by calculating $Pr_{(X,Y)\sim\omega^*}(X = Y)$.

$$\begin{aligned}
Pr_{(X,Y)\sim\omega^*}(X=Y) &= \sum_{x\in\Omega} \min\{\mu(x),\nu(x)\} &(11)\\
&= \sum_{x\in A}\nu(x) + \sum_{x\in\bar{A}}\mu(x) &(12)\\
&= \nu(A) + \mu(\bar{A}) &(13)\\
&= \nu(A) + (1-\mu(A)) &(14)\\
&= 1 - (\mu(A)-\nu(A)) &(15)\\
&= 1 - D_{TV}(\mu,\nu) &(16)
\end{aligned}$$

So $Pr_{(X,Y)\sim\omega^*}(X \neq Y) = D_{TV}(\mu,\nu)$

2 Stochastic Dominance

The idea of proof comes from our teacher's lecture notes in previous year.
Following the idea of that lecture note, I will first prove the proposition about monotone coupling, because the other two questions are actually applications of this proposition.

2.1 Monotone Coupling

2.2 Binomial Distribution

Sufficiency:
If $p \geq q$, suppose $X \sim \text{Binom}(n, p)$, $Y \sim \text{Binom}(n, q)$. We define such a coupling ω of these two Binomial Distributions where we do n trials and for these n trials we independetly pick a real r in $[0, 1]$ uniformly at random and every trial is independent. Then let $X = x$ where x is the number of these trials with $r \leq p$ and let $Y = y$ where y is the number of these trials with $r \leq q$.
By our definition we can see clearly $Pr_{(X,Y)\sim\omega}(X \geq Y) = 1$. So there exists a monotone coupling of $\text{Binom}(n, p)$ and $\text{Binom}(n, q)$. By the proposition we have proven above, $\text{Binom}(n, p) \succeq \text{Binom}(n, q)$.
Necessity:
Prove by contradiction. If $p < q$, consider the case where $a = n$. $Pr(X \geq a) = Pr(X = n) = p^n < q^n = Pr(Y = n) = Pr(Y \geq a)$. This violates $\text{Binom}(n, p) \succeq \text{Binom}(n, q)$. So $p \geq q$.

2.3 Random Graph

Suppose $G \sim \mathcal{G}(n, p)$, $H \sim \mathcal{G}(n, q)$. Consider such a coupling ω of $\mathcal{G}(n, p)$ and $\mathcal{G}(n, q)$ where we generate G and H simultaneously. For each pair of vertices $\{i, j\}$ we independently pick a real r in $[0, 1]$ uniformly at random. Let G have edge $\{i, j\}$ iff $r \leq p$ and let H have edge $\{i, j\}$ iff $r \leq q$.
For any $p, q \in [0, 1]$ satisfying $p \geq q$, H is a subgraph of G, so ω is a monotone coupling. So by the proposition we have proven $Pr_{G\sim\mathcal{G}(n,p)}(G \text{ is connected}) \geq Pr_{H\sim\mathcal{G}(n,q)}(H \text{ is connected})$

3 Total Variation Distance is Non-Increasing

Let $X_0 \sim \mu_0$ and $Y_0 \sim \pi$. For any $t \geq 0$, we can couple the distributions of random variables X_t and Y_t such that $Pr(X_t \neq Y_t) = \Delta(t)$. This coupling is feasible because we have proven, in problem 1, that an optimal coupling exists.
Then we can construct a coupling of the distributions of X_{t+1} and Y_{t+1} with this coupling. We define

$$\begin{cases} X_{t+1} = Y_{t+1}, \quad \text{if} \quad X_t = Y_t \\ X_{t+1} \sim \mu_{t+1}, Y_{t+1} \sim \pi \quad \text{if} \quad X_t \neq Y_t \end{cases}$$

Then by the coupling lemma again we have

$$\Delta(t+1) \leq Pr(X_{t+1} \neq Y_{t+1}) \tag{17}$$

By our construction of coupling at t and $t+1$ we have

$$Pr(X_{t+1} \neq Y_{t+1}) \tag{18}$$
$$= Pr(X_{t+1} \neq Y_{t+1} | X_t = Y_t) Pr(X_t = Y_t) + Pr(X_{t+1} \neq Y_{t+1} | X_t \neq Y_t) Pr(X_t \neq Y_t) \tag{19}$$
$$\leq Pr(X_t \neq Y_t) \tag{20}$$

That inequality is true because all posibilities must be in $[0, 1]$.
So it follows That

$$\Delta(t+1) \leq Pr(X_{t+1} \neq Y_{t+1}) \leq Pr(X_t \neq Y_t) \leq \Delta(t), \forall t \geq 0$$

And that is true for any $t \geq 0$ by induction. The induction process is actually how the Markov Chain X_t and Y_t forms. Since any step of this induction process is the same as our construction of coupling and proof above, we do not write it again.

Acknowledgements

The inspiration of constructing the optimal coupling is from the lecture notes of a MIT course 6.896 Probability and Computation. Actually my construction is not entirely like its demonstration, but the core ideas are the same.

In Problem 2 I actually learnt our teacher's lecture notes 5 from last year and then solved these questions. I think it will be extremely difficult to solve these questions without knowing the theorem about monotone coupling.