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

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

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

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

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

Sufficiency:
If a monotone coupling of μ and ν and exists, denoted by ω. Then

$$\begin{aligned}
\forall a \in \Omega, Pr_{Y\sim\nu}(Y \geq a) &= Pr_{(X,Y)\sim\omega}(Y \geq a) && (17)\\
&= Pr_{(X,Y)\sim\omega}(X \geq Y \wedge Y \geq a) + Pr_{(X,Y)\sim\omega}(X < Y \wedge Y \geq a) && (18)\\
&= Pr_{(X,Y)\sim\omega}(X \geq Y \wedge Y \geq a) && (19)\\
&\leq Pr_{(X,Y)\sim\omega}(X \geq a) && (20)\\
&= Pr_{X\sim\mu}(X \geq a) && (21)
\end{aligned}$$

Hence $Pr_{X\sim\mu}(X \geq a) \geq Pr_{Y\sim\nu}(Y \geq a), \forall a \in \Omega$.
Necessity:
We prove by showing that there is a method of constructing such a monotone coupling. The basic idea is to construct this coupling greedily.
Firstly since Ω is a finite set of integers, there exists a smallest element a_0. By the stochastic dominance of μ over ν, $\mu(a_0) < \nu(a_0)$. So in our coupling $\Pr(X = Y = a_0)$ should be $\mu(a_0)$. We can use $\{a_0, a_1, a_2, a_3 \ldots, a_n\}$ to denote Ω. Then it follows that $\Pr(X = a_0, Y > a_0) = 0$. So we have assigned valid values for all events in the coupling with $X = a_0$.
Now we consider the case with $X = a_1$. It is always possible to assign $\Pr(X = Y = a_1) = \min\{\mu(a_1), \nu(a_1)\}$. Then if $\min\{\mu(a_1), \nu(a_1)\} = \mu(a_1)$, we can assign $\Pr(X = a_1, Y \neq a_1) = 0$ and we are done for cases with $X = a_1$. If $\min\{\mu(a_1), \nu(a_1)\} = \nu(a_1)$, Let $\Pr(X = a_1, Y = a_0) = \mu(a_1) - \nu(a_1)$, which is possible because $\Pr(X \leq a_1) < \Pr(Y \leq a_1)$. Assign other cases with 0 and we are also done for $X = a_1$.
Then we can always carry out such a process step by step.
To be specific, for the case $X = a_k$, first assign $\Pr(X = a_k, Y > a_k) = 0$ and $\Pr(X = a_k, Y = a_k) = \min\{\mu(a_k), \nu(a_k)\}$, if $\min\{\mu(a_k), \nu(a_k)\} = \mu(x_k)$ let $\Pr(X = a_k, Y \neq a_k) = 0$ and we are done. If not, we assign values for $\Pr(X = a_k, Y = a_l), l < k$ by a descending order of l. First we maximize $\Pr(X = a_k, Y = a_{k-1})$, then we maximize $\Pr(X = a_k, Y = a_{k-2})\ldots$ We are done whenever $\sum_{j=0}^{n} \Pr(X = a_k, Y = a_{k-j}) = \mu(a_k)$ for some n, This process is always possible by the stochastic dominance of μ over ν and by our construction process from the smallest a_0 to the largest a_n.

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

$$\Delta(t+1) \leq Pr(X_{t+1} \neq Y_{t+1}) \tag{22}$$

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

$$\begin{aligned} & Pr(X_{t+1} \neq Y_{t+1}) && (23) \\ = & Pr(X_{t+1} \neq Y_{t+1} | X_t = Y_t) Pr(X_t = Y_t) + Pr(X_{t+1} \neq Y_{t+1} | X_t \neq Y_t) Pr(X_t \neq Y_t) && (24) \\ \leq & Pr(X_t \neq Y_t) && (25) \end{aligned}$$

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

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

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.

4 Acknowledgements

4.1

For Problem 1 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.

4.2

For 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.

4.3

For Problem 3 I referred to the lecture note 7 of a UC Berkeley course CS294 Partition Functions(do not know why such a course contains materials about our course). I cannot justify that my works are different from that lecture note because I cannot figure out another easy approach to solve this problem.