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

3 Total Variation Distance is Non-Increasing

4 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.