Lecture 23

* Hamming Cost

Another interesting setting where a comprehensive sol^n is avaible is the following:

$\mathcal{X} = \mathcal{Y}$, $c(x,y) = \mathbb{1}_{\{x \neq y\}}$ } P2

Lemma (Maximal coupling) For $c(x,y) = \mathbb{1}_{\{x \neq y\}}$,

$$c(P, Q) = d_{TV}(P, Q)$$

$$\left(= \sup_{h: \mathcal{X} \to [-1,1], \text{ continuous}} \mathbb{E}_Q[h(X)] - \mathbb{E}_P[h(X)] \right)$$

Proof. We will only prove for discrete $\mathcal{X}$.

$$\mathbb{P}(X=Y) = \sum_x \mathbb{P}(X=Y=x) \leq P(A^c) + Q(A) \quad \forall A \subseteq \mathcal{X}.$$

$$\leq 1 - d_{TV}(P,Q).$$

$$\Rightarrow c(P,Q) \geq d_{TV}(P,Q).$$

For the other direction, let $A = \{x : P(x) \geq Q(x)\}$ so that $d_{TV}(P,Q) = P(A) - Q(A)$.

Let $$\mathbb{P}_1(x,y) = \frac{\min\{P(x), Q(x)\} \mathbb{1}_{\{x=y\}}}{\sum_{x'} \min\{P(x'), Q(x')\}}$$

and $$\mathbb{P}_2(x,y) = \frac{(P(x) - Q(x))_+ (Q(y) - P(y))_+}{\sum_{x',y'} (P(x') - Q(x'))_+ (Q(y') - P(y'))_+}$$

Note that

$$\mathbb{P}_1(x,y) = \frac{\min\{P(x), Q(x)\}}{1-d_{TV}(P,Q)} \mathbb{1}_{\{x=y\}}$$

$$\mathbb{P}_2(x,y) = \frac{(P(x)-Q(x))(Q(y)-P(y))}{d_{TV}^2(P,Q)} \mathbb{1}_{\{x\in A,\, y\in A^c\}}.$$

Let $\mathbb{P}(x,y) = (1-d_{TV}(P,Q))\,\mathbb{P}_1(x,y) + d_{TV}(P,Q)\,\mathbb{P}_2(x,y)$.

Thus, $\mathbb{P}(X=Y) = [1-d_{TV}(P,Q)]\,\mathbb{P}_1(X=Y)$

$$= 1-d_{TV}(P,Q).$$

(Note that this is not a deterministic coupling). ■

A transportation cost inequality is an upper bound for $c(P,Q)$.

For $c(x,y) = \mathbb{1}_{\{x\neq y\}}$, the so-called Pinsker's inequality showing $d_{TV}(P,Q) \le \sqrt{2D(P\|Q)}$ is an example.

$\boxed{B}$ Link between transportation cost inequalities and measure concentration

Lemma (Transportation Lemma)

$$\psi_{Z-\mathbb{E}Z}(\lambda) \le \frac{v\lambda^2}{2} \quad \forall\ \lambda>0$$

iff

$$\mathbb{E}_Q Z - \mathbb{E}_P Z \le \sqrt{2v\,D(Q\|P)} \text{ for all } Q \ll P.$$

Proof.

Recall the Gibb's variational formula:

$$\log \mathbb{E}_P[e^{\lambda Z}] = \max_{Q \ll P} \lambda \mathbb{E}_Q[Z] - D(Q\|P),$$

i.e., $$\psi_{Z-\mathbb{E}Z}(\lambda) = \max_{Q \ll P} \lambda\left(\mathbb{E}_Q Z - \mathbb{E}_P Z\right) - D(Q\|P).$$

Thus, if $\psi_{Z-\mathbb{E}Z}(\lambda) \le \frac{v\lambda^2}{2}$, $\forall Q \ll P$

$$D(Q\|P) \geq \lambda\left(\mathbb{E}_Q Z - \mathbb{E}_P Z\right) - \frac{v\lambda^2}{2}.$$

Maximizing the right-side over $\lambda \in \mathbb{R}$, we get

$$D(Q\|P) \geq \frac{\left(\mathbb{E}_Q Z - \mathbb{E}_P Z\right)^2}{2v}.$$

Conversely,

$$\mathbb{E}_Q Z - \mathbb{E}_P Z \le \sqrt{2v\, D(Q\|P)}, \qquad \forall Q \ll P,$$

implies that every $\lambda > 0$,

$$\begin{aligned}
\psi_{Z-\mathbb{E}Z}(\lambda) &\le \sup_{Q \ll P} \lambda\sqrt{2v\, D(Q\|P)} - D(Q\|P) \\
&\le \lambda^2 \sup_x \sqrt{2v}\, x - x^2 \\
&= \frac{\lambda^2 v}{2}.
\end{aligned}$$

■

The claim can be extended to any "nice" function g:

$$\psi_{Z-\mathbb{E}Z}(\lambda) \le \underset{\forall \lambda > 0}{g(\lambda)} \Leftrightarrow \mathbb{E}_Q Z - \mathbb{E}_P Z \le g^{-1}\left(D(Q\|P)\right).$$

C McDiarmid Inequality via Transportation Lemma

Suppose $f: \mathcal{X}^n \to \mathbb{R}$ satisfies BDP with constants $c = (c_1, \dots, c_n)$.

Let P be a product distribution on $\mathcal{X}^n$ and Q be any dist. on $\mathcal{X}^n$. Consider $\mathbb{P} \in \mathcal{P}(P, Q)$. Then,

$$\begin{aligned} \mathbb{E}_Q[f(X)] - \mathbb{E}_P[f(X)] &= \mathbb{E}[f(X) - f(Y)] \\ &\le \mathbb{E}\left[\sum_{i=1}^n c_i \mathbb{1}_{\{X_i \neq Y_i\}}\right] \\ &= \sum_{i=1}^n c_i \mathbb{P}(X_i \neq Y_i) \end{aligned}$$

Thus,

$$\mathbb{E}_Q[f(X)] - \mathbb{E}_P[f(X)] \le \sqrt{\sum_{i=1}^n c_i^2} \cdot \sqrt{\sum_{i=1}^n \mathbb{P}(X_i \neq Y_i)^2}.$$

Therefore, by Transportation Lemma, it suffices to show

$$\boxed{\begin{gathered} \exists \text{ a coupling } \mathbb{P} \in \mathcal{P}(P,Q) \text{ such that,} \\ \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i)^2 \le \frac{1}{2} \cdot D(Q \| P) \quad \forall Q \ll P \end{gathered}} \qquad (\#)$$

to get $\psi_{Z - \mathbb{E}Z}(\lambda) \le \left(\sum_{i=1}^n c_i^2\right) \lambda^2/8 \quad \forall \lambda > 0,$

which in turn leads to

$$\mathbb{P}\left(f(X) > \mathbb{E}f(X) + t\right) \le \exp\left(-2t^2 / \sum_{i=1}^n c_i^2\right).$$

Hence, Transportation Lemma allows us to reduce conc. bounds to Transportation Cost Inequality such as $(\#)$.

D Pinsker's inequality

$$d_{TV}(P,Q) = \sup_A P(A) - Q(A)$$

Lemma (Pinsker's inequality)

For two distributions P and Q on $\mathcal{X}$ s.t. $Q \ll P$. Then,

$$d_{TV}(P,Q)^2 \leq \frac{1}{2} D(Q\|P).$$

Proof. Note that for $A^* \stackrel{\text{def}}{=} \{\frac{dQ}{dP} \geq 1\}$,

$$d_{TV}(P,Q) = Q(A^*) - P(A^*) = \mathbb{E}_Q Z - \mathbb{E}_P Z$$

where $Z = \mathbb{1}_{A^*}$. Note that by Hoeffding's lemma,

$$\psi_{Z-\mathbb{E}Z}(\lambda) \leq \frac{\lambda^2}{8}.$$

Thus, by Transportation Lemma,

$$\mathbb{E}_Q Z - \mathbb{E}_P Z \leq \sqrt{\frac{1}{2} D(Q\|P)}.$$ ∎

E Marton's Transportation Cost Inequality

Theorem. For $X = (X_1, \dots, X_n) \sim P = P_1 \otimes \dots \otimes P_n$ and Q s.t. $Q \ll P$, let $Y = (Y_1, \dots, Y_n) \sim Q$. Then, there exists a coupling $\mathbb{P}$ of P and Q s.t.

$$\sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i)^2 \leq \frac{1}{2} D(Q\|P).$$

Proof. For $n=1$, maximal coupling lemma gives

$$\mathbb{P}(X \neq Y)^2 = d^2(P,Q) \leq \frac{1}{2} D(Q\|P).$$

We can complete the proof by induction.

Given $Y^k = y^k$, consider the maximal coupling
$\mathbb{P}_{X_{k+1} Y_{k+1} | X^k Y^k} \in \mathcal{P}\left(P_{X_{k+1}}, Q_{Y_{k+1}|Y^k = y^k}\right)$.

Then, $\mathbb{P}\left(X_{k+1} \neq Y_{k+1} \,\middle|\, X^k = x^k, Y^k = y^k\right)$

$$= d_{TV}\left(P_{X_{k+1}}, Q_{Y_{k+1}|Y^k = y^k}\right)$$

$$\leq \sqrt{\frac{1}{2} D\left(Q_{Y_{k+1}|Y^k = y^k} \| P_k\right)}$$

$$\Rightarrow \mathbb{P}\left(X_{k+1} \neq Y_{k+1}\right) \leq \mathbb{E}_{Q_{Y^k}}\left[\sqrt{\frac{1}{2} D\left(Q_{Y_{k+1}|Y^k} \| P_k\right)}\right]$$

$$\leq \sqrt{\frac{1}{2} \mathbb{E}_{Q_{Y^k}}\left[D\left(Q_{Y_{k+1}|Y^k} \| P_k\right)\right]}.$$

Thus, $$\sum_{i=1}^{n} \mathbb{P}\left(X_i \neq Y_i\right)^2 \leq \frac{1}{2} \sum_{i=1}^{n} D\left(Q_{Y_i} \| P_i \,\middle|\, Q_{Y^{i-1}}\right)$$

$$= \frac{1}{2} D\left(Q_{Y^n} \| P\right).$$

$\blacksquare$

Note that our constructed coupling $\mathbb{P}_{X^n Y^n}$ satisfies $\mathbb{P}_{X^k Y^k} = \mathbb{P}_{X_k} \cdot \mathbb{P}_{Y_k | Y^{k-1}} \cdot \mathbb{P}_{X^{k-1} Y^{k-1}}$.