Lecture 25

Marton's Conditional Transportation Cost Inequality

$$\min_{(X,Y) \in \mathcal{P}(P^n, Q)} \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i \mid X)^2 + \mathbb{P}(X_i \neq Y_i \mid Y)^2\right]$$

$$\leq 2D(Q \| P^n).$$

for all $Q \ll P^n = P_1 \times \ldots \times P_n$.

Claim 1: It suffices to show the inequality for $n=1$.

Suppose that $\forall n \leq k \; \exists \, (X^n, Y^n) \in \mathcal{P}(P^n, Q)$ s.t.

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i \mid X)^2 + \mathbb{P}(X_i \neq Y_i \mid Y)^2\right]$$

$$\leq 2D(Q \| P^n).$$

Then, we will get a coupling for $n = k+1$.

We begin by using the construction for $n=1$.

$$\exists \; (X_{k+1}, Y_{k+1}) \in \mathcal{P}\left(P_{k+1}, Q_{Y_{k+1} \mid Y^k = y^k}\right)$$

s.t.

$$\mathbb{E}\left[\left(\mathbb{P}(X_{k+1} \neq Y_{k+1} \mid Y^{k+1})^2 + \mathbb{P}(X_{k+1} \neq Y_{k+1} \mid X_{k+1}, Y^k)^2\right) \mid Y^k\right]$$

$$\leq 2D\left(Q_{Y_{k+1} \mid Y^k = y^k} \,\|\, P_{X_{k+1}}\right).$$

We set

$$P_{X_{k+1} Y_{k+1} \mid X^k Y^k} = P_{X_{k+1} Y_{k+1} \mid Y^k}.$$

Thus, the Markov relation $(X_{k+1}, Y_{k+1}) - Y^k - X^k$,
$\mathbb{P}(X_{k+1} \neq Y_{k+1} | X_{k+1}, Y^k) = \mathbb{P}(X_{k+1} \neq Y_{k+1} | X^{k+1}, Y^k)$.

Further, by Jensen's inequality,
$$\mathbb{E}[\mathbb{P}(X_{k+1} \neq Y_{k+1} | X^{k+1})^2] \leq \mathbb{E}[\mathbb{P}(X_{k+1} \neq Y_{k+1} | X^{k+1}, Y^k)^2].$$

Combining the bounds above:
$$\mathbb{E}[\mathbb{P}(X_{k+1} \neq Y_{k+1} | X^{k+1})^2 + \mathbb{P}(X_{k+1} \neq Y_{k+1} | Y^{k+1})^2]$$
$$\leq 2 D(Q_{Y_{k+1}|Y^k} \| P_{X_{k+1}} | Q_{Y^k}).$$

Furthermore, by the Markov condition above
$\mathbb{P}(X_i \neq Y_i | Y^k) = \mathbb{P}(X_i \neq Y_i | Y^{k+1})$ $\forall$ $i \leq k$ since
$Y_{k+1} - Y^k - X^k$.

Also, $P_{X_{k+1}|X^k Y^k}(x_{k+1} | x^k, y^k) = P_{X_{k+1}}(x_{k+1})$, i.e.,
X_{k+1} is indep. of (X^k, Y^k) whereby
$\mathbb{P}(X_i \neq Y_i | X^k) = \mathbb{P}(X_i \neq Y_i | X^{k+1})$ $\forall$ $i \leq k$.

Combining the bounds above, we get
$$\sum_{i=1}^{k+1} \mathbb{E}[\mathbb{P}(X_i \neq Y_i | X^{k+1})^2 + \mathbb{P}(X_i \neq Y_i | Y^{k+1})^2]$$
$$\leq 2 [D(Q_{Y^k} \| P_{X^k}) + D(Q_{Y_{k+1}|Y^k} \| P_{X_{k+1}|X^k} | Q_{Y^k})].$$

∎

Claim 2: Inequality holds for $n = 1$

Define

$$d_2(Q,P) \overset{\text{def}}{=} \sqrt{\sum_x \frac{(P(x)-Q(x))_+^2}{P(x)}}$$

$$\equiv \left\| \left(1-\frac{dQ}{dP}\right)_+ \right\|_2$$

Lemma. $\min_{(X,Y)\in \mathcal{P}(P,Q)} \mathbb{E}\left[\mathbb{P}(X\neq Y|X)^2 + \mathbb{P}(X\neq Y|Y)^2\right]$

Optimal transportation cost

$$= d_2^2(Q,P) + d_2^2(P,Q).$$

→ Before we prove this, we see how we use it.

Let $r(x)$ denote $Q(x)/P(x)$. Then,

$$d_2^2(Q,P) + d_2^2(P,Q) = \mathbb{E}_P\left[(1-r(X))_+^2 + \frac{(r(X)-1)_+^2}{r(X)}\right]$$

Let $h(t) = (1-t)\log(1-t) + t$, $t<1$. Note that

$h(t) \geq t^2/2$ if $0 \leq t < 1$, and

$h(-t) \geq \frac{t^2}{2(1+t)}$ if $t \leq 0$.

Thus,
$$(1-r(x))_+^2 \leq 2h((1-r(x))_+) \text{ and}$$

$$(r(x)-1)_+^2/r(x) \leq 2h(-(r(x)-1)_+),$$

which gives

$$d_2^2(Q,P) + d_2^2(P,Q) \leq 2\,\mathbb{E}_P\left[h((1-r(X))_+) + h(-(r(X)-1)_+)\right]$$

$$= 2\sum_x P(x)\left[r(x)\log r(x) + (1-r(x))\right]$$

$$= 2D(Q\|P)$$

Optimal transportation cost inequality

Proof of Lemma. Given $(X,Y) \in \mathcal{P}(P,Q)$,

$$\mathbb{P}(X=Y \mid X=x) = \frac{\mathbb{P}(X=x,\, Y=x)}{P(x)} \leq \min\left\{1,\, \frac{Q(x)}{P(x)}\right\}.$$

Therefore,
$$\mathbb{E}\left[\mathbb{P}(X\neq Y\mid X)^2\right] \geq \mathbb{E}_P\left[\left(1-\frac{Q(X)}{P(X)}\right)_+^2\right]$$
$$= d_2^2(Q,P)$$

and similarly,
$$\mathbb{E}\left[\mathbb{P}(X\neq Y\mid Y)^2\right] \geq d_2^2(P,Q).$$

Next, we exhibit a coupling which attains $d_2^2(Q,P)+d_2^2(P,Q)$.

In fact, this coupling is the same as the one we used in the maximal coupling lemma.

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

$$\mathbb{P}_2(x,y) = \frac{(P(x)-Q(x))_+\,(Q(x)-P(x))_+}{d_{TV}^2(P,Q)}$$

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

We have
$$\mathbb{P}(X\neq Y\mid X=x) = \left[\overbrace{\sum_y \frac{(Q(y)-P(y))_+}{d_{TV}(P,Q)}}^{1}\right]\cdot \frac{(P(x)-Q(x))_+}{P(x)}$$

and
$$\mathbb{P}(X\neq Y\mid Y=y) = \frac{(Q(y)-P(y))_+}{P(y)}.$$

Thus, $\mathbb{E}\left[\mathbb{P}(X\neq Y\mid X)^2 + \mathbb{P}(X\neq Y\mid Y)^2\right] = d_2^2(Q,P) + d_2^2(P,Q).$

■

What we covered in this course

* Chernoff bounds : Hoeffding, Bernstein, Bennett,
  Martingale method : Azuma, McDiarmid

* Efron-Stein inequality

* Entropy method, Herbst's inequality and log-Sobolev,
  modified log-Sobolev inequality
  - binary LSI
  - Gaussian LSI
  - conc. of (a,b)-weakly self bounding
  - $Z = f(X)$, $Z_i' = \min_{x_i'} f(X^{i-1}, x_i', X_{i+1}^n)$

    Suppose $\sum_{i=1}^n (Z - Z_i')^2 \leq v$ a.s.

    Then, $$\mathbb{P}(Z > \mathbb{E}Z + t) \leq e^{-t^2/2v}.$$

* Isoperimetry and concentration

  $$(\#) \quad f(x) - f(y) \leq \sum_{i=1}^n c_i(x) \mathbb{1}_{\{x_i \neq y_i\}}$$

  Let $v_\infty = \sup_x \sum_{i=1}^n c_i(x)^2$

  $$\mathbb{P}(f(X) > \mathbb{E}f(X) + t) \leq 2 . e^{-t^2/4v_\infty}.$$

* Transportation method

  Assume $(\#)$. Let $v = \sum_{i=1}^n \mathbb{E}[c_i(X)^2]$.

  $$\mathbb{P}(f(X) < \mathbb{E}f(X) - t) \leq e^{-t^2/2v}.$$

* Extra classes : Hypercontractivity, Analysis of LASSO

* Not covered : Boolean analysis, Gaussian conc., Gaussian Iso.