Lecture 2

Review: * P vs Q: $P_e^*(\frac{1}{2},\frac{1}{2}) = \frac{1}{2}\left(1 - d(P,Q)\right)$

Thus, we need to take as many samples n as needed to make $d(P^n, Q^n)$ constant

* $\boxed{d(P^n, Q^n) \le n\, d(P,Q)}$ (#)

Agenda: * Kullback-Leibler divergence $D(P\|Q)$

- Data processing inequality
- Pinsker's inequality and improvement over (#)
- Fano's inequality **Bonus: A new proof !!**

A Kullback-Leibler Divergence

$$D(P\|Q) = \begin{cases} \sum_x P(x)\log\frac{P(x)}{Q(x)}, & \text{iff } supp(P) \subseteq supp(Q) \\ \infty, & \text{o.w.} \end{cases}$$

$$= \begin{cases} \int f(x)\log\frac{f(x)}{g(x)}\,\mu(dx), & \text{iff } supp(f) \subseteq supp(g) \\ \infty, & \text{o.w.} \end{cases}$$

$$= \begin{cases} \mathbb{E}_Q\left[\frac{dP}{dQ}\log\frac{dP}{dQ}\right], & \text{if } P \ll Q \\ \infty, & \text{o.w.} \end{cases}$$

(a) Data Processing Inequality

Let $W: \mathcal{X} \to \mathcal{Y}$ be a fixed channel.

Denote by PW the distribution $\sum_{x} P(x) W(y|x)$.

$\rightarrow$ $d(PW, QW) \le d(P, Q)$ (triangle inequality)

$\rightarrow$ $D(PW, QW) \le D(P, Q)$

Pf. Follows from the log-sum inequality

$$\sum a_i \log \frac{a_i}{b_i} \ge \sum a_i \log \frac{\sum a_i}{\sum b_i} \quad \text{for } a_i, b_i \ge 0.$$

(b) Chain rules

$$\rightarrow d(P_{X_1 \dots X_n}, Q_{Y_1 \dots Y_n}) \le \sum_{i=1}^{n} d(P_{X^i}, P_{X^{i-1}} Q_{Y_i|X^{i-1}})$$

$$\rightarrow D(P_{X_1 \dots X_n}, Q_{Y_1 \dots Y_n}) = \mathbb{E}\left[\log \frac{P_{X_1 \dots X_n}(X^n)}{Q_{Y_1 \dots Y_n}(X^n)}\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} \log \frac{P_{X_i|X^{i-1}}(X_i|X^{i-1})}{Q_{Y_i|Y^{i-1}}(X_i|X^{i-1})}\right]$$

$$= \sum_{i=1}^{n} \underbrace{\mathbb{E}_{x^{i-1} \sim P_{X^{i-1}}}\left[D(P_{X_i|X^{i-1}=x^{i-1}} \| Q_{Y_i|Y^{i-1}=x^{i-1}})\right]}_{=: D(P_{X_i|X^{i-1}} \| Q_{Y_i|Y^{i-1}} | P_{X^{i-1}})}$$

$$= \sum_{i=1}^{n} D(P_{X^i} \| Q_{Y_i|Y^{i-1}} P_{X^{i-1}})$$

(c) Pinker's inequality

$$\boxed{d^2(P,Q) \le \frac{1}{2\ln 2} D(P\|Q)}$$

How this improves over (#)

$$d^2(P^n, Q^n) \le \frac{1}{2\ln 2} \cdot D(P^n \| Q^n)$$

$$= \frac{n\, D(P\|Q)}{2\ln 2}$$

$$\Rightarrow \boxed{d(P^n, Q^n) \le \sqrt{\frac{n\, D(P\|Q)}{2\ln 2}}} \quad (\#\#)$$

[If $D(P\|Q)$ is of the same order as $d^2(P,Q)$, then (##) is an improvement over (#).]

Proof of Pinsker's inequality

Step 1. For any $A \subseteq \mathcal{X}$,

$$D(P\|Q) \ge D(P(A) \| Q(A))$$

$\rightarrow P(A) \log \frac{P(A)}{Q(A)} + (1-P(A)) \log \frac{1-P(A)}{1-Q(A)}$

by the data processing inequality.

Step 2. $D(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$

Suffices: $\quad p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q} \ge 2(p-q)^2$

Proof: $f(p,q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q} - 2(p-q)^2$

$$\frac{df}{dq} = -\frac{p}{q} + \frac{(1-p)}{1-q} + 4(p-q)$$

$$= (p-q) \underbrace{\left[4 - \frac{1}{q(1-q)} \right]}_{\ge 0}$$

$\Rightarrow \frac{df}{dq} \ge 0$ iff $p \ge q \Rightarrow f(p,q) \ge f(q,q) = 0$. ■

(d) Convexity of $D(P\|Q)$

$D(P\|Q)$ is convex in (P,Q). (Proof uses only log-sum inequality)

B Fano's inequality

Recall that

$$P_e^*\left(\tfrac{1}{2},\tfrac{1}{2}\right) \geq \frac{1}{2}\left(1-d(P,Q)\right)$$

$$\geq \frac{1}{2}\left(1-\sqrt{\frac{1}{2\ln 2}D(P\|Q)}\right)$$

This bound allows us to quantize the difficulty of hypothesis testing in terms of "distance" $D(P\|Q)$.

The next result provides a similar bound for M-ary hypothesis testing.

Problem. $\mathcal{H}_m : X \sim P_m, \quad m=1,\dots,M$

$d: X \to \{1,\dots,M\}$ be a randomized map.

$$P_e^*(\text{unif}) = \inf_d \frac{1}{M}\sum_{m=1}^{M} P_m\left(d(X)\neq m\right)$$

Theorem (Fano's inequality)

$$P_e^*(\text{unif}) \geq 1 - \frac{\frac{1}{M}\sum_{m=1}^{M} D\left(P_m \Big\| \frac{1}{M}\sum_{m'=1}^{M} P_{m'}\right)+1}{\log M}$$

Remark. Think of $U\sim \text{unif}[M]$ as input to a channel which then produces the output $X \sim P_u$. The quantity

$$\frac{1}{M}\sum_{m=1}^{M} D\left(P_m \,\Big\|\, \frac{1}{M}\sum_{m'=1}^{M} P_{m'}\right)$$

is then called the <u>mutual information</u> between U and X, denoted $I(U \wedge X)$. Note the following properties of $I(U \wedge X)$:

$\rightarrow$ $I(U \wedge X) = D(P_{UX} \| P_U P_X)$

(for discrete U, X) $= H(X) - \underbrace{H(X|U)}_{\mathbb{E}_{(X,U)\sim P_{XU}}\left[\log \frac{1}{P_{X|U}(X|U)}\right]}$

$\rightarrow$ $I(U \wedge X) \le \max\limits_{u,u'} D(P_u \| P_{u'})$, using the convexity of $D(P\|Q)$.

$\rightarrow$ $I(U \wedge XY) = I(U \wedge X) + \underbrace{I(U \wedge Y|X)}_{\sum_x P_X(x)\, \underbrace{I(U \wedge Y | X=x)}_{I \text{ under } P_{UY|X=x}}}$

This follows from the chain rule of KL divergence.

<u>Proof of Fano's inequality</u>. Restrict first to a deterministic d. $U \sim \mathrm{unif}\{1, \ldots, M\}$; Let $Q_{UX} = P_U P_X$

Let $B = \mathbb{1}(U = d(X))$. Then,

$I(U \wedge X) = D(P_{UX} \| Q_{UX}) \ge D(P_B \| Q_B)$ (by data processing inequality)

Denote $\mathbb{E}_P[B] = p$, $\mathbb{E}_Q[B] = q$.

Then, the right-side above equal

$$p\log p + (1-p)\log(1-p) + p\log\frac{1}{q} + (1-p)\log\frac{1}{1-q}$$

$$\geq p\log\frac{1}{q} - h(p)$$

Note that $p = P(U = d(X))$ and

$$q = Q(U = d(X)) = \frac{1}{M}\sum_{m=1}^{M} Q_X(D_m)$$

$$\leq \frac{1}{M}$$

$$\Rightarrow I(U \wedge X) \geq (1 - P_e(d))\log M - h(p)$$

$$\Leftrightarrow P_e(d) \geq 1 - \frac{I(U\wedge X) + h(p)}{\log M}$$

$$\geq 1 - \frac{I(U \wedge X) + 1}{\log M},$$

which completes the proof for deterministic d.

For a randomized d, note that $\exists\, v$ s.t.

$$\underbrace{P_e(d)}_{\text{randomization using } V} = \mathbb{E}_V[P_e(d_V)] \geq P_e(\underbrace{d_v}_{\text{deterministic rule}})$$

$\blacksquare$