Information and Coding Theory **Autumn 2014**

Lecture 5: October 14, 2014

Lecturer: Madhur Tulsiani Scribe: Gökalp Demirci

1 Pinsker's inequality and its applications to lower bounds

1.1 Chain rule for KL-divergence

We first derive a chain rule for KL-divergence which will be helpful in the proof of Pinkser's inequality. In this lectute, we use capital letters like $P(X)$ to denote a distribution for the random variable X and lowercase letters like $p(x)$ to denote the probability for a specific element x. Let $P(X,Y)$ and $Q(X,Y)$ be two distributions for a pair of variables X and Y. We then have

$$\begin{aligned}
D(P(X,Y)||Q(X,Y)) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{q(x,y)} \\
&= \sum_{x,y} p(x)p(y|x) \log \frac{p(x)}{q(x)} \frac{p(y|x)}{q(y|x)} \\
&= \sum_{x} p(x) \log \frac{p(x)}{q(x)} \sum_{y} p(y|x) + \sum_{x} p(x) \sum_{y} p(y|x) \log \frac{p(y|x)}{q(y|x)} \\
&= D(P(X)||Q(X)) + \sum_{x} p(x) D(P||Q|X=x) \\
&= D(P(X)||Q(X)) + D(P||Q|X) \qquad (1)
\end{aligned}$$

Here $P(X)$ and $Q(X)$ denote the marginal distributions for the first variable. Also, $D(P||Q|X)$ denotes the expected (according to P) value of $D(P(Y|X=x)||Q(Y|X=x))$.

Note that if $P(X,Y) = P_1(X)P_2(Y)$ and $Q(X,Y) = Q_1(X)Q_2(Y)$, then $D(P||Q) = D(P_1||Q_1) + D(P_2||Q_2)$.

1.2 Pinsker's inequality

Lemma 1.1 (Pinsker's inequality) *Let P and Q be two distributions defined on the universe U. Then,*

$$D(P||Q) \geq \frac{1}{2\ln 2} \cdot {||P-Q||_1}^2. \qquad (2)$$

Proof: A special case:

$$P = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

and

$$Q = \begin{cases} 1 & \text{w.p. } q \\ 0 & \text{w.p. } 1-q \end{cases}$$

We assume $p \geq q$ (other case is similar), and let

$$f(p,q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} - \frac{1}{2 \ln 2} \left(2(p-q)\right)^2.$$

Since

$$\frac{\partial f}{\partial q} = -\frac{p-q}{\ln 2} \left(\frac{1}{q(1-q)} - 4 \right) \leq 0,$$

and $f = 0$ when $q = p$, we conclude that $f(p,q) \geq 0$ where $q \leq p$. Thus, we have that $D(P||Q) \geq \frac{1}{2 \ln 2} {||P-Q||_1}^2$ for this special case.

<u>General case:</u> Let P and Q be distributions on U. Let $A \subset U$ be

$$A = \{x \mid p(x) \geq q(x)\}.$$

and P_A and Q_A be

$$P_A = \begin{cases} 1 & \text{w.p. } \sum_{x \in A} p(x) \\ 0 & \text{w.p. } \sum_{x \notin A} p(x) \end{cases}$$

$$Q_A = \begin{cases} 1 & \text{w.p. } \sum_{x \in A} q(x) \\ 0 & \text{w.p. } \sum_{x \notin A} q(x) \end{cases}$$

Then,

$$\begin{aligned} ||P-Q||_1 &= \sum_x |p(x) - q(x)| \\ &= \sum_{x \in A} p(x) - q(x) + \sum_{x \notin A} q(x) - p(x) \\ &= \left| \sum_{x \in A} p(x) - \sum_{x \notin A} q(x) \right| + \left| \left(1 - \sum_{x \in A} p(x)\right) - \left(1 - \sum_{x \notin A} q(x)\right) \right| \\ ||P-Q||_1 &= ||P_A - Q_A||_1 \end{aligned} \tag{3}$$

Define a random variable Z as $Z = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$. We have that $D(P||Q) = D(P(Z)||Q(Z)) + D(P||Q|Z)$. Since $D(P(Z)||Q(Z)) = D(P_A||Q_A)$ and $D(P||Q|Z) \geq 0$, we have

$$\begin{aligned} D(P||Q) &\geq D(P_A||Q_A) \\ &\geq \frac{1}{2 \ln 2} \cdot {||P_A - Q_A||_1}^2 && \text{(use the special case)} \\ &= \frac{1}{2 \ln 2} \cdot {||P-Q||_1}^2 && \text{(use equation 3)} \end{aligned}$$

.

$\blacksquare$

1.3 Coin tossing

We will now use Pinsker's inequality to derive a lower bound on the number of samples neede to distinguish two coins with slightly differing biases. You can use Chernoff bounds to see that this bound is optimal. The optimality will also follow from a much more general result known as Sanov's theorem which we will derive in the next few lectures. Suppose we are given one of the following two coins (think of 1 as "heads" and 0 as "tails"):

$$P = \begin{cases} 1 & \text{w.p. } \frac{1}{2} - \varepsilon \\ 0 & \text{w.p. } \frac{1}{2} + \varepsilon \end{cases}$$

or

$$Q = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ 0 & \text{w.p. } \frac{1}{2} \end{cases}$$

Suppose we have an algorithm $A(x_1, x_2, ...x_m) \to \{0, 1\}$ that takes the output of m independent coin tosses, and makes a decision about which coint the tosses came from. Suppose that A outputs 0 to indicate the coin with distribution P and 1 to indicate the coin with distribution Q. Let us say that A identifies both coins with probability at least 9/10, i.e.,

$$\mathop{\mathbb{P}}_{x \in P^m}[A(x) = 0] \geq \frac{9}{10} \quad \text{and} \quad \mathop{\mathbb{P}}_{x \in Q^m}[A(x) = 1] \geq \frac{9}{10}$$

The goal is to derive a lower bound for m. We will be able to derive a lower bound without knowing anything about A. We first rewrite the above conditions as

$$\mathop{\mathbb{E}}_{x \in P^m}[A(x)] \leq \frac{1}{10} \quad \text{and} \quad \mathop{\mathbb{E}}_{x \in Q^m}[A(x)] \geq \frac{9}{10},$$

which gives

$$\mathop{\mathbb{E}}_{x \in Q^m}[A(x)] - \mathop{\mathbb{E}}_{x \in P^m}[A(x)] \;\geq\; \frac{8}{10}.$$

The following lemma shows that $\|P^m - Q^m\|_1$ must be large for the above condition to be true.

Lemma 1.2 *Let $\tilde{P}, \tilde{Q}$ be any distributions on U. Let $f : U \to [0, B]$. Then*

$$\left| \mathop{\mathbb{E}}_{\tilde{P}}[f(x)] - \mathop{\mathbb{E}}_{\tilde{Q}}[f(x)] \right| \;\leq\; \frac{B}{2} \cdot \left\| \tilde{P} - \tilde{Q} \right\|_1. \tag{4}$$

Proof:

$$\begin{aligned} \left| \sum_x \tilde{p}(x) f(x) - \sum_x \tilde{q}(x) f(x) \right| &= \left| \sum_x (\tilde{p}(x) - \tilde{q}(x))\, f(x) \right| \\ &= \left| \sum_x (\tilde{p}(x) - \tilde{q}(x)) \left(f(x) - \frac{B}{2} \right) + \frac{B}{2} \left(\sum_x \tilde{p}(x) - \tilde{q}(x) \right) \right| \\ &\leq \sum_x |\tilde{p}(x) - \tilde{q}(x)| \left| f(x) - \frac{B}{2} \right| \\ &\leq \frac{B}{2} \cdot ||\tilde{P} - \tilde{Q}||_1 \end{aligned} \tag{5}$$

$\blacksquare$

Taking $f = A$, $\tilde{P} = P^m$ and $\tilde{Q} = Q^m$, the above lemma gives

$$\|P^m - Q^m\|_1 \;\geq\; 2\cot \left| \mathop{\mathbb{E}}_{x\in P^m}[A(x)] - \mathop{\mathbb{E}}_{x\in Q^m}[A(x)] \right| \;=\; \frac{8}{5}\,.$$

Using the chain rule for KL-divergence and Pinsker's inequality, we get

$$m \cdot D(P||Q) = D\left(P^m||Q^m\right) \geq \frac{1}{2\ln 2} \cdot \left(\frac{8}{5}\right)^2 \quad \Rightarrow \quad m \geq \frac{1}{2\ln 2 \cdot D(P||Q)} \cdot \left(\frac{8}{5}\right)^2$$

Finally, it remains to give an upper bound on $D(P||Q)$, which can be obtained by writing it out as

$$\begin{aligned}
D(P||Q) \;&=\; \left(\frac{1}{2} - \varepsilon\right)\log\left(\frac{\frac{1}{2}-\varepsilon}{\frac{1}{2}}\right) + \left(\frac{1}{2}+\varepsilon\right)\log\left(\frac{\frac{1}{2}+\varepsilon}{\frac{1}{2}}\right) \\
&=\; \frac{1}{2}\log\left((1-2\varepsilon)(1+2\varepsilon)\right) + \varepsilon\log\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right) \\
&\leq\; \frac{\varepsilon}{\ln 2}\ln\left(1 + \frac{4\varepsilon}{1-2\varepsilon}\right) \\
&\leq\; \frac{4\varepsilon^2}{\ln 2}\frac{1}{1-2\varepsilon} && (\text{used } \ln 1 + x \leq e^x) \\
D(P||Q) \;&\leq\; \frac{8\varepsilon^2}{\ln 2} && \left(\text{assumed } \varepsilon \leq \frac{1}{4}\right)
\end{aligned}$$

Pluggin in this upper bound, we get

$$m \;\geq\; \frac{1}{2\ln 2 \cdot D(P||Q)} \cdot \left(\frac{8}{5}\right)^2 \;\geq\; \frac{4}{25\varepsilon^2}\,.$$

As discussed earlier, one can show that this bound is upto constants using the Chernoff bound. In the next lecture, we'll see another application of Pinsker's inequality on multi-armed bandits problem.