Spring 2016

6.441 - Information Theory

Homework 2

Due: Thur, Feb 18, 2016 (in class)

`Prof. Y. Polyanskiy`

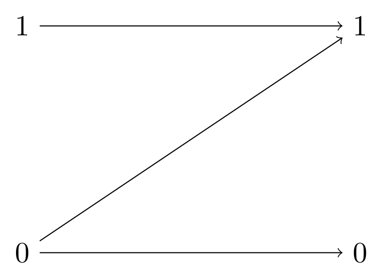
1 Reading (optional)

1. Read [1, Chapter 2]

2 Exercises

NOTE: Each exercise is 10 points. Only 3 exercises per assignment will be graded. If you submit more than 3 solved exercises please indicate which ones you want to be graded.

1 Consider the following *Z-channel* given by $P_{Y|X}[1|1] = 1$ and $P_{Y|X}[1|0] = P_{Y|X}[0|0] = 1/2$.



1. Find the capacity
$$C = \max_X I(X;Y)\,.$$

2. Find $D(P_{Y|X=0}||P_Y^*)$ and $D(P_{Y|X=1}||P_Y^*)$ where P_Y^* is the capacity-achieving output distribution, or *caod*, i.e., the distribution of Y induced by the maximizer of $I(X;Y)$.

2 Consider the following *multiple-access channel*

$$Y = X_1 + X_2\,,$$

where $X_i \in \{0,1\}$ and $Y \in \{0,1,2\}$ (addition over $\mathbb{Z}$). Compute $\max_{X_1,X_2} I(X_1,X_2;Y)$ over:

a) all distributions P_{X_1,X_2};

b) all *product* distributions $P_{X_1,X_2} = P_{X_1}P_{X_2}$.

Let (X_1^*, X_2^*, Y^*) be the random variables achieving the maximum for case b (thus $X_1^* \perp\!\!\!\perp X_2^*$). Is it true that

$$D(P_{Y|X_1=a,X_2=b}||P_{Y_*}) \le I(X_1^*, X_2^*; Y^*)$$

for all $a, b \in \{0,1\}$? Does this violate the capacity saddle point theorem?

3 For any Gaussian random variable $X_{\mathtt{G}}$ and any random variable Y with finite variance, show that

$$I(X_{\mathtt{G}};Y_{\mathtt{G}}) \le I(X_{\mathtt{G}};Y)$$

where $Y_{\mathtt{G}}$ is jointly Gaussian with $X_{\mathtt{G}}$ with the same mean and variance as Y and $E[X_{\mathtt{G}}Y_{\mathtt{G}}] = E[X_{\mathtt{G}}Y]$. Does the claim also hold if $Y_{\mathtt{G}}$ is Gaussian but not jointly Gaussian with $X_{\mathtt{G}}$?

4 Consider a binary symmetric random walk X_n on $\mathbb{Z}$ that starts at zero. In other words, $X_n = \sum_{j=1}^n B_j$, where $(B_1, B_2, \ldots)$ are independent and equally likely to be ± 1. Question: When $n \gg 1$ does knowing X_{2n} provide any information about X_n? Namely, prove or disprove:

$$\lim_{n\to\infty} I(X_n; X_{2n}) = 0$$

(Hint: lower-semicontinuity and central-limit theorem).
Bonus: Try to prove an asymptotically tight lower bound on $I(X_n; X_{2n})$.

5 (Combinatorial meaning of entropy)

1 Fix $n \ge 1$ and $0 \le k \le n$. Let $p = \frac{k}{n}$ and define $T_p \subset \{0,1\}^n$ to be the set of all binary sequences with p fraction of ones. Show that if $k \in [1, n-1]$ then

$$|T_p| = \sqrt{\frac{1}{np(1-p)}} \exp\{nh(p)\} C(n,k)$$

where $C(n,k)$ is bounded by two universal constants $C_0 \le C(n,k) \le C_1$, and $h(\cdot)$ is the binary entropy. Conclude that for all $0 \le k \le n$ we have

$$\log |T_p| = nh(p) + O(\log n)\,.$$

Hint: Stirling's approximation:

$$\mathrm{e}^{\frac{1}{12n+1}} \le \frac{n!}{\sqrt{2\pi n}(n/\mathrm{e})^n} \le \mathrm{e}^{\frac{1}{12n}}, \quad n \ge 1$$

2 Let $Q^n = \mathrm{Bern}(q)^n$ be iid Bernoulli distribution on $\{0,1\}^n$. Show that

$$\log Q^n[T_p] = -nd(p\|q) + O(\log n)$$

3* (optional) More generally, let $\mathcal{X}$ be a finite alphabet, $\hat{P}, Q$ distributions on $\mathcal{X}$, and $T_{\hat{P}}$ a set of all strings in $\mathcal{X}^n$ with composition $\hat{P}$. If $T_{\hat{P}}$ is non-empty (i.e. if $n\hat{P}(\cdot)$ is integral) then

$$\log |T_{\hat{P}}| = nH(\hat{P}) + O(\log n)$$
$$\log Q^n[T_{\hat{P}}] = -nD(\hat{P}\|Q) + O(\log n)$$

and furthermore, both $O(\log n)$ terms can be bounded as $|O(\log n)| \le |\mathcal{X}| \log(n+1)$. (Hint: show that number of non-empty $T_{\hat{P}}$ is $\le (n+1)^{|\mathcal{X}|}$.)

6 (Effective de Finetti) We will show that for any distribution P_{X^n} invariant to permutation and $k < n$ there exists a distribution Q_{X^k}, which is a mixture of iid's, approximating P_{X^k} well:

$$\mathrm{TV}(P_{X^k}, Q_{X^k}) \le \sqrt{\frac{k^2 \log_e |\mathcal{X}|}{2n}} \tag{1}$$

Follow the steps:

1. Show the identity (here P_{X^k} is arbitrary)

$$D(P_{X^k} \| \prod P_{X_j}) = \sum_{j=1}^{k-1} I(X^j; X_{j+1})$$

2. Show that there must exist some $t \in [k, n]$ such that

$$I(X^{k-1}; X_k | X_{t+1}^n) \leq \frac{k \log |\mathcal{X}|}{n-k}$$

(Hint: expand $I(X^{k-1}; X_k^n)$ via chain rule, use permutation invariance and $H(X^k) \leq k \log |\mathcal{X}|$.)

3. Show from 1 and 2 that

$$D(P_{X^k|T} \| \prod P_{X_j|T} | P_T) \leq \frac{k^2 \log |\mathcal{X}|}{n-k}$$

where $T = X_{t+1}^n$.

4. By Pinsker inequality

$$\mathbb{E}_T[\mathrm{TV}(P_{X^k|T}, \prod P_{X_j|T})] \leq \sqrt{\frac{k^2 \log_e |\mathcal{X}|}{2(n-k)}}$$

Conclude the proof of (1) by convexity of total variation.

References

[1] T. Cover and J. Thomas, *Elements of Information Theory,* Second Edition, Wiley, 2006