Spring 2016
6.441 - Information Theory
Homework 1
Due: Tue, Feb 9, 2016 (in class)
`Prof. Y. Polyanskiy`

1 Reading (optional)

1. Read [2, Chapter 1]
2. Read [3]

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 Bob is to eat all the cookies from a jar containing three peanut butter, two chocolate, and one oatmeal cookies. He decides to proceed completely randomly. Denote by X and Y the flavor of the first and the second cookie he eats.

1. Find $H(X)$, $H(Y)$, $H(X,Y)$, $H(Y|X)$, $H(X|Y)$, $I(X;Y)$, $D(P_{Y|X=\text{chocolate}}||P_{Y|X=\text{oatmeal}})$ and $D(P_{Y|X=\text{oatmeal}}||P_{Y|X=\text{chocolate}})$.
2. Now, what if Y denotes the flavor of the last cookie Bob eats?
3. How much information is provided by the sequence in which the cookies are eaten?

2 Let $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ be the Gaussian distribution on $\mathbb{R}^n$ with mean $\mathbf{m} \in \mathbb{R}^n$ and covariance matrix $\mathbf{\Sigma}$.

1. Under what conditions on $\mathbf{m}_0, \mathbf{\Sigma}_0, \mathbf{m}_1, \mathbf{\Sigma}_1$ is
$$D(\ \mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)\ ||\ \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)\ ) < \infty$$
2. Compute $D(\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})||\mathcal{N}(0, \mathbf{I}_n))$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.
3. Compute $D(\ \mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1)\ ||\ \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)\ )$ for a non-singular $\mathbf{\Sigma_0}$. (Hint: think how Gaussian distribution changes under shifts $\mathbf{x} \mapsto \mathbf{x} + \mathbf{a}$ and non-singular linear transformations $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. Apply data-processing to reduce to previous case.)

3 Recall that $d(p||q) = D(\text{Bern}(p)||\text{Bern}(q))$ denotes the binary divergence function:
$$d(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}\,.$$

1. Prove for all $p, q \in [0,1]$
$$d(p||q) \geq 2(p-q)^2 \log e\,.$$
Note: Proof by drawing is NOT accepted.

2. Apply data processing inequality to prove the *Pinsker-Csiszár inequality*:

$$\mathrm{TV}(P,Q) \le \sqrt{\frac{1}{2\log e} D(P||Q)}\,,$$

where $\mathrm{TV}(P,Q)$ is the *total variation* distance between probability distribution P and Q:

$$\mathrm{TV}(P,Q) \triangleq \sup_E \left(P[E] - Q[E]\right)\,,$$

with the supremum taken over all events E.

4 (Information lost in erasures) Let X, Y be a pair of random variables with $I(X;Y) < \infty$. Let Z be obtained from Y by passing the latter through an erasure channel, i.e., $X \to Y \to Z$ where

$$P_{Z|Y}(z|y) = \begin{cases} 1-\delta, & z = y\,, \\ \delta, & z = ? \end{cases}$$

where ? is a symbol not in the alphabet of Y. Find $I(X;Z)$.

5 1. Someone arranged a set of n points in $\mathbb{R}^3$ in such a way that any of its projections on xy, xz and yz-planes has cardinality m. Obviosly, $m \le n$. Show that also

$$n \le m^{\frac{3}{2}} \tag{1}$$

(Hint: Han's inequality)

2. Show that when $\sqrt{m}$ is integer there exists a configuration achieving (1) with equality.
3. More generally, prove Shearer's lemma: For n points in $\mathbb{R}^3$ let m_1, m_2, m_3 denote the number of distinct points projected onto the xy, xz and yz-plane, respectively. Then:

$$n \le \sqrt{m_1 m_2 m_3}\,. \tag{2}$$

Comments: This is an example of an information-theoretic proof of a combinatorial result.

6 Let (X,Y) be uniformly distributed inside the unit circle $\{(x,y) : x^2 + y^2 \le 1\}$.

1. Are they independent? Explain your answer.
2. Compute $I(X;Y)$.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication", *Bell Syst. Tech. J.*, pp. 379-423, 623-656, vol. 27, Jul.-Oct. 1948.

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

[3] Solomon W. Golomb, Elwyn Berlekamp, Thomas M. Cover, Robert G. Gallager, James L. Massey, and Andrew J. Viterbi, *Claude Elwood Shannon (1916–2001)*, NOTICES OF THE AMS, Vol. 49, No. 1, 2002