ECE 5630 - Homework Assignment 3

March 9th 2020

Due to: Thursday, March 19th, 2020 (at the beginning of the lecture)

Instructions: Submission in pairs is allowed. Prove and explain every step in your answers.

Upcoming Prelim: All the questions in this homework sheet are at prelim level. Use them for practice. Additional exercises at this level can be found in the 'Problems' section in Chapter 2 of:

T. M. Cover and J. Thomas, "Elements of Information Theory", *Wiley*, 2nd edition, 2006, New York, NY, US.

1) **Entropy (full) chain rule:** Let $(X_1,\dots,X_k)\sim P_{X_1,\dots,X_n}$. Show that:
   a) If $(X_1,\dots,X_k)$ is discrete, then its Shannon entropy decomposes as $H(X_1,\dots,X_k)=\sum_{i=1}^{k}H(X_i|X_1,\dots,X_{i-1})$, where $H(X_1|X_0)=H(X_1)$.
   b) If $(X_1,\dots,X_k)$ is jointly continuous, then its differential entropy decomposes as $h(X_1,\dots,X_k)=h(X_k)+\sum_{i=1}^{k-1}h(X_{k-i}|X_k,\dots,X_{k-(i-1)})$.

2) **Properties of mutual information:** Let $(X,Y,Z)\sim P_{X,Y,Z}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})$. Establish the following relations:
   a) <u>KL divergence chain rule:</u> For any $Q_{X,Y}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, we have $D_{\mathsf{KL}}(P_{X,Y}\|Q_{X,Y})=D_{\mathsf{KL}}(P_X\|Q_X)+D_{\mathsf{KL}}(P_{Y|X}\|Q_{Y|X}|P_X)$.
   b) <u>Relation to conditional KL divergence:</u> $I(X;Y)=D_{\mathsf{KL}}(P_{Y|X}\|P_Y|P_X)$, where $P_{X,Y}=P_XP_{Y|X}$ and P_Y is its Y-marginal.

      Hint: Use section (a).
   c) <u>Symmetry:</u> $I(X;Y)=I(Y;X)$.
   d) <u>More data $\implies$ more information:</u> $I(X;Y)\leq I(X;Y,Z)$.
   e) <u>Mutual information and functions:</u> $I(X;Y)\geq I\big(X;f(Y)\big)$ for any deterministic function f. Furthermore, if f is continuous and one-to-one, then $I\big(X;f(X)\big)=H(X)$ for discrete X, and $I\big(X;f(X)\big)=\infty$ for continuous X. Do <u>not</u> use mutual information Data Processing Inequality (DPI) in your proof.

      Hint: You may use f-divergence DPI.

3) **Entropy of a sum:** Let $(X,Y)\sim P_{X,Y}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, where $\mathcal{X}=\{x_1,\dots,x_r\}$ and $\mathcal{Y}=\{y_1,\dots,y_s\}$, and define $Z=X+Y$.
   a) Show that $H(Z)\leq H(X)+H(Y)$, for arbitrarily correlated (X,Y), and that $\max\{H(X),H(Y)\}\leq H(Z)$, when (X,Y) are independent.
   b) Show that $H(Z|X)=H(Y|X)$. Argue that if (X,Y) are independent, then $H(Y)\leq H(Z)$ and $H(X)\leq H(Z)$. Thus the summation of *independent* random variables increases uncertainty.
   c) Give an example of dependent random variables for which $H(X)>H(Z)$ and $H(Y)>H(Z)$.

4) **Information inequalities:** Let $(X, Y, Z) \sim P_{X,Y,Z} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$. Prove the following inequalities and find (necessary and sufficient) conditions for equality.
   a) $H(X, Y|Z) \geq H(X|Z)$.
   b) $I(X, Y; Z) \geq I(X; Z)$.
   c) $H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$.
   d) $I(X; Z|Y) \geq I(Z; Y|X) - I(Z; Y) + I(X; Z)$.

5) **Shannon entropy on infinite alphabets:** Let $X \sim P \in \mathcal{P}(\mathbb{N})$, where $\mathbb{N} = \{1, 2, \ldots\}$ is the set on natural numbers.
   a) Prove that $H(P) \leq \log(\pi^2/6) + 2\mathbb{E}_P\big[\log(X)\big]$.
   Hint: Relate entropy to KL divergence and use the fact that $q(n) = \frac{6}{\pi^2 n^2}$ is a valid PMF on $\mathbb{N}$.
   b) Provide an example of a distribution P such that $H(P) = \infty$.

6) **Convexity/concavity of mutual information:** For $(X, Y) \sim P_{X,Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. The mutual information $I(X; Y)$ is a functional of $P_{X,Y}$. With the decomposition $P_{X,Y} = P_X P_{Y|X}$, the mutual information can be equivalently represented as a functional of the pair $(P_X, P_{Y|X})$. In this question we focus on the latter representation, and henceforth use the notation $I(P_X, P_{Y|X})$ in place of $I(X; Y)$. Prove the following:
   a) For fixed P_X, $I(P_X, P_{Y|X})$ is convex in $P_{Y|X}$.
   b) For fixed $P_{Y|X}$, $I(P_X, P_{Y|X})$ is concave in P_X. Follow these steps:
      (i) Let $P_X^{(1)}, P_X^{(2)} \in \mathcal{P}(\mathcal{X})$ and consider $X_i \sim P_X^{(i)}$, for $i = 1, 2$. Let $\Phi \sim P_\Phi = \mathsf{Ber}(\alpha)$, $\alpha \in [0, 1]$, be independent of (X_1, X_2), and define $\Theta = \Phi + 1$ and $X := X_\Theta$. Express the distribution of X, denoted by P_X, in terms of $P_X^{(1)}, P_X^{(2)}, \alpha$.
      (ii) Consider the triplet (Θ, X, Y) where Y is obtained by passing X through the transition kernel $P_{Y|X}$. Argue that $\Theta \to X \to Y$ forms a Markov chain.
      (iii) Show that $I(X; Y) \geq I(X; Y|\Theta)$ and deduce the desired concavity result.

7) **Mutual information of sums:** Let $Z_1, Z_2, Z_3, \ldots$ be an i.i.d. sequence of $\mathsf{Ber}\left(\frac{1}{2}\right)$ random variables. Define

$$X_i := \sum_{j=1}^{i} Z_j\,, \quad 1 \leq i \leq n.$$

   Find $I(X_1; X_2, X_3, \ldots, X_n)$.

8) **KL divergence and L^2 norm:** Let $P, Q \in \mathcal{P}([0, 1])$ with PDFs p and q, respectively. Assume that $0 < c_1 \leq p(x), q(x) < c_2 < \infty$ for all $x \in [0, 1]$. Show that the KL divergence is equivalent to the L_2 distance between the two PDFs. That is, there exists positive constants k_1 and k_2 such that

$$k_1 \int (p(x) - q(x))^2 \mathsf{d}x \leq D_{\mathsf{KL}}(P\|Q) \leq k_2 \int (p(x) - q(x))^2 \mathsf{d}x.$$

   Hint: Find the range of values $p(x)/q(x)$ can take. Establish second-order polynomial upper and lower bounds on $\log(y)$ using Taylor's theorem.