ECE 5630 - Homework Assignment 4

April 14th 2020

Due to: Friday, April 24th, at 4pm.

Instructions: Submission in pairs is allowed. Prove and explain every step in your answers. HW sheets are to be submitted via Canvas.

Assumption: In this homework sheet we assume that alphabets are countable throughout.

1) **Monotonicity of entropy for stationary processes:** Let $\{X_i\}_{i=1}^{\infty}$ be a stationary sequence of random variables, i.e., the joint distribution of any finite tuple is invariant to translations. Namely, stationarity means that for any indices $1 \leq i_1 < i_2 < \ldots < i_k$, where $k \in \mathbb{N}$, and a shift $s \in \mathbb{Z}$ such that $i_1 + s \geq 1$, we have $P_{X_{i_1},\ldots,X_{i_k}} = P_{X_{i_1+s},\ldots,X_{i_k+s}}$. Recall that for $n \in \mathbb{N}$ we denote $X^n := (X_1, \ldots, X_n)$. Prove that:
   a) For any $i, n \in \mathbb{N}$ with $1 \leq i \leq n$, we have $H(X_n|X^{n-1}) \leq H(X_i|X^{i-1})$.
   b) For any $n \in \mathbb{N}$, we have
   $$\frac{H(X^n)}{n} \leq \frac{H(X^{n-1})}{n-1}.$$
   c) For any $n \in \mathbb{N}$, we have
   $$\frac{H(X^n)}{n} \geq H(X_n|X^{n-1}).$$

2) **Entropy in bytes:** Let $P \in \mathcal{P}(\mathcal{X})$ and denote by p the associated PMF. The units of the entropy $H_a(P) = -\sum_{x\in\mathcal{X}} p(x)\log_a p(x)$ are bits if the logarithm is to the base of $a = 2$ and bytes if the base is $a = 256$. Express $H_{256}(P)$ in terms of $H_2(X)$.

3) **A measure of correlation:** Let X_1 and X_2 be identically distributed, but not necessarily independent. Assume that X_1 is not a constant, i.e., $H(X_1) > 0$. Define
   $$\rho := 1 - \frac{H(X_2|X_1)}{H(X_1)}$$
   and show that
   a) $\rho = \frac{I(X_1;X_2)}{H(X_1)}$ (there is no typo in the definition of ρ above).
   b) $0 \leq \rho \leq 1$.
   c) Find a necessary and sufficient condition for $\rho = 0$.
   d) Find a sufficient condition for $\rho = 1$.

4) **Random questions:** One wishes to learn the value of a random variable $X \sim P_X \in \mathcal{P}(\mathcal{X})$. A question $Q \sim P_Q \in \mathcal{P}(\mathcal{Q})$ is asked at random according to P_Q. This results in a answer $A := a(X,Q)$, where $a : \mathcal{X} \times \mathcal{Q} \to \mathcal{A}$ is a deterministic answer function that attaches an answer $a(x,q)$ to any value-question pair $(x,q) \in \mathcal{X} \times \mathcal{Q}$. Suppose that X and the

question Q are independent (modeling the fact that the inquirer has no prior knowledge about X when asking Q). With respect to this model, $I(X;Q,A)$ is the information the question-answer pair (Q,A) conveys about X.

a) Show that $I(X;Q,A) = H(A|Q)$ and interpret this result.

b) Now suppose that two i.i.d. questions $Q_1, Q_2 \sim P_Q$ are asked, eliciting answers $A_1 := A(X,Q_1)$ and $A_2 := A(X,Q_2)$. Show that two questions are less valuable than twice the value of a single question in the sense that $I(X;Q_1,A_1,Q_2,A_2) \le 2I(X;Q_1,A_1)$.

5) **Joint letter-typical set:** Let $P_{X,Y} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ be a distribution with $\left|\operatorname{supp}(P_{X,Y})\right| < \infty$ and denote by $p_{X,Y}$ its PMF. For $n \in \mathbb{N}$ and $\epsilon > 0$ recall the definition of the joint letter-typical set

$$\mathcal{T}_\epsilon^{(n)}(P_{X,Y}) := \left\{(x^n,y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\nu_{x^n,y^n}(a,b) - p_{X,Y}(a,b)\right| < \epsilon p_{X,Y}(a,b),\ \forall (a,b) \in \mathcal{X}\times\mathcal{Y}\right\},$$

where $\nu_{x^n,y^n}(a,b) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{(x_i,y_i)=(a,b)\}}$, for $(a,b) \in \mathcal{X}\times\mathcal{Y}$, is the empirical frequency of the pair $(x^n,y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. Prove the following properties:

a) If $(x^n,y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{X,Y})$ then $x^n \in \mathcal{T}_\epsilon^{(n)}(P_X)$ and $y^n \in \mathcal{T}_\epsilon^{(n)}(P_Y)$.

b) For any $(x^n,y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{X,Y})$, we have

   (i) $2^{-n(1+\epsilon)H(P_{X,Y})} \le P_{X,Y}^{\otimes n}(\{x^n,y^n\}) \le 2^{-n(1-\epsilon)H(P_{X,Y})}$.

   (ii) $2^{-n(1+\epsilon)H(P_X)} \le P_X^{\otimes n}(\{x^n\}) \le 2^{-n(1-\epsilon)H(P_X)}$.

   (iii) $2^{-n(1+\epsilon)H(P_Y)} \le P_Y^{\otimes n}(\{y^n\}) \le 2^{-n(1-\epsilon)H(P_Y)}$.

c) If $(X_1,Y_1),(X_2,Y_2),\ldots$ are i.i.d. according to $P_{X,Y}$, then

$$\lim_{n\to\infty} P_{X,Y}^{\otimes n}\left(\mathcal{T}_\epsilon^{(n)}(P_{X,Y})\right) = 1$$

.

d) The cardinality of $\mathcal{T}_\epsilon^{(n)}(P_{X,Y})$ is bounded as

$$(1-\delta)2^{n(1-\epsilon)H(P_{X,Y})} \le \left|\mathcal{T}_\epsilon^{(n)}(P_{X,Y})\right| \le 2^{n(1+\epsilon)H(P_{X,Y})}$$

where the lower bound holds for any $\delta > 0$ and n large enough.

6) **Mismatch letter-typicality:** Let $n \in \mathbb{N}$, $\epsilon > 0$, $X^n \sim P^{\otimes n}$ and $Q \ll P$.

a) Prove that

$$(1-\epsilon)2^{-n(\mathsf{D}_{\mathsf{KL}}(Q\|P)+\delta(\epsilon))} \le P^{\otimes n}\left(\mathcal{T}_\epsilon^{(n)}(Q)\right) \le 2^{-n(\mathsf{D}_{\mathsf{KL}}(Q\|P)-\delta(\epsilon))},$$

where $\lim_{\epsilon\to 0}\delta(\epsilon) = 0$ and the lower bound holds for any n large enough. In your answer, provide an explicit expression for $\delta(\epsilon)$.

b) Deduce that for $P_{X,Y} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ with marginals P_X and P_Y, we have

$$(1-\epsilon)2^{-n(I(X;Y)+\tilde{\delta}(\epsilon))} \le \left(P_X^{\otimes n} \otimes P_Y^{\otimes n}\right)\left(\mathcal{T}_\epsilon^{(n)}(P_{X,Y})\right) \le 2^{-n(I(X;Y)-\tilde{\delta}(\epsilon))}.$$

What is $\tilde{\delta}(\epsilon)$ in this case?

7) **Discrete memoryless channel without feedback:** Consider the communication over a noisy channel scenario as described by the induced distribution on $\mathcal{M} \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{M}$:

$$P^{(c_n)}_{M,X^n,Y^n,\hat{M}}(m, x^n, y^n, \hat{m}) = P_M(m)\mathbb{1}_{\{x^n = f_n(m)\}} P^{\otimes n}_{Y|X}(y^n|x^n)\mathbb{1}_{\{\hat{m} = g_n(y^n)\}},$$

where $P_M \in \mathcal{P}(\mathcal{M})$ is a message distribution and $c_n := (f_n, g_n)$ is a code (encoder-decoder pair). Assume that (henceforth we omit the subscripts of $P^{(c_n)}$ as they are merely uppercase versions of the arguments of the PMF):

(i) The channel is memoryless, i.e., there exists a (single-letter) transition kernel $P_{Y|X}$ such that $P^{(c_n)}(y_i|m, x^i, y^{i-1}) = P_{Y|X}(y_i|x_i)$, for all $i = 1 \ldots, n$.

(ii) The channel is without feedback, i.e., $P^{(c_n)}(x_i|m, x^{i-1}, y^{i-1}) = P^{(c_n)}(x_i|m, x^{i-1})$, for all $i = 1 \ldots, n$.

Prove that $P^{(c_n)}(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$.

8) **Capacity of binary erasure channel:** Consider the binary erasure channel (BEC) in which a fraction $\alpha \in [0, 1]$ of the transmitted bits are lost (erased) as depicted in Figure 1. More precisely, the BEC of parameter α is specified by the tuple $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, where $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, \mathrm{e}\}$ and $P_{Y|X}$ is described by the relation:

$$Y = \begin{cases} X, & \text{w.p. } 1-\alpha, \\ \mathrm{e}, & \text{w.p. } \alpha. \end{cases}$$

Find a closed form expression that depends only on α for the capacity $\max_{P_X} I(X; Y)$ of this BEC.

Hint: Consider the function $E = \mathbb{1}_{\{Y=e\}}$ and show that $I(X; Y) = I(X; Y, E) = I(X; Y|E)$.

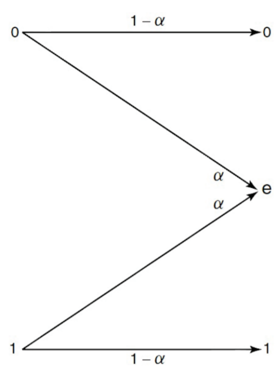

Fig. 1: Binary erasure channel.

9) **Capacity of noisy typewriter:** Suppose we have a malfunctioning typewriter that we model as a channel from the keystroke X_{in} to the typed symbol Y_{out}. Specifically, let $\mathcal{X}_{\text{in}} = \mathcal{Y}_{\text{out}} = \{A, B, C, \ldots, Z\}$ and define $\mathsf{con} : \mathcal{X}_{\text{in}} \to \mathcal{Y}_{\text{out}}$ as the function that (circularly) maps any letter of the alphabet to the next one, e.g., $\mathsf{con}(A) = B$ and $\mathsf{con}(Z) = A$. The noisy typewriter channel is described by the relation

$$Y_{\text{out}} = \begin{cases} X_{\text{in}}, & \text{w.p. } \frac{1}{2}, \\ \mathsf{con}(X_{\text{in}}), & \text{w.p. } \frac{1}{2} \end{cases}$$

In words, the keystroke X_{in} is either typed unaltered with probability $\frac{1}{2}$ or is transformed to the next letter of the alphabet with probability $\frac{1}{2}$ (see Figure 2). Find the capacity $\max_{P_X} I(X;Y)$ of the noisy typewriter channel.

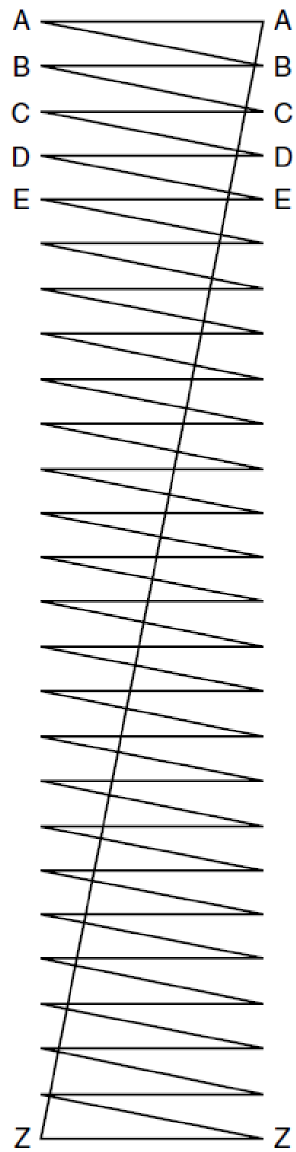

Fig. 2: Noisy typewriter.