EE2340/EE5847: Information Sciences/Information Theory **2021**

Homework 1: 3rd Mar 2021

Instructor: Shashank Vatedka

Instructions: *You are encouraged to discuss and collaborate with your classmates. However, you* must explicitly *mention at the top of your submission who you collaborated with, and all external resources (websites, books) you used, if any. Copying is NOT permitted, and solutions must be written independently and in your own words.*

Please scan a copy of your handwritten assignment as pdf with filename <your ID>_HW<homework no>.pdf. Example: EE19BTECH00000_HW1.pdf.

For programming questions, create separate files. Please use the naming convention <your ID>_HW<homework no>_problem<problem no>.. Example: EE19BTECH00000_HW1_problem1.c.* You may upload c,cpp,py or m files only. No other format will be allowed.

Finally, upload your submission as separate files on Google classroom. Please do not upload zip files.

The following videos might be useful for problem 1: `https://www.youtube.com/watch?v=ew_CbNkxzYg` and `https://www.youtube.com/watch?v=ywduYSla88k`.

Exercise 1.1 (Jointly typical sets)**.** Let p_{XY} be any joint distribution on $\mathcal{X} \times \mathcal{Y}$. For any $\epsilon > 0$ and positive integer n, define the jointly typical set

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

where $\mu_{x^n y^n}(a,b)$ is the number of locations $i \in \{1, 2, \ldots, n\}$ for which $x_i = a$ and $y_i = b$.

Let X^n, Y^n be jointly distributed such that $X_i, Y_i \sim p_{XY}$ for all i whereas (X_i, Y_i) is independent of all other (X_j, Y_j) for all $j \neq i$.

1. Prove that $\lim_{n\to\infty} \Pr[(X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}(p_{XY})] = 0$
2. Show that for all $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and all $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(p_{XY})$,

$$\mathbb{E}_{XY}[g(X,Y)](1-\epsilon) \leqslant \frac{1}{n}\sum_{i=1}^{n} g(x_i, y_i) \leqslant \mathbb{E}_{XY}[g(X,Y)](1+\epsilon)$$

3. Use the above to obtain upper and lower bounds on $|\mathcal{T}_\epsilon^{(n)}(p_{XY})|$.

Exercise 1.2 (Implementing compression)**.** The shared folder contains files with characters randomly drawn from the set $\{a, b, c, d, e\}$. You must write a program to compress and decompress this file.

1. Write a program to compute the empirical type of the file, i.e., the fraction of occurrence of various symbols in the file. Also compute the entropy.
2. Using the above type, design the Shannon, Huffman and Shannon-Fano-Elias codes for this file. You can compute these on paper. Show all the steps.
3. Use the designed codes to compress the attached file.

4. Decode the compressed file, and verify that you get back the original file
5. Find the length of the compressed sequence, and compare this with the entropy.

Exercise 1.3. Repeat the same as in Problem 2 for the corresponding file in the folder for problem 3.