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

Homework 2:

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 a **single zip file** which includes all your programs and pdf file. The zip file should have filename <your ID>_HW<homework no>.zip. Example: EE19BTECH00000_HW_1.zip.

Recap

Recall that for any pair of random variables (X, Y) with joint distribution p_{XY}, the entropy is

$$H(X) = -\sum_{x\in\mathcal{X}} p_X(x) \log_2 p_X(x),$$

measures the uncertainty in X, the conditional entropy

$$H(X|Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p_{XY}(x, y) \log_2 p_{X|Y}(x|y)$$

measures the residual uncertainty in X after observing Y, and the mutual information

$$I(X; Y) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p_{XY}(x, y) \log_2 \frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)}$$

measures the information that X gives about Y, or Y gives about X. Likewise, the joint entropy

$$H(X, Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p_{XY}(x, y) \log_2 p_{XY}(x, y)$$

measures the total uncertainty in (X, Y).

The entropy is the minimum compression that you can achieve for fixed-length compression of a memoryless source.

The maximum rate of reliable communication over a discrete memoryless channel (DMC) $p_{Y|X}$ is equal to the capacity of the channel:

$$C = \max_{p_X} I(X; Y).$$

Exercise 2.1 (Entropy and mutual information).

1. Find the entropy of a random variable uniformly distributed over a finite set $\mathcal{X}$
2. Find the differential entropy of a Gaussian random variable with mean μ and variance σ^2
3. Find the differential entropy of an exponential random variable with mean μ
4. Let X be a Bernoulli(1/2) random variable, and Y be a Bernoulli(p) random variable (for arbitrary $0 < p < 1$) independent of X. Compute the entropy of $Z = X \oplus Y$, where $\oplus$ denotes the binary XOR operation. Also compute the mutual information between X and Y, and also between X and Z.
5. For the previous subproblem, compute the joint entropy $H(X, Y, Z)$.
6. Let X, Y be independent Gaussian random variables with zero mean and variances σ_x^2 and σ_y^2 respectively. Let $Z = X + Y$. Compute the mutual information between X and Y, and also between X and Z.

Exercise 2.2. A fair coin is flipped till the first head occurs. Let X be the number of flips required. Find the entropy of X.

Exercise 2.3. Consider the joint distribution given in the excel sheet. For this, compute $H(X), H(Y)$, $H(X,Y), H(Y|X), H(X|Y)$, and $I(X;Y)$.

Exercise 2.4. Consider the binary entropy function

$$H_2(p) = -p \log_2 p - (1-p) \log_2 (1-p).$$

for $p \in [0, 1]$.

1. Compute the first and second derivatives of $H_2(p)$ with respect to p. Use this to prove that $H_2(p)$ is concave.
2. Use the first and second derivatives to prove that $H_2(p)$ attains the maximum at $p = 1/2$.
3. Write a program and plot $H_2(p)$ as a function of p. Attach your plot as a jpg/png/pdf file. You need not submit your program.

Exercise 2.5. Consider the channel with transition probabilities given in the excel sheet. For this, compute the capacity.

Exercise 2.6. A first-order time homogeneous Markov chain defined over $\mathcal{X} = \{1, 2 \ldots, k\}$ can be described by a $k \times k$ transition probability matrix A, where the (i, j)th entry

$$A_{ij} = \Pr[X_t = i | X_{t-1} = j].$$

1. If $\underline{p}$ denotes a k-length probability vector, with the ith entry $p_i = \Pr[X_{t-1} = i]$, then prove that

   $$\underline{q} = A\underline{p}$$

   is a probability vector with $q_j = \Pr[X_t = j]$.
2. For the transition probabilities given in the excel sheet, compute the stationary distribution $\underline{\pi}$, i.e., the probability vector that satisfies

   $$\underline{\pi} = A\underline{\pi}.$$

3. Suppose that $X_1, X_2, \ldots$ is a first-order time homogeneous Markov chain with transition probabilities as given in the excel sheet, and initialized with the stationary distribution. Compute

$$\bar{H} = \lim_{n \to \infty} \frac{H(X_1, \ldots, X_n)}{n}.$$

Exercise 2.7. In the previous assignment, you wrote a program to compute the empirical pmf/frequency of letters in a text file. Modify that program to first compute the frequency of all ASCII characters (all 2^8 possible ASCII characters including newline, space, carriage return) in the file, and then compute the entropy of the text file (be careful with $0 \log 0$ errors). Also count the total number of characters in the file. Your program should print the number of ASCII characters in the file, and the entropy. Attach the program as a separate file.

Note that typically each ASCII character is represented using 1 byte (8 bits). This should give a lower bound on the compressed file size if the characters are iid:

$$\text{optimal compressed filesize in bytes} \ = \ \text{number of characters in file} \ \times \frac{\text{entropy}}{8}. \tag{2.1}$$

Compute the entropy for the two text files: (1) the file that you were given for Exercise 1.6 and the attached file2.txt (a copy of the book "War and Peace" by Leo Tolstoy: you may ignore any non-ASCII characters, if any). Also compress them using zip and compare the filesizes with what you have from the above.

Is there a significant difference between the zipped file and the computed optimal compressed filesize for the two cases? Why?

Let us now model the file as a time homogeneous first order Markov source. In addition to computing the empirical pmf, also compute the empirical probabilities of pairs of letters $\hat{p}_{X_1,X_2}$. In other words, $\hat{p}_{X_1,X_2}(a,a)$ denotes the number of occurrences of aa in the file, $\hat{p}_{X_1,X_2}(a,b)$ denotes the number of occurrences of ab in the file, and so on. Compute the entropy $H(X_1, X_2)$ using this method and thereby compute $H(X_2|X_1)$. If you replace the entropy in equation (2.1) with this conditional entropy, then is this closer to the compression rate achieved using zip? If so, why do you think so? If not, why not?