Information and coding theory

Project I

Practical informations

Each project should be executed by groups of two students. We expect each group to provide:

- A *brief* report (in PDF format) collecting the answers to the different questions.
- The scripts you have implemented.

The report and the scripts should be submitted as a tar.gz (or zip) file on Montefiore's submission plateform (`http://submit.montefiore.ulg.ac.be`) before the deadline. **You must concatenate your sXXXXXX ids** (e.g., s000007s123456) **as group, archive and report names.**

Questions

Information measures

Exercises by hand

Let $\mathcal{X} \in \{X_1, X_2, X_3, X_4\}$ and $\mathcal{Y} \in \{Y_1, Y_2, Y_3, Y_4\}$ be two independent discrete random variables that take on values in $\{0, 1, 2, 3\}$. The marginal probability distributions of these two random variables are :

value	0	1	2	3
$P(\mathcal{X} = X_i)$	$1/4$	$1/4$	$1/4$	$1/4$
$P(\mathcal{Y} = Y_i)$	$1/2$	$1/4$	$1/8$	$1/8$

Let us also consider a discrete random variable $\mathcal{S} = \mathcal{X} + \mathcal{Y}$, and a binary random variable $\mathcal{Z}$ such that $\mathcal{Z} = 0 \Leftrightarrow \mathcal{X} \neq \mathcal{Y}$.

1. Give the joint distribution $P(\mathcal{X}, \mathcal{Y})$ and the marginal distributions of $\mathcal{S}$ and $\mathcal{Z}$.

Calculate

2. $H(\mathcal{X})$, $H(\mathcal{Y})$, $H(\mathcal{S})$, $H(\mathcal{Z})$
3. $H(\mathcal{X}, \mathcal{Y})$, $H(\mathcal{X}, \mathcal{S})$, $H(\mathcal{Y}, \mathcal{Z})$, $H(\mathcal{S}, \mathcal{Z})$
4. $H(\mathcal{X}|\mathcal{Y})$, $H(\mathcal{Z}|\mathcal{X})$, $H(\mathcal{S}|\mathcal{X})$, $H(\mathcal{S}|\mathcal{Z})$
5. $H(\mathcal{X}, \mathcal{Y}|\mathcal{S})$, $H(\mathcal{S}, \mathcal{Y}|\mathcal{X})$
6. $I(\mathcal{X}; \mathcal{Y})$, $I(\mathcal{X}; \mathcal{S})$, $I(\mathcal{Y}; \mathcal{Z})$, $I(\mathcal{S}; \mathcal{Z})$
7. $I(\mathcal{X}; \mathcal{Y}|\mathcal{S})$, $I(\mathcal{S}; \mathcal{Y}|\mathcal{X})$

Computer-aided exercises

In Python or Julia. Inputs of the functions are matrices of the probabilities associated to (combinations of) values taken by the input random variables (i.e., $\mathcal{X}$ and/or $\mathcal{Y}$ and/or $\mathcal{Z}$).

8. Write a function *entropy* that computes $H(\mathcal{X})$, the entropy of a discrete random variable $\mathcal{X}$, given[1] its probability distribution $P_{\mathcal{X}} = (p_1, p_2, \dots, p_n)$ where $p_i = P(\mathcal{X} = X_i)$. What are the key parts of your implementation? Intuitively, what is measured by the entropy?

9. Let $\mathcal{X}$ and $\mathcal{Y}$ be two discrete random variables. Write a function *joint_entropy* that computes $H(\mathcal{X}, \mathcal{Y})$, the joint entropy of $\mathcal{X}$ and $\mathcal{Y}$. What are the key parts of your implementation? Compare this function with the *entropy* function, what do you notice?

10. Let $\mathcal{X}$ and $\mathcal{Y}$ be two discrete random variables. Write a function *conditional_entropy* that computes $H(\mathcal{X}|\mathcal{Y})$, the conditional entropy of $\mathcal{X}$ given $\mathcal{Y}$. What are the key parts of your implementation? Compare intuitively and theoretically $H(\mathcal{X}, \mathcal{Y})$ and $H(\mathcal{X}|\mathcal{Y})$.

11. Let $\mathcal{X}$ and $\mathcal{Y}$ be two discrete random variables. Write a function *mutual_information* that computes $I(\mathcal{X}; \mathcal{Y})$, the mutual information between $\mathcal{X}$ and $\mathcal{Y}$. What are the key parts of your implementation? What can you deduce from the influence of one variable on the other?

12. Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be three discrete random variables. Write functions *cond_joint_entropy* and *cond_mutual_information* that compute $H(\mathcal{X}, \mathcal{Y}|\mathcal{Z})$ and $I(\mathcal{X}; \mathcal{Y}|\mathcal{Z})$ respectively. To do so, use the *joint_entropy* and *mutual_information* functions. Explain briefly the key parts of your implementation. Draw a Venn diagram that summarizes the relationship between $H(\mathcal{X}, \mathcal{Y}|\mathcal{Z})$ and $I(\mathcal{X}; \mathcal{Y}|\mathcal{Z})$.

13. First, generate samples with realizations of $\mathcal{X}, \mathcal{Y}, \mathcal{S}, \mathcal{Z}$ and use them to derive the different required probability distributions. Then, using implemented functions, verify and compare your results of questions 2 to 7. Give your observations.

Designing informative experiments

4			14
			22
		1	9
15	21	9	

4	7	3	14
9	8	5	22
2	6	1	9
15	21	9	

Figure 1: Example of a Fubuki grid and its solution.

The goal of this logic-based game is to fill in a 3×3 grid with the following rules:

(a) each digit $(1, 2, \dots, 9)$ can only appear once in the grid;

[1] In practice, you will pass the appropriate probability distribution(s) as argument(s) of the functions.

(b) the sum of a row/column should be equal to the given corresponding number (e.g., for the first row $4 + 7 + 3 = 14$).

Typically, an *unsolved Fubuki* (see left grid on Figure 1) is a partially completed grid where each given digit is called a *clue*. The *solutions* is shown in blue on the right figure.

Let us model this problem as follows. Let us associate a random variable $\mathcal{X}_i$ to each *square* of the grid (as shown by Figure 2). Let us also denote the row and column numbers by r_i and c_i.

Justify all your answers.

14. What is the entropy of the following subgrid (just one row and only rule (a))?

4		

15. What is the entropy of the following subgrid (just one row and rules (a) and (b))?

4			14

16. What is the effect of taking into account a row/column constraint on the entropy of a square and on the entropy of the (sub)grid?

For the sake of simplicity, let us now assume (assumption A) that you only have to consider the initial state of the grid (i.e., the clues and constraints r_i and c_i) to find the value of a square and you do not need to take into account other square values (i.e., independently of other solutions). In other words, it means that you still have to follow rule (a) given the clues only but you can forget rule (b) and not take into account other solutions for rule (a).

$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{X}_3$	r_1
$\mathcal{X}_4$	$\mathcal{X}_5$	$\mathcal{X}_6$	r_2
$\mathcal{X}_7$	$\mathcal{X}_8$	$\mathcal{X}_9$	r_3
c_1	c_2	c_3	

Figure 2: Grid model

17. What is the entropy of a single square under assumption A?

18. What is the entropy of the *unsolved Fubuki grid* under assumption A?

19. It is obvious that assumption A will not allow to find the right solution of a Fubuki grid. First, answer questions 14 and 15 under assumption A and explain (theoretically) the impact of assumption A on the results in terms of information theory. Then, explain what would imply (in terms of information theory and computational complexity) to not consider assumption A for questions 17 and 18.

For the following questions, Assumption A is no longer assumed.

20. Using information theory, how would you proceed to solve the Fubuki? In particular, explain which squares you would fill in first? Justify.
21. Let us assume that you can choose *one* additional clue (i.e., revealing the correct digit in an empty square). Which one would you choose and why?
22. Let us assume that you can choose *sequentially* exactly k clues. Design a strategy (using information theory) that determines the sequence of k clues to reveal.
23. Let us assume that you can choose *simultaneously* exactly k clues. Design a strategy (using information theory) that determines the next k clues to reveal (at once).