Master in applied sciences
Computer science and engineering
Year 2018-2019

# Information and coding theory

# Project 1

**INFO8003-1**
Louis Wehenkel

Olivier Moitroux - s152122
Pierre Hockers - s152764

Liège

Février 2019

# Exercises by hand

The source entropy is given by

$$H(S) \triangleq E\{h(s)\} = -\sum_{i=1}^{n} p_i \log_2(P_i) \tag{1}$$

While the units are not present in the scan, the entropy is indeed expressed in Shannon.

## 1

We can determine $p(x_i) = \sum_{j=1}^{4} p(x_i, y_j)p(y_j) = \frac{1}{4}$, $\forall i = 1, \ldots, 4$. The same formula can be used to determine $p(y) = [\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}]$. This way,

a) $$H(x) = -p(x_1)\log_2 p(x_1) - \ldots - p(x_4)\log_2 p(x_4) \quad (1)$$
$$= \frac{1}{4}\cdot\left(-\log_2 \frac{1}{4}\right) = \boxed{2}$$

b) $$H(y) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{8}\log_2\frac{1}{8} \quad (2)$$
$$= 7/4 = \boxed{1{,}75}$$

c) $$P(w=1) = \sum_{i=1}^{4} P(x_i, y_i) = \frac{1}{8} + \frac{1}{8} + \frac{1}{16} = \frac{5}{16} \quad (3)$$
$$P(w=0) = 1 - \frac{5}{16} = \frac{11}{16}$$

$$H(w) = -\frac{5}{16}\log_2\frac{5}{16} - \frac{11}{16}\log_2\frac{11}{16} = \boxed{0{,}89604} \quad (4)$$

d) $$P(z=1) = P(w=0) = \frac{1}{16} + \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{4} = 11/16 \quad (5)$$
$$P(z=0) = 1 - \frac{11}{16} = 5/16$$

$$H(z) = H(w) = -\frac{5}{16}\log_2\frac{5}{16} - \frac{11}{16}\log_2\frac{11}{16}$$
$$= \boxed{0{,}89604} \quad (6)$$

## 2

a)

$$H(\mathcal{X},\mathcal{Y}) \triangleq -\sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i \cap y_j) \log P(x_i \cap y_j)$$

$$= -\frac{12}{8}\log_2\frac{1}{8} + \frac{4.1}{32}\left(-\log_2\frac{1}{32}\right) - \frac{6.1}{16}\log_2\frac{1}{16} \quad (7)$$

$$= \boxed{3,375}$$

b) $P(w,x) = P(w|x).P(x)$

e.g.: $P(W=1, x=x_1) = P(W=1|x=x_1).P(x=x_1)$

$$= 1/8 \quad (8)$$

$P(W=0, x=x_1) = P(x_1,y_2)+P(x_1,y_3)+P(x_1,y_4)$

$$= \frac{1}{16}+\frac{1}{32}+\frac{1}{32}$$

$$= 1/8$$

This way, by using (8):

| JOINT | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $P(w)$ |
|---|---|---|---|---|---|
| $w_1$ | 1/8 | 1/8 | 1/16 | 0 | 5/16 |
| $w_0$ | 1/8 | 1/8 | 3/16 | 1/4 | 11/16 |
| $P(x)$ | 1/4 | 1/4 | 1/4 | 1/4 | 1 |

$$H(\mathcal{X},w) = -\sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i \cap w_j) \log P(x_i \cap w_j)$$

$$= -\frac{4.1}{8}.\log_2\frac{1}{8} - \frac{1}{16}\log_2\frac{1}{16} - \frac{3}{16}\log_2\frac{3}{16} - \frac{1}{4}\log_2\frac{1}{4}$$

$$= \boxed{2,70282} \quad (9)$$

c) $P(w,y) = P(w|y).P(y)$

By using the same technique as in b), we can determine the following table:

| JOINT | $y_1$ | $y_2$ | $y_3$ | $y_4$ | $P(w)$ |
|---|---|---|---|---|---|
| $w_1$ | 1/8 | 1/8 | 1/16 | 0 | 5/16 |
| $w_0$ | 3/8 | 1/8 | 1/16 | 1/8 | 11/16 |
| $P(x)$ | 1/2 | 1/4 | 1/8 | 1/8 | 1 |

$$H(y,w) = -\frac{4.1}{8}\log_2\frac{1}{8} - \frac{3}{8}\log_2\frac{3}{8} - \frac{2.1}{16}\log_2\frac{1}{16} \quad (10)$$

$$= \boxed{2,53064}$$

d) $P(w=1, z=1) = P(w=1 \mid z=1) \cdot P(z=1)$

$= 0 \cdot \frac{11}{16}$

$P(w=0, z=1) = 1 \cdot \frac{11}{16} = \frac{11}{16}$

$P(w=1, z=0) = 1 \cdot (1-\frac{11}{16}) = \frac{5}{16}$

$P(w=0, z=0) = 0 \cdot (1-\frac{11}{16}) = 0$

$H(w, z) = -\frac{11}{16} \log_2 \frac{11}{16} - \frac{5}{16} \log_2 \frac{5}{16}$

$= \boxed{0,896038}$ (11)

## 3

$H(\mathcal{X}|\mathcal{Y})$ is the result of averaging $H(\mathcal{X}|\mathcal{Y} = y)$ over all possible values $y$ that $\mathcal{Y}$ may take (weighted sum with weights being $p(y)$). This way, one could have used explicitly:

$$H(\mathcal{Y}|\mathcal{X}) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 \frac{p(x, y)}{p(y)} \quad (2)$$

a) $H(X|Y) = H(X,Y) - H(Y) = (7) - (2)$ (12)

$= 3,375 - 1,75 = \boxed{1,625}$

b) $H(W|X) = \underset{(9)}{H(W,X)} - \underset{(1)}{H(X)} = 2,70282 - 2$

$= \boxed{0,70282}$ (13)

c) $H(Z|W) = -p(Z=1|W=1) \log_2 p(\ldots) - \ldots - p(z=1|w=0) \log_2 p(\ldots)$

$= H(Z,W) - H(W)$

$= (11) - (6)$

$= \boxed{0}$ (14)

d) $H(W|Z) = H(Z,W) - H(Z)$

$= \boxed{0}$ (15)

## 4

4.a). $H(X, Y|W)$

$H(X,Y|W) = H(X,Y,W) - H(W)$ (16) And we know by (4) that $\boxed{H(W) = 0,89604}$

And $H(X,Y,W) = -\sum_{m=1}^{4}\sum_{m=1}^{4}\sum_{i=1}^{\bullet} P(X_m, Y_m, W=i) \log P(X_m, Y_m, W=i)$.

But it is easily realised that once we have the X and Y variable set, the W value is immediately determined. Therefore, 2 cases emmerge: either the value of W is the one matching the X, Y pair, and $P(X_m, Y_m, W=i) = P(X_m, Y_m)$, or the value of W doesn't and $P(X_m, Y_m, W=i) = 0$ and should be discarded.

This allows us to write

$$H(X,Y,W) = -\sum_{m=1}^{4}\sum_{m=1}^{4} P(X_m, Y_m) \log P(X_m, Y_m) = H(X,Y) \quad (17)$$

And we know from (7) that $\boxed{H(X,Y) = 3,375}$

$$\Rightarrow H(X,Y|W) = 3,375 - 0,89604 = 2,47896 \quad (18)$$

.b) $H(W,Z|X) = H(W,Z,X) - H(X)$ (19)

And we know from (1) that $\boxed{H(X) = 2}$

$$H(W,Z,X) = -\sum_{m}^{4}\sum_{m}^{1}\sum_{i}^{1} P(W_m, Z_i, X_m) \log P(W_m, Z_i, X_m)$$

It is easily spottable that when $W=1$, $Z=0$, and vice versa. Therefore, 2 cases emmerge: if W and Z have values that are different from one another, then simply having one of the 2 values is enough and the $2^{nd}$ is not bringing any new information, ergo $P(W_m, Z_i, X_m) = P(Z_i, X_m) = P(W_m, X_m)$.

If W and Z have the same value (0 or 1), then the case is impossible and $P(W_m, Z_i, X_m) = 0$

We can therefore write $H(W,Z,X) = -\sum\sum\sum P(X,Z,W) \log P(X,Z,W)$

$$= -\sum\sum P(X,W) \log P(X,W) = \boxed{H(X,W) = 2,702} \text{ via } (9)$$

$$= -\sum\sum P(X,Z) \log P(X,Z) = H(X,Z)$$

$$\Rightarrow H(W,Z|X) = H(W,Z,X) - H(X) \quad (20)$$

$$= 2,702 - 2 = 0,702$$

## 5

a) $I(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i \cap y_j) \log_2 \frac{P(x_i \cap y_j)}{P(x_i)P(y_j)}$

$= \frac{1}{8} \log \frac{1/8}{\frac{1}{4}\cdot\frac{1}{2}} + \dots + 0 \cdot \log \frac{0}{\frac{1}{4}\cdot\frac{1}{8}}$ (21)

$= H(X) - H(X|Y)$

$= (1) - (12)$

$= \boxed{0,375}$

b) $I(X,W) = \sum_{i=1}^{m}\sum_{j=1}^{m} P(X,W) \log_2 \frac{P(X,W)}{P(X)P(W)}$

$= H(X) - H(X|W)$

$= H(W) - H(W|X)$ (22)

$= (4) - (13)$

$= \boxed{0,1932}$

c) $I(Y,Z) = H(Y) - H(Y|Z)$

$= H(Y) - (H(Y,Z) - H(Z))$

$= H(Y) + H(Z) - H(Y,Z)$

| JOINT | $y_1$ | $y_2$ | $y_3$ | $y_4$ | $P(z)$ |
|---|---|---|---|---|---|
| $z=1$ | 3/8 | 1/8 | 1/16 | 1/8 | 11/16 |
| $z=0$ | 1/8 | 1/8 | 1/16 | 0 | 5/16 |
| $P(w)$ | 1/2 | 1/4 | 1/8 | 1/8 | 1 |

$H(Y,Z) = -\frac{3}{8} \log_2 \frac{3}{8} - \dots$ (23)

$= H(Y,W) = 2,53064$

$\Rightarrow I(Y,Z) = (1) + (2) - (23) = \boxed{0.7154}$ (24)

d) $I(W,Z) = H(W) - H(W|Z)$

$= (4) - 0$ (25)

$= \boxed{0,89604}$

Indeed, an alternative to $I(\mathcal{X},\mathcal{Y}) = I(\mathcal{Y},\mathcal{X}) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y) log \frac{p(x,y)}{p(x)p(y)}$ is to consider:

$$
\begin{aligned}
I(\mathcal{X},\mathcal{Y}) &= H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) && (3)\\
&= H(\mathcal{Y}) - (H(\mathcal{X},\mathcal{Y}) - H(\mathcal{X})) && (4)\\
&= H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X},\mathcal{Y}) && (5)
\end{aligned}
$$

that is easier to compute thanks to previous results.

# 6

6.a). $I(X;Y|W) = H(X|W) - H(X|Y,W)$

$H(X|W) = H(X,W) - H(W)$

$H(X,W) = 2{,}702$, from (9) $\quad H(W) = 0{,}83604$ from (4)

$\Rightarrow H(X|W) = 2{,}702 - 0{,}83604 = 1{,}806$

$$
\begin{aligned}
H(X|Y,W) &= H(X,Y|W) - H(X|W) \\
&= H(X,Y|W) - 1{,}806 \\
&= 2{,}47896 - 1{,}806 \qquad \text{from (12)} \\
&= 0{,}67296
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow I(X;Y|W) &= H(X|W) - H(X|Y,W) \\
&= 1{,}806 - 0{,}67296 \\
&= 1{,}13304
\end{aligned}
$$

6.b) $I(W;Z|X) = H(W|X) - H(W|Z,X)$

$\qquad = \underset{\text{from (13)}}{0{,}70282} - H(W|Z,X)$

Once again, $H(W|Z,X) = -\sum P(W|Z,X) \log P(W|Z,X)$, and we know that the case $Z = W$ ($= 0$ or $1$) is impossible (therefore $P(W|Z,X) = 0$ in this case). And in the remaining cases where $W \neq Z = \begin{cases}0\\1\end{cases}$ the probability that $W = \begin{cases}1\\0\end{cases}$ when knowing that $Z = \begin{cases}0\\1\end{cases}$ is always of 1. Therefore $H(W|Z,X) = 0$

~~and we have two cases: either W = Z and P(W|Z,X) = 0, or from the double info of the value of Z we get a probability of 1 if Z ≠ W~~

$$
\begin{aligned}
I(W;Z|X) &= 0{,}70282 - 0 \\
&= 0{,}70282
\end{aligned}
$$

# Implementaion

## 7 Computing the entropy

The function `entropy()` takes as an input the prior probability distribution of the variable we want to compute the entropy of. It basically implements the definition of the entropy:

$$H(\mathcal{X}) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{6}$$

where n corresponds to the number of values that $\mathcal{X}$ can take. The entropy corresponds to the quantity of information that is given by a source of information. The more the source emits different information, the bigger the entropy. In other words, the entropy indicates the quantity of information that is required for the receptor to determine what the source emits. By providing more information on what the source emit, the entropy will decrease.

## 8 Computing the joint entropy

`joint_entropy()` computes $H(\mathcal{X}, \mathcal{Y})$ that is given by:

$$H(\mathcal{X}, \mathcal{Y}) \triangleq -\sum_{i=1}^{m} \sum_{j=1}^{n} p(x_i, y_j) \log_2 p(x_i, y_j) \tag{7}$$

It takes as a parameter a numpy array that stores the joint probability of the two random variables regardless of the orientation of the array as $H(\mathcal{X}, \mathcal{Y}) = H(\mathcal{Y}, \mathcal{X})$. We can notice that

$$H(\mathcal{X}_1, \ldots, \mathcal{X}_k) \leq \sum_{i=1}^{k} H(\mathcal{X}_k) \tag{8}$$

whith an equality when considering independent random variables. The joint entropy of two random variables turns out to be greater than the entropy of a single one but less than the sum of the entropy of the two random variables considered.

## 9 Computing the conditional entropy

`conditional_entropy` simply relies on the result

$$H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{X}, \mathcal{Y}) - H(\mathcal{Y}) \tag{9}$$

If the two random variables are independant, we can expect

$$H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{X}) \tag{10}$$

but it is not the case here. We can however notice that $H(W|Z) = H(Z|W) = 0$ which means that the value of W/Z is completely determined by Z/W.

In order to detect implicitely wether to compute $H(\mathcal{X}|\mathcal{Y})$ or $H(\mathcal{Y}|\mathcal{X})$ from the joint distribution, one needs also to provide as a parameter the array of prior probabilities of the random variables used for conditioning. This way, when computing $H(\mathcal{X}|\mathcal{Y})$, one needs to provide both the joint distribution of $\mathcal{X}$ and $\mathcal{Y}$ as well as the prior distribution of $\mathcal{Y}$. Let's note that additional information never increases entropy:

$$H(\mathcal{X}|\mathcal{Y}) \leq H(\mathcal{X}) \tag{11}$$

## 10 Computing the mutual information

`mutual_information()` computes the mutual information between $\mathcal{X}$ and $\mathcal{Y}$ by implementing its definition:

$$I(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{12}$$

Alternatively, we could have used the same development as done by hand in order to exploit previous methods. $I(\mathcal{X}, \mathcal{Y})$ is equal to 0 if and only if $\mathcal{X}$ and $\mathcal{Y}$ are two indenpendent random variables.

## 11

`cond_joint_entropy()` computes

$$H(X, Y|W) = H(X, Y, W) - H(W) \tag{13}$$

The same principle applies for H(W, Z|X)
`cond_mutual_information()` computes

$$\begin{align} I(X, Y|W) &= H(X|W) - H(X|Y, W) \tag{14} \\ &= H(X|W) - (H(X, Y|W) - H(X|W)) \tag{15} \\ &= 2.H(X|W) - H(X, Y|W) \tag{16} \end{align}$$

# Computer-aided exercises

## 12

The output of the script is available in the appendix.

# Design informative experiments

## 13 Entropy of a single square independently of others

By making abstraction of the other squares, a square can take 9 possible values with equal probabilities of 1/9. This way, the entropy of a single square is

$$H = -\sum_{i=1}^{9} \frac{1}{9} \log_2(\frac{1}{9}) = -\log_2(1/9) = 3,1699 \text{ Shannon} \tag{17}$$

## 14 Entropy of subgrid

The given subgrid has 6 remaining squares to fill. That means that each square can take 6 different values (with equal probabilities of 1/6). In this way,

$$H = -\sum_{i=1}^{6} \frac{1}{6} \log_2(1/6) = -\log_2(1/6) \tag{18}$$
$$= 2,5849 \text{ Shannon} \tag{19}$$

## 15 Entropy of unsolved sudoku

The entropy of the unsolved sudoku is the sum of the entropy of each square in the grid. In order to evaluate the entropy of a square(random variable), we need to count the number of allowed remaining values it can take. To evaluate this number, we count the number of values that can be "seen" from the square, in

- The row of the current square
- The column of the current square
- The subgrid to which the current square belong

These "seen" numbers are stored in a list called "seen" in the code.

Let $k_i$ be the number of allowed remaining values for the square i. Of course, $k_i$ = 9-seen_values around square i. Finally, the entropy of the unsolved sudoku is:

$$H = -\sum_{i=1}^{81} k_i \frac{1}{k_i} \log_2(1/k_i) \tag{20}$$
$$= 82.57192489469372 \text{ Shannon} \tag{21}$$

In practice, we will simply ignore the cases that are not filled with a number, which is equivalent to making their probability equal to 0.

## 16 Methodology to fill the sudoku

In order to fill the sudoku, a simple method would be to find the case with the lowest entropy (where we have the most information) in the grid. This case is supposed to be solvable, as no uncertainty is supposed to be associated with it. A simple look around enables us to find what value should fill this case. We then repeat the process and see what is the next case with minimum entropy, etc..

## 17 Choosing one additional clue

At a first glance, we could want to reveal the square that give the highest amount of information (highest entropy) and that corresponds to the area where we have the least amount of information. However, this is not the optimal methodology: it is better to choose the clue which *reduces the most the entropy of the entire grid.* Indeed, revealing a clue can affect the entropy of other squares ! It is thus better to select the clue that yields to the highest reduction of the sum of the entropy of all the squares, a.k.a the entropy of the grid.

## 18 Choosing sequentially more than one clue

1. Reveal the square (random variable) that reduces the most the entropy of the grid (sum of entropy of individual squares) if it was revealed.
2. Update the probabilities in the grid
3. Re-compute the entropy of each square/random variable
4. Go back to 1 until solved

## 19 Choosing simultaneously more than one clue

We want to reveal $k$ squares that will bring together the most information. For this reason, we want to reveal the $k$ unrevealed squares that will together bring down the global entropy of the grid the most. Here we don't want to select the squares that have individually the biggest entropy. Choosing the k squares that brings us closer to winning the game (a.k.a bring more total information) is the key. We thus enumerate all the associations of k squares, computes the entropy of the grid if they were unrevealed together and reveal the combination that yields to the greatest reduction of the entropy of the grid.

# Appendix

Output script q12

```
Q1)
H(x) = 2.0
H(y) = 1.75
H(W) = 0.8960382325345574
H(Z) = 0.8960382325345574

Q2)
H(x, y) = 3.375
H(x, w) = 2.702819531114783
H(y, w) = 2.5306390622295662
H(w, z) = 0.8960382325345574

Q3)
H(x|y) = 1.625
H(w|x) = 0.7028195311147831
H(z|w) = 0.0
H(w|z) = 0.0

Q4
H(x, y|w) = 2.4789617674654427
H(w, z|x) = 0.7028195311147831

Q5)
I(x, y) = 0.375
I(x, w) = 0.1932187014197742
I(y, z) = 0.11539917030499097
I(w, z) = 0.8960382325345575

Q6)
I(x, y|z) = 1.134600829695009
I(w, z|x) = 0.7028195311147831
```