EISTI. **Année scolaire 2008-2009**

DÉPARTEMENT " INFORMATIQUE "

THÉORIE DE L'INFORMATION

Série d'exercices N^{o}2. **CORRIGÉ.**

PARTIE I. ENTROPIE D'UNE SOURCE. DÉFINITIONS ET PROPRIÉTÉS.

Exercice 1 (Calcul d'entropie. Exemple.)**.** Soit une source d'alphabet $\Omega = \{1, 2, 3, 5, 4\}$. Calculer son entropie pour les distributions de probabilités suivantes.

1. $P_1 = \{0.2, 0.2, 0.2, 0.2, 0.2\}$
2. $P_2 = \{0.05, 0.05, 0.05, 0.05, 0.8\}$
3. $P_3 = \{0.1, 0.2, 0.3, 0.15, 0.25\}$

Solution de l'exercice 1

Dans tous les cas on applique la définition de l'entropie :

$$H(X) = -\sum_{i=1}^{5} p_i \log(p_i)$$

1. $H(X_1) = -5 * 0.2 * \log(0.2) = -\log\left(\frac{1}{5}\right) = \log(5) \simeq 2.32.$
2. $H(X_2) = -4 * 0.05 * \log(0.05) - 0.8 \log(0.8) \simeq 1.12.$
3. $H(X_2) = -0.1 \log 0.1 - 0.2 \log 0.2 - 0.3 \log 0.3 - 0.15 \log 0.15 - 0.25 \log 0.25 \simeq 2.23.$

On remarquera que dans le premier cas l'entropie est la plus grande. *C'est une propriété générale de l'entropie : elle atteint son maximum lorsque tous les symboles d'un alphabet donné de taille n sont équiprobables. Elle est alors égale à* $\log(n)$.

Exercice 2.

1. On lance une pièce dont les deux cotés sont identiques : pile. Quelle est l'entropie associée à cette expérience ?
2. On lance un dé équilibré à 6 faces. Quelle est l'information moyenne apportée par l'observation de la parité du résultat ?
3. Un jeu de cartes contient 3 piques, 4 trèfles, 2 cœurs et 1 carreau. On tire une carte au hazard. Quelle est l'entropie de l'observation de la couleur de la carte ?

Solution de l'exercice 2

1. On lance une pièce dont les deux cotés sont identiques : pile. Quelle est l'entropie associée à cette expérience ?

   On remarque que "Pile" est un événement certain. Dans ce cas, il n'y a aucune incertitude sur l'issue de l'expérience. On a donc $H(X) = 0$.

2. On lance un dé équilibré à 6 faces. Quelle est l'information moyenne apportée par l'observation de la parité du résultat ?

   Soit
   $$Y = \begin{cases} 1, & \text{si le nombre est pair} \\ 0, & \text{si le nombre est impair} \end{cases}$$
   On a alors $P[Y = 0] = P[Y = 1] = 0.5$ donc, en applicant la propriété de l'entropie dans le cas d'une distribution uniforme on a immédiatement : $H = \log(2)$.

3. Un jeu de cartes contient 3 piques, 4 trèfles, 2 cœurs et 1 carreau. On tire une carte au hazard. Quelle est l'entropie de l'observation de la couleur de la carte ?

   Soit $Y \in \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$. On calcule la distributions de probabilités associée :

y	♠	♣	♡	♢
P(y)	3/10	4/10	2/10	1/10

   On en déduit alors l'entropie :
   $$H(Y) = -\frac{3}{10}\log\frac{3}{10} - \frac{4}{10}\log\frac{4}{10} - \frac{2}{10}\log\frac{2}{10} - \frac{1}{10}\log\frac{1}{10} \simeq 1.846$$

ATTENTION ! Dans les calculs on utilise **le logarithme de base 2 !**

PARTIE II. ENTROPIE D'UN COUPLE "ÉMETTEUR-RÉCEPTEUR".

Exercice 3. Nous reprenons ici l'exemple du TD1 (voir exercice 1). Soient X et Y deux variables aléatoires prenant leurs valeurs respectivement dans $\Omega_X = \{x_1, x_2, x_3\}$ et $\Omega_Y = \{y_1, y_2\}$ et ayant la matrice de probabilités conjointes suivante

$P(X,Y) =$

	y_1	y_2
x_1	0.25	0
x_2	0.1	0.3
x_3	0.1	0.25

Nous avons établi pour ce couple de variables aléatoires les distributions de probabilité suivantes :

Distribution marginale de X.

$$p(x_1) = 0.25 \quad p(x_2) = 0.1 + 0.3 = 0.4 \quad p(x_3) = 0.1 + 0.25 = 0.35$$

Distribution marginale de Y.

$$p(y_1) = 0.25 + 0.10 + 0.1 = 0.45 \quad p(y_2) = 0.3 + 0.25 = 0.55$$

Les distributions conditionnelles

$$P(X|Y) = \begin{pmatrix} \frac{5}{9} & 0 \\ \frac{2}{9} & \frac{6}{11} \\ \frac{2}{9} & \frac{5}{11} \end{pmatrix} \text{ et } P(Y|X) = \begin{pmatrix} 1 & 0 \\ \frac{1}{4} & \frac{3}{4} \\ \frac{2}{7} & \frac{5}{7} \end{pmatrix}$$

1. Calculer $H(X)$,$H(Y)$,$H(X,Y)$,$H(X|Y)$, $H(Y|X)$.
2. On définit l'information mutuelle de X et Y ou encore le gain d'information

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

   cette quantité représente la diminution de l'incertitude sur X lorsqu'on a observé Y. Calculer $I(X|Y)$ et $I(Y|X)$. Verifier la propriété de symétrie de l'information mutuelle :

   $$I(X|Y) = I(Y|X)$$

Solution de l'exercice 3

1. Calculer $H(X)$,$H(Y)$,$H(X,Y)$,$H(X|Y)$, $H(Y|X)$. Nous utilisons les définitions :

   $$H(X) = -0.25 \log 0.25 - 0.4 \log 0.4 - 0.35 \log 0.35 \simeq 1.558$$

   $$H(Y) = -0.45 \log 0.45 - 0.55 \log 0.55 \simeq 0.99$$

   Pour l'entropie conjointe on fait la somme le long de toute la matrice $P(X,Y)$.

   $$H(X,Y) = -\sum_{i=1}^{3}\sum_{j=1}^{2} p(x_i, y_j) \log p(x_i, y_j) \simeq 2.185$$

   ATTENTION ! Pour l'entropie conditionnelle moyenne on utilise deux matrices : $P(X,Y)$ et $P(X|Y)$:

   $$H(X|Y) = -\sum_{i=1}^{3}\sum_{j=1}^{2} p(x_i, y_j) \log p(x_i|y_j) \simeq 1.19$$

   $$H(Y|X) = -\sum_{i=1}^{3}\sum_{j=1}^{2} p(x_i, y_j) \log p(y_j|x_i) \simeq 0.6266$$

2. On définit l'information mutuelle de X et Y ou encore le gain d'information

   $$I(X|Y) = H(X) - H(X|Y) = 1.5588 - 1.19 = 0.3688 = I(Y|X) = H(Y) - H(Y|X) = 0.99 - 0.626$$

   On a bien

   $$I(X|Y) = I(Y|X)$$

PARTIE III. ENTROPIE : UN JEU D'ESPION !

Le chiffrement de César consiste à décaler l'alphabet de k postions de façon cyclique et de remplacer chaque lettre d'un massage clair par une lettre correspondante de l'alphabet décalé. La clé secrète de ce chiffre est un entier $1 \leq k \leq 25$ qui représente le décalage de l'alphabet. Nous allons dans un premier temps étudier la sécurité de ce chiffre et ensuite apprendre la méthode d'analyse des fréquences qui a permis de la casser.

Exercice 4 (Rendons à César ce qui est à César : son chiffre !)**.**

1. Montrer que pour les messages de longueur $l = 1$ le chiffre de César est parfaitement sûr au sens de Shannon.
2. Montrer que pour les messages de longueur $l \geq 2$ le chiffre n'est plus parfaitement sûr. Pour cela analysez l'exemple suivant. Soient le message $m = AB$ et le chiffré $c = DM$. Montrer que $P[M = m|C = c] = 0$ tandis que $P[M = m] \neq 0$.

Solution de l'exercice ??

1. Montrer que pour les messages de longueur $l = 1$ le chiffre de César est parfaitement sûr au sens de Shannon. L'ensemble de clairs possibles $\mathcal{M}$ et celui des chiffrés, $\mathcal{C}$, ainsi que celui des clés $\mathcal{K}$ coïncident avec l'alphabet latin et on a

   $$\forall m \in \mathcal{M}, P[M = m] = \frac{1}{26}$$

   De plus, la fonction de chiffrement établit un lien entre le clair, le chiffré et la clé :

   $$c = (m + k)\ mod\ 26$$

   Alors pour tout $c \in \mathcal{C}$ et pour tout $m \in \mathcal{M}$ on a

   $$P[M = m|C = c] = \frac{P[M = m\ et\ C = c]}{P[C = c]} = \frac{P[M = m\ etK = c - m]}{P[C = c]} = \frac{P[M = m] \cdot P[K = c - m]}{P[C = c]} =$$

   car le choix du message et de la clé sont indépendants et

   $$P[C = c] = P[K = k] = \frac{1}{26}$$

2. Montrer que pour les messages de longueur $l \geq 2$ le chiffre n'est plus parfaitement sûr. Pour cela analysez l'exemple suivant. Soient le message $m = AB$ et le chiffré $c = DM$. Montrer que $P[M = m|C = c] = 0$ tandis que $P[M = m] \neq 0$.

   Il est facile de constater qu'il n'existe aucun décalage de l'alphabet qui transforme AB en DM (les lettres A et B sont consécutives dans l'alphabet, elles devraient l'être dans l'alphabet décalé.) Donc $P[M = m \; et \; C = c] = 0$. Nous avons alors $P[M = m|C = c] = 0 \neq P[M = m]$. Donc le chiffrement de César avec $l > 1$ ne vérifie pas la condition de Shannon de système parfaitement sûr.

Exercice 5 (Cryptanalyse du chiffre de César)**.** La méthode d'analyse des fréquences a été inventé par le savant arabe AL-Kindi au IX-ème siècle. On suppose que l'on connaît la langue du texte clair et que l'on dispose du message chiffré. Dans le ca de chiffre de César on cherche à déterminer le paramètre k, clé du chiffre. La méthode consiste à comparer l'histogramme d'occurrences des caractères du chiffré avec la table des fréquences d'occurrence des caractères de la langue du texte clair. Voici la table des fréquences de la langue française.

Lettre	Fréquence %	Lettre	Fréquence %
A	8.4	N	7.13
B	1.06	O	5.26
C	3.03	P	3.01
D	4.18	Q	0.99
E	17.26	R	6.55
F	1.12	S	8.08
G	1.27	T	7.07
H	0.92	U	5.74
I	7.34	V	1.32
J	0.31	W	0.04
K	0.05	X	0.45
L	6.01	Y	0.3
M	2.96	Z	0.12

1. Votre mission, si vous l'acceptez, consiste à déchiffrer le message secret de votre binôme. Chacun de vous va composer un message de son choix. Pour le chiffrer, utiliser `http://www.bibmath.net/crypto/substi/cryptcesar.php3`l'applet java sur ce site. Envoyez le chiffré obtenu à votre voisin (par e-mail ou chat). Ensuite chacun cherchera à trouver la clé et le message clair par analyse fréquentielle.

   La clé peut être trouvée par méthode exhaustive en utilisant la même applet. Cela donne la réponse pour chaque message des élèves.

2. Et maintenant, déchiffrez ceci :

UHGCHNK. GHNL OHNL IKHIHLHGL MKHBL PTZHGL IHNK ITKMBK TN STGBSBUTK. OHMKX TOBHG OT ITKMBK ET HN OHNL OHNEXS. NG OKTB OTNMHNK

OXNM MHNCHNKL OHEXK ATNM IHNK OHBK LT IKHBX. NG SHFUB FTKVATBM XG SBZSTZTGM LNK ET KHNMX.

Solution du message chiffré est

BONJOUR. NOUS VOUS PROPOSONS TROIS WAGONS POUR PARTIR AU ZANIZIBAR. VOTRE AVION VA PARTIR LA OU VOUS VOULEZ. UN VRAI VAUTOUR VEUT TOUJOURS VOLER HAUT POUR VOIR SA PROIE. UN ZOMBI MARCHAIT EN ZIGZAGANT SUR LA ROUTE.

On peut la trouver avec la méthode exhaustive. Le piège pour la méthode d'analyse des fréquences est que les lettres les plus fréquentes du message ne sont pas celles du français.