Correction séance 7 (15 novembre)

Exercice 1. (Chiffre de César)
1. Les lignes correspondantes du carré de Vigenère donnent le tableau

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J

Ainsi, le C est codé en M, le E est codé en O... On trouve donc que C'EST PAS FAUX est codé par M'OCD ZKC PKEH.

2. Dans un espoir de garder la surprise du message à décoder, on va décoder un autre message que celui sur la feuille de TD (la méthode est bien-sûr la même). Regardons le message LMTQKQWCA TIAB KWCZAM. Sachant qu'il a été codé avec la clé I, on considère le tableau

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H

Qui nous apprend que la lettre L code la lettre D, donc la première lettre du message en clair est un D. De même la deuxième est un E, etc... On trouve que le message clair est DELICIOUS LAST COURSE.

Notons que la lettre A code la lettre S dans ce cas : en fait décoder un message qui a été codé via la clé I revient à encoder un message avec la clef S. Ça sera plus clair avec la question suivante.

3. La lettre A correspond à $0 \in \mathbb{Z}/26\mathbb{Z}$. Le codage par la lettre d'indice i envoie la lettre A sur la lettre en question. L'indice de A devient donc $0 + i = i \in \mathbb{Z}/26\mathbb{Z}$. De même, la lettre B est envoyé sur la suivante de la lettre d'indice i, donc la lettre d'indice $i+1 \in \mathbb{Z}/26\mathbb{Z}$. Et ainsi de suite on obtient que la transformation associée au codage est l'addition de i à l'indice des lettres (dans $\mathbb{Z}/26\mathbb{Z}$).

En particulier, on obtient que le décodage s'effectue en soustrayant i à l'indice, autrement dit en ajoutant $26 - i$: le décodage par la i-ème lettre est équivalent au codage par la $26 - i$-ème lettre, c'est ce qu'on a vu à la question précédente.

4. On pose $j := ai + b$. Pour que le chiffrement affine de clé a, b soit décodable, il faut et il suffit que la transformation affine associée soit inversible. On a

$$j = ai + b[26] \Leftrightarrow j - b = ai[26]$$

Si a n'est pas inversible, il existe i, i' tels que $ai + b = ai' + b[26]$, dans ce cas, le chiffrement affine est indécodable. Par exemple si $a = 2$ et $b = 1$, alors le message ANANAS et NANANS sont tous deux encodés en AAAAAL.
Si a est inversible dans $Z/26\mathbb{Z}$, alors il existe un $a' \in \mathbb{Z}$ tel que $aa' \equiv 1[26]$. On a alors $a'j - a'b = i[26]$. Le décodage existe et correspond alors à un chiffrement affine de clé $a', -a'b$.

5. Une fois encore dans un soucis de conserver la surprise, on décode un autre message : UXFFL JSDAEBJS, dont on suppose qu'il a été codé avec la même clef, à savoir $a = 9$, $b = 3$. On a $9 \cdot 3 = 27 \equiv 1[26]$, l'inverse de 9 modulo 26 est donc 3. D'après la question précédente, le décodage correspond à un chiffrement affine de clé $a' = 3, b' = -3 \cdot 3 = -9$.

Lettre codée	U	X	F	F	L	J	S	D	A	E	B	J	S
Indice j	20	23	5	5	11	9	18	3	0	4	1	9	18
Image $i = a'j + b'$	25	8	6	6	24	18	19	0	17	3	20	18	19
Lettre claire	Z	I	G	G	Y	S	T	A	R	D	U	S	T

6. Les lettres les plus fréquentes du message crypté sont respectivement le Z et le D, avec 19 et 18 occurrences respectivement. La lettre la plus fréquente suivante est le V, avec 13 occurrences.

On fait donc l'hypothèse raisonnable que l'une des lettres Z, D code le E, et l'autre code le A :

- Si Z code A et D code E, alors on a

$$\begin{cases} a \cdot 0 + b \equiv -1[26] \\ a \cdot 4 + b \equiv 5[26] \end{cases} \Leftrightarrow \begin{cases} b \equiv -1[26] \\ a \cdot 4 \equiv 6[26] \end{cases}$$

  Attention, comme 4 n'est pas inversible modulo 26, il y a plusieurs solutions à $4a \equiv 6[26]$:

$$4a \equiv 6[26] \Leftrightarrow 4a = 6 + 26k \Leftrightarrow 2a = 3 + 13k \Leftrightarrow 2a \equiv 3[13] \Leftrightarrow a \equiv 8[13]$$

  On a donc $a \equiv 8[26]$ ou $a \equiv 21[26]$. Le premier cas est impossible car alors a ne serait pas inversible modulo 26, on prend donc $a \equiv 21[26]$. On calcule $a' = a^{-1} = 5[26]$ et $b' = -a'b = 5[26]$. Le premier mot serait alors décodé par FAMUGH : raté.

- Si Z code E et D code A, alors on a

$$\begin{cases} a \cdot 4 + b \equiv -1[26] \\ a \cdot 0 + b \equiv 3[26] \end{cases} \Leftrightarrow \begin{cases} a \cdot 4 \equiv -4[26] \\ b \equiv 3[26] \end{cases}$$

  Attention : comme 4 n'est pas inversible modulo 26, on ne peut pas déduire que $a \equiv -1[26]$! Il y a une autre possibilité :

$$4a \equiv -4[26] \Leftrightarrow 4a = -4 + 26k \Leftrightarrow 2a = -2 + 13k \Leftrightarrow 2a \equiv -2[13] \Leftrightarrow a \equiv -1[13]$$

  On a donc $a \equiv -1[26]$ ou $a \equiv 12[26]$. Le second cas est impossible car alors a ne serait pas inversible modulo 26, on prend donc $a \equiv -1[26]$. On calcule $a' = a^{-1} = -1[26]$ et $b' = -a'b = 3[26]$, le premier mot serait alors décodé par DEMAIN, c'est un mot en français ! En décodant le reste du texte, on obtient :

DEMAIN DES L'AUBE, A L'HEURE OU BLANCHIT LA CAMPAGNE
JE PARTIRAI. VOIS-TU, JE SAIS QUE TU M'ATTENDS.
J'IRAI PAR LA FORET, J'IRAI PAR LA MONTAGNE.
JE NE PUIS DEMEURER LOIN DE TOI PLUS LONGTEMPS.

C'est la première strophe d'un célèbre poème de Victor Hugo.

Exercice 2. (Chiffre de Vigenère)
1. Comme indiqué, on répète la clé pour obtenir RATRATRATRA (de même longueur que le message). On obtient le codage suivant :

S	Y	N	A	P	T	O	S	O	M	E
↓R	↓A	↓T	↓R	↓A	↓T	↓R	↓A	↓T	↓R	↓A
J	Y	G	R	P	M	F	S	H	D	E

2. On retrouve le chiffre de César en prenant pour clé un mot d'une seule lettre.

3. a) Soit x la variable aléatoire donnant la valeur d'une lettre choisie au hasard dans le texte. Pour toute lettre l, on a $P(x = l) = \frac{n_l}{n}$: C'est une bonne vieille expérience de Bernoulli avec $p = \frac{n_l}{n}$.

On répète deux fois l'expérience de Bernoulli ci-dessus (avec remise→ répétitions indépendantes). On a donc deux variables x_1 et x_2. La probabilité d'avoir les deux lettres égales à l est alors $P(x_1 = l \text{ et } x_2 = l) = \frac{n_l^2}{n^2}$ (loi Binomiale).

La formule des probabilités totales nous donne alors

$$IC = \sum_{l=\text{A}}^{\text{Z}} P(x_1 = l \text{ et } x_2 = l) = \sum_{l=\text{A}}^{\text{Z}} \frac{n_l^2}{n^2}$$

b) Dans un texte totalement aléatoire, la fréquence de chaque lettre est $\frac{1}{26}$. On obtient alors dans ce cas

$$IC = \sum_{l=\mathtt{A}}^{\mathtt{Z}} \frac{1}{26^2} = \frac{26}{26^2} = \frac{1}{26} \approx 0,03846$$

c) Un chiffrement par permutation ne fait qu'intervertir les indices des lettres, sans changer leurs fréquences respectives. Le calcul de l'indice de coïncidence du texte crypté correspond alors a une simple permutation dans l'indice de la somme.

d) Dans un texte en français assez long, on choisi par hasard un sous-texte indépendant des différentes lettres lettres, de sorte que la fréquence de chaque lettre soit identique dans le sous-texte. Ainsi, l'indice de coïncidence n'est pas modifié dans le sous-texte. En pratique dans un texte plus court, on s'attend à ce que l'indice de coïncidence ne soit "pas trop modifié".

e) Chaque lettre de la forme x_{ik} pour $i \geqslant 0$ est codé par la première lettre de la clef : c'est un chiffre de César (il en va de même des sous-texte de la forme "x_{ik+j}"$_{i\geqslant 0}$ pour $j < k$).