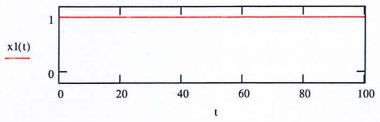
Théorie de l'information. Examen 2013-2014.

Ordinateur interdit. Aucun document papier autorisé à l'exception d'une feuille A4 Recto-Verso aide-mémoire. Calculatrice autorisée. Durée : 2h00.

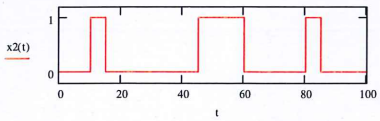
Exercice 1. Signal & information. (4 points)

1. Pour chaque signal numérique x_i ($i = 1 \dots 4$) suivant, dont on donne la trajectoire (l'évolution temporelle), indiquer la *quantité moyenne d'information* H_i ($i = 1 \dots 4$) qu'il contient :

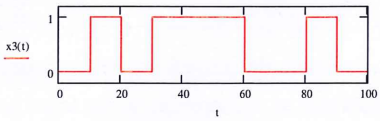
1.1. Signal **constant** $x_1(t)$:



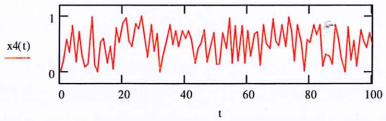
1.2. Signal aléatoire **carré** (succession de valeurs 0 et 1) $x_2(t)$ avec les probabilités : $p(0) = 0.75$ et $p(1) = 0.25$:



1.3. Signal aléatoire **carré** (succession de valeurs 0 et 1) $x_3(t)$ avec la distribution de probabilités *uniforme* : $p(0) = p(1) = 0.5$:



1.4. Signal aléatoire **bruit blanc** $x_4(t)$, de distribution *uniforme*, quantifié sur 8 bits :



2. En déduire les *taux maximum de compression* sans perte d'information $t_1, \dots, t_4$ (exprimés en %) correspondants autorisés par rapport à une dynamique de codage standard de 8 bits pour la valeur des signaux.

Exercice 2. Codage de Huffman. (5 points)

Soit le message suivant (séquence ADN de longueur 40) codé à partir d'un alphabet constitué de 4 symboles (marqueurs) C, T, G, A :
Séquence (**espaces** pour lisibilité mais **à ignorer**) *:* GAGGAGCTCA GGGGAGCGCA TAGGAGTGCA AGGAGTGGTG

1. Calculer *l'entropie* de la source d'information constituée par la séquence.
2. Construire un *arbre de Huffman* en vue de la compression sans pertes des données.
3. En déduire un *code de Huffman binaire* associé à chaque symbole.
4. Donner la *longueur moyenne* du code de Huffman construit.
5. Indiquer (en %) le *taux de compression t* réalisé par rapport à un codage standard (codage ASCII fixe 8 bits par symbole).
6. La *longueur moyenne* du code binaire obtenu atteint-elle la valeur de l'*entropie* ? Pourquoi ?

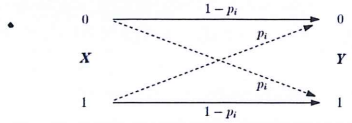
Exercice 3. Canal de transmission symétrique bruité. (3 points)

Une source d'information binaire émet les symboles 0 et 1 avec les probabilités $P(0) = p = 0.1$ et $P(1) = q = 1 - p = 0.9$. Ces bits sont transmis à un récepteur à travers un canal bruité illustré ci-dessous, avec la probabilité d'erreur de transmission $p_i = 10^{-1} = 0.1$. En notant X et Y les symboles respectivement émis et reçus, calculer les caractéristiques suivantes, en **valeurs numériques** :

1. Les *entropies* $H(X)$ et $H(Y)$.

2. Les *entropies* $H(X,Y)$, $H(Y|X)$.

3. L'*information mutuelle moyenne* $I(X,Y)$.



4. La *capacité* C du canal en *absence de bruit*.

Exercice 4. Code systématique. (5 points)

Soit le code linéaire $C_{n,k} = C_{6,3}$ qui au vecteur d'information (bits i_1 à i_3)

$$\mathbf{i}^T = (i_1 \quad i_2 \quad i_3)$$

associe le mot de code (bits i_1 à i_3 et c_4 à c_6)

$$\mathbf{c}^T = (i_1 \quad i_2 \quad i_3 \quad c_4 \quad c_5 \quad c_6) \qquad \text{avec} \quad \begin{cases} c_4 = i_1 \\ c_5 = i_1 + i_2 \\ c_6 = i_1 + i_2 + i_3 \end{cases}.$$

1. Donner la *matrice génératrice* directement systématique de ce code.
2. Soit le mot-source $\mathbf{i}^T = (1 \quad 0 \quad 1)$, quel est le *mot de code* associé ?
3. Soit le message $\mathbf{m}^T = (1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1)$. Est-ce un *mot du code* ?
4. Donner la *table de code* obtenue par construction systématique du code.
5. En déduire la *distance minimale* d de ce code.
6. Donner les capacités de *détection* et de *correction* de ce code.
7. Soit le mot reçu $\boldsymbol{\omega}^T = (1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0)$ entaché d'1 erreur. La *correction* de ce mot est-elle *garantie* ?
8. Si la correction est possible, donner le *mot-source* issu du décodage de $\boldsymbol{\omega}$.

Exercice 5. Chiffre de Vigenère. (3 points)

Le **chiffrement de César** (50 av. J.C.) est un chiffrement symétrique mono-alphabétique : le texte chiffré s'obtient en remplaçant chaque lettre du texte clair original par une lettre décalée dans l'alphabet (décalage fixe et circulaire, toujours dans le même sens) ; la clé de chiffrement est unique et constituée par la valeur du décalage. Le déchiffrement se fait selon le même principe avec la même clé mais en inversant le sens du décalage.

Exemple de chiffrement de César avec une clé égale à 2 : (casse, espaces, accents … ignorés) Clair : CESAR ; Chiffré : EGUCT

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

Le **chiffrement de Vigenère** (1586) est un chiffrement symétrique poly-alphabétique qui s'inspire du chiffrement de César : le chiffrement se fait par un décalage renseigné dans une table composée de 26 alphabets, écrits dans l'ordre, mais décalés de ligne en ligne d'un caractère (voir plus bas).

Pour coder un message (par exemple le clair INFORMATION), on choisit une clé unique qui est un mot de longueur arbitraire (par exemple la clé THEORIE). On écrit ensuite la clé sous le message clair, en répétant si nécessaire la clé pour que sa la longueur égale celle du clair :

I	N	F	O	R	M	A	T	I	O	N
T	H	E	O	R	I	E	T	H	E	O

Pour chiffrer on utilise la table des 26 alphabets (voir ci-dessous), à l'intersection de la ligne de la lettre à coder avec la colonne de la lettre de la clé (ou l'inverse). Ainsi, pour coder la 1ère lettre du clair (I) avec la 1ère lettre clé (T), on extrait de la table la lettre B du chiffré, à l'intersection de la ligne I et de la colonne T; on continue ainsi pour les lettres suivantes du clair : la 2nde lettre du clair (N) avec la lettre clé correspondante (H) donne la lettre U du chiffré. On obtient finalement le chiffré : BUJCIUEMPSB :

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

Le déchiffrement se fait selon le même principe avec la même clé en utilisant toujours la table des 26 alphabets : on regarde pour chaque lettre de la clé répétée, la ligne correspondante et on y cherche la lettre chiffrée. La lettre de la colonne que l'on trouve ainsi est la lettre déchiffrée.

1. *Quel avantage majeur* sur le chiffrement de César, le *chiffrement de Vigenère* présente-t-il ?

2. *Quelles conditions* doit-on imposer sur le *choix de la clé de cryptage de Vigenère* pour se protéger d'une attaque destinée à *casser* le chiffrement par *analyse fréquentielle* ?

3. *Quelle clé de chiffrement de Vigenère* choisir pour ramener le chiffrement de Vigenère à celui de César ?

4. En particulier, *quelle clé de chiffrement de Vigenère* permet de réaliser un chiffrement de César de clé de *décalage* 2 ?

5. Au XVIII^{ème} siècle dans ses correspondances avec le comte Axel de Fersen, la reine Marie-Antoinette utilisait le chiffrement de Vigenère, mais en chiffrant une lettre sur deux du message, une lettre sur deux étant laissée en clair. Quel était l'*intérêt de ne chiffrer qu'une lettre sur deux* ?

6. Par rapport au chiffrement de Vigenère de la totalité des caractères d'un message, la *sûreté du chiffrement* est-elle *accrue* ou *réduite* par le fait de ne chiffrer qu'une lettre sur deux ? (Justifier)

ANNEXE. Table de logarithme de base 2

p	$\log_2(p)$	$plog_2(1/p)$
0.01	-6.6438	0.0664
0.02	-5.6438	0.1128
0.03	-5.0588	0.1517
0.04	-4.6438	0.1857
0.05	-4.3219	0.2160
0.06	-4.0588	0.2435
0.07	-3.8365	0.2685
0.08	-3.6438	0.2915
0.09	-3.4739	0.3126
0.10	-3.3219	0.3321
0.11	-3.1844	0.3502
0.12	-3.0588	0.3670
0.13	-2.9434	0.3826
0.14	-2.8365	0.3971
0.15	-2.7369	0.4105
0.16	-2.6438	0.4230
0.17	-2.5563	0.4345
0.18	-2.4739	0.4453
0.19	-2.3959	0.4552
0.20	-2.3219	0.4643
0.21	-2.2515	0.4728
0.22	-2.1844	0.4805
0.23	-2.1202	0.4876
0.24	-2.0588	0.494134
0.25	-2.00	.50
0.26	-1.9434	0.5052
0.27	-1.8889	0.5100
0.28	-1.8365	0.5142
0.29	-1.7858	0.5179
0.30	-1.7369	0.5210
0.31	-1.6896	0.5237
0.32	-1.6438	0.5260
0.33	-1.5994	0.5278

p	$\log_2(p)$	$plog_2(1/p)$
0.34	-1.5563	0.5291
0.35	-1.5145	0.5301
0.36	-1.4739	0.5306
0.37	-1.4344	0.5307
0.38	-1.3959	0.5304
0.39	-1.3584	0.5297
0.40	-1.3219	0.5287
0.41	-1.2863	0.5273
0.42	-1.2515	0.5256
0.43	-1.2175	0.5235
0.44	-1.1844	0.5211
0.45	-1.1520	0.5184
0.46	-1.1202	0.5153
0.47	-1.0892	0.5119
0.48	-1.0588	0.5082
0.49	-1.0291	0.5042
0.50	-1.00	0.50
0.51	-0.9714	0.4954
0.52	-0.9434	0.4905
0.53	-0.9159	0.4854
0.54	-0.8889	0.4800
0.55	-0.8624	0.4743
0.56	-0.8365	0.4684
0.57	-0.8109	0.4622
0.58	-0.7858	0.4558
0.59	-0.7612	0.4491
0.60	-0.7369	0.4421
0.61	-0.7131	0.4350
0.62	-0.6896	0.4275
0.63	-0.6665	0.4199
0.64	-0.6438	0.4120
0.65	-0.6214	0.4039
0.66	-0.5994	0.3956

p	$\log_2(p)$	$plog_2(1/p)$
0.67	-0.5777	0.3871
0.68	-0.5563	0.3783
0.69	-0.5353	0.3693
0.70	-0.5145	0.3602
0.71	-0.4941	0.3508
0.72	-0.4739	0.3412
0.73	-0.4540	0.3314
0.74	-0.4344	0.3214
0.75	-0.4150	0.3112
0.76	-0.3959	0.3009
0.77	-0.3770	0.2903
0.78	-0.3584	0.2795
0.79	-0.3400	0.2686
0.80	-0.3219	0.2575
0.81	-0.3040	0.2462
0.82	-0.2863	0.2347
0.83	-0.2688	0.2231
0.84	-0.2515	0.2112
0.85	-0.2344	0.1992
0.86	-0.2175	0.1871
0.87	-0.2009	0.1747
0.88	-0.1844	0.1622
0.89	-0.1681	0.1496
0.90	-0.1520	0.1368
0.91	-0.1360	0.1238
0.92	-0.1202	0.1106
0.93	-0.1046	0.0973
0.94	-0.0892	0.0839
0.95	-0.0740	0.0703
0.96	-0.0588	0.0565
0.97	-0.0439	0.0426
0.98	-0.0291	0.0285
0.99	-0.0145	0.0143