CORRIGE DE L'EXAMEN DU 30/01/02

EXERCICE 1

1. Numérotons chacun des trois dés par les chiffres 1, 2 et 3 afin de les différencier. Notons alors X la variable aléatoire égale au numéro du dé qui est pipé et Y la variable correspondant au chiffre figurant sur la face apparente du dé lancé après stabilisation.

La quantité d'information moyenne apportée par Y sur X s'écrit :

$$I(Y;X) = H(X) - H(X/Y) \quad (1)$$
$$I(Y;X) = H(Y) - H(Y/X) \quad (2)$$

La formule (2) semble plus facile à appliquer car on connaît par hypothèse la loi de Y sachant X. Déterminons la loi de Y. Notons j le numéro du dé choisi au hasard. Alors :

$$P\{Y=1\} = P\{(Y=1 \cap X=j) \cup (Y=1 \cap X \neq j)\}$$
$$= P\{Y=1 \cap X=j\} + P\{Y=1 \cap X \neq j\}$$

Or $P\{Y=1 \cap X=j\} = P\{Y=1/X=j\} \times P\{X=j\} = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9}$

Et $P\{Y=1 \cap X \neq j\} = P\{Y=1/X \neq j\} \times P\{X \neq j\} = \frac{1}{6} \times \frac{2}{3} = \frac{1}{9}$

Soit $P\{Y=1\} = \frac{2}{9} + \frac{1}{9} = \frac{3}{9} = \frac{1}{3}$

En procédant de même pour les cinq autres valeurs, on trouverait :

$$\forall k \in \{2,3,4,5,6\} \quad P\{Y=k\} = \frac{2}{15}$$

On a alors :

$$H(Y) = -\frac{1}{3}\log\frac{1}{3} - 5 \times \frac{2}{15}\log\frac{2}{15} = 2{,}466 \text{ Sh}$$

De plus, si m et n représentent les deux autres valeurs possibles de X (différentes de j) :

$$H(Y/X) = P\{X=j\} \times H(Y/X=j) + P\{X=m\} \times H(Y/X=m)$$
$$+ P\{X=n\} \times H(Y/X=n)$$

$$H(Y/X=j)=-\frac{2}{3}\log\frac{2}{3}-5\times\frac{1}{15}\log\frac{1}{15}=1{,}69\ \text{Sh}$$

$$H(Y/X=m)=H(Y/X=n)=\log 6=2{,}585\ \text{Sh}$$

Comme chaque dé a la même probabilité (1/3) d'être pipé, on a :

$$P\{X=j\}=P\{X=m\}=P\{X=n\}=\frac{1}{3}$$

Et par conséquent :

$$H(Y/X)=\frac{1}{3}\times(1{,}69+2\times 2{,}585)=2{,}286\ \text{Sh}$$

Et finalement :

$$\boxed{I(X;Y)=2{,}466-2{,}286=0{,}18\ \text{Sh}}$$

2. L'incertitude moyenne qui règne sur l'identification du dé pipé est l'entropie de la variable aléatoire X. Comme X possède une loi uniforme sur 3 valeurs, on a :

$$\boxed{H(X)=\log 3=1{,}585\ \text{Sh}}$$

EXERCICE 2

1. Les lancers successifs étant indépendants, on a :

$$\forall k\in IN^*$$
$$P\{X=k\}=P\{\text{obtenir "face" lors des k - 1 premiers jets et "pile au k}^{\text{ème}}\}$$

soit :

$$P\{X=k\}=\left(\frac{1}{2}\right)^{k-1}\times\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^{k}$$

L'entropie s'écrit :

$$H(X)=-\sum_{k\geq 1}\left(\frac{1}{2}\right)^{k}\times\log\left(\frac{1}{2}\right)^{k}=\sum_{k\geq 1}\left(\frac{1}{2}\right)^{k}\times k$$

Si on pose $p=\frac{1}{2}$, on a :

$$H(X)=p\sum_{k\geq 1}p^{k-1}\times k=p\times\frac{d}{dp}\sum_{k\geq 1}p^{k}=p\times\frac{d}{dp}\frac{p}{1-p}=p\times\frac{(1-p)-p(-1)}{(1-p)^2}$$

soit :

$$\boxed{H(X)=\left(\frac{p}{(1-p)^2}\right)_{p=1/2}=2\,\mathrm{Sh}}$$

2. Chaque réponse peut être modélisée par une variable aléatoire X_n à deux valeurs. L'information moyenne apportée par X_n sur X est :

$$I(X_n;X)=H(X_n)-H(X_n/X)$$

Or, la connaissance de X entraîne celle de X_n, donc $H(X_n/X)=0$ et on a :

$$I(X_n;X)=H(X_n)$$

La maximisation de l'information moyenne apportée par X_n sur X équivaut à la maximisation de $H(X_n)$. X_n étant à deux valeurs possibles, le maximum de l'entropie est obtenu pour une loi uniforme, c'est à dire pour des réponses "oui" et "non" de probabilités égales.

Notons $X_1, X_2, ...$ les questions successives (en fait il s'agit plutôt des réponses aux questions car les questions sont déterministes). On notera :

$$\{X_k=1\}=\{\text{la réponse à la question k est "oui"}\}$$
$$\{X_k=0\}=\{\text{la réponse à la question k est "non"}\}$$

La première question à poser est "la valeur de X est-elle 1?" car les deux réponses possibles sont de probabilités égales puisque la probabilité que X prenne la valeur 1 est 1/2. Si la réponse à cette question est "oui", alors l'interrogation prend fin, sinon les valeurs restantes possibles pour X sont 2 et les valeurs supérieures ou égales à 3. On a :

$$P\{X_2=1\}=P\{X=2\}=P\{X\geq 3\}=\frac{1}{4}.$$

Il semble donc que la deuxième question à poser soit "la valeur de X est-elle 2?". Nous devons néanmoins vérifier que l'entropie de X_2 sachant $X_1=0$ est maximale. Pour ce faire calculons la loi de X_2 sachant $X_1=0$.

$$P\{X_2 = 1 / X_1 = 0\} = P\{X = 2 / X \neq 1\} = \frac{P\{X = 2 \cap X \neq 1\}}{P\{X \neq 1\}} = \frac{P\{X = 2\}}{P\{X \neq 1\}}$$

et :

$$P\{X_2 = 0 / X_1 = 0\} = P\{X \geq 3 / X \neq 1\} = \frac{P\{X \geq 3 \cap X \neq 1\}}{P\{X \neq 1\}} = \frac{P\{X \geq 3\}}{P\{X \neq 1\}}$$

Or $P\{X = 2\} = P\{X \geq 3\}$, donc $P\{X_2 = 1 / X_1 = 0\} = P\{X_2 = 0 / X_1 = 0\}$

Donc la deuxième question à poser est bien "la valeur de X est-elle 2?".

En procédant de même pour les rangs suivants, on établit sans peine qu'en cas de réponse négative à la question de rang k, la question de rang k + 1 est "la valeur de X est-elle k + 1?". Il s'agit d'un raisonnement par dichotomie consistant chaque fois à diviser les valeurs restantes possibles pour X en deux sous-ensembles de probabilités égales.

3. Compte tenu des questions posées, leur nombre sera égal à k si la valeur de X est k. Donc le nombre moyen de questions à poser pour découvrir X est :

$$\sum_{k \geq 1} k \times P\{X = k\} = E[X]$$

Comme X suit une loi géométrique de paramètre $p = \frac{1}{2}$, on a $E[X] = \frac{1}{p} = 2$

Le nombre moyen de questions à poser en utilisant la stratégie décrite à la question précédente est 2.

4. En termes de codage de source, l'élaboration de la séquence correspond à un codage de Shannon-Fano.

EXERCICE 3

1. En calculant les éléments binaires de contrôle, on obtient :

Mots source	Mots code
00	00000
01	01011
10	10101
11	11110

2. La matrice de contrôle s'écrit :

Soit $\boldsymbol{G}$ la matrice génératrice et $\boldsymbol{H}^{\boldsymbol{T}}$ la matrice de contrôle :

$$\boldsymbol{G}=\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\\ \mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\\ \mathbf{1} & \mathbf{1}\end{pmatrix} \qquad \boldsymbol{H}^{\boldsymbol{T}}=\begin{pmatrix}\mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0}\\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1}\end{pmatrix}$$

3. Les dimensions respectives de la matrice de contrôle et des mots reçus étant 3x5 et 5x1, le nombre de valeurs possibles pour le syndrome est $\mathbf{2^3 = 8}$. La valeur "000" du syndrome correspondant à l'absence d'erreur, il reste 7 configurations différentes. On peut en réserver 5 pour corriger toutes les configurations d'une erreur, deux sont alors disponibles pour corriger deux configurations de deux erreurs.

4. La table de décodage s'obtient simplement en calculant les syndromes des configurations d'erreurs. Ainsi :

syndromes	séquences z de poids minimum
000	00000
001	00001
010	00010
011	01000
100	00100
101	10000
110	00110
111	10010

Toutes les configurations d'une erreur sont corrigées, les deux configurations de deux erreurs corrigées sont, d'une part une erreur sur les troisième et quatrième éléments binaires et d'autre part une erreur sur les premier et quatrième éléments binaires.

5. Envisageons tout d'abord le cas avec codage décodage :

Soient les événements :

D_C = {le mot reçu est correctement interprété}

E_0 = {aucune erreur sur le mot transmis}

E_1 = {une erreur sur le mot transmis}

E_2 = {une des deux configurations de deux erreurs corrigées par la table apparaît sur le mot}

On a alors : $P\{D_C\} = P\left\{\bigcup_{i=0}^{2} E_i\right\} = \sum_{i=0}^{2} P\{E_i\}$

Or $P\{E_0\} = \left(1 - 10^{-2}\right)^5 = 0{,}951$

$P\{E_1\} = 5 \times 10^{-2} \times \left(1 - 10^{-2}\right)^4 = 0{,}048$

$P\{E_2\} = 2 \times \left(10^{-2}\right)^2 \times \left(1 - 10^{-2}\right)^3 = 1{,}94 \times 10^{-4}$

D'où :

$P\{D_C\} \approx 0{,}9992$

La probabilité d'erreur en utilisant le codage et la table de décodage est :

$\boxed{P\{\text{erreur}\} \approx 8 \times 10^{-4}}$

Envisageons le cas sans codage

$P\{D_C\} = \left(1 - 10^{-2}\right)^2 = 0{,}9801$

La probabilité d'erreur est donc à peu près égale à $\boxed{2 \times 10^{-2}}$

L'amélioration apportée par le codage est appréciable.