Fiche de Théorie de l'Information

I. Mesure de l'information

Notations

- X v.a discrète à valeur dans $x_1, \dots, x_n \; avec \; p_i = P(X = x_i)$
- Y v.a discrète à valeur dans $y_1, \dots, y_m \; avec \; q_i = P(Y = y_i)$
- $p_{ij} = P(X = x_i \cap Y = y_i)$

1. Entropie d'une source simple S

Source discrète simple (ou sans mémoire ou indépendante) : émet des symboles de manière indépendante avec la même loi de probabilité d'un symbole à l'autre

Entropie d'une variable aléatoire/source discrète simple :

Quantité moyenne délivrée par la source par symbole (unité : Sh)

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i = incertitude \; moyenne$$

Inégalité de Gibbs :

Pour deux v.a X et Y :

$$\sum_{i=1}^{n} p_i \log_2\left(\frac{q_i}{p_i}\right) \leq 0 \; ie - \sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i)$$

On a égalité quand $\forall i \in 1, \dots, n \; p_i = q_i$

On appelle divergence de Kullback-Leibler : D(P||Q) = $\sum_{i=1}^{n} p_i \log_2\left(\frac{p_i}{q_i}\right)$

Propriétés :

- H(S) est une fonction continue sur la distribution des probabilités
- Pour un alphabet de N symboles : $0 \leq H(S) \leq \log_2(N)$
- L'entropie est maximale pour la loi uniforme $1/N$

Redondance d'une source :

Ecart relatif à l'entropie maximale permise par la taille N de son alphabet

$$R(S) = 1 - \frac{H(S)}{\log_2 N}$$

Extension de la source :

L'extension d'ordre k de S, noté S^k émet des messages (mots) en nombre N^k composés de k symboles. On a donc $1 \; symbole \; étendu = 1 \; mot \; de \; k \; lettre$ et $H(S^k) = k \, H(S)$

2. Entropie entre 2 v.a et information mutuelle

Entropie conjointe du couple

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log_2 p_{ij}$$

Si X et Y sont indépendantes : $H(X,Y) = H(X) + H(Y)$

Si X = Y $H(X,Y) = H(X) = H(Y)$

On a $0 \leq H(X,Y) \leq H(X) + H(Y)$

Entropie conditionnelle

$$H(X|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log_2(P(X = x_i|Y = y_j))$$

Entropie conditionnelle de X sachant $Y = y_j$: $H(X|Y = y_j) = -\sum_{i=1}^{n}(X = x_i|\ Y = y_j)log_2(P(X = x_i\,|\,Y = y_j))$

Entropie conditionnelle de X sachant Y moyennée sur les valeurs de Y :$H(X|Y) = +\sum_{j=1}^{n} q_j H(X, Y = y_j)$

On a $H(X|Y) = H(X,Y) - H(Y)$ et $0 \leq H(X|Y) \leq H(X)$

Si X et Y indépendantes : $H(X|Y) = H(X)$

Si X=Y $H(X|Y) = 0$

Information mutuelle :

Quantité d'information moyenne partagée par X et Y

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

Si X et Y indépendantes : $I(X,Y) = 0$

Si X= Y $I(x,Y) = H(X) = H(Y)$

On a: $0 \leq I(X,Y) \leq H(X)\ \ et$ à $\leq I(X,Y) \leq H(Y)$

II. Codage de source (compression)

1. Caractérisation d'un codage

Codage de source : application des messages (lettre ou extension) de l'alphabet $s_1, \dots s_N$ de la source vers les mots-codes $C_i = c_1 c_2 \dots c_{l_i}$ où on note l_i la **longueur** du mot-code C_i ie le nombre de symbole composant le mot et $\forall i\ c_i \in alphabet\ de\ la\ sortie\ U, de\ taille\ Q$

Compacité ν

Longueur moyenne des mot-codes $\nu = \sum_{i=1}^{N} p(s_i).\, l_{s_i}$

Lorsque le code est sans perte, $H(U) = \frac{H(S)}{\nu}$ donc $\nu_{\min} = \frac{H(S)}{\log_2 Q} \leq \nu$

On note $H_Q = \frac{H(S)}{\log_2 Q} = -\sum_{i=1}^{N} p_i \log_Q p_i$ l'entropie en base Q de la source S

Efficacité et redondance du code

L'efficacité du code est $\eta = \frac{\nu_{\min}}{\nu} = \frac{H(S)}{\nu \log_2(Q)} = \frac{H(U)}{\log_2(Q)}$

La redondance du code ou taux de codage est $R(U) = 1 - \eta$

2. Qualités requises pour les codes

Un code doit pouvoir être décodé sans ambiguïté.

Code à décodage unique :

A chaque suite distincte de message de la source correspond une suite distincte de symboles en sortie du codeur. Il est donc :

- Régulier : un même mot code ne peut pas être associé à 2 mots différents
- Déchiffrable : à la réception d'une suite de symbole, il faut pouvoir trouver le début et la fin des mots-codes

Code instantané (ou irréductible)

Code à codage unique tel que le décodage d'un mot-code est possible dès la fin de sa réception, sans attendre la mot-code suivant.

Condition du préfixe

Un code est irréductible ssi aucun mot-code n'est le préfixe d'un autre

Cela revient à représenter le code par un arbre où les mots-codes sont uniquement des feuilles

Inégalité de Kraft :

Il existe un code instantané pour coder N messages avec des mots de longueurs $l_1, l_2, \dots, l_N$ construit avec des symboles de l'alphabet de sortie (taille Q) ssi : $\sum_{i=1}^{N} Q^{-l_i} \leq 1$

Si l'inégalité de Kraft est une égalité, on dit que le code est **optimal absolu**. On a alors $\nu_{\min} = \nu$ et $\forall i\ p_i = Q^{-n_i}$

3 CN d'optimalité :

- Les mot-codes les plus courts ont les probabilités les plus grandes
- Les 2 mots les plus longs sont de même longueur
- Parmi les mot-codes de longueur maximale, deux mots ne diffèrent que par leur dernier symbole

3. Codage de source

Théorème du codage de source (1er th. De Shannon) :

Il existe un procédé de codage instantané dont la compacité des mot-codes est aussi proche que l'on veut de sa borne minimale

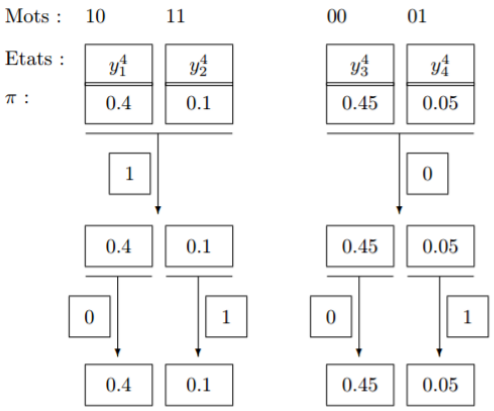
Repose sur 2 principes :

1. Pour une source S, on peut toujours construire un code instantané tel que :
$$\frac{H(S)}{\log_2 Q} = H_Q \leq \nu \leq \frac{H(S)}{\log_2 Q} + 1 = H_D + 1$$
2. Pour une source S, en appliquant l'extension d'ordre k, on peut toujours trouver un code vérifiant la condition du préfixe tel que
$$\frac{H(S)}{\log_2 Q} = H_D \leq \nu \leq \frac{H(S)}{\log_2 Q} + \frac{1}{k} = H_D + \frac{1}{k}$$

On peut donc choisir k suffisamment grand pour vérifier le th. De codage de source

Codage de Shannon-Fano

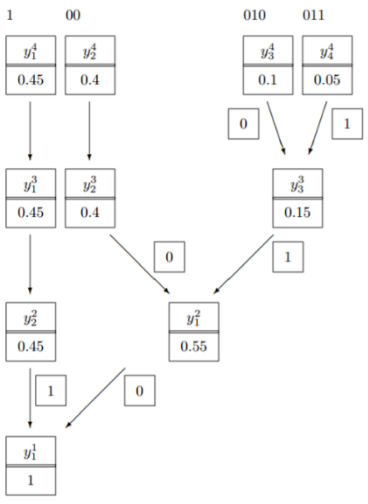
On cherche à maximiser l'entropie à la sortie du codeur en ayant les mêmes probabilités de part et d'autre des nœuds de l'arbre d'un même niveau. Ce code n'est pas optimal.



Codage de Huffman

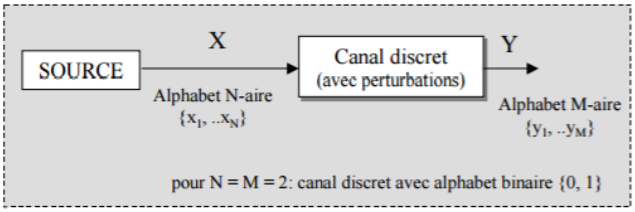
1. Ordonner les messages dans l'ordre des probabilités décroissantes
2. Additionner les Q messages de probabilité les plus faibles, qui forment un nouvel élément
3. S'il ne reste qu'un élément, on passe à 4. Sinon, on retourne à 1 avec le nombres d'élément réduit de Q-1
4. Attribuer 0, …Q-1 comme symbole aux transitions, jusqu'à atteindre le dernier niveau

Ce code est optimal



III. Capacité et codage canal

Notations :



1. Caractéristiques d'un canal

Un canal est **discret** si les alphabets d'entrée et de sortie ne contiennent qu'un nombre fini de symboles.

Un canal est **sans mémoire** si la sortie y_k ne dépend que de l'entrée x_k

Matrice de transition

$$[P(Y/X)] = \begin{bmatrix} p(y_1/x_1) & \cdots & p(y_M/x_1) \\ & p(y_j/x_i) & \\ p(y_1/x_N) & & p(y_M/x_N) \end{bmatrix}$$

Permet de faire le lien entre les entrées et les sorties

$\forall i,j\ m_{ij} = P(Y = y_i|X = x_j)$ et $P_Y = [P(Y|X)]^T P_X$

La somme des valeurs d'une ligne vaut 1

Canal uniforme par rapport à l'entrée

Les symboles sont tous affectés de la même manière par les erreurs : les lignes sont identiques à une permutation près.

Dans ce cas, l'entropie conditionnelle ne dépend pas de la loi de l'entrée ie $H(Y|X) = H(Y|X = x_i)$

Canal uniforme par rapport à la sortie

Les colonnes sont identiques à une permutation près.

Une loi uniforme en entrée induit une loi uniforme en sortie

Canal symétrique

Le canal est uniforme en sortie et en entrée et N=M

Vraisemblance

La vraisemblance des informations y en sortie de canal $p(y|C)$ où C est l'entrée.

Le décodage est optimal quand la vraisemblance est maximale

2. Capacité d'un canal discret

Capacité

La capacité d'un canal est l'information maximale pouvant traverser ce canal pour toute les distribution de probabilité possibles en entrée : $C = \max_{P(X)} I(X,Y) = \max_{P(X)} H(X) - H(X|Y)$

On a $C \leq \log_2 N$

Pour un canal symétrique :

- H(Y|X) est indépendante de H(Y) donc $C = \max_{P(X)} H(Y) - H_{cst}$
- H(Y) est maximal pour une loi d'entrée uniforme et vaut alors $\log_2 M$
- Donc $C = \log_2 M + \sum_{j=1}^{M} p(y_j|x_i) \log_2 \left(p(y_j|x_i)\right)$ ie C est maximale pour une loi d'entrée uniforme

3. Codage canal ou codage de détection et correction d'ereur

Pour pouvoir détecter les erreurs (voire les corriger) on insère des symboles de redondance en plus. Plus on ajoute de bit de contrôle, plus la fiabilité du code augmente.

Théorème de codage canal (2nd théorème de Shannon)

$R < C \Leftrightarrow il\ existe\ un\ codage\ canal\ de\ longueur\ n, de\ redondance\ R\ et\ de\ capacité\ C\ \ tel\ que\ P(erreur) < \epsilon$

Rendement du code

Le rendement du code correspond à la quantité d'information réelle sur l'information totale envoyée. Pour un code de longueur n contenant des mots codés sur $k \leq n$ bits, on a $R = \frac{k}{n}$

La **distance minimale** d'un code est le nombre de bit minimum qui diffèrent entre deux mots différents du code

Un **codage par bloc est linéaire** si les 2^k mots du code forment un sous espace vectoriel de F_2^n

Matrice génératrice d'un code C

Les mots c du code vérifient $c = mG, où\ m \in F_2^k\ est\ une\ séquance\ de\ k\ bits\ d'info$

$$\mathbf{G} = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & & \\ g_{k1} & \cdots & g_{kn} \end{bmatrix}$$ est la matrice génératrice du code C

$c \in C \Leftrightarrow c \in Im(G)$

On peut toujours écrire G sous **forme systématique** c-à-d : $G = [I_{k*k} P_{k*(n-k)}]$

Matrice de contrôle de parité

La matrice de contrôle de parité est la matrice $(n-k)*n$ définie par $GH^T = 0_{k*(n-k)}$

Si G est sous forme systématique, $H = [P^T_{k*(n-k)}\ I_{(n-k)*(n-k)}]$

$c \in C \Leftrightarrow c \in Ker(H)\ ie\ cH^T = 0$

Capacité de détection et de correction

Pour tout code linéaire par bloc de distance minimale d_{min},

- On peut détecter $d_{\min} - 1$ erreurs
- On peut corriger $\left\lfloor \frac{d_{\min}-1}{2} \right\rfloor$ erreurs

Détection et correction d'erreur

Si l'on reçoit la séquence r en sortie du canal, on calcule le **syndrome** $s = rH^T$

- Si s est nul, r est un mot du code
- Si s est non nul, il y a une erreur au moins. On cherche alors le mot de code c le plus proche tel que $r = c - e$ et donc $eH^T = rH^T$ avec e l'erreur