Théorie de la transmission de l’information

Introduction

1. Terminologie

• **Information :** Supposons que dans une certaine situation donnée puissent se réaliser N événements différents équiprobables. La probabilité de réalisation est événements est donc :

$$p = \frac{1}{N}$$

La réalisation d'un événement parmi les N possibles signifie que l'on obtient une information. Cette information sera d'autant plus grande que sa probabilité sera plus petite. Par définition l'information obtenue dans ce cas sera :

$$i = \log\left(\frac{1}{p}\right) = -\log\left(p\right) = \log\left(N\right)$$

Le logarithme est introduit afin d'assurer la propriété d'additivité à L'information

- **Message :** signal qui correspond à une certaine réalisation particulière parmi l'ensemble des idées, images et données devant être transmises à un correspondant.

- **Source d'information :** mécanisme servant à choisir, parmi l'ensemble de messages possibles et d'une manière imprévisible pour l'observateur un certain message particulier destiné à être transmis. Les réalisations particulières de la sources sont appelées *symboles*.

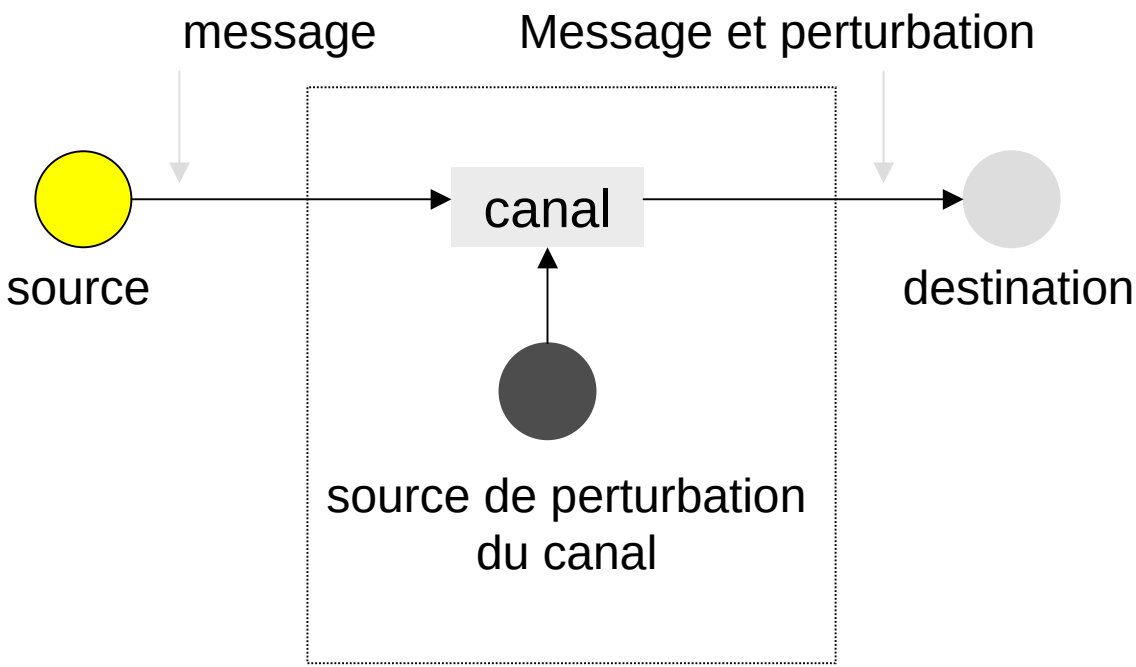
- **Canal :** totalité des moyens destinés à la transmission du message.

- **Codage :** transformation d'une séquence de signaux discrets (qui contient une information) dans une séquence devant être transmise par le canal.

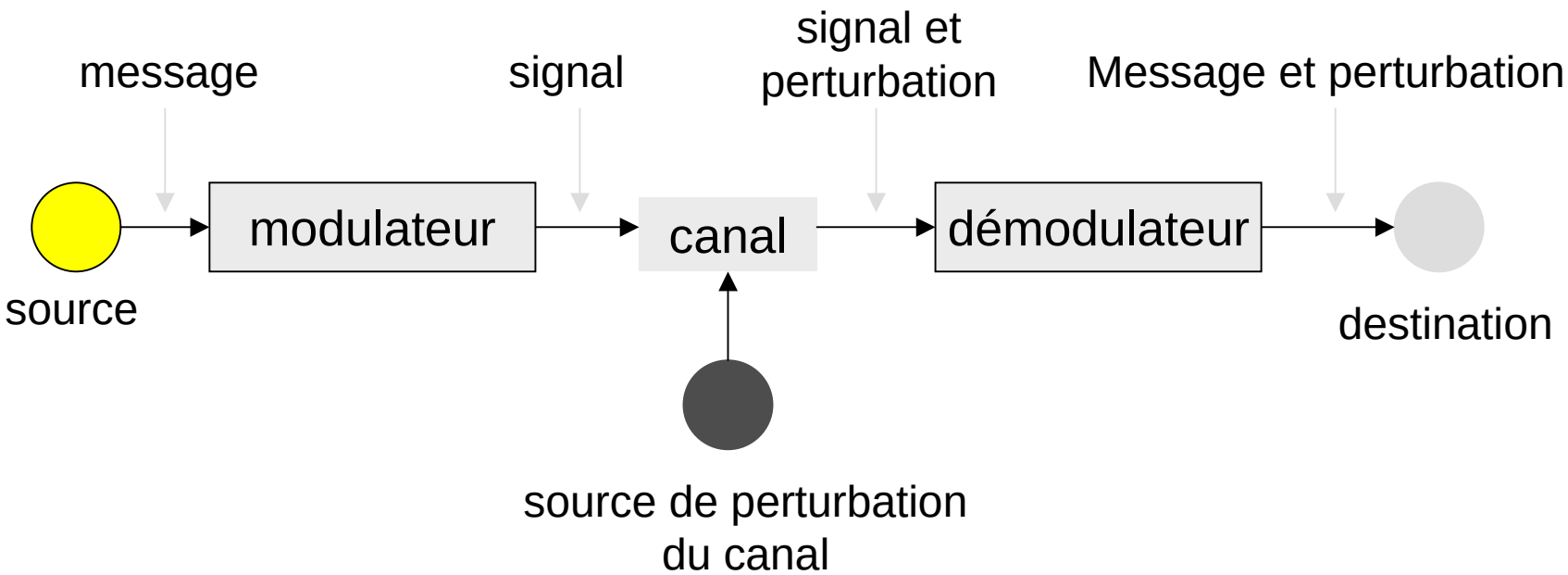
- **Décodage :** transformation de la séquence des signaux réceptionnés dans une séquence qui estime la séquence transmise à l'origine

2. Modèle d’un système de transmission de l’information

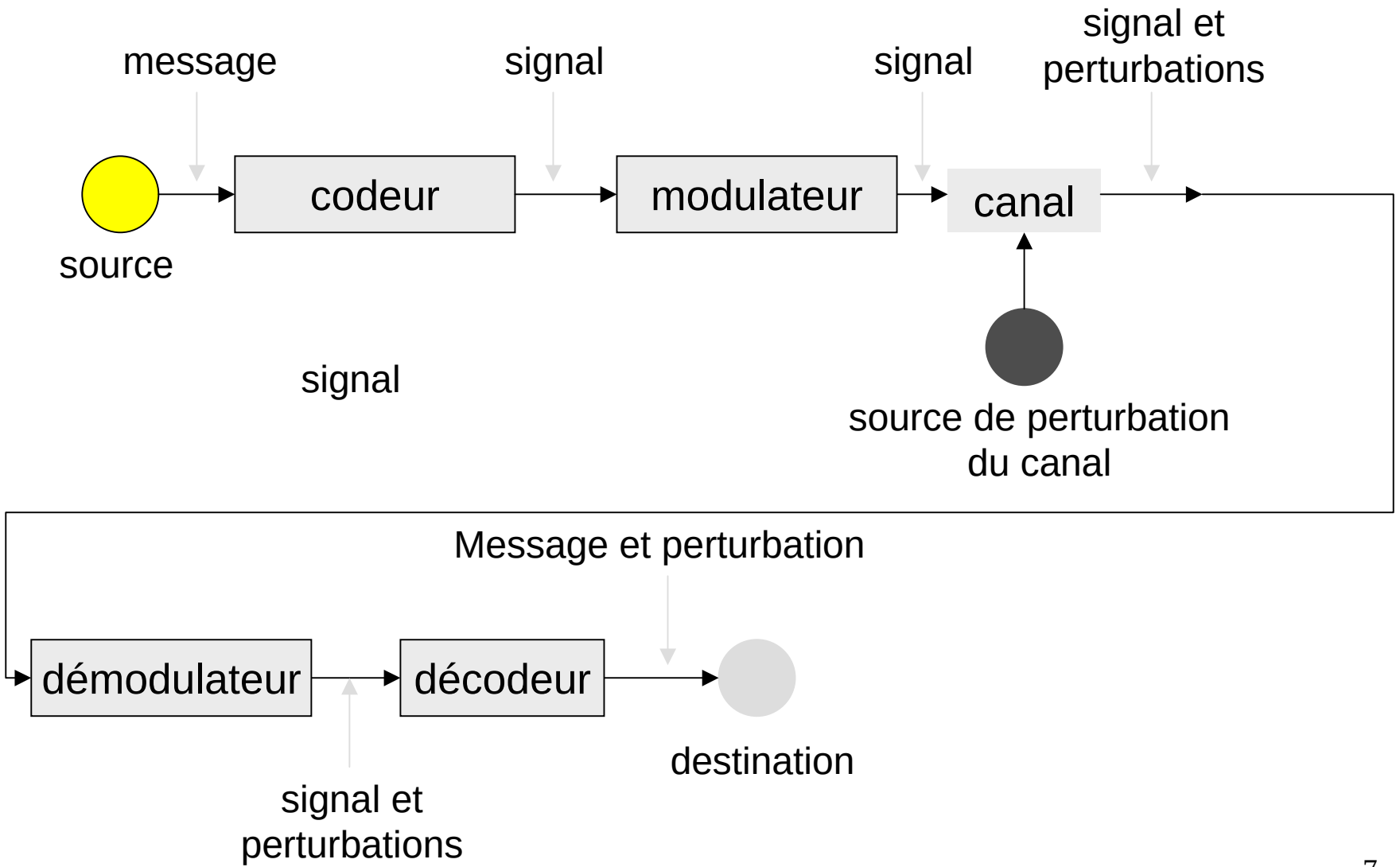
2.1 Transmission directe



2.2. Transmission avec modulation-démodulation



2.3 Transmission avec modulation-démodulation et codage décodage



Buts du codage

• **Codage source** : transformer la séquence générée par la source en une séquence plus courte.

• **Codage canal** : transformer la séquence devant être transmise par le canal à perturbations, dans une séquence plus longue, par appoint de symboles supplémentaires destinés à réaliser, à la réception, une opération de détection ou correction d'erreurs.

• **Cryptage :** transformer la séquence générée par la source en une autre séquence à partir de laquelle un intercepteur non autorisé ne puisse par extraire l'information transmise.

3. Qualité de la transmission

3.1. Critère de fidélité (Mean Squared Error : MSE)

$$\varepsilon = \frac{1}{T}\int_0^T \left(x(t) - y(t)\right)^2 dt$$

avec :
x(t) : signal transmis
y(t) : signal réceptionné

3.2. Rapport signal sur bruit (Signal to Noise Ratio : SNR)

$$\xi = \frac{\int_0^T \left(y(t)\right)^2 dt}{\int_0^T \left(\eta(t)\right)^2 dt}$$

avec :
n(t) : signal perturbateur
y(t) : signal réceptionné

3.3. Signal numérique

Probabilité de réceptionner un signal erroné

Mesure de l’information des signaux discrets

La mesure de l'information est donnée par la mesure de l'incertitude d'un système d'événements, c'est à dire du choix aléatoire d'un événement dans un ensemble d'événements possibles et distincts.
C'est une mesure objective.

1. Mesure de l'information dans le cas discret

Soit un ensemble fini d'événements possibles que l'on désigne sous le nom *d'espace des échantillons* et que l'on note :

$$[X] = [x_1, x_2 ...., x_n]$$

où $\bigcup_{i=1}^{n} x_i = E$: événement certain et

$\bigcap_{i=1}^{n} x_i = \Phi$: événement impossible

à chaque éléments de X est associée une probabilité donnée par la matrice :

$$[P_X] = [p(x_1), p(x_2) \ldots, p(x_n)] \quad \text{où} \quad \sum_{i=1}^{n} p(x_i) = 1$$

La mesure de l'incertitude sur la réalisation d'un événement x_i, notée par $U(x_i)$, est une fonction $F(p_i)$ de la probabilité à priori $p_i=p(x_i)$ de la réalisation de cet événement :

$$U(x_i) = F(p_i)$$

et représente l'incertitude initiale (a priori) sur la réalisation de l'événement x_i

Lorsque x_i se réalise, cette incertitude est écartée et l'on peut dire qu'on a obtenu une information $i(x_i)$ sur la réalisation de x_i. Cette information peut être définie comme étant :

- l'information obtenue sur x_i par réalisation de x_i ou comme
- l'annulation de l'incertitude sur la réalisation de x_i, après que x_i soit réalisé

En conséquence :

$$i\left(x_i\right)=U\left(x_i\right)$$

donc :

$$i\left(x_i\right)=F\left(p_i\right)$$ L'information est une mesure de l'incertitude

Supposons que dans le processus d'observation des événements xi, des perturbations interviennent. Dans ce cas il n'y aura pas toujours de correspondance entre les événements réalisés x_i et ceux observés y_i.
On notera :

$$\left[Y\right]=\left[y_1, y_2 \ldots., \ y_m\right] \quad \text{avec} \quad m \neq n$$

La mesure de l'incertitude $U(x_i/y_i)$ sur la réalisation de l'événement x_i, lorsque y_i est observé est une fonction $F[p(x_i/y_i)]$ de la probabilité conditionnelle de x_i conditionné par y_i :

$$U\left(x_i / y_i\right)=F\left[p\left(x_i / y_i\right)\right]$$

Cette fonction représente l'incertitude à posteriori sur la réalisation de x_i, lorsque y_i se réalise (après avoir observé y_i il reste une incertitude sur l'événement qui s'est réellement produit).

à partir de : $i(x_i) = U(x_i)$ et $U(x_i / y_i) = F[p(x_i / y_i)]$

On définit : $i(x_i ; y_i) = U(x_i) - U(x_i / y_i)$

qui représente l'information obtenue sur la réalisation de x_i, chaque fois que y_i est observé.

Dans une interprétation équivalente, $i(x_i;y_i)$ représente :

- l'information obtenue sur x_i quand on observe y_i ou comme
- la diminution de l'incertitude x_i, due à la réception de y_i

Sans perturbation : $U(x_i / y_i) = 0$ $\quad i(x_i ; y_i) = U(x_i)$

Avec de fortes perturbations : $U(x_i / y_i) = U(x_i)$ $\quad i(x_i ; y_i) = 0$

Cas général : $i(x_i ; y_i) = F[p(x_i)] - F[p(x_i / y_i)]$

2. Spécification de la fonction de mesure

La fonction de mesure doit être additive

Supposons l'événement x_i constitué de deux événements indépendants x_{i1} et x_{i2}, soit :

$$x_i = x_{i1} \cap x_{i2}$$

En postulant que l'information est additive, il vient :

$$i(x_i) = i(x_{i1}) + i(x_{i2}) \quad \text{ou} \quad U(x_i) = U(x_{i1}) + U(x_{i2})$$

$$\Rightarrow F[p(x_i)] = F[p(x_{i1})] + F[p(x_{i2})]$$

$$F[p(x_{i1}).p(x_{i2})] = F[p(x_{i1})] + F[p(x_{i2})] \text{ (indépendance de } x_{i1} \text{ et } x_{i2})$$

La solution de cette équation est donnée par :

$$\boxed{F(p) = -\lambda \log(p) \quad \lambda > 0}$$

Soit : $i(x_i) = -\lambda \log\ (p(x_i))$: information propre de x_i

et $i(x_i ; y_i) = -\lambda \log\ (p(x_i)) + \lambda (\log\ p(x_i / y_i))$

ou : $i(x_i ; y_i) = \lambda \log \left(\frac{p(x_i / y_i)}{p(x_i)} \right)$: information mutuelle associée à x_i et y_i

3. Unité de mesure de l'information

On a convenu de choisir, comme unité d'information, l'information que l'on obtient par le choix aléatoire d'un seul événement parmi deux événements équiprobables.
Dans ce cas :

$$[X] = [x_1, x_2] \text{ et } [P_X] = \left[\frac{1}{2}, \frac{1}{2}\right]$$

$$i(x_1) = i(x_2) = -\lambda \log \left(\frac{1}{2}\right) = 1$$

Si on choisit le logarithme à base 2 alors $\lambda=1$

$$i(x_1)=i(x_2)=\log_2(2)=1\ bit$$

De manière générale

$$i(x_i)=-\log_2[p(x_i)]$$

et

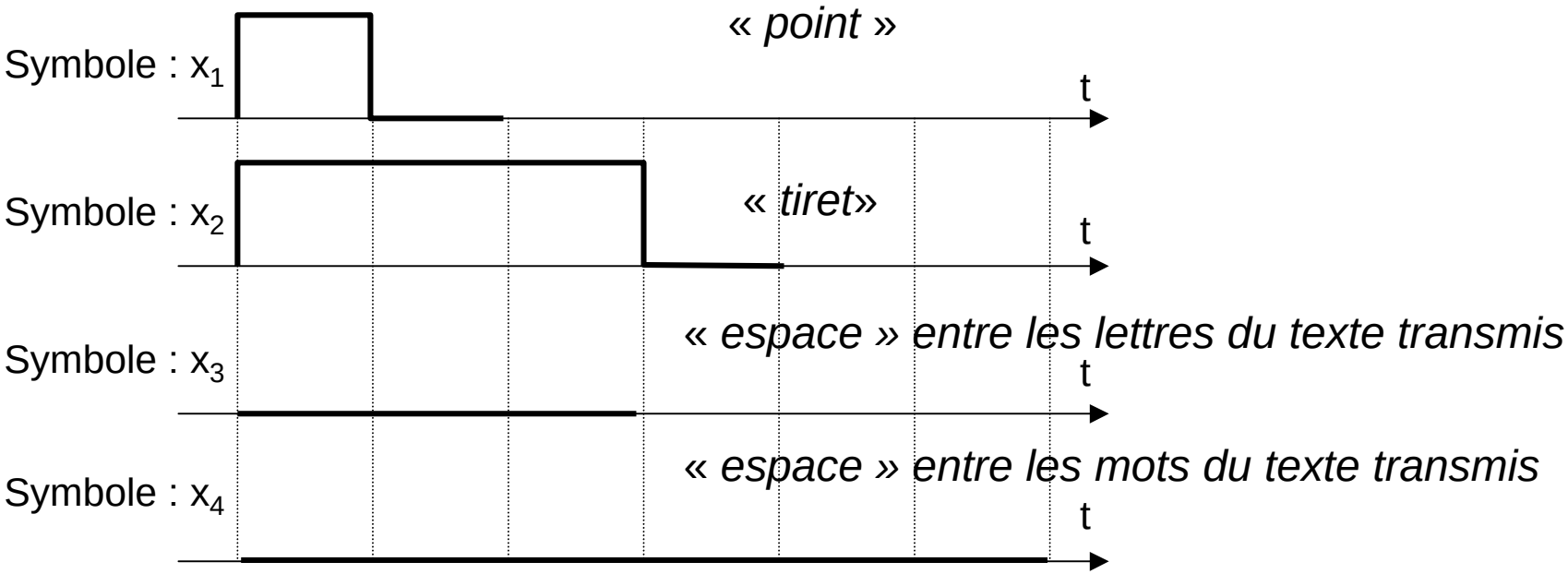
$$i(x_i;y_i)=\log_2\left(\frac{p(x_i/y_i)}{p(x_i)}\right)$$

4. Sources discrètes

Définition : suite de variables aléatoires discrètes : x_1, x_2,… x_n

Symbole ou lettre : élément fondamental irréductible contenant une information, c'est à dire une réalisation particulière de la source d'information.

Exemple du code Morse :



Alphabet : totalité des symboles, $[X]=[x_1, x_1,\ldots x_D]$
Pour le Morse le code est composé de 4 symboles. Un convertisseur A/N à N niveaux de quantification comportera N symboles.

Mot : succession finie de symboles.

Langue : totalité des mots construits à partir d'un seul alphabet

Source discrète sans mémoire : source pour laquelle la probabilité d'apparition d'un symbole ne dépend pas des symboles précédents.

$$p\left(x_{in} / x_{in-1}, x_{in-2} \ldots\ldots \right) = p\left(x_{in}\right)$$

Source discrète à mémoire : source pour laquelle la probabilité d'apparition d'un symbole dépend du symbole précédent ou d'une suite de symboles antérieurs.

Source stationnaire : source pour laquelle les probabilités d'apparition des différents symboles ne dépendent pas de l'origine des temps mais seulement de leurs positions relatives.

$$p\left(x_{in}\right) = p\left(x_{in+k}\right) \forall\ k \in Z$$

Source ergodique : source stationnaire à mémoire finie, pour laquelle toutes les suites de symboles sont équivalentes en probabilité (chaque symbole a même probabilité d'apparition quelque soit la suite).

Source à débit contrôlable : source générant des messages en réponse à une commande externe à la source (Ex : le système télégraphique).

Source à débit non contrôlable : source générant des messages à débit fixe, non contrôlable (Ex : échantillonnage d'un signal analogique en respectant le théorème de Shannon.

Source discrètes à contraintes fixes : source pour laquelle certains symboles ne peuvent être utilisés qu'en des conditions bien déterminées (Ex : en code Morse les symboles « espace »).

Source discrètes à contraintes probabilistes : source à mémoire. Dans un état donné, la source peut générer n'importe quel symbole avec une certaine probabilité qui dépend des symboles générés précédemment.

Source de Markov : Elle joue un rôle important dans les problèmes de communication. Elle se caractérise par :

$$p\left(x_{in} / x_{in-1}, x_{in-2} \ldots\ldots\right) = p\left(x_{in} / x_{in-1}\right)$$

où

$$x_{in}, x_{in-1}, x_{in-2} \ldots\ldots \in [X] = [x_1, x_2 \ldots. \; x_D]$$

La probabilité que la source génère un symbole quelconque $x_{in}=x_j$ au moment n, si elle a généré le symbole x_j au moment n-1, ne dépend pas des symboles générés aux moments antérieurs n-2, n-3, …

La probabilité $p\left(x_{jn} / x_{in-1}\right) = p_{ij}$ s'appelle *probabilité de transition* de l'état i à l'état j

On a évidemment : $\sum_{j=1}^{D} p_{ij} = 1$

La probabilité qu'au moment n la source se trouve dans l'état j est

$$p(x_{jn}) = \sum_{j=1}^{D} p(x_{jn}, x_{in-1}) = \sum_{j=1}^{D} p(x_{in-1}) p(x_{jn} / x_{in-1})$$

Ou : $p(x_{jn}) = \sum_{i=1}^{D} p(x_{in-1}) p_{ij}$

En posant :

$$P_n = \begin{bmatrix} p(x_{in}) \\ p(x_{2n}) \\ . \\ . \\ p(x_{Dn}) \end{bmatrix} \quad P_{n-1} = \begin{bmatrix} p(x_{in-1}) \\ p(x_{2n-1}) \\ . \\ . \\ p(x_{Dn-1}) \end{bmatrix} \quad T = \begin{bmatrix} p_{11} & p_{12} & . & . & p_{1D} \\ p_{21} & p_{22} & . & . & p_{2D} \\ . & & & & . \\ . & & & & . \\ p_{D1} & p_{D1} & & & p_{DD} \end{bmatrix}$$

Les relations s'écrivent : $P_n = T^T P_{n-1}$

Si la source est stationnaire les probabilités ne dépendant pas du temps

$$P_n = P_{n-1}$$

$$P_1 = T^T P_0 \qquad P_2 = T^T P_1 = T^T \left[T^T P_0 \right] = \left(T^T \right)^2 P_0$$

$$\boxed{P_n = \left(T^T \right)^n P_0}$$

5. Entropie

Considérons une suite stationnaire ergodique et sans mémoire, dont l'alphabet est :

$$[X] = [x_1, x_2 ...., \ x_n]$$

générée avec les probabilités :

$$[P_X] = [p(x_1), p(x_2) ..., \ p(x_n)]$$

L'information propre par symbole est :

$$i(x_i) = -\log\ (p(x_i))$$

L'information propre moyenne sur tous les symboles est :

$$\boxed{H(X) = -\sum_{i=1}^{n} p(x_i)\ \log\ (p(x_i))}$$

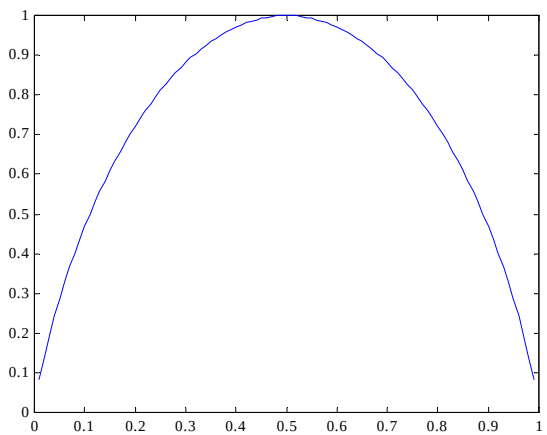
H(X) s'appelle entropie de la source, elle représente l'incertitude moyenne à priori que l'on a sur les événements [X].

L'entropie est la quantité d'information qu'apporte en moyenne une réalisation de X(n), sa valeur est maximale pour $p(x_1)= p(x_2)= ….p(x_n)$.

H(x) représente le nombre moyen de bits qu'il faut pour coder la source.

Exemple : soit une source binaire, $X=[x_1,x_2]$ et $p=p(x_1)$. L'entropie de cette source est :

$$H(X)=-p\log_2(p)-(1-p)\log_2(1-p)$$



L'entropie est positive ou nulle. Pour p=0 ou p=1 elle est nulle la source est totalement prédictible. Elle est maximale pour p=1/2 lorsque les deux symboles sont équiprobables.

6. Débit d'information et redondance d'une source

Si la durée moyenne d'un symbole est τ, alors le débit d'information de la source sera :

$$H_t(X) = \frac{H(X)}{\tau}$$

Pour indiquer l'écart entre l'entropie d'une source et sa valeur maximale, on définit la redondance comme la différence entre la valeur maximale possible de l'entropie d'une source et sa valeur réelle :

$$R_t = H_{\max}(X) - H(X)$$

La redondance relative s'exprime par :

$$\rho_s = \frac{H_{\max}(X) - H(X)}{H_{\max}(X)} = 1 - \frac{H(X)}{H_{\max}(X)} = 1 - \frac{H(X)}{\log(n)}$$

7. Entropie de la source de Markov

La valeur moyenne de l'information contenue dans le symbole à l'état i, est donnée par :

$$H_i = -\sum_{j=1}^{D} p_{i,j} \log(\ p_{i,j})$$

Si on considère tous les états de la source (au nombre de D), l'entropie de la source sera donnée par la moyenne des entropies de chaque état :

$$H = -\sum_{j=1}^{D} p(x_i) H_i = -\sum_{i=1}^{D} \sum_{j=1}^{D} p(x_i).\, p_{i,j} \log(\ p_{i,j})$$

Exercices : 1, 2, 3, 4, 5

Codage source sans perturbation

1. Introduction

• En général l'alphabet de la source diffère de l'alphabet du canal. Le but du codage de source est de permettre le passage de l'alphabet de la source à celui du canal.

• Afin que l'efficacité soit maximale, on réalise *une adaptation statistique de la source au canal,* soit :

$$C = \max(\ \ H(X)) = \log(\ \ D\ )$$

Ou D est le nombre de symboles de l'alphabet du canal. Le but du codage de source est donc de transformer la source primaire en une source a entropie maximale.

2. Codes à décodage unique

Soit une source $[S]=[s_1, s_2,\ldots, s_N]$

dont les probabilités sont : $[P]=[p(s_1), p(s_2),\ldots, p(s_N)]$

Soit $[X]=[x_1, x_2,\ldots, x_D]$ l'alphabet du code (donc du canal)

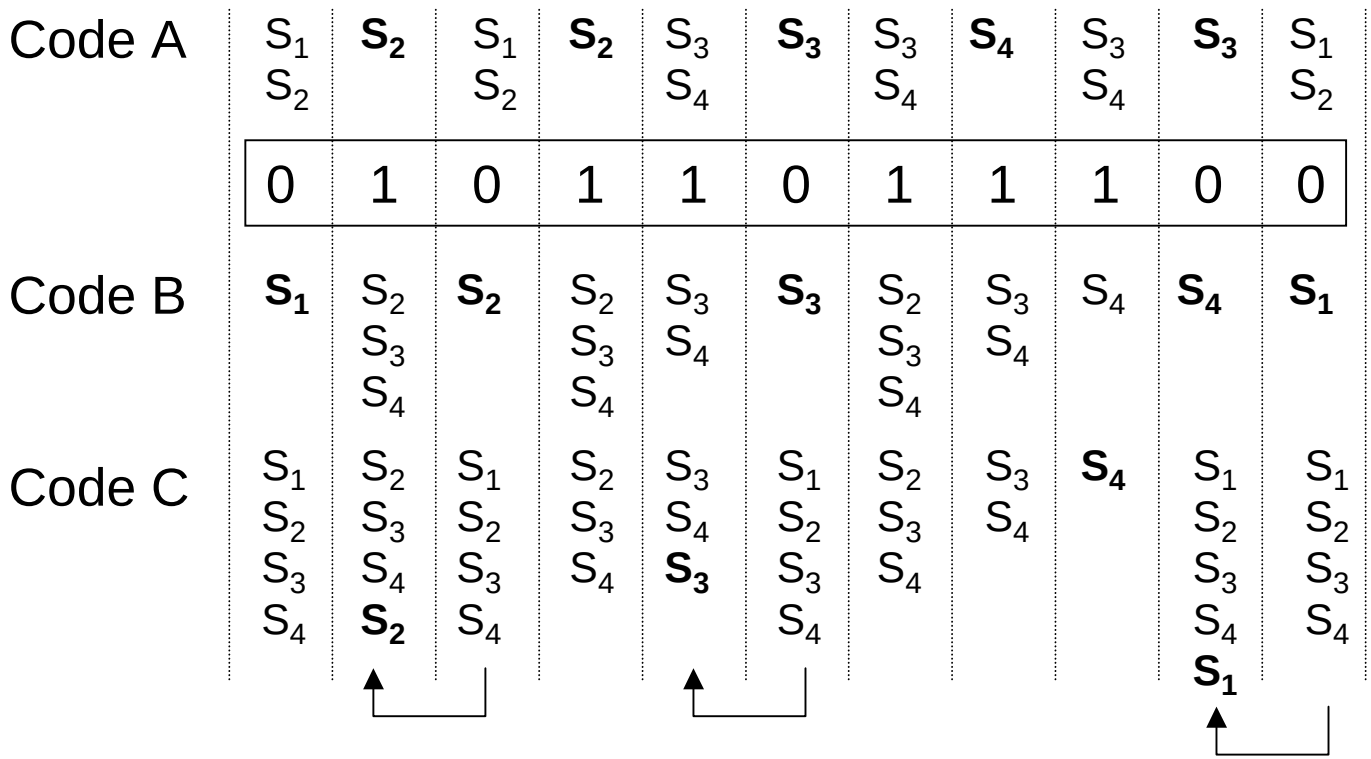
Avec ces lettres on forme un nombre N de mots-code :
$[C]=[c_1, c_2,\ldots, c_N]$

Les mots-code sont des successions finies de lettres de l'alphabet [X], le codage établi une relation bijective entre les symboles s_k de S et les mots C_k de C. On peut cependant à partir de X former des mots qui n'ont pas de correspondant dans S. Les mots auxquels correspondent des symboles de S, s'appellent *mots-code*.

Si les mots-code sont choisis convenablement on peut construire un code à décodage unique qui n'a pas besoin de signe séparateur entre les mots.

Exemple de code à décodage unique

S_k	Code A	Code B	Code C	Code D
S_1	00	0	0	0
S_2	01	10	01	10
S_3	10	110	011	110
S_4	11	1110	0111	111



Code non instantané (retard de décodage)

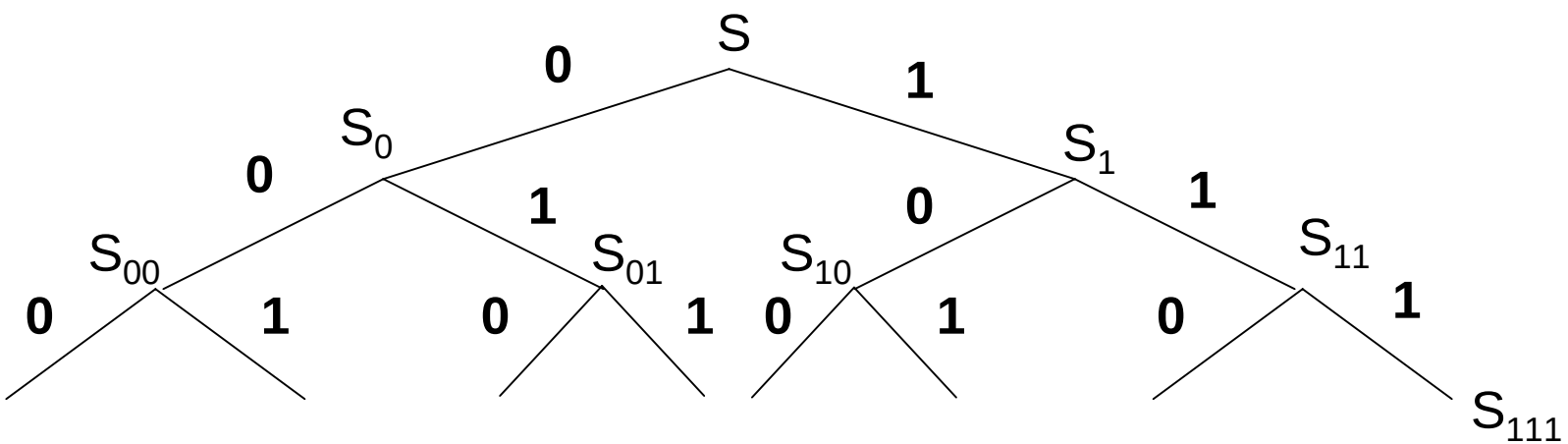
Conditions pour qu'un code soit instantané

Préfixe : Soit $c_i=x_{i1}, x_{I2}, \ldots, x_{im}$ un mot du vocabulaire d'un code. La suite de lettres $x_{i1}, x_{I2}, \ldots, x_{ik}$ avec k<m, s'appelle le préfixe du mot c_i.

La condition nécessaire et suffisante pour qu'un code soit instantané est qu'aucun mot du code ne soit le préfixe d'un autre mot du code.

Algorithme de construction d'un code binaire instantané

$$S=\left[\underbrace{s_1\, s_2,\ldots,\ s_k}_{S_0},\underbrace{s_{k+1},\ldots\ldots,\ s_N}_{S_1}\right]\quad S_1=\left[\underbrace{s_{k+1},s_{k+2},\ldots\ s_l}_{S_{10}},\underbrace{s_{l+1},\ldots\ldots,\ s_N}_{S_{11}}\right]\ldots\ldots$$



3. Longueur moyenne d'un mot code

Soit t_i le temps de transmission du mot-code c_i. Si le coût de transmission est fonction linéaire du temps de transmission, alors le coût moyen par message est :

$$\overline{C} = \sum_{i=1}^{N} t_i \; p(c_i) = \sum_{i=1}^{N} t_i \; p(s_i)$$

Si toutes les lettres x_i de l'alphabet [X] ont la même durée τ de transmission alors :

$t_i = l_i \,.\, \tau$ l_i : longueur du mot-code c_i

Si on considère pour simplifier que τ=1

$$t_i = l_i$$

$$\overline{C} = \sum_{i=1}^{N} p(s_i).l_i = \overline{l}$$

Le coût moyen de transmission est égal à la longueur moyenne d'un mot

4. Limite inférieure de la longueur moyenne d'un mot code

Soit une source $[S]=[s_1, s_2,\ldots, s_N]$

dont les probabilités sont : $[P]=[p(s_1), p(s_2),\ldots, p(s_N)]$

Soient les mots-code : $[C]=[c_1, c_2,\ldots, c_N]$

dont les probabilités sont les mêmes que les messages de la source : $[P_c]= [P]=[p_1=p(s_1), p_2= p(s_2),\ldots, p_N= p(s_N)]$

Les longueurs des mots-code sont : $[L]=[l_1, l_2,\ldots, l_N]$

L'alphabet du code est : $[X]=[x_1, x_2,\ldots, x_D]$

Entropie de la source est :

$$H(S)=H(C)=-\sum_{i=1}^{N} p(s_i)\log\left(p(s_i)\right)$$

Entropie de l'alphabet du code [X] est :

$$H(X)=-\sum_{i=1}^{N} p(x_i)\log\left(p(x_i)\right)$$

$$H(S)=H(C)=\bar{l}.H(X)$$

$$Max\left[H(S)\right]=Max\left[H(S)\right]\Rightarrow p(x_1)=p(x_2)=.....=p(x_D)=\frac{1}{D}$$

$$H(S)\leq \log(\ D\ )$$

donc :

$$H(S)=H(C)=\bar{l}.H(X)\leq \bar{l}.\log(\ D\ )$$

$$\boxed{\bar{l}\geq \frac{H(S)}{\log(\ D\ )}=\bar{l}_{\min}}$$

5. Capacité, efficacité et redondance du code

Capacité : $C = \max(H(X)) = \log(D)$

Efficacité du code : $\eta = \frac{\bar{l}_{\min}}{\bar{l}}$

$$\begin{cases} \bar{l}_{\min} = \dfrac{H(S)}{\log(D)} \\ \bar{l} = \dfrac{H(S)}{H(X)} \end{cases} \Rightarrow \boxed{\eta = \frac{H(S)}{\bar{l}\log(D)} = \frac{H(X)}{\log(D)}}$$

Redondance :

$$\boxed{\rho = 1 - \eta = \frac{\bar{l}\log(D) - H(S)}{\bar{l}\log(D)} = \frac{\log(D) - H(X)}{\log(D)}}$$

Exemple : [S]=[s_1, s_2, s_3 , s_4] et : [P]=[1/2, 1/4;, 1/8, 1/8]

$$H(S)=-\sum_{i=1}^{4} p(s_i)\log\left(p(s_i)\right)=\frac{7}{4}$$ bit/symbole

Soit [X]=[0,1] et le code suivant :

$$s_1 \rightarrow 00$$
$$s_2 \rightarrow 01$$
$$s_3 \rightarrow 10$$
$$s_4 \rightarrow 11$$

$$\bar{l}=\sum_{i=1}^{4} p(s_i).l_i=2.(\sum_{i=1}^{4} p(s_i))=2$$

$$\eta=\frac{H(X)}{\bar{l}\log(D)}=\frac{7/4}{2\log(2)}=\frac{7}{8}=0,875 \quad (D=2)$$

$$\rho=1-\eta=1-\frac{7}{8}=\frac{1}{8}=0.125$$

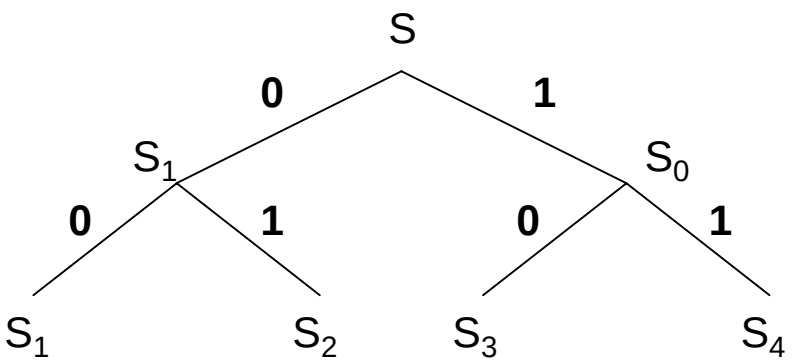
Soit [X]=[0,1] avec un autre le code :

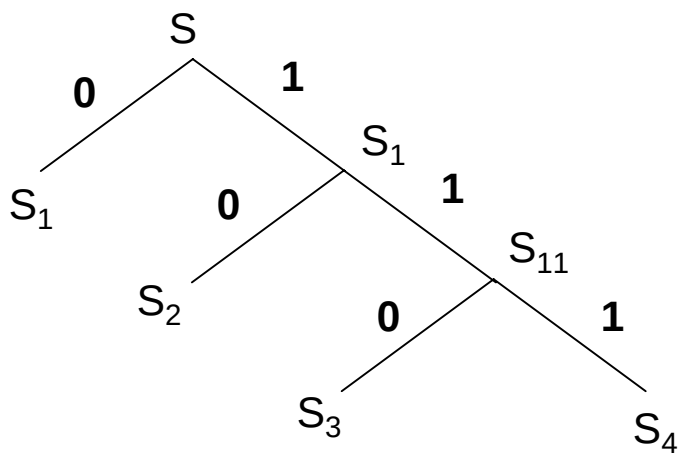
$$s_1 \rightarrow 0$$
$$s_2 \rightarrow 10$$
$$s_3 \rightarrow 110$$
$$s_4 \rightarrow 111$$

$$\bar{l} = \sum_{i=1}^{4} p(s_i).l_i = 1.75$$

$$\eta = \frac{7/4}{1.75 \log(2)} = 1$$

$$\rho = 1 - \eta = 0$$





6. Codes optimaux absolus

$$\frac{H(S)}{\log(D)} = \bar{l}_{\min} \Rightarrow H(S) = H(C) = \bar{l}_{\min} \log(D)$$

Ceci est vrai si $p(x_1) = p(x_2) = \ldots p(x_D) = 1/D$

Dans ce cas : $\eta = \frac{H(X)}{\log(D)} = 1$: **code optimal absolu**

Les lettres de l'alphabet étant considérées comme indépendantes

$$p(s_i) = p(c_i) = \left(\frac{1}{D}\right)^{l_i} = D^{-l_i}$$

Comme : $\sum_{i=1}^{N} p(s_i) = 1$ alors : $\sum_{i=1}^{N} D^{-l_i} = 1$

C 'est une condition nécessaire et suffisante pour qu'il existe un code absolu

$$\sum_{i=1}^{N} D^{-l_i} \leq 1$$: Inégalité de Mc Millan (code irréductible)

7. Premier théorème de Shannon

On a vu que : $p(s_i) = \left(\frac{1}{D}\right)^{l_i} \Rightarrow l_i = \frac{-\log\left(p(s_i)\right)}{\log(D)}$

Etudions ce qui arrive lorsque les probabilités d'apparition des messages à coder sont arbitraires.

Dans ce cas $r_i = \frac{-\log\left(p(s_i)\right)}{\log(D)}$ n'est généralement pas un nombre entier

La longueur du mot c_i du code [C] est alors choisie comme suit :

$$\frac{-\log\left(p(s_i)\right)}{\log(D)} \leq l_i \leq \frac{-\log\left(p(s_i)\right)}{\log(D)} + 1$$

l_i est le nombre entier le plus proche de r_i :

Il faut vérifier si les longueurs l_i satisfont l'inégalité de Mc Millan autrement dit si on peut former un code absolu (irréductible).

$$\frac{-\log\left(p(s_i)\right)}{\log(D)} \leq l_i \Leftrightarrow \log\left(p(s_i)\right) \leq l_i \log(D) = \log(D^{l_i})$$

ou encore : $D^{-l_i} \leq p(s_i)$

d'où : $\sum_{i=1}^{N} D^{-l_i} \leq \sum_{i=1}^{N} p(s_i) = 1$

Il existe donc un code absolu (irréductible) ayant des mots de longueurs l_i à partir de :

$$\frac{-\log\left(p(s_i)\right)}{\log(D)} \leq l_i \leq \frac{-\log\left(p(s_i)\right)}{\log(D)} + 1$$

On obtient :

$$\frac{-\sum_{i=1}^{N} p(s_i)\log\left(p(s_i)\right)}{\log(D)} \leq \bar{l} \leq \frac{-\sum_{i=1}^{N} p(s_i)\log\left(p(s_i)\right)}{\log(D)} + 1$$

soit : $\frac{H(S)}{\log(D)} \leq \bar{l} \leq \frac{H(S)}{\log(D)} + 1$

Ceci est vrai pour toute source sans mémoire. En particulier pour une source fictive [S^n] où chaque symbole est constitué à partir d'une succession de n symboles de la source [S]. Dans ce cas on montre que l'entropie de la source [S^n] est :

$$H(S^n) = nH(S)$$

De cette façon au lieu de coder symbole par symbole, on fait un codage par groupe de n symboles, on note $\bar{l}_n$ la longueur moyenne d'un mot-code on a alors :

$$\frac{H(S^n)}{\log(D)} \leq \bar{l}_n \leq \frac{H(S^n)}{\log(D)} + 1$$

ou encore

$$\frac{H(S)}{\log(D)} \leq \frac{\bar{l}_n}{n} \leq \frac{H(S)}{\log(D)} + \frac{1}{n}$$

à la limite si n est très grand on a :

$$\lim_{n \to \infty} \left(\frac{\bar{l}}{n} \right) = \frac{H(S)}{\log(D)} = \bar{l}$$

ou $\bar{l}$ est la longueur moyenne d'un mot code du code [C]

$$\Rightarrow \frac{H(S)}{\bar{l}} = \log(D)$$

$\frac{H(S)}{\bar{l}}$ peut être amenée aussi proche que l'on veut de la capacité du code log(D)
Ceci constitue le premier théorème de Shannon

Bien sur en pratique n aura toujours une valeur finie et on essaiera de construire des codes dont l'efficacité est proche de 1.

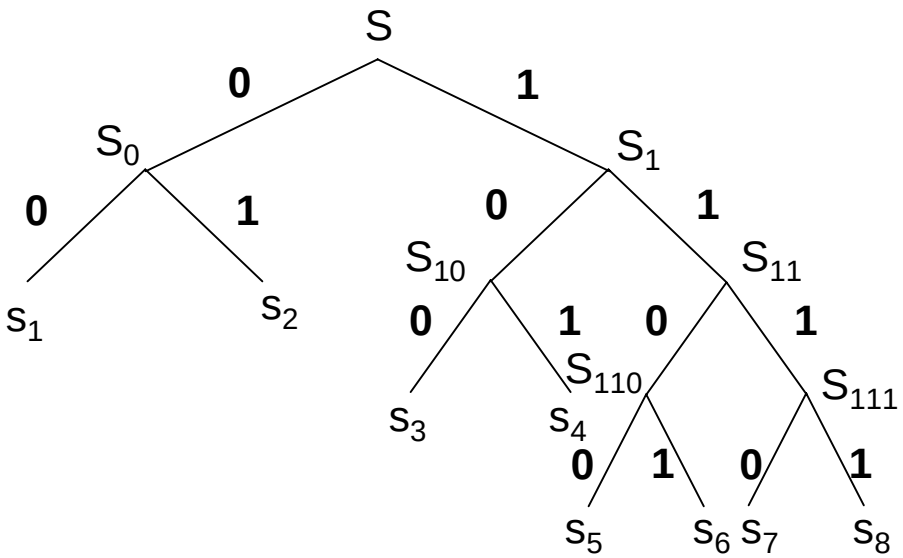
8. Codage de Shannon-Fano

Soit une source $[S]=[s_1, s_2,\ldots, s_N]$ qui peut être divisée en deux ensembles S_0 et S_1 dont les probabilités sont $p(S_0)=p(S_1)= ½$, en supposant à nouveau que S_0 et S_1 puissent être divisées en deux ensembles S_{00}, S_{01} et S_{10}, S_{11} avec les probabilités sont $p(S_{00})=p(S_{01})=p(S_{10})=p(S_{11})= 1/4$, et ainsi de suite jusqu'à ce que les ensembles en question ne contiennent plus qu'un élément **alors le codage sera absolu**

Exemple : [S]=[s_1, s_2, s_3 , s_4, s_5, s_6, s_7 , s_8] et

[P]=[1/4, 1/4, 1/8, 1/8 1/16, 1/16, 1/16, 1/16]

s_k	$P(s_k)$					c_k	l_k
s_1	0,25	0	0			00	2
s_k	0,25		1			01	2
s_3	0,125	1	0	0		100	3
s_4	0,125			1		101	3
s_5	0,0625		1	0	0	1100	4
s_6	0,0625				1	1101	4
s_7	0,0625			1	0	1110	4
s_8	0,0625				1	1111	4



$$H(S) = -\sum_{i=1}^{8} p(s_i)\log\left(p(s_i)\right) = 2,75$$

La longueur moyenne d'un mot-code est :

$$\bar{l} = \sum_{i=1}^{8} l_i \log\left(p(s_i)\right) = 2,75 \quad \bar{l}_{\max} = \frac{H(S)}{\log(D)} = 2,75 \ (D=2)$$

$$\boxed{\eta = 1}$$

9. Codage binaire de Huffman

Pour toute source discrète sans mémoire X(n), il existe un code instantané optimal représentant exactement cette source et uniquement décodable

Algorithme de Huffman

• On construit un arbre en partant des noeuds terminaux.

• On part de deux listes $\{x^1, \ldots, x^{Lx}\}$ et {px(1), …,px(Lx)},

• On sélectionne les deux symboles les moins probables, on crée deux branches dans l'arbre et on les étiquettes par les deux symboles binaires 0 et 1.

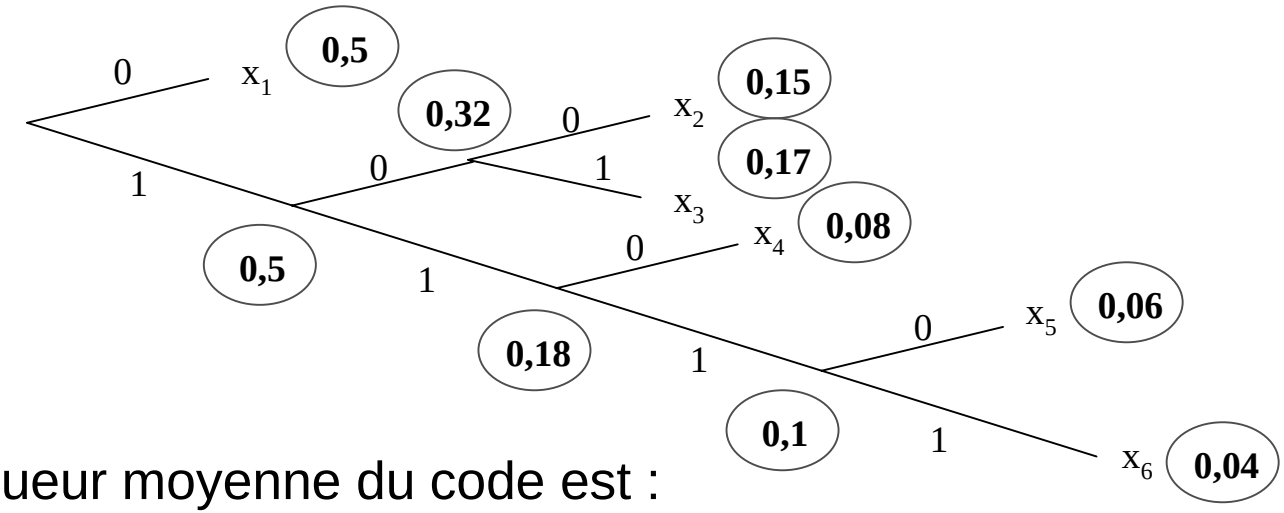
• On actualise les deux listes en rassemblant les deux symboles utilisés en un nouveau symbole et en lui associant comme probabilité la somme des deux probabilités sélectionnées.

• On recommence les deux étapes précédentes tant qu'il reste plus d'un symbole dans la liste.

cet algorithme est l'algorithme optimal
(Longueur moyenne des mots la plus faible)

Exemple:

Symboles	x_1	x_2	x_3	x_4	x_5	x_6
Probabilités	0,5	0,15	0,17	0,08	0,06	0,04



La longueur moyenne du code est :

$$\bar{l} = \sum_{i=1}^{Lx} px\,(x)\, l(c^i) = 2,1\ bits$$

L'entropie est de 2,06 valeur très proche de la longueur moyenne obtenue par Huffman.

$$\eta = \frac{H(S)}{\bar{l}\log(2)} = \frac{2,06}{2,1} = 0,981$$

Exercices : 6, 7

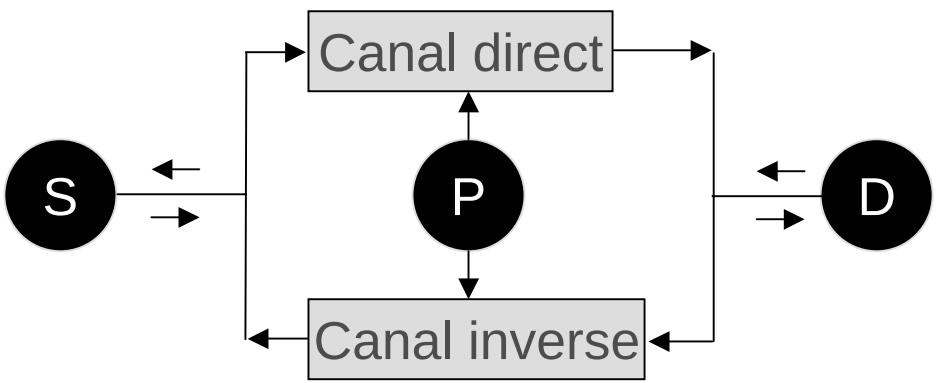
Codage canal pour canaux à perturbations

1. Introduction

• Le codage source transforme une source quelconque en une source à entropie maximale afin d'obtenir une efficacité aussi haute que possible.

• En présence d'un canal avec perturbations on ajoute quelques symboles appelés « *symboles de contrôle* » avant transmission afin d'indiquer au destinataire la présence d'erreurs voire de lui donner la possibilité de les corriger, on parle alors de *codes détecteurs d'erreurs* et de *codes correcteurs d'erreurs*.

• Lorsqu'il s'agit de détection d'erreurs, il faut prévoir un canal de retour permettant la répétition du message, ce canal peut être de faible capacité.



2. Classification des codes correcteurs d'erreurs

Lorsque le processus de détection ou de correction opère sur des blocs de n symboles, on dit qu'on a affaire à *des codes en blocs*, la suite des n symboles constituant un mot.

Lorsque le traitement a lieu de manière continue, on dit qu'on a affaire à des *codes convolutifs*.

Il existe deux types de codes en blocs :

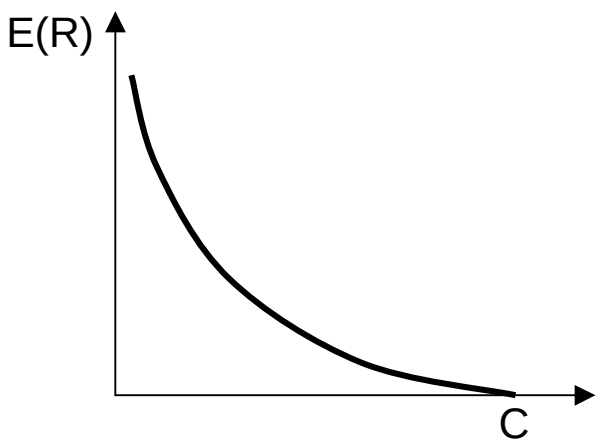
- les codes groupe sont ceux pour lesquels les mots sont considérés comme étant des éléments dans un espace vectoriel, à savoir des vecteurs,
- Les codes cycliques sont ceux considérés comme étant des éléments d'une algèbre, à savoir les polynômes.

3. Théorème de Shannon pour les canaux à perturbations

C'est un théorème d'existence :
Pour une source à débit d'information R bit/s et un canal de capacité C bit/s si R<C, il y aura un code ayant des mots d'une longueur n, de sorte que la probabilité d'erreur de décodage P_E soit :

$$P_E = 2^{-nE(R)}$$

Où :
n est la longueur du mot-code et E(R) est une fonction non-négative appelée exposant de l'erreur



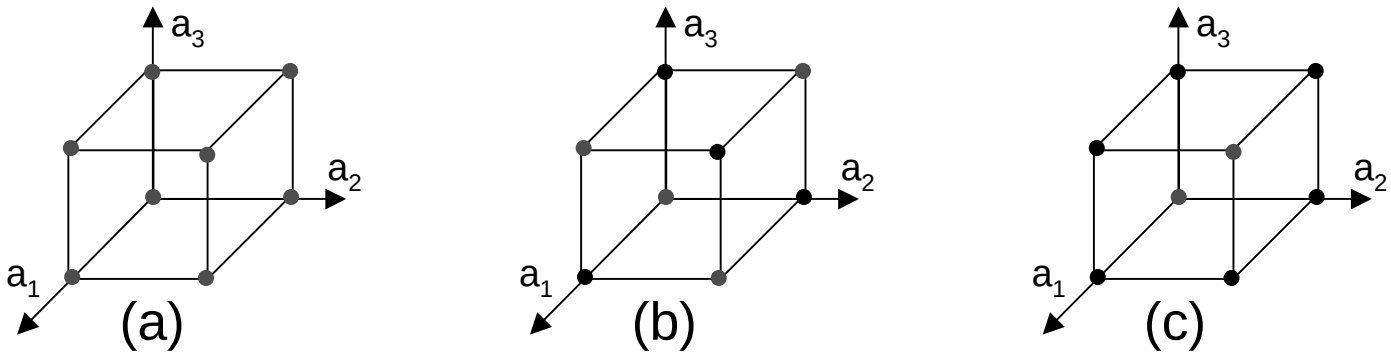
Quel que soit le niveau des perturbations d'un canal, on peut toujours y passer des messages avec une probabilité d'erreur aussi faible que l'on veut.
En pratique si R<0.5C, il existe des codes qui permettent P_E très faible.

4. Codes-groupe

Code en blocs où les n symboles constituant un mot sont considérés comme étant un vecteur dans un espace n-dimensionnel.
Les composantes d'un mot-code sont représentés sous forme matricielle : W=[a_1, a_1,… a_n];
On se restreint aux codes binaires $a_i \in (0,1)$
Il y a donc la possibilité de créer $N = 2^n$ mots-codes
Afin de détecter la présence d'erreurs, on procède comme suit :
On partage l'ensemble W en deux sous ensembles complémentaires V et F,
On attribue un sens à tous les éléments de V (ce sont donc des mots codes) tandis que les éléments de F sont dépourvus de sens.
Supposons Card(V)=2^k=S avec k<n
L'information moyenne transmise par mot-code est I=log(S)=k
tandis que l'information moyenne par symbole est i_k=k/n
Il y a bien sûr une perte d'information moyenne (si V=W alors I=n et i_k=1)

Exemple pour n=3

Card(W)=2^3=8, en rouge les mots-code, en blanc les mots dépourvus de sens.



(a) Aucune détection et donc correction n'est possible,
(b) Les erreurs sont détectées mais pas corrigées
(c) Les erreurs sont détectées et peuvent être corrigées

On notera v_i les mots-codes et v'_i les mots réceptionnés.

$v_i=[a_{i1}, a_{i2},\ldots.., a_{in}]$

$v'_i=[a'_{i1}, a'_{i2},\ldots.., a'_{in}]$

Si $v_i= v'_i$ la transmission est sans erreur

Les mots-codes comme éléments des classes voisines

On considère le sous ensemble V (sous espace vectoriel de W qui a une structure de groupe) et on attribue un sens à ces éléments v. Par rapport à V, on peut former les classes voisines comme suit :

- la première classe est formée d'éléments de v de V ayant un sens commençant avec l'élément nul
- dans le deuxième classe, on choisit comme élément un des mots (dépourvus de sens) ayant le plus petit nombre de composantes « 1 »qui ne figurent pas dans la première classe, on note $\varepsilon 1$ cet élément
- le restant des éléments de la deuxième classe sera formé en additionnant modulo 2 l'élément $\varepsilon 1$ aux éléments de la première classe comme suit:

0	v_1	v_2	……	V_{s-1}
$\varepsilon 1$	$v_1+\varepsilon 1$	$v_2+\varepsilon 1$	……	$V_{s-1}+\varepsilon 1$
$\varepsilon 2$	$v_1+\varepsilon 2$	$v_2+\varepsilon 2$	……	$V_{s-1}+\varepsilon 2$
.				

On continue l'opération jusqu'à ce que tous les éléments de W soient traités

Exemple : soit card(W)=2^3=8 mots dont 2^1 ont un sens. Dans ce cas les classes voisines sont données dans le tableau suivant :

000	111
001	110
010	101
100	011

A la réception du mot 100, on décidera que le mot transmis était 000, il y a donc eu une erreur sur le 1er bit.