L'INFORMATION TRANSPORTEE PAR LE SIGNAL

L'information est une notion intuitive mais qu'il est difficile à priori de quantifier. Il est cependant normal d'admettre qu'un message transporte d'autant plus d'information qu'il est difficile à deviner. Il est par exemple plus difficile de deviner le tirage du Loto que le résultat d'un lancé de dés. Ce dernier résultat est donc un message moins porteur d'information que le premier .C'est à partir d'exemples simples de ce type que s'est bâtie peu à peu la théorie.

FORMULE FONDAMENTALE : ENTROPIE D'UNE SOURCE

Nous supposerons d'abord que les messages susceptibles d'être reçus sont en nombre fini n et équiprobables ,c'est le cas des deux exemples précédents.

La quantité d'information q reçue lors de la réception d'un message est d'autant plus grande que le message est difficile à deviner, donc que n est grand.

Mathématiquement q est donc une fonction monotone croissante de n .

Mais d'autre part il est normal de penser que l'information est une grandeur additive , par exemple le résultat d'un tir de carte peut être annoncé en 2 fois, c'est un cœur, puis c'est un valet. Alors La quantité d'information totale est la somme de la quantité d'information associé à la connaissance de la couleur (cœur) plus celle de la carte dans la couleur (un valet) .

Mais il y a seulement n_1=4 couleurs possibles , ces 4 événements sont équiprobables . Donc

Quantité d'information apportée par la connaissance de la couleur = $q(n_1)$

Mais la couleur étant connue il y a n_2=8 cartes possibles d'égale probabilité. Donc :

Quantité d'information apportée par la valeur de la carte = q(n2)

Mais toutes les cartes sortent avec une égale probabilité, or leur nombre est $n=n_1n_2$, la quantité d'information apportée par la connaissance exacte de la carte tirée est donc $q(n_1n_2)$.

La fonction q(n) monotone croissante doit donc satisfaire à la condition :

$$q(n_1n_2)=q(n_1)+q(n_2)$$

Cette équation a une solution évidente q(n)=log(n)

La base des logarithmes peut être quelconque .

Pour des logarithmes en base 2 l'unité a été nommée à l'origine le bit, pour éviter une confusion avec les bits informatifs il faut mieux lui donner son nom officiel le Shannon en l'honneur du chercheur américain qui le premier à bâti cette théorie. Un shannon est ainsi la quantité d'information portée par le résultat d'un jeu de pile ou face .

Pour des logarithmes de base 10 on parle de Décit ou Hartley.

Le Nat est une unité plus théorique pour laquelle le logatithme est naturel (base e).

Information et probabilité :

La définition précédente ne s'applique que pour des événements équiprobables .Que devient elle s'ils ne le sont pas ?

Dans l'expérience précédente divisons le jeu de carte en 2 paquets inégaux. Groupe 1 :les cœur (n_1=8 cartes) et groupe 2 les autres.(n_2=24 cartes) Le résultat du tir est énoncé en deux fois, le numéro du groupe puis la carte dans le groupe . Pour un valet de cœur on énonce Groupe 1 puis carte 5 dans le groupe .Alors :

Q= quantité d'information associé à la connaissance du groupe + q(numéro dans le groupe)

Il y a au total n_1+n_2 possibilités équiprobables et dans chaque groupe les résultats sont également équiprobables :

Soit $q(n_1+n_2)=q(\text{connaissance du groupe , ici 1 }) + q(n_1)$

C'est à dire $q(groupe\,1) = q(n_1+n_2) - q(n_1) = \log(n_1+n_2) - \log n_1 = -\log(\frac{n_1}{n_1+n_2})$

Mais $\frac{n_1}{n_1+n_2}$ est la probabilité de tirer une carte du groupe 1

Les deux événements groupe 1 et groupe 2 ne sont pas équiprobables mais l'information liée à leur connaissance est déterminée par leur probabilité :

C'est l'expression la plus générale $\boxed{q = -\log(p)}$

Système de transmission de l'information Entropie d'une source

Tout système de transmission de l'information est constitué :

D'une source d'information qui crée les messages
Un canal d'information qui les transporte. Un tel canal est souvent perturbé par le bruit.
Un récepteur qui reçoit et identifie les messages.

Les messages transmis par une source sont de probabilités très variables donc d'information liée différentes. La quantité d'information associé à un message ne caractérise aucunement la source . Dans un texte en français la lettre e est très courante , elle ne transporte que peu d'information, d'ailleurs si un e est effacé par une tache le lecteur le reconstitue immédiatement . Un w au contraire est plus rare et sa perte n'est pas évidente à combler .

Ce qui caractérise une source ce n'est pas l'information liée à tet ou tel message mais la moyenne de l(information de tous les messages transmis. C'est l'entropie de la source définie par :

$$\boxed{H = -\sum_i p_i.\log(p_i)}$$

Le nom Entropie vient du fait que l'entropie en thermodynamique statistique est définie par la même formule .

Théorème : L'entropie d'une source est maximale si les messages sont équiprobables .

Ce théorème est évident pour deux messages de probabilités p et 1-p , en effet la fonction
$H = -p.\log(p) - (1-p).\log(1-p)$
est maximale H=1 pour p=1/2

Lors d'un lancé de pièce de monnaie, jeu de pile ou face , l'entropie est maximale 1 shannon si la pièce n'est pas truquée et les deux résultats équiprobables .

Dans le cas d'un texte ,26 lettres sont possibles, si elles étaient de même probabilité la source aurait une entropie $H = -\log_2(26) = -\frac{\log_{10} 26}{\log_{10} 2} = 4{,}7 shannons$

En réalité les lettres n'étant pas équiprobables l'entropie réelle est légèrement inférieure à 4

Cas d'événements liés

Soient deux événements A et B , on définit en théorie des probabilités :

-La probabilité de voir se réaliser les deux événements p(AB)
-Les probabilités de l'un ou l'autre p(A) et p(B)
-La probabilité de voir B se réalisé si A l'est déjà, que l'on note p(A/B) (probabilité de A si B est réalisé)

Pour une source dont les messages sont constitués par l'association d'un événement A et d'un événement B l'entropie sera bien sûr définie par :

$$H(AB) = -\sum_A \sum_B p(AB).\log p(AB)$$

On peut aussi considérer que le message est A lorsque B a été reçu :, il lui est associé une entropie conditionnelle H(A/B)

Pour la calculer on fixe d'abord la valeur de A soit $A=A_i$ Alors la quantité moyenne d'information est :

$$-\sum_B p(B/A=A_i).\log p(B/A=A_i)$$

il faut ensuite moyenner sur les valeurs de A :

$$H(B/A) = -\sum_A p(A)\sum_B p(B/A).\log p(B/A) = -\sum_A\sum_B p(A).p(B/A).\log p(B/A) = \sum_A\sum_B p(AB)\log p(B/A)$$

Il est bien connu que p(AB)=p(A).p(B/A)=p(B).p(A/B)
Il en résulte en prenant les logarithmes que

H(AB)=H(A)+H(B/A)=H(B)+H(A/B)

Il faut enfin noter que si les événements A et B sont indépendants , la connaissance de A n'apporte aucune information sur B , donc :

H(B/A)=H(B)

Sinon A apporte de l'information sur B et :

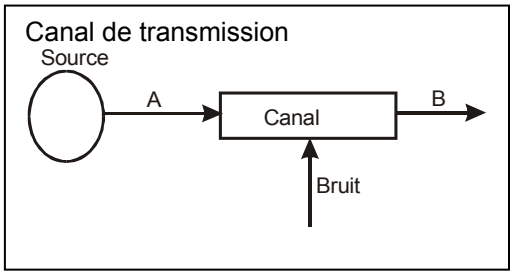
H(B/A) < H(B)

Dans ce cas H(AB)<H(A)+H(B)

INFORMATION TRANSPORTEE PAR UN CANAL.

CAPACITE D'UN CANAL DE TRANSMISSION NUMERIQUE

Définition de la capacité d'un canal



Un système de transmission est constitué d'une source qui applique à l'entrée du canal des messages A .Le message à la sortie est noté B , il peut être différent de A si du bruit perturbe la transmission.

En absence de tout bruit H(B/A) est nul car A étant connu il n'y a plus aucune incertitude sur B et dans ce cas B=A donc H(B)=H(A).

Si au contraire le bruit est si fort que B ne dépend plus de A H(B/A)=H(B) mais l'information transmise est nulle .

Dans une situation intermédiaire la connaissance de A donne de l'information sur B et H(B/A)<H(B) .

La quantité moyenne d'information transmise par le canal est donc :

$I_A(B)=H(B)-H(B/A)$

Mais on a vu que p(A).p(B/A)=p(B).p(A/B) soit en prenant le logarithme :

H(A)+H(B/A)=H(B)+H(A/B)

En intervertissant les termes

$I_A(B)=H(B)-H(B/A)=H(A)-H(A/B)=I_B(A)=I_{AB}$ (1)

Cette expression peut se mettre sous des formes différentes :
En remplaçant H(B/A) par sa valeur

H(B/A)=H(AB)-H(A)

Il vient

I(AB)=H(B)+H(A)-H(AB)

C'est à dire :

$$I(AB) = \overline{-\log p(B)} - \overline{\log p(A)} + \overline{\log p(AB)} = \overline{\overline{-\log pA)} - \log p(B) + \log p(AB)}$$

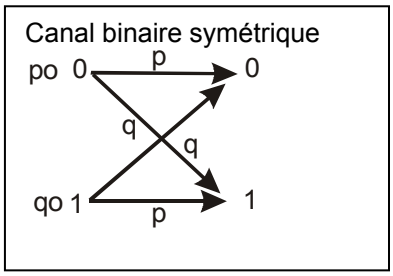
ou encore : $$I(AB) = \overline{\log \frac{p(AB)}{p(A)p(B)}} \qquad (2)$$

En reprenant l'expression initiale sous la forme I(AB)=H(A)-H(A/B)
Le premier terme H(A) est la quantité moyenne d'information liée à la connaissance à priori de A
Le second est la quantité moyenne d'information liée à la connaissance de A lorsque l'on connaît B
Si on définit :

$$I(A_iB_j) = \log \frac{probabilité\ à\ postériori\ de\ A = A_i\ si\ B = B_j}{probabilité\ à\ priori\ de\ A = A_i}$$

alors $$I(AB) = \overline{I(A_iB_j)} = \sum_i \sum_j p(A_i, B_j).I(A_iB_j) \qquad (3)$$

Canal binaire symétrique



Le message d'entrée est constitué de bits 0 ou 1 de probabilités respectives p_0 et q_0 . $p_0+q_0=1$

p et q sont les probabilités de bonne transmission et d'erreur , bien sûr p+q=1

Alors : $$p(B=0/A=0) = p(B=1/A=1) = p$$
$$p(B=1/A=0) = p(B=0/A=1) = q$$

Nous calculerons l'information transmise en utilisant la formule (1) I(AB)=H(B)-H(B/A)
Le second terme est facile à calculer

$$H(B/A) = -\sum_A p(A) \sum_B p(B/A).\log p(B/A)$$

pour a=0 p(A)=p0 et le second terme Σ donne :

$$-p(0/0)\log p(0/0) - p(1/0).\log p(1/0) = -(p.\log p + q\log q)$$

pour a=1 on trouve de la même façon le même résultat. Finalement puisque $p_0+q_0=1$

$$H(B/A) = -(p\log p + q\log q)$$

reste à calculer $H(B) = -\sum_B p(B)\log p(B)$

mais p(B)=p(A).p(B/A) soit :

pour B=0 $$p(B=0) = p(A=0).p(B=0/A=0) + p(A=1).p(B=0/A=1) = p_0p + q_0q$$

de même pour B=1 $p(B=1) = p_0q + q_0p$

Finalement

$$I(AB) = p\log p + q\log q - (p_0p + q_0q)\log(p_0p + q_0q) - (p_0q + q_0p)\log(p_0q + q_0p)$$

Mais cette quantité d'information transmise est maximale si l'information entrée est elle même maximale ce qui implique que les deux signaux d'entrée soient équiprobables $p_0=q_0=1/2$ .La quantité précédente est alors une caractéristique du canal seul, c'est la **capacité du canal**

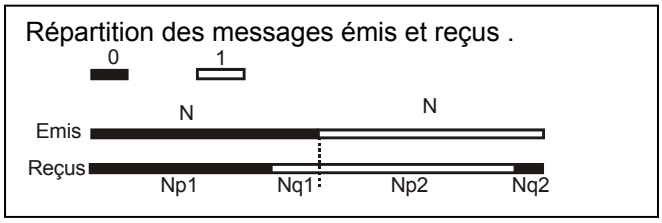
$$C = p\log p + q\log q + 1 = q\log q + (1-q)\log(1-q) + 1$$

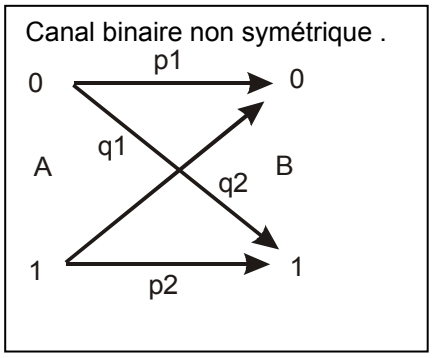
Il est facile de vérifier que pour q=0 la dérivée dC/dq est infinie négative, la capacité d'un canal décroît donc très vite lorsque le taux d'erreur croit .Pour q=1% C ne vaut que 0,92 shannon par bit

Canal binaire non symétrique

Le calcul est un peu plus laborieux et il est plus commode de faire appel à l'expression (3) de I(AB) .

Si 2N bits ont été envoyés (avec N très grand) , c'est à dire N (0) et N (1) (équiprobables à l'entrée puisque l'on calcule la capacité du canal) .Mais parmi les N (0) Np_1 donneront en sortie des (0) et Nq_1 des (1) .De même les N (1) donneront Np_2 (1) et Nq_2 (0) . (figure ci dessous) :





Ce diagramme permet de calculer directement les termes de l'expression (3)

$$p(A=0\ si\ B=0)=\frac{Np_1}{Np_1+Nq_2}=\frac{p_1}{p_1+q_2} \qquad avec \quad p(A=0\ et\ B=0)=\frac{Np_1}{2N}=\frac{p_1}{2}$$

de même pour les 3 autres cas

$$p(A=0\ si\ B=1)=\frac{q_1}{p_2+q_1} \qquad avec \quad p(A=0\ et\ B=1)=\frac{q_1}{2}$$

$$p(A=1\ si\ B=1)=\frac{p_2}{p_2+q_1} \qquad avec \quad p(A=1\ et\ B=1)=\frac{p_2}{2}$$

$$p(A=1\ si\ B=0)=\frac{q_2}{q_2+p_1} \qquad avec \quad p(A=1\ et\ B=0)=\frac{q_2}{2}$$

D'ou l'expression de la capacité puisque p(A=0)=p(A=1)=1/2

$$C=\frac{p_1}{2}\log\frac{\frac{p_1}{p_1+q_2}}{\frac{1}{2}}+\frac{q_1}{2}\log\frac{\frac{q_1}{p_2+q_1}}{\frac{1}{2}}+\frac{p_2}{2}\log\frac{\frac{p_2}{p_2+q_1}}{\frac{1}{2}}+\frac{q_2}{2}\log\frac{\frac{q_2}{q_2+p_1}}{\frac{1}{2}}$$

on pourra vérifier que pour p_1=3/4 q_1=1/4 p_2=9/10 q_2=1/10 C=0,3436 shannon par bit

Codage des messages

Les messages émis par la source sont codés sous forme de mots de plusieurs caractères choisis dans un alphabet de q caractères (q=2 pour un canal binaire)
Un message m_j de probabilité p_j est codé par un mot de n_j caractères
L'entropie de la source est :

$$H=-\sum_i p_i.\log p_i$$

la longueur moyenne des mots est :

$$\bar{n}=\sum_i n_i p_i$$

Mais un caractère choisi dans un alphabet de q peut transporter une quantité d'information log q si tous les caractères sont équiprobables . Pour transporter une information moyenne H il devrait suffire de H/log(q) caractères . C'est la longueur moyenne minimale possible .

Le **rendement de codage** est défini par :

$$\eta=\frac{H}{\bar{n}.\log q}$$

$\rho=1-\eta$ est la **redondance** du code .

Une source doit transmettre 4 messages m_0 m_1 m_2 m_3 de probabilités respectives ½ 1/5 1/5 et 1/10
Pour un canal binaire le plus simple est de coder les messages par leur numéro écrit en binaire ce qui correspond au tableau A suivant :

Tableau A

message	probabilité	Mot code
M0	½	00
M1	1/5	01
M2	1/5	10
M3	1/10	11

L'entropie de la source est

$$H = \frac{1}{2}\log\frac{1}{2} + 2.\frac{1}{5}\log\frac{1}{5} + \frac{1}{10}\log\frac{1}{10} = 1{,}76096 sh$$

tous les mots codes ont même longueur

$\bar{n} = 2$

et l'alphabet comporte 2 caractères q=2

le rendement est donc $\eta = \frac{1{,}76}{2} = 88\%$

Pour l'améliorer on peu imaginer de coder les messages les plus courants avec un mot court , par exemple (tableau B) :

Tableau B

message	probabilité	Mot code
M0	½	0
M1	1/5	01
M2	1/5	110
M3	1/10	111

La longueur moyenne des messages a changé :

$$\bar{n} = 1.\frac{1}{2} + 2.\frac{1}{5} + 3.(\frac{1}{5} + \frac{1}{10}) = 1{,}8$$

l'entropie étant évidemment la même , le rendement devient

$$\eta = \frac{1{,}76}{1{,}8} = 97{,}8\%$$

Le résultat est si bon que l'on est tenté d'aller encore plus loin dans la réduction de longueur des mots. C'est ce qui est fait sur la tableau C .

Tableau C

message	probabilité	Mot code
M0	½	0
M1	1/5	1
M2	1/5	10
M3	1/10	11

La longueur moyenne tombe alors à

$$\bar{n} = 1.(\frac{1}{2} + \frac{1}{5}) + 2.(\frac{1}{5} + \frac{1}{10}) = 1{,}3$$

mais alors le rendement est supérieur à 1 ce qui est impossible. En effet ce code n'est pas déchiffrable, la suite de messages

$m_1m_1m_2m_1m_3m_0m_3m_2$

est codée 111011101110 et peut être lue comme 1 1 1 0 1 1 1 … soit m_1 m_1 m_1 m_0 m_1 ..

les messages sont indiscernables l'un de l'autre. La première idée qui vient à l'esprit est , comme c(est le cas en Morse, d'introduire un blanc entre les mots (tableau D)

Tableau D

message	probabilité	Mot code
M0	½	0-
M1	1/5	1-
M2	1/5	10-
M3	1/10	11-

Mais le blanc est un nouveau caractère et il faut considérer que l'alphabet utilisé comporte 3 caractères q=3 .Alors

La longueur moyenne devient :

$$\bar{n} = 2.(\frac{1}{2} + \frac{1}{5}) + 3.(\frac{1}{5} + \frac{1}{10}) = 2{,}3$$

le nombre minimal de caractères par mot est

$$\overline{n_{\min i}} = \frac{1{,}76}{\log_2 3} = 1{,}11$$

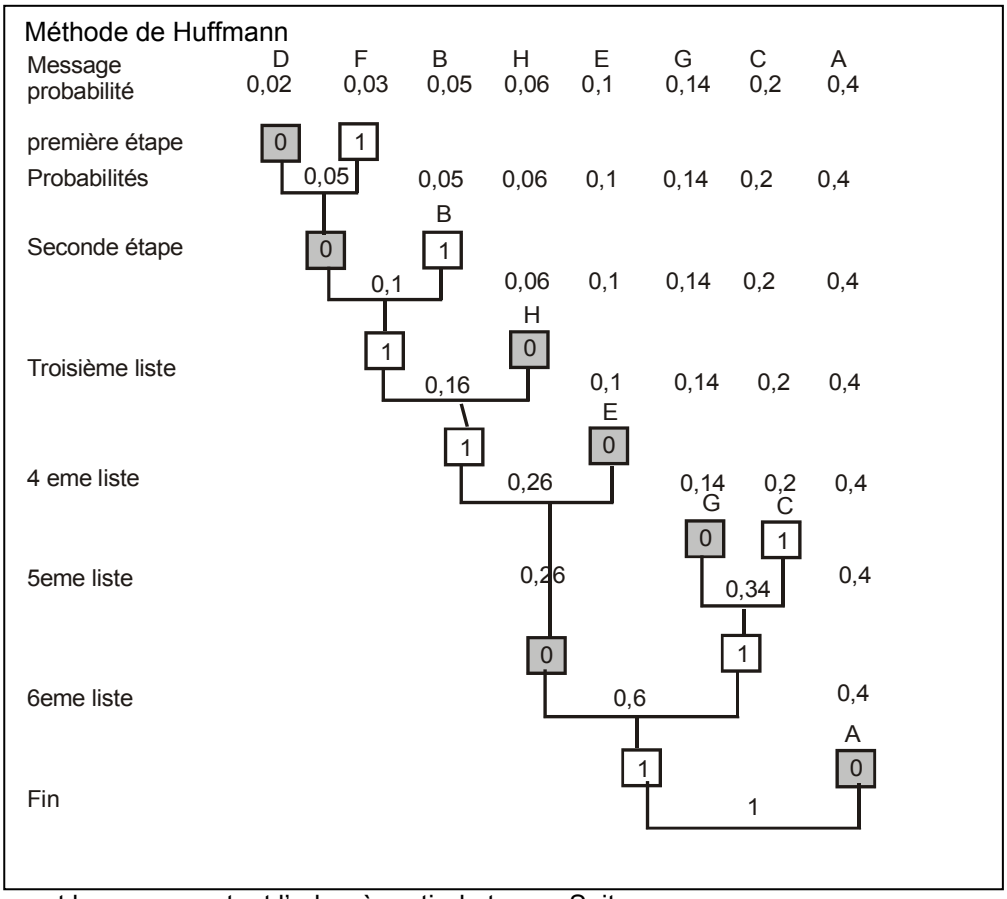
le rendement obtenu $\eta = \frac{1{,}11}{2{,}3} = 0{,}48$ est lamentable !

Il est possible de ne pas utiliser de caractère de séparation si aucun message n'est le début d'un autre , c'est la **condition du préfixe** . Les mots codes qui satisfont à cette condition peuvent être fabriqués grâce à un arbre de codage .

Méthode de Huffmann

Les messages sont rangés par ordre de probabilité décroissantes.

Un O et un 1 est attribué aux deux messages de plus faible probabilité, ils sont alors regroupés et leur probabilités ajoutées . On obtient alors une seconde liste plus courte pour laquelle la même opération est effectuée. Un exemple est représenté ci dessous .



Les codes sont lus en remontant l'arbre à partir du tronc . Soit

Message	Mot code	Message	Mot code
A	0	E	100
B	10011	F	101101
C	111	G	110
D	101100	H	1010

Par construction ces mots obéissent à la condition du préfixe . On pourra vérifier que le rendement de codage est excellent H=2,4437 sh $\bar{n} = 2{,}55 \quad \eta = 95{,}9\%$

Décodage

La source émet des messages codés en mots constitués de caractères binaires (bits)

$$U_J = \left[x_{1j} x_{2j} \ldots\ldots x_{nJ}\right]$$

mais le récepteur identifie des mots qui peuvent être différents , à cause du bruit du canal , soit :

$$V_J = \left[y_{1J} y_{2J} \ldots\ldots y_{NJ}\right]$$

le problème est de déterminer , au reçu d'un mot Vk ,quel est le mot Uj émis qui est le plus probable. C'est un problème d'optimisation de la probabilité :

Prob(Uj émis si Vk reçu)

Un paramètre qui intervient dans le calcul de cette probabilité est la **distance de Hamming**. C'est le nombre de bits différents entre deux mots .

$$d(uv) = \sum_i x_{iJ} \otimes y_{ik}$$

Par exemple entre 01110101 et 01011100 la distance de Hamming est 2 car il y a 2 bits différents , le 3eme et le dernier . La distance de Hamming a les propriétés topologiques d'une distance :

$$\begin{cases} d(u,v) \geq 0 \\ Si \quad d(u,v) = 0 \quad alors \quad u = v \\ d(u,v) \leq d(u,w) + d(w,v) \end{cases}$$

IL existe des codes qui permettent de détecter une erreur ou même de la corriger. Un cours complet doit être réservé à ce sujet, nous nous limiterons à des généralités.

Si l'on choisi des mots code émis qui ont entre eux des distances de Hamming supérieures à 2 et paires , l'introduction d'une erreur est immédiatement détectée car le mot reçu ne figure plus dans la liste des mots émis possibles . Cette propriété est mise à profit dans le système des bits de parité.

Bits de parité

Si les mots à transmettre ont n bits on leur ajoute avant émission un bit supplémentaire appelé bit de parité de façon que le mot de n+1 bits construit contienne un nombre pair de 1 .

Plaçons nous par exemple dans le cas n=5 et supposons que le taux d'erreur soit de 1/1000 $q=10^{-3}$ Sur 10^5 mots émis , environ 500 sont faux .

Ajoutons alors un 6eme bit , le nombre de mots erronés passe à 600 environ , mais la plupart ne contiennent qu'une erreur et sont détectés car ils renferment alors un nombre impair de 1 . Evaluons combien de mots sont dans ce cas .

La probabilité pour que l'erreur soit en 1ere position est q , la probabilité pour que le second bit soit exact est 1-q , de même pour le 3eme etc.. Ainsi la probabilité d'avoir le 1er bit faux et les 5 autres justes est $q(1-q)^5$, mais il y a 6 positions possibles pour l'erreur. La probabilité d'une seule erreur est finalement : $6q(1-q)^5$ soit $5{,}97\ 10^{-3}$. C'est à dire pour 100000 mots envoyés 597 erreurs uniques . Or il y avait au total environ 600 mots faux, seuls 3 d'entre eux ne sont pas détectés.

On peut d'ailleurs calculer la probabilité de double erreur. 2 bits faux et 4 justes mais C_2^6 combinaisons possibles (2 bits parmi 6) soit $6.\frac{6-1}{2}q^2.(1-q)^4 = 1{,}49.10^{-5}$ Pour 100000 mots moins de 2 ne sont pas détectés car ont 2 erreurs .

La probabilité de 3 erreurs est complètement négligeable.

Codes de Hamming

Si la distance entre deux mots du code envoyé est un multiple de 3 ,un mot erroné par 1 bit faux est non seulement détecté mais aussi corrigé, car sa distance de Hamming ne vaut 1 qu'avec un seul mot possible .

Pour d=5n on pourra corriger des erreurs doubles , d=7n des erreurs triples etc…

Les codes les plus connus sont les codes cycliques (Reed Salomon) ou BCH (Bose-Chandhuri - Hocquenheim) Consulter par exemple l'excellent ouvrage ci dessous .

Error Control Coding
SHULIN Edit Prentice Hall 1983

Le premier code correcteur fut décrit par Hamming, nous le citons ici par curiosité.

Soit un mot contenant m bits qui contiennent l'information à transmettre .Pour corriger les erreurs simples nous y ajouterons k bits supplémentaires de contrôle. Le mot a alors une longueur m+k , l'erreur s'il y en a peut prendre m+k+1 positions (la position 0 correspond à une absence d'erreur) Or les k bits de contrôle doivent permettre de déterminer cette position d'erreur. Avec k bits on peut décrire 2^k positions. Pour que la correction soit possible il faut donc que

Nombre de bits de contrôle

k=nombre de bits de controle	Nombre de bits utiles
1	O
2	1
3	4
4	11
5	26
6	57

$m+k+1 <= 2^k$

Pour m fixé c'est une condition sur k

Le tableau ci contre montre que le nombre de bits de contrôle diminue très vite en valeur relative si le mot s'allonge. Cependant pour des mots très longs la probabilité d'erreur multiple (paquets d'erreurs) devient importante. Pour décrire la méthode nous prendrons n=4 k=3 c'est à dire des mots de 7 bits .

Les bits de contrôle peuvent être placés n'importe ou dans le mot , plaçons les d'abord à la fin. Le mot à transmettre est alors de la forme :

[a1a2a3a4a5a6a7] avec a5a6a7=k1k2k3 les bits de contrôle.

A partir de ces 3 bits de contrôle il nous faut fabriquer le nombre qui est la position de l'erreur, qui en binaire s'écrit [e2e1e0]

Le bit e0 doit être égal à 1 si l'erreur se trouve en position 1 3 5 ou 7

e1 vaut 1 si l'erreur est en 2 3 6 ou 7

e2 vaut 1 si l'erreur est en 4 5 6 ou 7 ..

Position de l'erreur	mot [e2e1e0]
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

Si nous définissons les 3 bits e par :

$$\begin{cases} e2 = a4 \otimes a5 \otimes a6 \otimes a7 \\ e1 = a2 \otimes a3 \otimes a6 \otimes a7 \\ e0 = a1 \otimes a3 \otimes a5 \otimes a7 \end{cases}$$

ils seront nuls en absence d'erreur et pour une erreur unique donneront sa position.

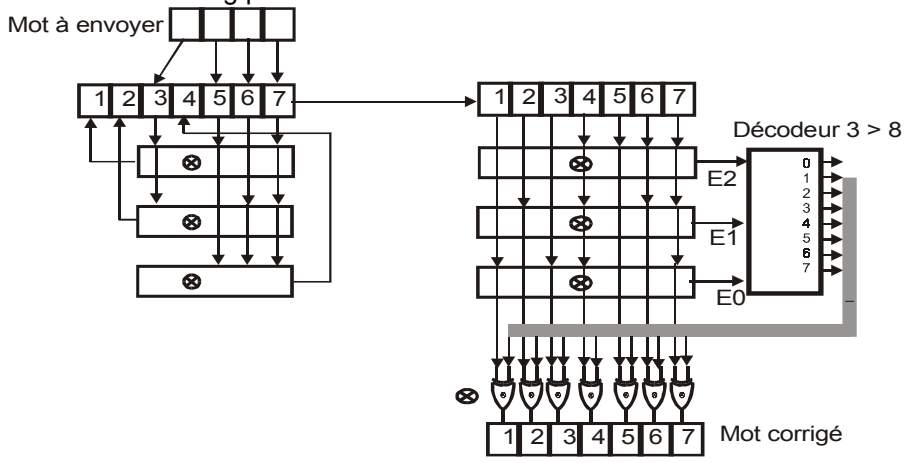
Les bits a1a2a3a4 étant connus ces équations permettrent de calculer les bits de contrôle pour chaque mot .

On notera que dans chaque équation il y a à la fois des bits connus et inconnus ,en plaçant les bits de contrôle en position 1 2 et 4 chaque équation ne contient plus qu'un seul bit de contrôle et la résolution est simplifiée. En effet dans ce cas les bits de contrôle sont donnés par :

$$\begin{cases} a1 = a3 \otimes a5 \otimes a7 \\ a2 = a3 \otimes a6 \otimes a7 \\ a4 = a5 \otimes a6 \otimes a7 \end{cases}$$

Il existe des circuits intégrés qui génèrent et décodent automatiquement de tels codes

Codeur décodeur de Hamming pour mots de 4+3 bits



INFORMATION DANS LES SYSTEMES CONTINUS

Nous avons vu qu'un signal continu , sous réserve qu'il ait un spectre limité (fréquence de coupure f_c) , peut être entièrement décrit par un nombre fini $N=2f_c$ d'échantillons chaque seconde. .Si Q est la quantité d'information transportée par l'un de ces échantillons, la quantité d'information transmise chaque seconde est donc $2f_cQ$. Nous pouvons donc nous limiter à définir l'information transmise par un échantillon analogique .

Pour être quantifié l'échantillon doit être compris entre deux limites d'amplitude définies que nous noterons ±A .Lorsque la quantification avec un pas q=Δv est effectuée l'échantillon est transformé en un nombre parmi 2A/q possibles, et l'on peut appliquer les résultats précédents.

L'entropie de la source , c'est à dire la quantité moyenne d'information attachée à la transmission d'un échantillon numérique est :

$$-\sum_J p(v_J)\Delta v Log\big(p(v_J)\Delta v\big)$$

en effet $v=n_j$ (nombre correspondant à v voisin de v_j) si $v_j-q/2<v<v_j+q/2$ et la probabilité correspondante est , au premier ordre $p(v_j)\Delta v$

Pour obtenir l'entropie de l'échantillon analogique il suffit de faire tendre le pas de quantification vers zéro.

$$H = Lim_{\Delta v\to 0}\left[-\sum_J p(v_J)\Delta v Log\big(p(v_J)\Delta v\big)\right]$$

Ce passage à la limite pose cependant quelques problèmes en effet l'expression précédente se développe en :

$$H = Lim_{\Delta v\to 0}\left[-\sum_j p(v_J)\Delta v_J Log\big[p(v_J)\big]-\sum_j p(v_J)\Delta v_J Log\big[\Delta v\big]\right]$$

le premier terme tend naturellement vers l'intégrale : $\int p(v)\log p(v)dv$

mais le second contient $\log(\Delta v)$ qui tend vers $-\infty$ lorsque $\Delta v\to 0$.Il n'est pas possible de lever totalement cette difficulté car ce terme , suivant l'expression de p(x), peut converger ou non. On renonce donc à une définition absolue de l'entropie pour une définition relative

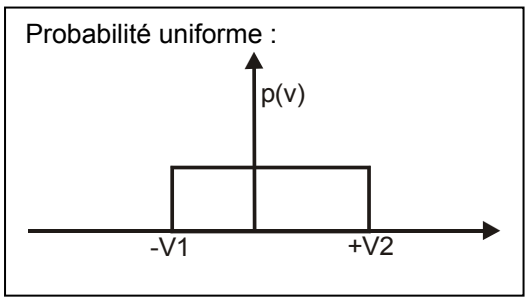
$$H = --\int_{-\infty}^{+\infty} p(x).\log p(x)dx$$

Exemples fondamentaux

1° Densité de probabilité uniforme

La probabilité de trouver l'amplitude v dans un intervalle Δv est indépendante de v et proportionnelle à Δv p(v).Δv = K Δv

L'amplitude étant comprise dans l'intervalle $-v_1, v_2$ p=1 si $\Delta v=v_1+v_2$



Donc $K=v_1+v_2$

Et

$$p(v) = \frac{1}{v_1+v_2}$$

donc l'entropie :

$$H = -\int_{-V_2}^{V_2} \frac{1}{v_1+v_2}\log\frac{1}{v_1+v_2}dv = \log(v_1+v_2)$$

2° Signal de puissance imposée

La puissance du signal est fixée et nous allons chercher pour quelle densité de probabilité l'entropie est elle maximale .

Nous avons :

$$\begin{cases} P = \overline{x^2} = \sigma^2 = \int_{-\infty}^{\infty} x^2 . p(x) dx \\ I_1 = \int_{-\infty}^{\infty} p(x) dx = 1 \\ I_2 = H = -\int_{-\infty}^{\infty} p(x) . \log p(x) dx \end{cases}$$

Considérons $I = I_2 + \lambda I_1 + \mu P = H + Cte$

Le maximum de I donne aussi celui de H .

$$I = \int p(\_ \log p + \lambda + \mu x^2) dx$$

En dérivant par rapport à p qui est l'inconnue :

$$\frac{dI}{dp} = \int (\mu x^2 + \lambda - \log p - 1) dv$$

dérivée qui est nulle pour $p = e^{\lambda - 1} . e^{\mu x^2}$

Les coefficients λ et μ sont calculés en reportant cette forme de p dans les équations précédentes :

$$\begin{cases} \int_{-\infty}^{\infty} e^{\lambda - 1} . e^{\mu x^2} dx = 1 \\ \int_{-\infty}^{\infty} x^2 e^{\lambda - 1} . e^{\mu x^2} dx = \sigma^2 \end{cases}$$

mais

$$\begin{cases} \int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi} \\ \int_{-\infty}^{\infty} y^2 . e^{-y^2} dy = \frac{\sqrt{\pi}}{2} \end{cases}$$

En identifiant les termes il vient $\mu = \frac{-1}{2\sigma^2} \quad e^{\lambda - 1} = \frac{1}{\sigma\sqrt{2\pi}}$

Soit la densité de probabilité cherchée, c'est une gaussienne.

$$\boxed{p(x) = \frac{1}{\sigma\sqrt{2\pi}} . \exp(-\frac{x^2}{2\sigma^2})}$$

Ce résultat est fondamental , **pour une puissance déterminée le signal gaussien est celui qui contient le maximum d'information .**

Calculons son entropie.

$$H = -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} . \log \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} dx$$

en prenant bien sûr des logarithmes naturels :

$$H = -\left[\frac{1}{\sigma\sqrt{2\pi}} . \log \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{x^2}{2\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \right]$$

mais

$$\int_{-\infty}^{\infty} p(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = 1 \quad et \quad \int_{-\infty}^{\infty} x^2 . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma^2$$

d'ou finalement l'expression très simple :

$$H = Log\left(\sigma\sqrt{2\pi}\right) + \frac{1}{2}$$

mais $$\frac{1}{2} = Log\sqrt{e}$$

L'entropie se met enfin sous la forme :

$$\boxed{H_{nats} = \log\left[\sigma\sqrt{2\pi e}\right]}$$

ou encore : $H_{nats} = Log\sqrt{2\pi eP}$ P étant la puissance du signal.

Pour obtenir le résultat en shannons il suffit de prendre les logarithmes à base 2 , alors

$$\sqrt{2\pi eP} = 2^{H_{sh}} \quad soit \quad P = \frac{2^{2H}}{2\pi e}$$

Pour un processus non gaussien l'expression précédente de la puissance est appelée puissance d'entropie.

Théorème de Shannon relatif à la capacité d'un canal continu perturbé par un bruit

Premier théorème de Shannon

Le signal y reçu à la sortie d'un canal peut , à cause du bruit, être différent du signal x introduit à l'entrée. Nous poserons y=x + b , b étant le bruit introduit par le canal.

Nous avons montré plus haut que la quantité d'information transmise par un canal avait pour expression :

$$I(x,y) = H(y) - H(y/x)$$

Mais y= x + b , lorsque x est connu toute l'information sur y est une information sur b ,soit

$$I(x,y) = H(y) - H(b/x)$$

mais le bruit est indépendant du signal d'entrée x donc $H\left(b/x\right) = H(b)$

et finalement : $I(x,y) = H(y) - H(b)$

Formule de Shannon

La quantité d'information transitant par un canal est maximale lorsque la quantité introduite à l'entrée est elle même maximale, pour une puissance donnée le signal x doit donc être gaussien.

Dans ce cas $H(x) = \log\sqrt{2\pi eP}$

Mais le signal est perturbé au maximum lorsque le bruit est lui même gaussien , si B est sa puissance :

$$H(b) = \log\sqrt{2\pi eB}$$

mais la somme de deux gaussiennes est une gaussienne dont la puissance est la somme des puissances. Soit

$$H(y) = \log\sqrt{2\pi e(P+B)}$$

En reportant dans l'expression établie dans le paragraphe précédent :

$$I(x,y) = \log\sqrt{2\pi e(P+B)} - \log\sqrt{2\pi eB} = \log\sqrt{1+\frac{P}{B}}$$

C'est la quantité moyenne d'information pour un échantillon , or le signal est décrit par 2F échantillons chaque seconde , la capacité du canal est donc

$$C = F.\log\left(1+\frac{P}{B}\right)$$

c'est la formule de Shannon .

Conséquences de la formule de Shannon

Nous désignerons par γ le rapport signal sur bruit P/B

1° Relation entre bande passante et durée de transmission

Pendant une durée T la quantité maximale d'information transmise est $Q = FT\log(1+\gamma)$

En conservant le même rapport signal bruit cette information peut être transmise en des temps différents T_1 ou T_2 si la bande passante du canal obéit à la relation :

$$F_1T_1 = F_2T_2$$

Durée de transmission et bande passante sont inversement proportionnelles l'une de l'autre. Ce résultat est bien connu, si l'on veut copier rapidement une bande magnétique dont la vitesse normale de lecture est 4,5 cm/s pour une bande passante de 5kHz, on peut multiplier la vitesse par 10 (45cm/sec) mais l'amplificateur doit avoir une bande passante 10 fois plus large (50kHz)

2° Relation entre durée de transmission et rapport signal sur bruit

La bande passante du canal est fixée. La quantité d'information transmise est proportionnelle au produit T par log(1+γ) on peut donc compenser un faible rapport signal bruit par de la durée de transmission .

Pour Q et F constantes :

$$T_1\log(1+\gamma_1) = T_2\log(1+\gamma_2)$$

pour γ grand , ce qui est le cas de bons canaux ,

$$\frac{T_1}{T_2} \approx \frac{\log\gamma_2}{\log\gamma_1}$$

La durée de transmission est inversement proportionnelle au logarithme du rapport signal bruit.

Par exemple si γ_1=50 ,pour une durée doublée $T_2=2T_1$, $\log\gamma_2=(1/2)\log\gamma_1$ d'ou γ_2=7

Pour transmettre la même quantité d'information un rapport signal bruit de 7 seulement suffit si la durée de transmission est doublée.

3° influence de la bande passante du canal

La puissance de bruit est le plus souvent proportionnelle à la bande passante :

B=AF , c'est par exemple le cas du bruit thermique B=KθF , θ étant la température

Alors :

$$C = F.\log(1+\frac{P}{kF})$$

La capacité augmente avec F mais il existe cependant une limite

$$C_{\lim ite} = \frac{P}{A} \quad ou\ en\ shannons \quad C_{\max Shnaanos} = 1{,}44\frac{P}{A}$$

Pour un bruit thermique :

$$C_{\max} = 1{,}44\frac{P}{K\theta}$$

Pour C=1 shannon par seconde $P \approx 10^{-23}.\vartheta$

A 300°K il faut pour transmettre 1 bit par seconde une puissance minimale de 3.10^{-21} watt

Conséquences :

Soit un système de transmission pour lequel on veut conserver le débit C mais on modifie bande passante et rapport S/B. Le bruit est d'origine thermique .

$$F_1.\log\left(1+\frac{P_1}{k\theta F_1}\right)=F_2\log\left(1+\frac{P_2}{k\theta F_2}\right)$$

en posant $\frac{P_1}{k\theta F_1}=\gamma_1$ cette expression peut se mettre sous la forme :

$$\frac{P_2}{P_1}=\frac{1}{\gamma_1}\frac{F_2}{F_1}\left[\left(1+\gamma_1\right)^{F2/F1}-1\right]$$

Pour un canal de bonne qualité , γ_1=1000 (30dB) augmentons, en modifiant le codage ,la largeur de bande de 50%, soit F2/F1=1,5.L'expression ci dessus donne alors

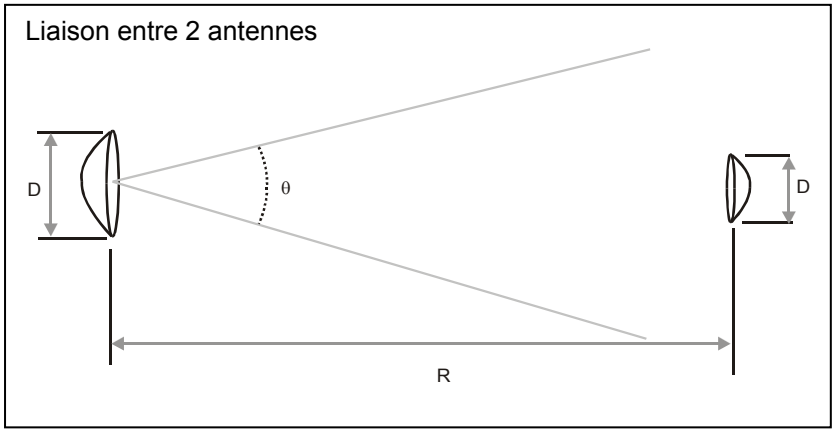
$$\frac{P_2}{P_1}=0{,}148$$

La puissance nécessaire peut être réduite dans un rapport 7 .Ce résultat est exploité dans la technique d'étalement de spectre bien connue en télécommunications.

Exercice: Portée d'une liaison spatiale.

La liaison doit être assuré entre deux antennes paraboliques ,l'une de grand diamètre est placée au sol ,l'autre sur un vaisseau spatial.

On rappelle qu'une antenne de diamètre D à un lobe d'ouverture angulaire λ/D ou λ est la longueur d'onde de la porteuse utilisée pour transmettre le message .



A une distance R l'antenne émettrice arrose une surface circulaire de diamètre Rθ, donc de surface

$$\pi\frac{R^2\theta^2}{4}$$

c'est à dire :

$$\pi\frac{R^2\lambda^2}{4d^2}$$

La puissance P de l'émetteur est répartie dans cette surface. Mais l'antenne réceptrice n'en capte qu'une très faible partie , environ dans le rapport des surfaces soit

$$P_2=P.\frac{\frac{\pi d^2}{4}}{\pi\frac{R^2\lambda^2}{4D^2}}=\left(\frac{Dd}{R\lambda}\right)^2$$

La portée maximale pour un débit de 1 bit/sec est obtenue lorsque la puissance reçue est égale à la limite calculée plus haut P_l=3 10^{-21}W, soit

$$R_{\lim}=\frac{dD\sqrt{P}}{\lambda\sqrt{P_{\lim}}}$$

Pour P=1MW (10^6W) D=100m d=10m λ=3cm (bande X) P_{lim}=3 10^{-21} il vient R=6.10^{17} mètres

C'est à dire environ 60 Années lumière , résultat bien sûr très optimiste et difficile à vérifier pour le moment !