Théorie de l'information

PAEL5 ~ 2 ème semestre 2000 - 2001
Enseignant : Laurent Koechlin

Ce poly sera disponible sur : http://webast.ast.obs-mip.fr/people/koechlin/

1. Information et transmission

2. Mesure quantitative de l'information

3.Vitesse de transmission d'un message

4. Information mutuelle, écart entropique

5. Transmission de l'information

6. Codage avec contrôle d'erreur

7. Imageurs optiques et information

1. Information et transmission de l'information.

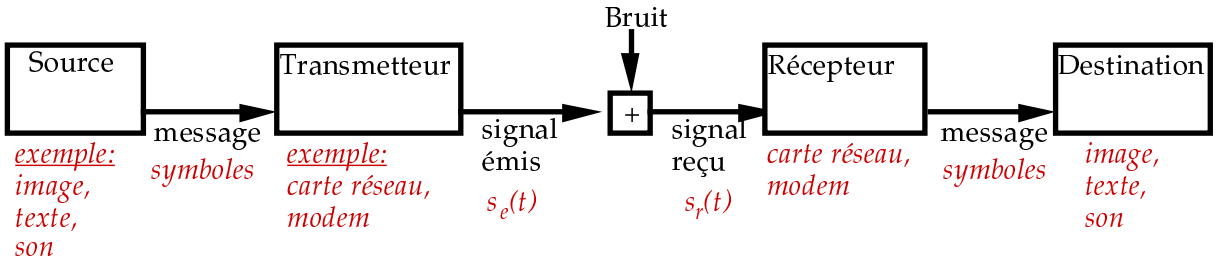

Diagramme adapté du livre de Shannon et Weaver "The mathematical theory of communication, 1939"

le signal émis s(t) est une fonction du temps, même s'il code une fonction de l'espace comme une image, ou de l'espace et du temps comme un film. On utilise alors le multiplexage : chaque pixel de l'image est transmis séquentiellement.

Les pixels sont les éléments d'image (picture element -> par contraction pixel). Une image codée en pixels est une matrice de valeurs numériques qui représentent chacune une petite zone de l'image, supposée de teinte uniforme. On associe au codage de l'image une table des couleurs qui à chaque teinte fait correspondre une valeur numérique.

1.1. Concept d'information.

Le concept d'information ne peut être dissocié de celui de transmission de l'information. l'information ne se "révèle" en tant que telle que lorsqu'elle est transmise. Par exemple, si vous avez un livre et que vous ne savez pas lire, ce livre ne contient pratiquement aucune information pour vous, de même un fichier ".doc" récent a une valeur informative quasi-nulle quand on cherche sans succès à l'ouvrir avec un "word" moins récent [1] .

Pour définir quantitativement l'information il faut s'intéresser à la transmission de celle-ci.

Weaver (1939) a décrit trois niveaux de qualification de la communication :

A - avec quelle *précision* (et quelle vitesse) les symboles sont transmis.
B - avec quelle *fidélité* ces symboles traduisent le message.
C - avec quelle *efficacité* le message transmis affecte le comportement du récepteur dans le sens voulu.

Il pourrait paraître prétentieux de dire que le seul but d'un message est de modifier le comportement de celui ou celle qui le reçoit, mais si le message transmis n'a aucun effet discernable (sans contrainte de délai), il n'y a aucun moyen de savoir s'il a été reçu.

On ne va s'intéresser dans la suite qu'au niveau A, les deux autres niveaux sortent du domaine de la physique.

2. Mesure quantitative de l'information

[1] ce n'est pas le cas avec tous les traitements de texte, heureusement il y en a de plus stables dans le temps.

2.1 Approche par l'informatique

Combien faut-il d'information pour décrire une image ? On peut associer intuitivement le nombre de pixels d'une image au nombre de "degrés de liberté" de cette image ou bien à la quantité d'information qu'elle contient. Il faut aussi tenir compte du nombre de couleurs possibles par pixel. Par exemple, pour stocker une image de 1000 par 1000 pixels codée sur 24 bits [(2)] par pixel (16 millions de couleurs), il faut en tout $24\ 10^6$ bits.

Exercice 2.1 : *Hors du temps*

1) Combien d'images différentes peut-il exister, codées en 1000 x 1000 pixels sur 24 bits ?

2) En supposant que l'on cherche à les explorer une à une avec un ordinateur hyper-rapide capable d'an analyser 10^{12} par seconde, combien de temps cela prendrait-il ?

3) Même question, toujours avec cet ordinateur hypothétique, mais en prenant simplement une petite image en noir ou blanc (1 bit par pixel) de 10 x 10 pixels. Comparer la durée de calcul avec l'âge de l'Univers (15 milliards d'années).

Le nombre d'états possibles que peut prendre une image (ou tout autre message) codé sur n bits est $N = 2^n$. En assimilant la quantité d'information H contenue dans l'image à l'occupation mémoire de celle-ci en bits on aurait :

$H = \text{lb}\, N$ (lb = log à base 2)

Cette mesure à l'avantage d'être additive, mais une expression du type :

$H = k \ln N$, avec k: constante positive arbitraire, est tout aussi valable.

<u>rappel</u> : $\log_a(x) = \ln(x) / \ln(a)$

2.1 Relation entre Information et Entropie.

L'expression $\mathbf{H = lb\ N}$ est identique (à une constante K près) à l'expression en thermodynamique de l'entropie d'un système ayant Ω états :

$S = k \ln \Omega$

Dans le cas de la thermo, k est la constante de Boltzmann :

$k_b = 1{,}380 \times 10^{-23}\ \text{J K}^{-1}$

Dans le cas de l'info, on utilise le log à base 2 et une constante k = 1.

Que ces deux expressions soient les mêmes n'est pas un hasard. Plus un système a d'états possibles, plus il faut d'information pour définir l'un de ses états. L'entropie d'un système correspond à la quantité d'information qu'on acquiert quand on passe de: "on ne sait rien sur ce système" à "on a la certitude que ce système est dans un état donné".

Dire qu'une image "contient H bits d' information" revient à dire : "l'information apportée par le fait qu'on a vu cette image particulière fait passer l'entropie du système de

---

[2] un bit est un chiffre en base 2. Il prend deux valeurs possibles : 0 ou 1.

H= lb(N) à H=0". Il n'y a plus qu'un seul état possible une fois qu'on a vu l'image.

Exercice 2.2 : *Mémoire à ADN*

Les généticiens parlent en "nombre de "bases" ADN. Sachant qu'une base ADN peut prendre 4 valeurs différentes (A,G,C,U), si l'on prend comme mesure de l'information contenue dans un code génétique le nombre de bases ADN, quelle est la constante k à laquelle cela correspond dans la formule :

$G = k \, lb\,(N)$?

où : G est la mesure d'information génétique en nombre de bases ADN,
N est le nombre d'états possibles d'un système à G bases ADN.

Exercice 2.3 : *Entropie thermodynamique et entropie de Shannon.*

En thermo, l'entropie s'exprime par $S = K_b \ln \Omega$,
En théorie de l'info, l'entropie par $H = lb\,\Omega$.
Quel est la valeur d'une unité d'entropie "thermo" exprimée en bits ?
On donne : constante de Boltzmann : $K_b = 1{,}38 \;\; 10^{-23}\ JK^{-1}$

2.3 Entropie d'un système à N états non équiprobables.

En supposant que l'on ait une probabilité $P = 1/N$ que l'image vue soit l'une particulière des N possibles, la mesure de l'information s'exprime :

$H = k \ln (N) = k \ln (1/P) = -K \ln(P)$

On a supposé jusqu'ici que les états étaient équiprobables, comme en thermo, mais c'est rarement le cas en théorie de l'information. Certaines images sont plus probables que d'autres dans des situations données. De même, parmi N séquences possibles de 10 symboles consécutifs tirées de la littérature française ($N \approx 27^{10}$ en prenant 27 symboles différents pour représenter les lettres et l'espace entre les mots), la séquence "o iuywrsdz" est beaucoup moins probable que "la moyenne". Parmi l'ensemble des 27 séquences possibles de une lettre ou espace, la fréquence de "e" est plus élevée que celle de "w".

C'est ici que Claude E. Shannon a fait une conjecture fondamentale : Si les N états possibles d'un système ont des probabilités différentes P_e d'être observées, telles que $\Sigma P_e = 1$, l'information nécessaire pour décrire un état de probabilité P_e vaut :

$I_e = -k \ln P_e$ avec k , constante positive arbitraire.

L'information <u>nécessaire pour décrire</u> un événement (état du système) est donc d'autant plus élevée que cet événement est rare. Elle est égale à celle <u>apportée par l'observation</u> de cet état.

Par exemple, les journalistes ne s'intéressent pas aux trains qui arrivent à l'heure. On illustre ce fait par l'expérience suivante :

Soit un ensemble de 256 boules, dont 255 blanches et une noire. Tirez une boule au hasard, il y a deux cas possibles : soit vous tirez une boule noire, soit vous en tirez une blanche. Le système a donc deux états (événements possibles), leurs probas sont respectivement 1/256 et 255/256. L'événement "tirage d'une boule noire" définit totalement l'état ultérieur du système car il ne restera que des boules blanches. Le tirage d'une boule

blanche, plus probable, apportera beaucoup moins d'information.

Propriétés de I_e :

Positivité: $I_e \geq 0$

$I_e = 0$ si l'événement est certain.

Additivité : Des événements indépendants apportent des informations qui s'ajoutent :

$I_{1 \text{ et } 2} = I_1 + I_2$

Exercice 2.4 : *Additivité de l'information apportée par 2 événements indépendants*

Montrer que si deux événements x_1 et x_2 de probabilités respectives $P(x_1)$ et $P(x_2)$ sont indépendants, l'information apportée par l'observation de chacun d'eux s'ajoute.

Généralisation : Soit un état observé avec la probabilité P_e , l'information h_e qu'apporte en moyenne son observation est modulée par sa fréquence d'apparition.

$h_e = -k\, P_e \ln P_e$

On voit que h_e vaut 0 quand $p_e = 0$ et aussi quand $p_e = 1$.

L'information H nécessaire pour décrire complètement un système de N états est obtenue en sommant les contributions individuelles h_e de chacun les états :

$$\mathbf{H = -k \sum P_e \ln P_e}$$

Si l'on choisit $k=1/\ln(2)$, on a $\mathbf{H = - \sum P_e \, lb \, P_e}$

Par exemple, calculons la valeur de H pour deux systèmes :

- un système A à 2 états équiprobables : $P_{a1} = P_{a2} = 1/2$,
- un système B à 2 états pour lequel $P_{b1} = 255/256$ et $P_{b2} = 1/256$.

$H_a = 1/2 + 1/2 = 1$

$H_b = - 255/256 \text{ lb } (255/256) - 1/256 \text{ lb } (1/256) = 0.00565 + 0.03125 = 0.0369$

L'état 2 du système B contient 8 fois plus d'information que l'un des états du système A, mais il apporte 32 fois moins à d'information en moyenne car il est très rarement observé. Au final, l'information contenue dans, (ou bien: nécessaire pour définir) B est 27 fois plus faible que celle de A.

Le système A pourrait correspondre à un pixel soit blanc soit noir dans une image totalement aléatoire en noir ou blanc, et le système B à un pixel soit blanc soit noir dans cette page de texte : la surface couverte par l'encre noire étant beaucoup plus petite que celle restée blanche, la probabilité qu'un pixel soit noir est faible.

Pour des états équiprobables ($P_e = 1/N$), on retrouve bien la première expression :

$H = -k \sum (1/N) \log (1/N) = k \log N$

En préliminaire au chapitre suivant, on peut voir qu'il y a besoin de moins d'information pour décrire une page de texte que pour décrire l'image correspondante. Cela rend possible la compression de l'information dans les transmissions par fax.

La transmission par fax est loin d'être optimale: un texte codé sous forme de suite de caractères (un par octet) est encore plus compact que le fax correspondant. On peut aller toujours plus loin dans la compression sans perte d'information : en tirant parti des probabilités relatives des différents caractères. C'est ce qu'utilisent les utilitaires du type "zip" ou "Stuffit".

3. Codage d'un message et vitesse de transmission

3,1 Codage optimal

Soit un message codé sur 4 lettres (les symboles A,B,C,D) dont les probabilités respectives sont 1/2, 1/4, 1/8, 1/8.

Un tel message pourrait être :

BADBAACABADACBA.....

La "source" du message est un système à 4 états possibles (les 4 symboles) et son entropie H vaut :

$H = - \sum P_e \, lb \, P_e = 1/2 + 1/2 + 3/8 + 3/8 = 7/4.$

Pour transmettre des messages de cette source efficacement, il faut réduire au minimum possible le nombre de bits par symbole. Dans le tableau suivant on voit deux exemples de codage, l'un à nombre fixe de bits par symbole (deux), l'autre à nombre variable de bits par symbole (de un à trois)

symbole	**A**	**B**	**C**	**D**	
P_e	1/2	1/4	1/8	1/8	
- lb P_e	1	2	3	3	
h = - P_e lb P_e	1/2	1/2	3/8	3/8	H = 7/4
codage fixe	00	01	10	11	
nb de bits	2	2	2	2	<bits/symb> = 2
codage optimal	0	10	110	111	
nb de bits	1	2	3	3	<bits/symb> = 7/4

Avec le codage de longueur variable illustré dans les deux dernières lignes du tableau, on a pris pour chaque symbole un nombre de bits égal à $- lb \, P_e$.

Ceci a permis d'égaliser le nombre moyen de bits par symbole à l'entropie de la source. Shannon a démontré que l'on ne peut pas faire mieux : le nombre moyen de bits par symbole est toujours égal ou supérieur à l'entropie de la source.

Exercice 3.1 : *Entropie d'une source à deux états de probabilités inégales.*

Soit une source à deux états possibles : "1" et "2", dont l'état "1" a un proba P1.

Calculer son entropie en fonction de P1. étudier et tracer la courbe.

3.2 Débit d'information d'une source.

Soit une source d'entropie H qui génère (ou dont on lit) un nombre r de symboles par seconde. le débit d'information de cette source sera :

$R = r H$

R est en bits s^{-1} ; r est en symboles s^{-1} ; H est en bits par symbole.

Exercice 3.2 : *Débit d'information d'un fax.*

On considère une feuille de papier dont on va transmettre en noir ou blanc une zone de 20x29 cm^2, avec des pixels de 0,2x 0,2 mm^2.

1) A combien de bits cela correspond-il si l'on n'a aucune information particulière sur la proportion de noir et de blanc dans la page ?

2) On veut transmettre cette page en 10 s, et on dispose d'une ligne à 28,8 k bits s^{-1} . Est-ce possible ?

3) En fait, on sait que cette page contiendra du texte écrit en noir sur fond blanc, et que le noir ne couvrira que 1% de la surface, contre 99% pour le blanc.

a) Quelle est la quantité d'information dans un pixel ?
b) et dans la page ?

4) on veut transmettre cette page en 10 s, toujours avec la ligne à 28,8 k bits s^{-1} .
Est-ce possible ?

Exercice 3.3 : *compression vidéo.*

1) calculez l'entropie d'une image codée sur 256 niveaux, de 400 x 500 pixels sans optimisation du codage.

2) trouver un codage qui tire parti du fait que la valeur de deux pixels consécutifs a une probabilité beaucoup plus forte d'être 0, 1 ou -1 que des valeurs supérieures.

Loi supposée de proba pour la variation de brillance d'un pixel au suivant : P(0) = 1/2 ; P(1) = 1/4 ; P(-1) = 1/4 ; probabilité des autres valeurs négligeable.
Calculez l'entropie.

3) Peut-on améliorer encore le codage en tenant compte de la faible variation d'un pixel

donné d'une image à la suivante ?

Exercice 3.4 : *l'esprit frappeur...*

Lors d'une séance de spiritisme, une voyante met en relation ses clients avec les esprits en utilisant le code suivant: un coup pour A, deux coups pour B, trois coups pour C, etc... jusqu'à Z, avec une seconde de silence pour séparer les lettres, soit:
A -> 10
B -> 110
C -> 1110
.....
Z -> 11111111111111111111111111110
L'esprit frappe un coup par seconde et la voyante facture ses consultations à la durée.

1) Quel est le nombre moyen de bits par symbole avec le code utilisé ?
2) En prenant une probabilité uniforme pour toutes les lettres de l'alphabet dans les messages transmis, quelle est l'entropie de la source ?
en fait l'esprit s'appelle *Zyquywkyxyzz*, mais on n'en tiendra pas compte.
3) Trouvez un système de codage meilleur. Donnez un exemple et expliquez pourquoi il serait plus efficace.
4) (*question carrément hors sujet*) Comment seriez-vous reçu par la voyante si vous lui proposiez d'adopter ce nouveau code ?

---

4. Information mutuelle, écart entropique.

4.1 Différence d'entropie entre deux sources.

On va s'intéresser à la différence d'entropie entre deux sources de même nombre d'états, mais de loi de probabilités différentes, et voir que l'entropie d'une source est maximale quand ses états sont équiprobables.

Soit H_m l'entropie d'une source, calculée en supposant (par manque d'information) que ses N états sont équiprobables, donc chacun de proba $Q_i = Q$ (constant) $=1/N$.

$H_m = -\Sigma\, Q_i \text{ lb } Q_i$, avec $Q_i = Q \;\forall\, i$, $\Sigma\, Q_i = 1$, ($H_m = \text{lb } N$)

Soit maintenant H l'entropie de cette même source, calculée en connaissant les probabilités individuelles de ses différents états :

$H = -\Sigma\, P_i \text{ lb } P_i$, P_i variable selon i, avec $\Sigma\, P_i = 1$.

La différence entre les deux entropies vaut :

$H_m - H = -\Sigma\, Q_i \text{ lb } Q_i + \Sigma\, P_i \text{ lb } P_i$

Comme les Q_i sont tous égaux : $H_m - H = -\text{lb } Q\, \Sigma\, Q_i + \Sigma\, P_i \text{ lb } P_i$

Comme $1 = \Sigma\, Q_i = \Sigma\, P_i$: $H_m - H = -\text{lb } Q\, \Sigma\, P_i + \Sigma\, P_i \text{ lb } P_i$

on peut factoriser $H_m - H$ en : $\mathbf{H_m - H = \Sigma\, P_i \text{ lb } (P_i / Q_i)}$

4.2 Distance entropique de Kullback - Lieber.

Soit toujours une source à N états de probabilités P_i , on appelle distance de **Kullback - Lieber** la valeur donnée par l'expression :

$K = \Sigma P_i \, lb \, (P_i / Q_i)$

Avec l'ensemble des Q_i formant une loi arbitraire de probabilité des états, différente de celle des P_i . C'est une généralisation de l'expression précédente : les Q_i peuvent maintenant être différents entre eux.

K représente la quantité d'information apportée sur une source par la connaissance exacte de sa répartition de proba P_i , par rapport à l'estimation de cette source faite en supposant une loi de proba erronée Q_i . K est une distance au sens mathématique du terme, car elle est toujours positive (ou nulle), et "relie" deux points de coordonnées formées respectivement par les Q_i et les P_i dans un espace à N dimensions.

Exercice 4.1 : *Démontrer que la distance de Kullback-Lieber est une valeur positive* $\forall Q_i$ et $\forall P_i$.
Indice: Démontrer d'abord que $\Sigma P_i \, lb \, Q_i \leq \Sigma P_i \, lb \, P_i \quad \forall Q_i$ et P_i
en définissant la différence des probas: $U_i \equiv Q_i - P_i$ dans l'expression de $lb \, Q_i$

---

5. Canaux de transmission

Rappel : la <u>probabilité conditionnelle</u> de Bayes :

$P(x \mid y) = P(x, y) / p(y)$

Proba de réalisation de x sachant que y est arrivé. — Proba de réalisation conjointe de x et de y. — Proba de réalisation de y.

Dans le cas où les événements x et y sont indépendants, on a
$P(x, y) = p(x) \, . \, p(y) \Rightarrow P(x \mid y) = P(x)$.

Un système de communication (canal) relie deux sources d'information entre elles : la source émettrice à l'entrée et la source réceptrice à la sortie du canal. Il s'agit ici de canaux discrets sans mémoire, mais en général avec bruit.

Discret :
A états dénombrables
(Non analogique)

Sans mémoire :
La probabilité d'un symbole (état)
ne dépend pas des états précédents

Exercice 5.1 : *Redondance du langage.*

Si le canal entre l'émetteur et le récepteur rajoute du bruit, celui-ci se manifestera par le fait qu'un état de la source émettrice ne va pas toujours donner le même état sur la source réceptrice. Voyons de quelle façon la connaissance de la source réceptrice donne de l'information sur la source émettrice. Voici par exemple un canal ayant 3 états en entrée et 4 états en sortie. C'est un canal discret car il relie une source émettrice et une source réceptrice qui ont un nombre fini d'états.

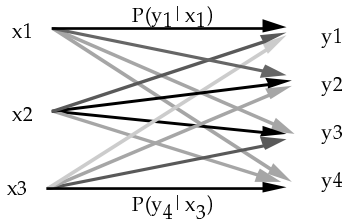

Figure 5,1 : exemple de canal discret avec bruit.

On note: x_e les m états de la source X "émetteur" , et P (x_e) leurs probas ;
y_r les n états de la source Y "récepteur", et P (y_r) et leurs probas.

5.1 Vecteurs et matrices associés aux canaux.

Soient:

X le vecteur formé avec pour composantes les m états x_e ;
Y le vecteur formé avec pour composantes les n états y_r .

et

P(**X**) le vecteur formé avec pour composantes les m probas P(x_e) ;
P(**Y**) le vecteur formé avec pour composantes les n probas P(y_r) .

La matrice de transmission du canal : **M** (**Y** | **X**) est formée avec pour éléments les probabilités conditionnelles P (y_r | x_e) : probabilité d'avoir l'état y_r en sortie (réception) sachant qu'on a émis x_e .

Le canal illustré en figure 5.1 a pour matrice de transmission une matrice (3,4):

$$\mathbf{M}(\mathbf{Y}|\mathbf{X}) = \begin{matrix} P(y_1|x_1) & P(y_2|x_1) & P(y_3|x_1) & P(y_4|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & P(y_3|x_2) & P(y_4|x_2) \\ P(y_1|x_3) & P(y_2|x_3) & P(y_3|x_3) & P(y_4|x_3) \end{matrix}$$

La somme de chaque ligne vaut 1 : Σ_r P (x_e | y_r) = 1 $\forall$ e , car pour un état émetteur x_e donné, on a la certitude d'avoir l'un quelconque des états possibles récepteur y_r .

D'après la loi de Bayes on a la relation :

$$P(y_r) = \Sigma_e P(x_e, y_r) = \Sigma_e P(x_e) P(y_r | x_e)$$

Ce qui s'écrit sous forme matricielle :

$$\mathbf{P}(\mathbf{Y}) = \mathbf{P}(\mathbf{X}) \ \mathbf{M}(\mathbf{Y}|\mathbf{X})$$

Le canal est sans mémoire si la matrice de transmission reste invariable après transmission d'un message.

La <u>matrice des probabilités conjointes</u> : **M** (**Y, X**) est définie avec pour éléments les probabilités conjointes : $P(y_r, x_e)$.

Si on calcule **M** (**X , X**) on obtient une matrice carrée avec les éléments du vecteur **P** (**X**) sur la diagonale et des 0 partout ailleurs. Il y a en effet une probabilité nulle d'avoir simultanément deux états différents x_e . Notons cette matrice **M** (**X**) ≡ **M** (**X , X**) .

On a la relation :

M (**Y, X**) = **M** (**X**) **M** (**Y** | **X**)

Par exemple : prenons un canal à deux états, dit "binaire", et donnons-lui arbitrairement la matrice de transmission suivante :

$$\mathbf{M}(\mathbf{Y}|\mathbf{X}) = \begin{matrix} P(y_1 | x_1) = 0{,}8 & P(y_2 | x_1) = 0{,}2 \\ P(y_1 | x_2) = 0{,}3 & P(y_2 | x_2) = 0{,}7 \end{matrix}$$

On obtient le schéma suivant :

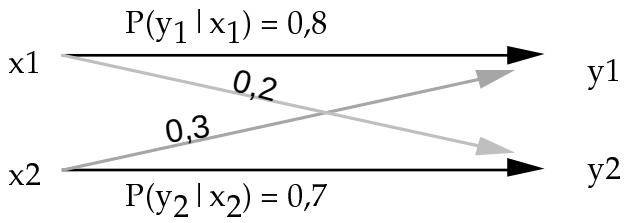

figure 5.2 : exemple de canal binaire.

Choisissons maintenant des probabilités arbitraires pour les états en émission:

$P(x_1) = 0{,}4$; $P(x_2) = 0{,}6$

Par la relation **P** (**Y**) = **P** (**X**) **M** (**Y** | **X**) on obtient :

$$\mathbf{P}(\mathbf{Y}) = \begin{pmatrix} 0{,}4 & 0{,}6 \end{pmatrix} \times \begin{matrix} 0{,}8 & 0{,}2 \\ 0{,}3 & 0{,}7 \end{matrix} = \begin{pmatrix} 0{,}5 & 0{,}5 \end{pmatrix}$$

et par la relation **M** (**Y, X**) = **M** (**X**) **M** (**Y** | **X**) on obtient la matrice des probabilités conjointes:

$$\mathbf{M}(\mathbf{Y}, \mathbf{X}) = \begin{matrix} P(x_1, y_1) & P(x_1, y_2) \\ P(x_2, y_1) & P(x_2, y_2) \end{matrix} = \begin{matrix} 0{,}4 & 0 \\ 0 & 0{,}6 \end{matrix} \times \begin{matrix} 0{,}8 & 0{,}2 \\ 0{,}3 & 0{,}7 \end{matrix} = \begin{matrix} 0{,}32 & 0{,}08 \\ 0{,}18 & 0{,}42 \end{matrix}$$

5.2 Entropie conjointe et entropie conditionnelle de deux sources liées.

On va maintenant pour tout type de canal étudier de quelle façon la connaissance de la source réceptrice apporte de l'information sur la source émettrice. On se place donc à partir de maintenant du point de vue "récepteur", c'est à dire qu'on va partir des $P(y_r)$ et s'intéresser aux $P(x_e \mid y_r)$.

La loi de Bayes donne:

$P(x_e, y_r) = P(y_r)\ P(x_e \mid y_r)$

$\Sigma_e P(x_e, y_r) = P(y_r)\ \Sigma_e P(x_e \mid y_r) = P(y_r)$.

<u>L'entropie conjointe</u> de X et Y se définit à partir de l'ensemble des couples : "un état de X et un état de Y"et des probabilités conjointes, c'est l'entropie d'un système à M N états :

$H(X, Y) = -\Sigma_e \Sigma_r P(x_e, y_r)\ \mathrm{lb}\ [\ P(x_e, y_r)\ ]$.

On peut l'exprimer en termes de probabilités conditionnelles :

$H(X, Y) = -\Sigma_e \Sigma_r P(x_e, y_r)\ \mathrm{lb}\ [\ P(y_r)\ P(x_e \mid y_r)\ ]$,

$H(X, Y) = -\Sigma_e \Sigma_r P(x_e, y_r)\ \mathrm{lb}\ P(y_r) \quad - \Sigma_e \Sigma_r P(x_e, y_r)\ \mathrm{lb}\ P(x_e \mid y_r)$

$\qquad = -\Sigma_r P(y_r)\ \mathrm{lb}\ P(y_r) \quad - \Sigma_e \Sigma_r P(x_e, y_r)\ \mathrm{lb}\ P(x_e \mid y_r)$.

Le premier terme de l'expression est l'entropie de la source réceptrice :

$H(Y) = -\Sigma_r P(y_r)\ \mathrm{lb}\ P(y_r)$,

le deuxième terme est appelé <u>entropie conditionnelle</u> :

$H(X \mid Y) = -\Sigma_e \Sigma_r P(x_e, y_r)\ \mathrm{lb}\ P(x_e \mid y_r)$,

on a donc :

$\mathbf{H(X, Y) = H(Y) + H(X \mid Y)}$.

Comme $P(x, y) < P(x)$, on a : $H(X \mid Y) <$ ou $= H(X)$

Donc toute connaissance dr récepteur Y ne peut que diminuer l'entropie de l'émetteur X.

La comparaison de H(X) et de $H(X \mid Y)$ permet d'évaluer en termes "entropiques" le rôle joué par le récepteur dans la connaissance de l'émetteur.

Si X et Y sont indépendants, on a $H(X, Y) = H(X) + H(Y)$, donc $H(X \mid Y) = H(X)$. La connaissance de l'état en réception n'apporte alors aucune information sur l'émetteur. C'est le cas le plus défavorable : la transmission ne marche pas.

Si X et Y sont parfaitement liés, on a $H(X) = H(X, Y) = H(Y)$, et $H(X \mid Y) = 0$. La connaissance de l'état en réception apporte toute l'information sur l'émetteur. C'est le cas le plus favorable : la transmission est parfaitement fidèle.

5.3 Information mutuelle de deux sources liées.

$I(X|Y) \equiv H(X) - H(X|Y)$

En reprenant les deux exemples précédents, on voit que si X et Y sont indépendants, $H(X|Y) = H(X)$, donc $I(X|Y) = 0$.

Si au contraire X et Y sont parfaitement liés, on a $I(X|Y) = H(X)$.

On voit que $H(X) > I(X|Y) > 0$.

$I(X|Y) = -\Sigma_e P(x_e)\, lb\, P(x_e) + \Sigma_e \Sigma_r P(x_e, y_r)\, lb\, P(x_e, y_r) / lb\, P(y_r)$.

or : $\Sigma_e P(x_e) = \Sigma_e \Sigma_r P(x_e, y_r)$.

On a donc : $\Sigma_e P(x_e)\, lb\, P(x_e) = \Sigma_e\, lb\, P(x_e)\, \Sigma_r P(x_e, y_r) = \Sigma_e \Sigma_r P(x_e, y_r)\, lb\, P(x_e)$

$I(X|Y) = -\Sigma_e \Sigma_r P(x_e, y_r)\, lb\, P(x_e) + \Sigma_e \Sigma_r P(x_e, y_r)\, lb\, P(x_e, y_r) / lb\, P(y_r)$.

$I(X|Y) = \Sigma_e \Sigma_r P(x_e, y_r)\, lb\, [\, P(x_e, y_r) / P(x_e)\, P(y_r)\,]$

On voit en passant que $I(X|Y) = I(Y|X)$.

En utilisant la relation : $H(X, Y) = H(Y) + H(X|Y)$, on trouve :

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

---

exercice 5.2 :

Pour le canal binaire défini dans la figure 5.2 :

avec $P(X) = (0,4 ; 0,6)$ on a vu qu'on obtenait $P(Y) = (0,5 ; 0,5)$.

Calculer $M(X|Y)$.

calculer les entropies conjointes et conditionnelles pour cette source émettrice.

Calculer l'information mutuelle entre la source émettrice et la source réceptrice.

5.4 Capacité d'un canal.

La capacité d'un canal est la valeur maximale de l'information transportée entre émetteur X et récepteur Y. La capacité d'un canal s'exprime en bits.

$C = \max_x I(X|Y) = \max_x I(Y|X) = \max_x [\, H(X) - H(X|Y)\,]$

"$\max_x$" veut dire : le maximum pour toutes valeurs possibles de X.

Prenons une source émettrice à deux états de loi de probabilité $P(X) = (\alpha ; 1 - \alpha)$, et un canal binaire dont la matrice de transmission serait :

$$\mathbf{M}(\mathbf{Y}|\mathbf{X}) = \begin{matrix} P(y_1 | x_1) = p & P(y_2 | x_1) = 1 - p \\ P(y_1 | x_2) = 1 - P & P(y_2 | x_2) = p \end{matrix}$$

On obtient le schéma suivant :

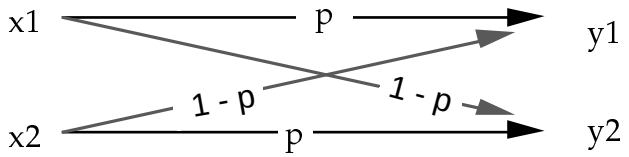

Figure 5,3 : canal binaire symétrique

$H(X|Y) = -\Sigma_e \Sigma_r P(x_e, y_r) \text{ lb } P(x_e | y_r)$,

$H(X|Y) = -\alpha p \text{ lb } p - (1 - \alpha)p \text{ lb } p - \alpha(1 - p) \text{ lb } (1 - p) - (1 - \alpha)(1 - p) \text{ lb } (1 - p)$,

$H(X|Y) = -p \text{ lb}p - (1 - p) \text{ lb } (1 - p)$.

Comme $H(X|Y)$ ne dépend pas de α :

$\max_x [H(X) - H(X|Y)] = \max_x [H(X)] - H(X|Y)$.

Or $\max_x [H(X)] = 1$, valeur atteinte pour $\alpha = 0{,}5$.

On a donc :

$C = 1 + -p \text{ lb}p + (1 - p) \text{ lb } (1 - p)$.

Situation optimale : C = 1 pour p = 0 ou 1

Situation la plus mauvaise : C = 0 pour P = 0,5.

Limitations dues à la d'un canal de transmission.

Un canal de capacité C ne pourra correctement transmettre le débit d'information d'une source de débit R que si $R \leq C$.

5.5 Canaux remarquables.

<u>canal sans perte :</u>

Un canal est sans perte lorsque tout signal émis par X est reçu par Y :

$P(x_e | y_r) = 0$ ou 1 .

les relations entre émetteur et récepteur sont de nature injective (voir figure).

Pour canal est sans perte : $H(X|Y) = 0$. En effet,

$H(X|Y) = -\Sigma_e \Sigma_r P(x_e, y_r) \text{ lb } P(x_e | y_r)$

$H(X|Y) = -\Sigma_e \Sigma_r P(y_r) P(x_e | y_r) \text{ lb } P(x_e | y_r)$

$H(X|Y) = -\Sigma_r P(y_r) \quad \Sigma_e P(x_e | y_r) \text{ lb } P(x_e | y_r) \quad = 0$

vaut 0 si $P(x_e \mid y_r) = 0$ ou 1

Capacité du canal :

$I(X \mid Y) = H(X) \Rightarrow C = \max H(x)$

Le canal est toujours adapté à la source émettrice, aucune information n'est perdue mais on a augmenté inutilement l'entropie à la réception.

$$\mathbf{M}(\mathbf{Y} \mid \mathbf{X}) = \begin{matrix} P(y_1, x_1) = 3/4 & 1/4 & 0 & 0 & P(y_5, x_1) = 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ P(y_1, x_3) = 0 & 0 & 0 & 0 & P(y_5, x_3) = 1 \end{matrix}$$

On remarque qu'il n'y a qu'un seul élément différent de 0 par colonne dans ma matrice de transmission.

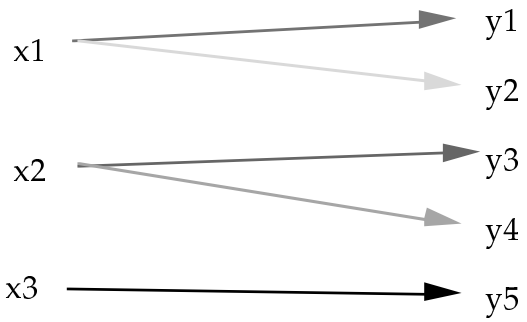

Figure 5,4 : exemple de canal sans perte

---

exercice 5.3 :

En prenant une loi de proba pour l'émission de (1/3,1/3,1/3) et le canal défini dans la figure 5.4, calculer l'entropie à l'émission et à la réception, et l'information mutuelle.

<u>Canal déterministe</u> :

Un canal est déterministe lorsque $P(y_r \mid x_e) = 1$ pour une valeur de y et 0 pour les autres. les relations entre émetteur et récepteur sont de nature surjective.

$H(Y \mid X) = 0$ puisque $P(y_r \mid x_e) = 0$ ou 1

$I(Y \mid X) = H(Y) - H(Y \mid X) = H(Y)$

$C = \max H(Y) = \text{lb } n$

$$\mathbf{M}(\mathbf{Y} \mid \mathbf{X}) = \begin{matrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{matrix}$$

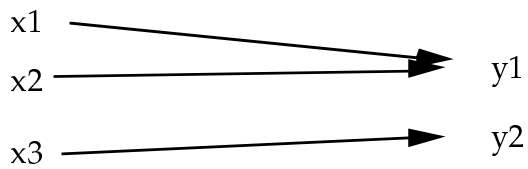

Figure 5,5: exemple de canal déterministe

<u>Canal sans bruit</u> :

Un canal sans bruit est un canal à la fois déterministe et sans perte : les relations entre

émetteur et récepteur sont de nature bijective.

$P(x_e \mid y_r) = 1$ ou $0 \Rightarrow$ sans perte
$P(y_r \mid x_e) = 1$ ou $0 \Rightarrow$ déterministe

$H(Y \mid X) = 0$ et $H(X \mid Y) = 0$

$C = \max H(X) = \max H(Y) = \text{lb } m$

$$\mathbf{M}(\mathbf{Y} \mid \mathbf{X}) = \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \qquad \begin{matrix} x1 \longrightarrow y1 \\ x2 \longrightarrow y2 \\ x3 \longrightarrow y3 \end{matrix}$$

Figure 5,5: exemple de canal sans bruit

5.6 Exemples concrets de canaux discrets ou non :

signal avec niveaux quantifiés et bruit analogique ajouté : canal "sans bruit" tant que l'amplitude du bruit ajouté ne dépasse pas la moitié de l'écart entre les niveaux.

encryptage et décryptage de message $\Rightarrow$ sans bruit car réversible

toute transformation mathématique inversible $\Rightarrow$ sans bruit

compression d'image JPEG : canal déterministe, donc avec bruit.

6. Codage avec contrôle d'erreur

6,1 Distance de Hamming

La distance de Hamming est le nombre de discordances bit à bit entre deux nombres exprimées en binaire. C'est une distance au sens mathématique.

exemple:
a = 1011000101001011
b = 1011<u>1</u>0101000<u>0</u>011 $\qquad d_H(a, b) = 2$

6.2 détection d'erreur par parité

La parité d'un nombre binaire est le nb de bits à 1 exprimé modulo 2 :
Exemple pour le nombre 21 = 10101_2
le nb de bits à 1 dans " 10101" est 3 ; 3 (modulo 2) = 1 $\Rightarrow$ parité = 1

Le codage associé à la parité permet de détecter un éventuelle erreur dans l'envoi d'un séquence de bits, à condition qu'il n'y ait pas plus de un bit qui change (distance de Hamming ≥ 1) entre le signal correct et le signal erroné:

signal correct pour "21" : 101011

Il y a n+1 bits, dont n de données et un de parité. On a introduit de la redondance: le nb de bits à l'émission est supérieur à ce qui est nécessaire. Mais en contrepartie, on est capable de détecter s'il y a eu une erreur lors de la réception.

exercice 6.1 :

Calculer la distance minimale de Hamming entre deux nombres (mots) quelconques différents codés en binaire.

Calculer la distance minimale de Hamming entre deux mots quelconques différents codés en binaire, auxquels on a rajouté la parité.

exercice 6.2 :

Comparer le nombre d'états et l'entropie d'une source sans parité avec une source formée à partir de la précédente à laquelle on a rajouté la parité.

A la réception, si le canal est bruité, on aura une probabilité non nulle pour que l'un des bits soit inversé.

exemple de signal reçu erroné : 100011

En recalculant la parité à partir des 5 bits de données reçus, on trouve 0 , ce qui est en désaccord avec la parité reçue. On a donc détecté une erreur de transmission.

6.3 Correction d'erreur

En rajoutant des bits de redondance on peut non seulement détecter des erreurs, mais aussi les corriger. Si, avec un codage redondant, on obtient une distance de Hamming minimale (d_{Hmin}) entre deux mots distincts de l'ensemble de départ,

On pourra détecter au plus $(d_{Hmin} - 1)$ erreurs.
On pourra détecter et corriger au plus $(d_{Hmin} - 1)/2$ erreurs.

Voici un exemple de codage avec correction d'erreur nécessitant lb(n) bits de redondance (check bits) pour n bits de données. Ce codage datant de 1950 est dû à Hamming. Il permet de détecter et corriger un bit erroné dans le mot de n bits transmis.

Le but du codage de Hamming est qu'à chaque position de bit de donnée du mot transmis corresponde une combinaison de "check bits" qui lui est propre.

Exemple : codons 1110 1001 en intercalant des "check bits" toutes les 2^n positions

a) On insère les bits de donnée en laissant vides les positions des check bits.

donn e	_	_	1	_	1	1	0	_	1	0	0	1
position	1	2	3	4	5	6	7	8	9	10	11	12

b) La contribution de chaque bit de donnée aux check bits va dépendre de sa position de façon unique.

Les positions des bits peuvent s'exprimer en somme de puissance de 2: chaque check bit ne contribue qu'à lui-même. Chaque bit de donnée contribue aux checks bits qui correspondent à la décomposition de sa position en puissance de 2.

pos.	décomposition binaire	
1	1	check bit

```
 2            2        check bit
 3            2 + 1
 4        4            check bit
 5        4      + 1
 6        4 + 2
 7        4 + 2 + 1
 8   8                 check bit
 9   8 +          + 1
10   8 +        2
11   8 +        2 + 1
12   8 +   4
```

Sur le check bit 1 on sommera la parité des bits, 3,5,7,9,11 = 1+1+0+1+0 = 1 ;
sur le check bit 2 la parité des bits 3,6,7,10,11 = 1+1+0+0+0 = 0 ;
sur le check bit 4 la parité des bits 5,6,7,12 = 1+1+0+1 = 1 ;
sur le check bit 8 la parité des bits 9, 10, 11, 12 = 1+0+0+1 = 1 .

Cela donne :

donn e	1	0	1	1	1	1	0	1	1	0	0	1
position	1	2	3	4	5	6	7	8	9	10	11	12

A la réception on recalcule les check bits et l'on compare (avec un ou exclusif) avec ceux reçus. Si un des bits est inversé, tous les check bits codant sa position auront changé. On saura ainsi quel bit a été inversé et il suffira de le rétablir.

6.4 Limitation de la portée des erreurs : code Gray

Dans le code binaire "naturel", le passage d'un nombre au suivant (de n à n+1) peut changer radicalement son expression. Par exemple en passant de 7 à 8 on passe de 0111 à 1000. Tous les bits changent de valeur en même temps. Si pour des raisons mécaniques ou électroniques cette transition n'est pas simultanée, on peut voir 1111 au lieu de 1000, et lire 15 au lieu de 8.

Dans le code binaire Gray, il y a toujours un seul bit qui change entre la représentation d'un nombre n et celle de n+1. Ceci permet de diminuer l'impact d'une erreur lors d'une lecture: Sl la valeur lue est en train de changer, une erreur éventuelle de lecture n'aura pour conséquence que de changer au plus d'une unité la valeur lue.

décimal	binaire	Gray
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110
5	101	111
6	110	101
7	111	100
8	1000	1100

La conversion binaire -> Gray : se fait par un "ou" exclusif entre le nombre binaire et lui-même décalé d'un cran vers la droite. Exemple de code en C :

```
Gray = bin ^  (bin >> 1);
```

Conversion Gray -> binaire. Exemple de code en C :

```
int bin = Gray ;      // au départ les variables bin et Gray sont en Gray
while (Gray != 0)
{
      Gray >>= 1;   // décalage des bits de 1 cran vers la droite
      bin ^= Gray;  // ou exclusif
}
```

7. Imageurs optiques et information

Lecture des données dans l'acquisition d'images

Selon le type de caméra, les données apparaissent comme une lecture séquentielle des pixels ou comme une liste des coordonnées des photons détectés: avec les détecteurs à comptage on a la position de chaque photon individuellement. Dans ce cas on peut reconstruire l'image $I(\mathbf{x})$ en faisant l'histogramme des coordonnées:

$$I(\mathbf{x}) = \sum_{1}^{p} \delta(\mathbf{x} - \mathbf{x}_i)$$

ici $\mathbf{x}$ (en gras) est un vecteur 2D qui représente la position (x,y) dans l'image.
$\mathbf{x}_i$ est la coordonnée d'arrivée du photon i dans l'image.
p est le nombre de photons.
δ est une distribution de Dirac : $\delta(\mathbf{x}) = 0$ pour $\mathbf{x} \neq \mathbf{0}$ et $\int \delta(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} = f(\mathbf{0})$

On pourra aussi directement faire la transformée de Fourier de l'image à partir des coordonnées de photons :

$$\hat{I}(\mathbf{u}) = \int \Sigma_i\ \delta(\mathbf{x} - \mathbf{x}_i) \exp(-2i\pi\mathbf{u}\mathbf{x})\, d\mathbf{x} \qquad = \Sigma_i\ \exp(-2i\pi\mathbf{u}\mathbf{x}_i)$$

Faire de cette façon peut conduire à un gain énorme en temps de calcul.

EN PRÉPARATION

7.1 Pixels et resels,
defs de la convolution et des relations d'imagerie
un imageur est-il un canal discret ?

relation entre convolution et bruit au sens entropique
à partir de quelle taille de pixel un imageur est-il un canal sans bruit ?

7.2 Interférométrie
Schéma d'un interféromètre
Imagerie par TF
Résolution max, taille de l'élément de résolution

La TF conserve l'entropie d'une source (car transformation inversible)
Entropie d'une source d'info "liste de visibilités"
Conséquences sur l'entropie d'une image reconstruite
champ d'une image reconstruite

matrices creuses de photons ...