Théorie de la transmission de l'information

Introduction

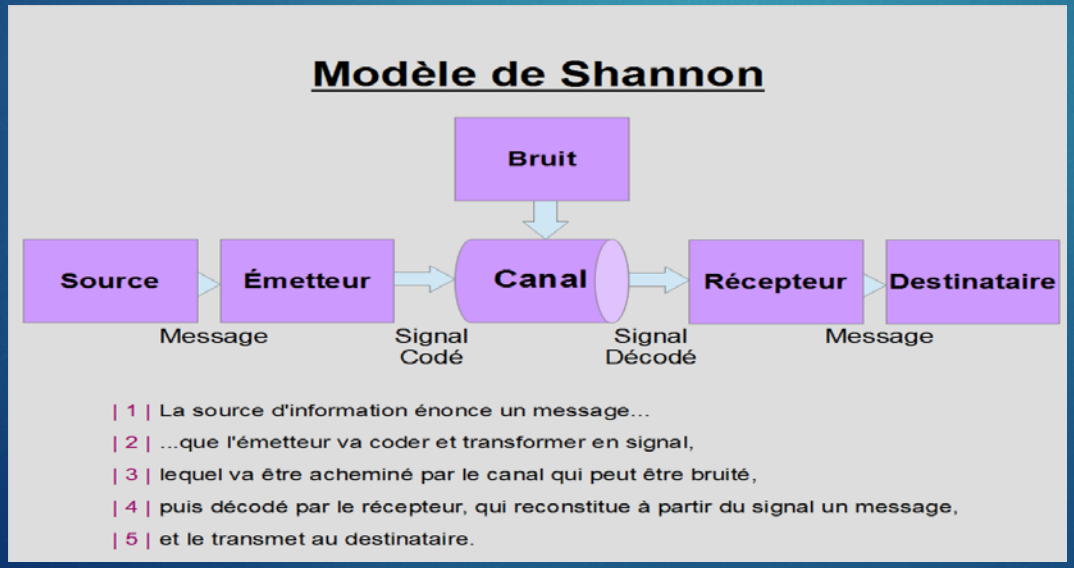
Claude SHANNON 1916-2001
Le père de la théorie de l'information

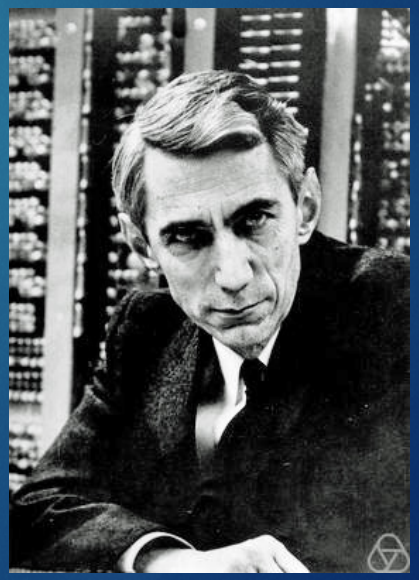
Ingénieur et chercheur, mathématicien

- **MIT (1958-1978)**
- **Laboratoires BELL (1941-1972)**
- **Très actif dans les service secrets américains durant la Seconde Guerre mondiale (cryptographie)**

Article fondateur de la théorie de l'information : 1948

A Mathematical Theory of Communication





Mesure de la quantité d'information : exemple élémentaire

Considérons *N* boîtes numérotées de 1 à *N*. Un individu A a caché au hasard un objet dans une de ces boîtes. Un individu B doit trouver le numéro de la boîte où est caché l'objet. Pour cela, il a le droit de poser des questions à l'individu A auxquelles celui-ci doit répondre sans mentir par OUI ou NON. Mais chaque question posée représente un coût à payer par l'individu B (par exemple un euro). Un individu C sait dans quelle boîte est caché l'objet. Il a la possibilité de vendre cette information à l'individu B. B n'acceptera ce marché que si le prix de C est inférieur ou égal au coût moyen que B devrait dépenser pour trouver la boîte en posant des questions à A. L'information détenue par C a donc un certain prix. Ce prix représente la quantité d'information représentée par la connaissance de la bonne boîte : c'est le nombre moyen de questions à poser pour identifier cette boîte. Nous la noterons I.

- Si $N = 1$, $I = 0$.

Il n'y a qu'une seule boîte. Aucune question n'est nécessaire.

- Si $N = 2$, $I = 1$. On demande si la bonne boîte est la boîte n° 1. La réponse OUI ou NON determine alors sans ambiguïté quelle est la boîte cherchée.

- Si $N = 4$, $I = 2$.

On demande si la boîte porte le n° 1 ou 2. La réponse permet alors d'éliminer deux des boîtes et il suffit d'une dernière question pour trouver quelle est la bonne boîte parmi les deux restantes.

- Si $N = 2^k$, $I = k$.

On écrit les numéros des boîtes en base 2. Les numéros ont au plus *k* chiffres binaires, et pour chacun des rangs de ces chiffres, on demande si la boîte cherchée possède le chiffre 0 ou le chiffre 1.

En *k* questions, on a déterminé tous les chiffres binaires de la bonne boîte. Cela revient également à poser *k* questions, chaque question ayant pour but de diviser successivement le nombre de boîtes considérées par 2 (méthode de dichotomie). On est donc amené à poser $I = \log_2(N)$, mais cette configuration ne se produit que dans le cas de *N* événements équiprobables.

Mesure de la quantité d'information : exemple élémentaire suite

Supposons maintenant que les boîtes soient colorées, et qu'il y ait *n* boîtes rouges. Supposons également que C sache que la boîte où est caché l'objet est rouge. Quel est le prix de cette information ? Sans cette information, le prix à payer est $\log_2(N)$. Muni de cette information, le prix à payer n'est plus que $\log_2(n)$. Le prix de l'information « la boîte cherchée est rouge » est donc $\log_2(N) - \log_2(n) = \log_2(N/n)$.

On définit ainsi la quantité d'information comme une fonction croissante de N/n avec :
N le nombre d'évènements possibles
n le nombre d'éléments du sous-ensemble délimité par l'information

Afin de mesurer cette quantité d'information, on pose :

$$I = log_2\left(\frac{N}{n}\right)$$

I est exprimé en bit (ou « logon », unité introduite par Shannon, de laquelle, dans les faits, bit est devenu un synonyme), ou bien en « nat » si on utilise le logarithme naturel à la place du logarithme de base 2.

Cette définition se justifie, car l'on veut les propriétés suivantes :

- l'information est comprise entre 0 et ∞ ;
- un évènement avec peu de probabilité représente beaucoup d'information (exemple : « Il neige en janvier » contient beaucoup moins d'information que « Il neige en août » pour peu que l'on soit dans l'hémisphère nord) ;
- l'information doit être additive.

1. Terminologie

- **Information :** Supposons que dans une certaine situation donnée puissent se réaliser N événements différents équiprobables. La probabilité de réalisation d'un événement 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 :

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

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 des 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*.

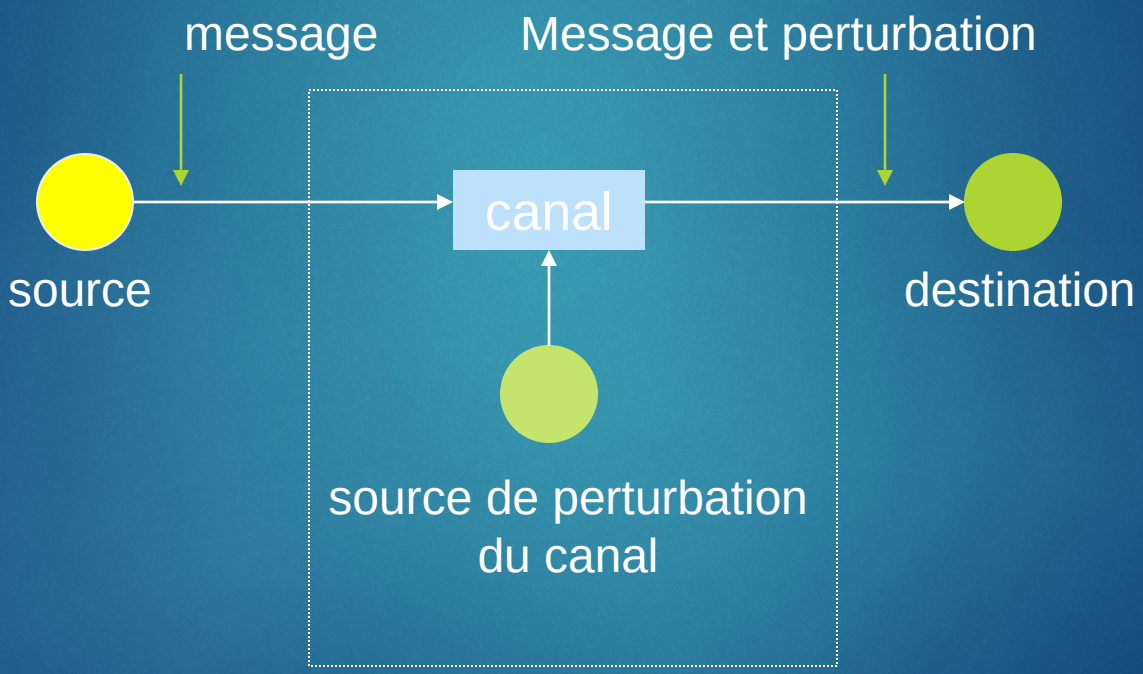
- **Canal :** totalité des moyens destinés à la transmission du message.

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

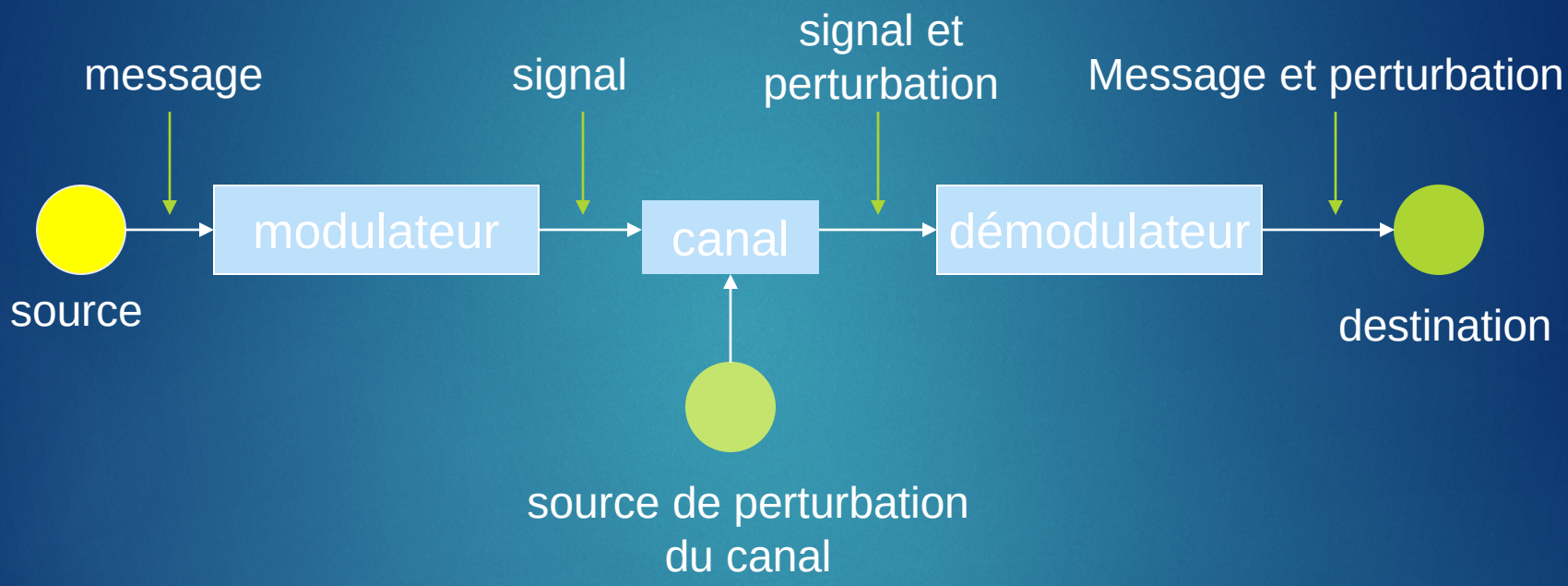
- **Décodage :** transformation de la séquence des signaux réceptionnés dans une séquence qui estime la séquence transmise à l'origine

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

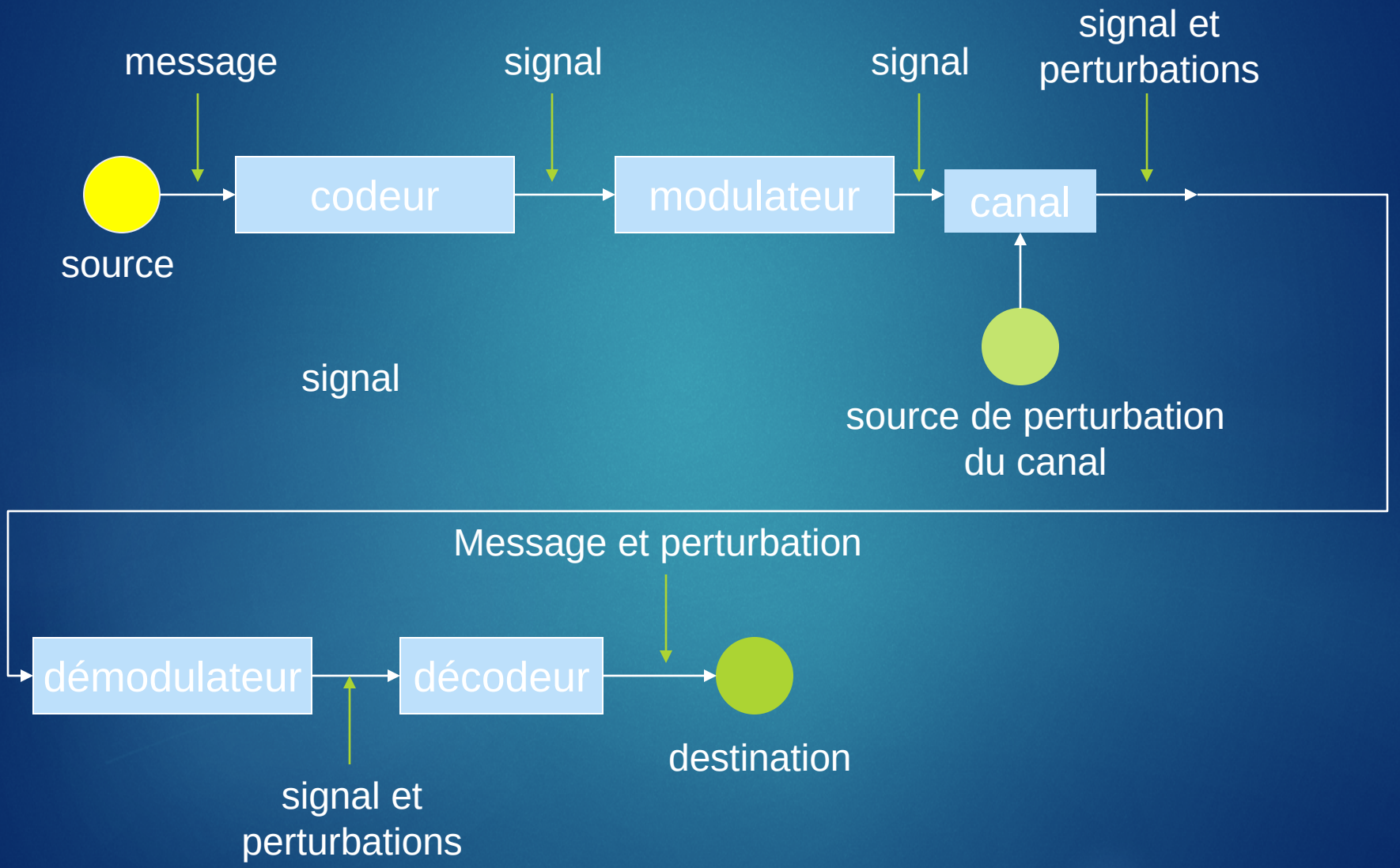
2.1 Transmission directe



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 :

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

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

$\bigcap_{i=1}^{n} x_i = \emptyset$: 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 :

$$i(x_i) = U(x_i)$$

donc :

$$i(x_i) = F(p_i)$$ 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 :

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

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 :

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

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 :

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

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 :

$[X] = [x_1, x_2]$ 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

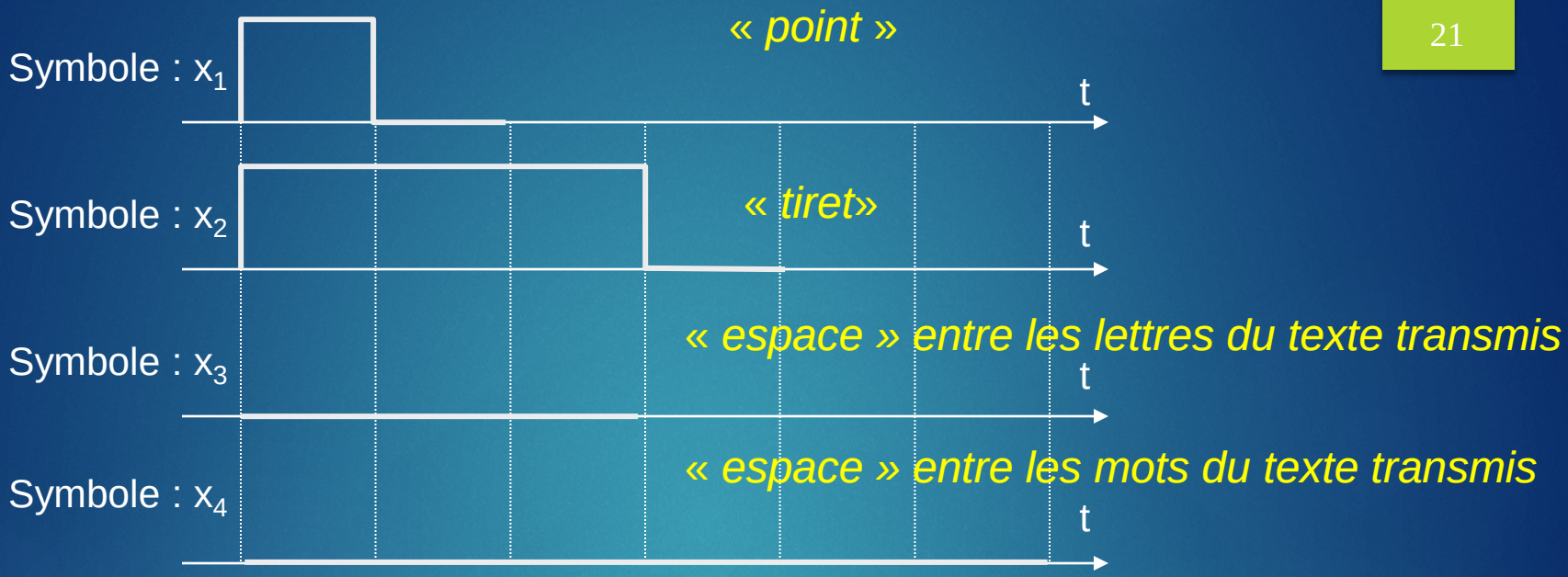
$$i(x_i;y_i)=\log_2\left(\frac{p(x_i/y_i)}{p(x_i)}\right)$$

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.

$$p(x_{in}/x_{in-1},x_{in-2}.....)=p(x_{in})$$

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.

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

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

$$p(x_{jn})=\sum_{i=1}^{D} p(x_{j_n},x_{i_{n-1}})=\sum_{i=1}^{D} p(x_{i_{n-1}})\,p(x_{j_n}/x_{i_{n-1}})$$

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

En posant :

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

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 :

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

générée avec les probabilités :

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

L'information propre par symbole est :

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

L'information propre moyenne sur tous les symboles est :

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

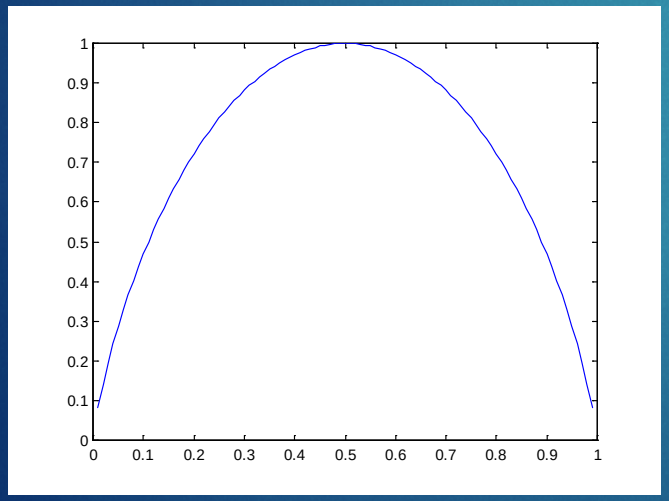
H(X) s'appelle entropie de la source, elle représente l'incertitude moyenne à priori que l'on a sur les événements [X].

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