Information
et
complexité

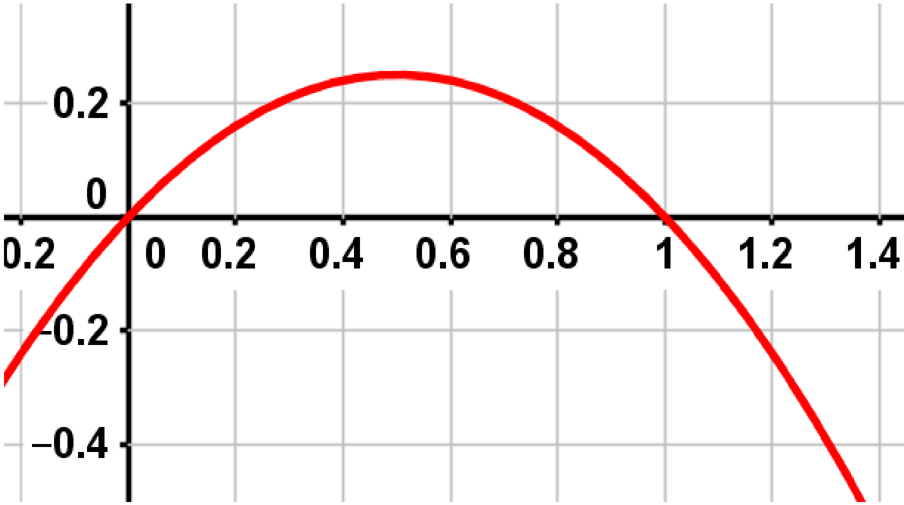
Un jeu de boîtes

Un objet est déposé dans une parmi N boîtes. L'individu V est prêt à le vendre un certain prix à l'individu A. Il accepte également que A lui pose des questions auxquelles il répondra par OUI ou NON. Chaque question coûte une certaine somme. Comment A peut-il choisir : payer de suite ou tenter les questions ?

La dichotomie

L'idée est d'éliminer à chaque question le plus grand nombre de boîtes. Exemple pour deux questions, la première éliminant une proportion x, la seconde $1 - x$. Le maximum est atteint pour $x = \frac{1}{2}$.

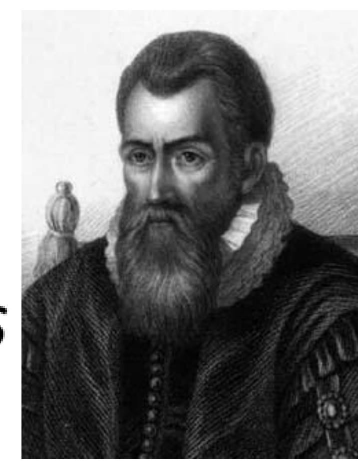
Pour éliminer une boîte sur deux, on les numérote en base 2 et on élimine chiffre par chiffre…



L'information vaut $\log_2 N$

Lorsque $N = 2^k$, k questions suffisent pour dévoiler le numéro de la boîte solution. Lorsque $2^{k-1} \leq N < 2^k$, c'est $k - 1$ questions qu'il suffit de poser. Ce nombre est, à 1 près, le nombre de chiffres de l'écriture de N en base 2. C'est pourquoi on l'assimile au logarithme en base 2.

Les logarithmes, introduits en Occident par John Napier (1550-1617), sont des fonctions transformant les produits en sommes.



John Napier

De l’information à l’entropie

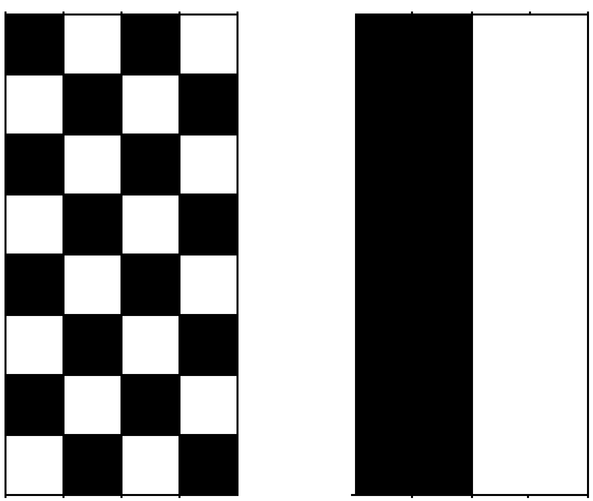
Si les boîtes sont de plusieurs couleurs, l’information « la boîte est une des n boîtes rouges » vaut $\log_2 n$. Muni de cette information, le prix à payer est $\log_2 N - \log_2 n$, soit $\log_2 \left(\frac{N}{n}\right)$. S’il y a *p* couleurs, le prix moyen de l’information sur la couleur de la boîte cherchée est :

$$H = \sum_{k=1}^{p} \frac{n_k}{n} \log_2 \frac{N}{n_k}$$

C’est *l’entropie de Shannon.*

Calculs d'entropie

Si l'information est dispensée par un message de N symboles pris parmi n symboles s_i ayant une fréquence d'apparition p_i, l'entropie est donnée par $H = -N \sum_{i=1}^{n} p_i \log p_i$



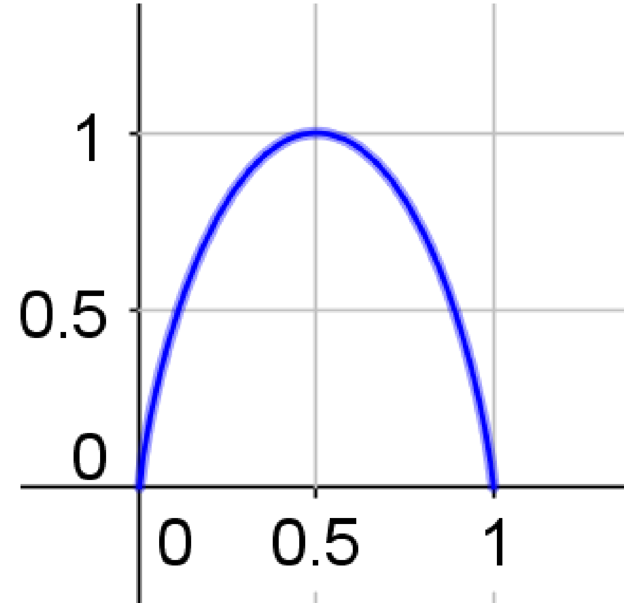
Ces deux pages ont la même entropie

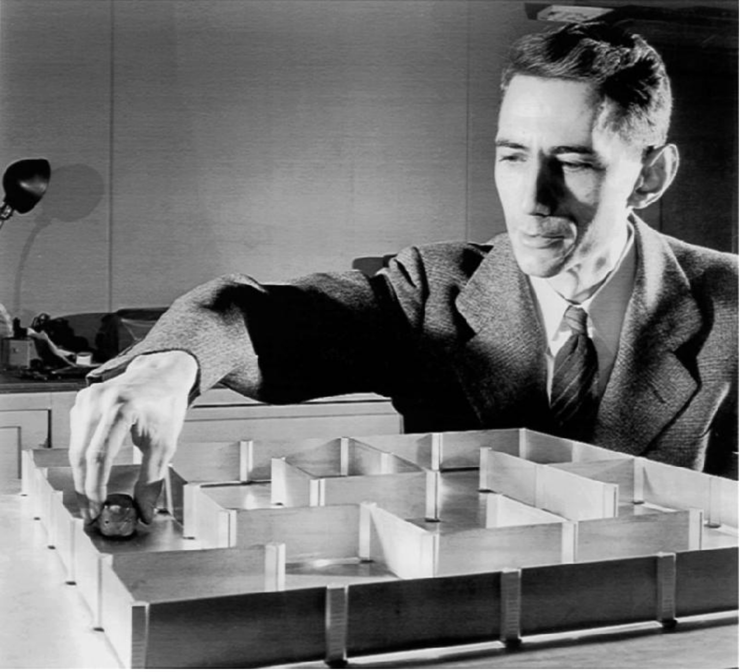
Message comportant des symboles de fréquence d'apparition égale : $H = N \log n$ (exemple $n = 256$)

Un seul symbole : entropie 0, deux symboles également répartis : entropie $N \log 2 = N$

Entropie et codage

Un message de N lettres choisies parmi a et b recèle une quantité d'information égale à $H = -N(x\log x + (1-x)\log(1-x))$, si la fréquence d'apparition de a est x. La longueur du message (N) est supérieure à la quantité d'information qu'il contient. Un **théorème de Shannon** énonce qu'il est possible de trouver un code rendant sa longueur aussi proche qu'on veut de *H.*

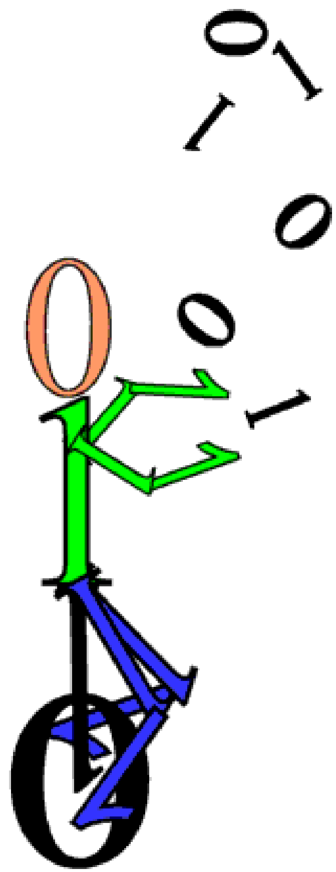


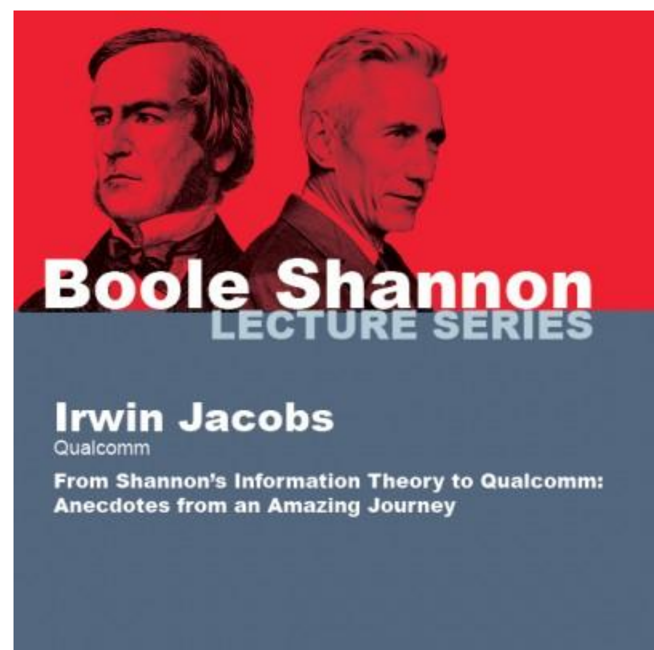


Bit

Claude Shannon

1916 - 2001





Améliorer le codage

Deux exemples :

Codage de Huffman (code lié au contenu)

e	s	a	n	t	i
00	100	110	010	011	1110

Codes correcteurs d'erreurs

http://xavier.hubaut.info/coursmath/app/codes.htm

Idée : on écrit des « mots » de 2^n lettres, 0 ou 1. Parmi ces mots, seuls 2^n par exemple appartiennent au langage. Si on rencontre un mot n'appartenant pas au langage, on le remplace par le mot le plus proche dans le langage.

Claude Berrou et les turbo codes

T	H	E	T	A
A	R	I	O	N
L	I	E	S	T
S	E	N	T	N
E	R	S	U	L

Horizontalement	Verticalement
1. Grecque	1. Ballet
2. Aéronef	2. Lavabo
3. Réfutent	3. Attaches
4. Chemin	4. Taille
5. Anneaux	5. Piliers

T	H	E	T	A
A	V	I	O	N
L	I	E	S	T
S	E	N	T	E
E	R	S	U	L

V	E	L	T	A
A	V	I	O	N
L	I	E	S	T
S	E	N	T	E
E	R	S	U	S

Horizontalement	Verticalement
1. Grecques	1. Ballet
2. Aéronef	2. Lavabo
3. Réfutent	3. Attaches
4. Chemin	4. Taille
5. Anneaux	5. Piliers

D	E	L	T	A
A	V	I	O	N
N	I	E	N	T
S	E	N	T	E
E	R	S	E	S

Complexité

Nous doutons que l'entropie de Shannon soit propre à définir la complexité – par exemple la complexité d'une suite de symboles. Voici quelques exemples de suites de chiffres :

000000000000000000000000000000000000

1357902468135790246813579024681357902

1234567891011121314151617181920212223

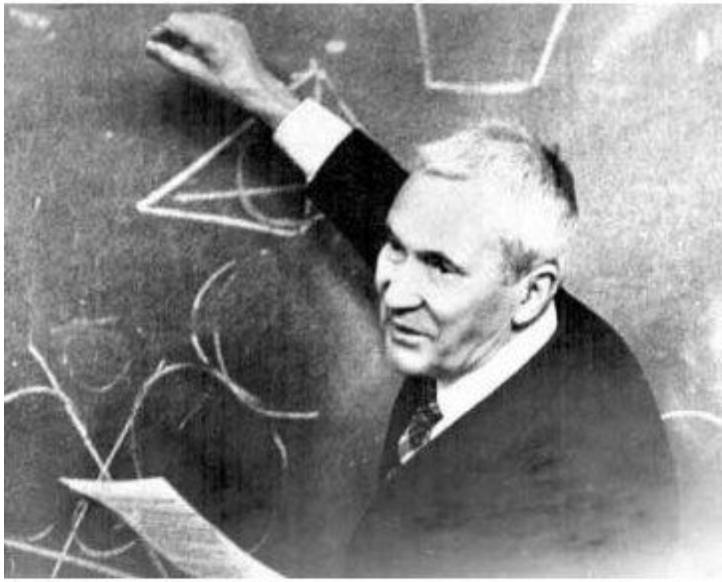
1415926535897932384626433832795028841

01101001100101101001011001101001

Complexité algorithmique

« La complexité d'une suite de symboles (ou de pixels, etc.) est la longueur du programme le plus court permettant de la réaliser ».

Evidemment, le programme dépend du langage utilisé, mais pour les langages universels on dispose d'interpréteurs.



Andrei Nikolaievitch Kolmogorov