Information Theory

Some Applications in Computer Science

Jean-Marc Vincent

MESCAL-INRIA Project
Laboratoire d'Informatique de Grenoble
Universities of Grenoble, France
Jean-Marc.Vincent@imag.fr
LICIA - joint laboratory LIG -UFRGS

This work was partially supported by CAPES/COFECUB

Motivations

Uncertainty

- knowledge on input data
- user's behavior
- losses and crashes
- ...

Complexity

- What is a complex system ?
- Could we compare the complexity of systems ?
- How could we reduce complexity, the abstraction process ?
- ...

Information

- definition of information
- measure of information
- structural properties
- ...

Applications (selected)

Coding

- Codes properties
- Kraft inequality and Shannon's theorems
- Optimal codes

Modeling

- MaxEnt principle
- Classical laws from constraints
- Gibbs distributions

Representation

- Aggregation quantification
- Large distributed systems trace analysis
- optimal visualization

Information

Information. n. (late Middle English (also in the sense 'formation of the mind, teaching'), via Old French from Latin informatio(n-), from the verb informare)

1. facts provided or learned about something or someone : a vital piece of information (law)
2. what is conveyed or represented by a particular arrangement or sequence of things : *genetically transmitted information*
   - Computing data as processed, stored, or transmitted by a computer.
   - (in information theory) a mathematical quantity expressing the probability of occurrence of a particular sequence of symbols, impulses, etc., as contrasted with that of alternative sequences.

Une machine à calculer "peut communiquer à des utilisateurs les résultats de ses calculs, c'est à dire de l'information" (De Broglie)

Oxford Dictionary

http://oxforddictionaries.com

Numerical Revolution

Every information could be represented by a finite sequence of 0 and 1.

Universal $\Rightarrow$ Universal machine to process information : the computer

1. Signs, pictograms, natural languages
2. Signals : audio, images, video,
3. physical word (simulations) : forces, flows, fields,...
4. . . .

Sequence of bits

- The sequence of bits 01010101010101010101 is represented by a simple algorithm

  `Repeat n times write "0" then "1"`

  The binary representation of a sequence with length $2n$ is

  $$\log_2(n) + C$$

  size of n (in bits) + size of the code of the algorithm

- The sequence of the binary digits of π is easily extracted from the formula (Bailey-Borwein-Plouffe1995)

$$\pi = \sum_{k=0}^{+\infty} \frac{1}{16k} \left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right).$$

  Size of the representation of the first n bits

  $$\log_2(n) + C$$

Kolmogorov 's Complexity (1974)

X^* set of finite sequences of bits.
Φ computable function

$$\begin{array}{ccc} \Phi : X^* & \longrightarrow & X^* \\ y & \longmapsto & x \end{array}$$

Definition (Complexity x relatively to Φ)

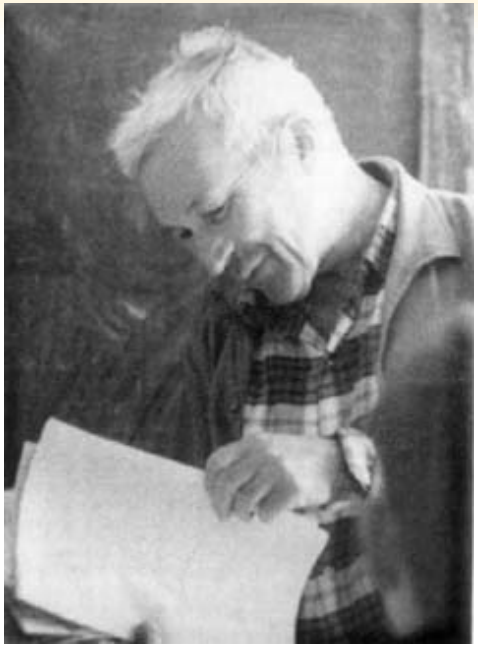
$$\mathcal{K}_\Phi(x) = \begin{cases} \inf\{length(y) \;\; \Phi(y) = x\} \\ +\infty \;\; si\; x \notin \Phi(X^*) \end{cases}$$

Definition (Complexité de Kolmogorov de la chaîne finie x)

$$\mathcal{K}(x) = \inf_{\Phi}\{\mathcal{K}_\Phi(x) + size(\Phi)\}.$$

Remark : $\mathcal{K}(x) \leqslant l(x)$.

Andreï Kolmogorov (1903-1987)



Mathématicien russe dont le nom est principalement attaché à l'axiomatisation du calcul des probabilités dans le cadre de la théorie des ensembles. Fils d'un agronome, Andreï Nikolaïevitch Kolmogorov est né à Tambov. Il entra à dix-sept ans à l'université de Moscou, où il suivit des cours de N. Lusin et de P. Urysohn ; chercheur associé à cette université depuis 1925, il y devint professeur en 1931 et directeur du département de mathématiques deux ans plus tard. En 1939, il fut élu à l'Académie des sciences de l'U.R.S.S.

Les premiers travaux de Kolmogorov portent sur les fonctions de variable réelle (séries trigonométriques, opérations sur les ensembles) ; en 1922, il a construit un exemple célèbre de fonction intégrable dont la série de Fourier est divergente en tout point, ce qui relançait le problème de la convergence des séries de Fourier. Quelques années plus tard, il étendit la sphère de ses recherches à la logique mathématique et aux problèmes de fondements. ¿ partir de 1925, en collaboration avec A. Khintchine, Kolmogorov a étudié les problèmes de convergence de séries d'éléments aléatoires, sur lesquels il a publié de nombreux articles devenus classiques. Son mémoire fondamental, Théorie générale de la mesure et théorie des probabilités (1929), donne la première construction axiomatique du calcul des probabilités fondée sur la théorie de la mesure ; il développa ses idées dans l'ouvrage Grundbegriffe der Wahrscheinlichkeitsrechnung (trad. angl. Foundations of the Theory of Probability, 1950), publié en 1933. Avec son ouvrage Méthodes analytiques de théorie des probabilités, Kolmogorov est un des fondateurs de la théorie des processus stochastiques. Il a étudié plus spécialement ceux qui sont connus de nos jours sous le nom de processus de Markov où deux systèmes d'équations aux dérivées partielles portent son nom ; cette théorie a d'importantes applications en physique (mouvement brownien, diffusion). Mentionnons aussi des recherches très importantes sur les processus aléatoires stationnaires, dont Wiener a souligné le rôle dans la théorie statistique de l'information sur laquelle s'appuie, pour une part, la cybernétique. Kolmogorov a également fait des recherches en topologie, géométrie, analyse fonctionnelle et approximation optimale des fonctions. Depuis 1950, il a publié des travaux sur la théorie de l'information, la mécanique classique et la théorie des automates finis. Il a consacré ses dernières années à des problèmes d'enseignement des mathématiques et a publié plusieurs ouvrages de pédagogie à l'usage des parents et des enseignants. Il termina sa vie à Moscou

$\mathcal{K}$-randomness

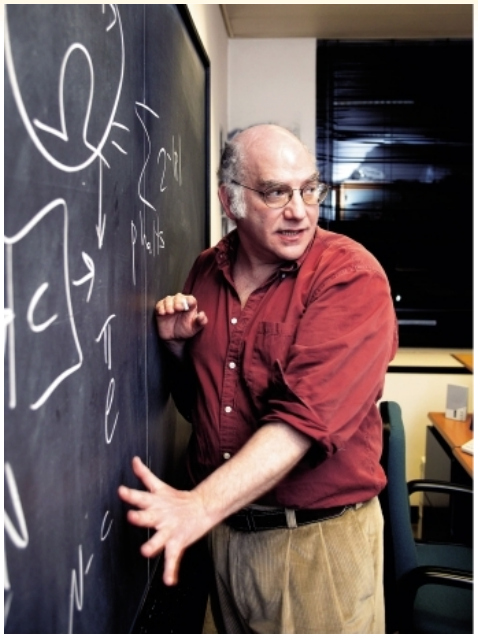
Definition

An infinite sequence $x = \{x_n\}$ is $\mathcal{K}$-random off there is c such that

$$\mathcal{K}(x_1 x_2 \cdots x_n) \geqslant n - c.$$

- sequences that could not be compressed
- linear complexity
- simultaneous discovery by Solomonov, Chaitin and Kolmogorov from1964 to 1967.
- coherent theory based on the Universal Turing Machine and the invariance theorem
- other definitions based on the same idea Chaitin-Levin 1966.

Gregory Chaitin (1947-)



Gregory John Chaitin (born 1947) is an Argentine-American mathematician and computer scientist.
Beginning in the late 1960s, Chaitin made important contributions to algorithmic information theory and metamathematics, in particular a new incompleteness theorem similar in spirit to Gödel's incompleteness theorem.

Chaitin has defined Chaitin's constant Ω, a real number whose digits are equidistributed and which is sometimes informally described as an expression of the probability that a random program will halt. Ω has the mathematical property that it is definable but not computable.
Chaitin's early work on algorithmic information theory paralleled the earlier work of Kolmogorov.
Chaitin also writes about philosophy, especially metaphysics and philosophy of mathematics (particularly about epistemological matters in mathematics). In metaphysics, Chaitin claims that algorithmic information theory is the key to solving problems in the field of biology (obtaining a formal definition of life, its origin and evolution) and neuroscience (the problem of consciousness and the study of the mind). Indeed, in recent writings, he defends a position known as digital philosophy. In the epistemology of mathematics, he claims that his findings in mathematical logic and algorithmic information theory show there are "mathematical facts that are true for no reason, they're true by accident. They are random mathematical facts". Chaitin proposes that mathematicians must abandon any hope of proving those mathematical facts and adopt a quasi-empirical methodology.
Chaitin's mathematical work is generally agreed to be correct, and has been cited, discussed and continued by many mathematicians. Some philosophers or logicians strongly disagree with his philosophical interpretation of it. Philosopher Panu Raatikainen argues that Chaitin misinterprets the implications of his own work and his conclusions about philosophical matters are not solid. The logician Torkel Franzén criticizes Chaitin's interpretation of Gödel's Incompleteness Theorem and the alleged explanation for it that Chaitin's work represents.
Chaitin is also the originator of using graph coloring to do register allocation in compiling, a process known as Chaitin's algorithm.

Leonid Levin (1948-)

Leonid Levin (born November 2, 1948, in Dnipropetrovsk USSR) is a computer scientist. He studied under Andrey Kolmogorov.
He obtained his first Ph.D. in 1972 at Moscow University. Later, he emigrated to the USA in 1978 and earned another Ph.D at the Massachusetts Institute of Technology in 1979.

He is well known for his work in randomness in computing, algorithmic complexity and intractability, foundations of mathematics and computer science, algorithmic probability, theory of computation, and information theory.
His life is described in a chapter in the book : Out of Their Minds : The Lives and Discoveries of 15 Great Computer Scientists.
Levin independently discovered a theorem that was also discovered and proven by Stephen Cook. This NP-completeness theorem, often called by inventors' names (see Cook-Levin Theorem) was a basis for one of the seven "Millennium Math. Problems" declared by Clay Mathematics Institute with a $ 1,000,000 prize offered. It was a breakthrough in computer science and is the foundation of computational complexity. Levin's journal article on this theorem was published in 1973 ; he had lectured on the ideas in it for some years before that time (see Trakhtenbrot's survey below), though complete formal writing of the results took place after Cook's publication.
He is currently a professor of computer science at Boston University, where he began teaching in 1980.

$\mathcal{K}$-random Sequences

Proposition (Almost all sequences are $\mathcal{K}$-random)

The amount of sequences x with length n and $\mathcal{K}(x) \geqslant n - c$ is lower boundered by

$$2^n(1 - 2^{-c})$$

Proof : the number of sequences with complexity $< n - c$ is

$$2^0 + 2^1 + \cdots + 2^{n-c-1} = 2^{n-c} - 1.$$

Show that a sequence is $\mathcal{K}$-random is, in general, undecidable

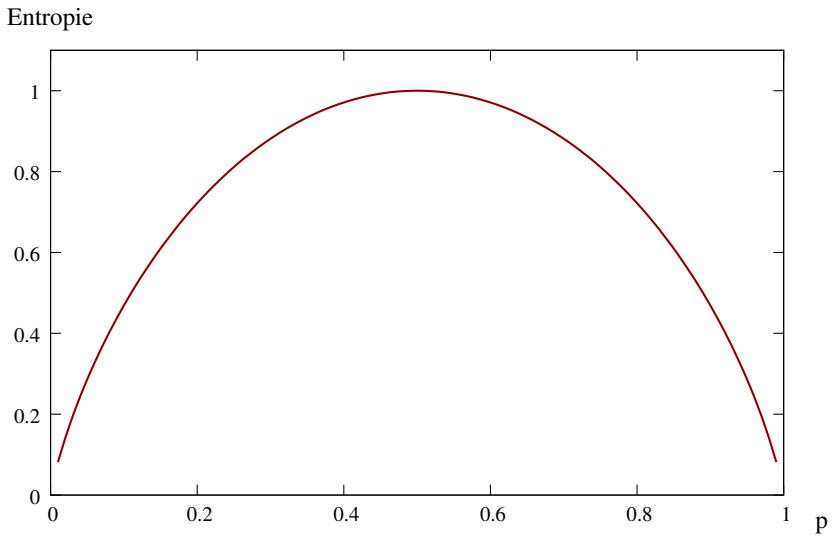
Biased coin tossing game complexity

$$x = \{x_n\}_{n\in\mathbb{N}} \qquad \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = p \neq \frac{1}{2}.$$

Proposition (Complexity and Entropy)

$$\mathcal{K}(x_1 x_2 \cdots x_n) \leqslant n[-p\log_2 p - (1-p)\log_2(1-p)] = n\mathcal{H}_2(p).$$

$\mathcal{H}_2(p)$ *is the* **entropy** *of the distribution* $(p, 1-p)$.



Biased coin tossing game : proof $p < \frac{1}{2}$

oOrder sequences of n bits by
1) increasing order of the number of "1" ;
2) lexicographical order on the sequence with the same number of "1".
Example : $n = 4$
0000 0001 0010 0100 1000 0011 0101 0110 1001 1010 1100 0111 1011 1101 1110 1111.
For this ordering Φ, give a sequence x of size n it is sufficient to provide the rank of the sequence in the ordering $r(x)$.

$$\mathcal{K}_\Phi(x) \sim \log_2 r(x); \quad \text{neglect } l(\Phi) \text{ and } \log_2 l(x) \ .$$

If k "1" in x then $r(x) \leqslant C_n^0 + C_n^1 + \cdots C_n^k$.
for $k < \frac{n}{2}$ and $\frac{k}{n} \overset{+\infty}{\sim} p$

$$\begin{aligned} \log_2 r(x) &\leqslant \log_2(C_n^0 + C_n^1 + \cdots C_n^k) \sim \log_2 C_n^k \text{ apply Stirling's formula} \\ &\sim \log_2(p^{-k}(1-p)^{-(n-k)}) \sim n\mathcal{H}_2(p). \end{aligned}$$

□

Entropie

Definition (Discrete entropy)

Let X a discrete random variable with probability distribution $p = (p_1, \cdots, p_K)$

$$\mathcal{H}(X) = \mathcal{H}(p) = \mathcal{H}(p_1, \cdots, p_K) = -\sum_{i=1}^{K} p_i \log p_i.$$

Structural properties

1. $\mathcal{H}(X) \geqslant 0$
2. $\mathcal{H}$ is maximal for $p = \frac{1}{K}$
3. $\mathcal{H}(X) = 0$ if for one i $p_i = 1$ and other $p_j = 0$ (deterministic system).
4. $\mathcal{H}(p_1, \cdots, p_K) = \mathcal{H}(p_1, \cdots, p_K, 0)$ extensibility property.

Proof : (1) $-x \log_2 x$ is non-negative on $[0, 1]$, (2) $\mathcal{H}$ is concave, (3) all terms in $\mathcal{H}$ are 0, then $p_i = 0$ or 1, p is a probability vector, all p_i are 0 except one 1, (4) clear.

Boltzman Interpretation (1877)

Modèle :

- n particles «undistinguishable» (n huge),
- K possible states for each particle (levels of energy),
- macro-state **observable**

$$\underline{n} = (n_1, \cdots, n_K);$$

Number of configurations represented by macro-state $\underline{n}$

$$\mathcal{C}(n, n_1, \cdots, n_k) = \frac{n!}{n_1! \cdots n_k!}.$$

Boltzman Interpretation (1877) (2)

Compare the probability of two macro-states $\underline{n}, \underline{n}'$

$$R(\underline{n}, \underline{n}') = \frac{\mathcal{C}(n, n_1, \cdots, n_k)}{\mathcal{C}(n, n'_1, \cdots, n'_k)} = \frac{\prod n'_i!}{\prod n_i!}.$$

For large n, $p_i = \frac{n_i}{n}$ (formule de Stirling)

$$R(\underline{n}, \underline{n}') \sim e^{n(\mathcal{H}(p) - \mathcal{H}(p'))}.$$

- One macro-état p^* will be observable (at the maximum of entropy)
- Exponential decreasing around p^*
- Without any constraints $p^* = \frac{1}{K}$ will be observed.
- With a constraint on energy average per particle E :

  $p_i^* = Ce^{-\lambda E_i}$ with C a normalization constant (partition function).

Ludwig Boltzmann (1844-1906)

Ludwig Eduard Boltzmann, physicien autrichien né le 20 février 1844 à Vienne (Autriche), mort le 5 septembre 1906 à Duino.
Il est considéré comme le père de la physique statistique, fervent défenseur de l'existence des atomes. Validant l'hypothèse de Démocrite selon laquelle la matière peut être considérée comme un ensemble d'entités indivisibles, Ludwig Boltzmann, à l'aide de son équation cinétique dite "de Boltzmann", théorise de nombreuses équations de mécanique des fluides.

Ludwig Boltzmann obtient son doctorat à l'université de Vienne en 1866, avec une thèse sur la théorie cinétique des gaz, dirigée par Josef Stefan, dont il devient ensuite assistant.
Il étudia successivement à Graz, Heidelberg et Berlin, où il suivit les cours de Bunsen, Kirchhoff et Helmholtz.
En 1869, il obtient une chaire de physique théorique à Graz, où il reste pendant 4 ans. En 1873, il accepte une chaire de mathématiques à Vienne, mais revient à Graz 3 ans plus tard, cette fois pour enseigner la physique expérimentale. Il devient membre étranger de la Royal Society en 1899.
Il entretint des échanges, parfois vifs, avec les physiciens à propos de ses travaux. Cela affecta particulièrement Boltzmann et entraina des crises de dépression qui l'ont conduit à une première tentative de suicide à Leipzig, puis à une seconde à Duino , près de Trieste, qui lui sera malheureusement fatale. Boltzmann meurt avant même d'avoir vu ses idées s'imposer.
Au Cimetière central de Vienne, la tombe de Boltzmann a une équation inscrite au-dessus de la statue du physicien. Cette épitaphe est l'équation $S = kln\omega$, laquelle exprime l'entropie S en fonction du nombre ω des états d'énergie équiprobables possibles, avec k la constante de Boltzmann.
Les conceptions atomistiques qui sont à la base des recherches de Boltzmann lui ont valu une vigoureuse hostilité de la part de ses confrères. Faute de développements nouveaux, ses résultats entraînèrent un certain discrédit sur ses travaux théoriques, jusqu'à ce que ceux-ci soient remis à l'honneur par les découvertes de Max Planck dans l'analyse du rayonnement du corps noir, puis celles d'Albert Einstein avec l'effet photoélectrique.

Conditional property

Let $X = (X_1, X_2)$ be a couple of random variables,
$p = (p_{i,j})$ on $[1, K] \times [1, L]$ be the joint distribution
$p_{i,.}$ and$p_{.,j}$ the marginal distributions.
Define the conditional distribution

$$\mathbb{P}(X_2 = j | X_1 = i) = \frac{p_{i,j}}{p_{i,.}}, \text{ and the conditional entropy knowing } X_1 = i$$

$$\mathcal{H}_{X_1=i}(X_2) = -\sum_j q_{i,j} \log q_{i,j}, \text{ and finally } \mathcal{H}_{X_1}(X_2) = -\sum_i p_{i,.} \mathcal{H}_{X_1=i}(X_2),$$

Theorem (Additivity property)

$$\mathcal{H}(X) = \mathcal{H}(X_1) + \mathcal{H}_{X_1}(X_2).$$

Conditional property

Proof :

$$\mathcal{H}(X) = -\sum_{i,j} p_{i,j} \log p_{i,j} = -\sum_{i,j} p_{i,.} q_{i,j} \log(p_{i,.} q_{i,j}) =$$

$$-\sum_{i=1}^{k} \left(p_{i,.} \log p_{i,.} \underbrace{\sum_{j=1}^{l} q_{i,j}}_{=1} \right) - \sum_{i=1}^{k} \left(p_{i,.} \underbrace{\sum_{j=1}^{l} q_{i,j} \log q_{i,j}}_{=\mathcal{H}_{X_1=i}(X_2)} \right) .$$

Information and Statistical independence

Theorem

A necessary and sufficient condition for X_1 and X_2 to be independent is

$$\mathcal{H}(X_1, X_2) = \mathcal{H}(X_1) + \mathcal{H}(X_2).$$

Proof :
$\Rightarrow$ X_1 and X_2 independent implies $q_{i,j} = p_{.,j}$ then $\mathcal{H}_{X_1}(X_2) = \mathcal{H}(X_2)$.
$\Leftarrow$

$$-\sum_{i=1}^{k} p_{i,.}\, q_{i,j} \log q_{i,j} \leqslant -\underbrace{\left(\sum_{i=1}^{k} p_{i,.}\, q_{i,j}\right)}_{=p_{.,j}} \log \left(\sum_{i=1}^{k} p_{i,.}\, q_{i,j}\right) = -p_{.,j} \log(p_{.,j}). \tag{1}$$

Corollary

For any vector $X = (X_1, X_2, \cdots, X_n)$

$$\mathcal{H}(X_1, X_2, \cdots X_n) \leqslant \sum_i \mathcal{H}(X_i),$$

with equality iff the X_i are mutually independent.

Uniqueness of Entropy

Theorem (Kintchine)

Let $\mathcal{H}$, defined for k and $p = (p_1, \cdots, p_k)$ probability distribution, satisfying :

(i.) *$\mathcal{H}$ is non-negative and continuous ;*

(ii.) *$\mathcal{H}(p_1, \cdots, p_k)$ is maximal for equirepartition*

(iii.) *If $(p_1, \cdots, p_k)$, $\{(q_{i,1}, \cdots, q_{i,l})\}_{1 \leqslant i \leqslant k}$ are $k+1$ distributions then, for $p_{i,j} = p_i q_{i,j}$,*

$$\mathcal{H}(p_{i,j}) = \mathcal{H}(p_i) + \sum_{i=1}^{k} p_i \mathcal{H}(q_{i,j}, \cdots, q_{i,l}).$$

(iv.) $$\mathcal{H}(p_1, \cdots, p_k, 0) = \mathcal{H}(p_1, \cdots, p_k).$$

then

$$\mathcal{H}(p_1, \cdots, p_k) = -C \sum_{i=1}^{k} p_i \log p_i,$$

for C arbitrary constant.

Aleksandr Yakovlevich Khinchin (1894-1959)

Aleksandr Yakovlevich Khinchin (Russian) was a Soviet mathematician and one of the most significant people in the Soviet school of probability theory. He was born in the village of Kondrovo, Kaluga Governorate, Russian Empire. While studying at Moscow State University, he became one of the first followers of the famous Luzin school. Khinchin graduated from the university in 1916 and six years later he became a full professor there, retaining that position until his death.

Khinchin's early works focused on real analysis. Later he applied methods from the metric theory of functions to problems in probability theory and number theory. He became one of the founders of modern probability theory, discovering the law of the iterated logarithm in 1924, achieving important results in the field of limit theorems, giving a definition of a stationary process and laying a foundation for the theory of such processes. Khinchin made significant contributions to the metric theory of Diophantine approximations and established an important result for simple real continued fractions, discovering a property of such numbers that leads to what is now known as Khinchin's constant. He also published several important works on statistical physics, where he used the methods of probability theory, and on information theory, queuing theory and mathematical analysis. (Wikipedia)

Coding

- Symbols $\mathcal{S} = \{s_1, \cdots, s_k\}$
- Code

$$\begin{array}{rcl} C : \mathcal{S} & \longrightarrow & \{0,1\}^* \\ s_i & \longmapsto & c(s_i) \text{ length } l_i \end{array}$$

- uniquely decodable (préfix property) ;
- **Kraft's Inequality** For a prefix code we have

$$\sum_{i=1}^{k} 2^{-l_i} \leqslant 1, \tag{2}$$

Reciprocally if (2) is satisfied for some $l = (l_1, \cdots, l_k)$ there is a prefix code with length l.

Proof of Kraft by a picture

Code complexity, Shannon's theorem

Random sources : $p_1, \cdots, p_k$;
average length of the code

$$L(c) = \sum_{i=1}^{k} p_i l_i;$$

$$L_{inf} = \inf_c L(c);\ \ c \text{ prefix code}$$

Theorem (Shannon 1948)

$$\mathcal{H}(p) \leqslant L_{inf} \leqslant \mathcal{H}(p) + 1.$$

$$\mathcal{H}(p) \leqslant L_{inf} \leqslant \mathcal{H}(p) + 1$$

Minimize $f(x_1, \cdots, x_k) = \sum_{i=1}^{k} p_i x_i$, under the constraint $\sum_{i=1}^{k} 2^{-x_i} \leqslant 1$.

The optimal point is $l_i^* = -\log_2 p_i$. At l^*

$$\sum_{i=1}^{k} 2^{-l_i^*} = \sum_{i=1}^{k} 2^{-(-\log_2 p_i)} = \sum_{i=1}^{k} p_i = 1$$

and

$$\sum_{i=1}^{k} p_i l_i^* = \sum_{i=1}^{k} p_i(-\log_2 p_i) = \mathcal{H}(p).$$

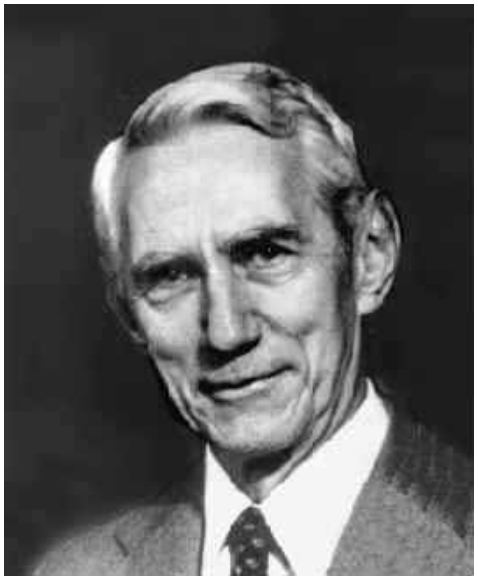
Then for all prefix code with length $l_1, \cdots, l_k$

$$\mathcal{H}(p) = f(l_1^*, \cdots, l_k^*) \leqslant f(l_1, \cdots, l_k) = L(h).$$

The reciprocal is based on the fact that $l_i^{sup} = \lceil l_i^* \rceil$, satisfies Kraft's inequality, and we have a prefix code, it's length

$$\sum_{i=1}^{k} p_i l_i^{sup} \leqslant \sum_{i=1}^{k} p_i(l_i^* + 1) = \mathcal{H}(p) + 1.$$

Claude Shannon (1916-2001)

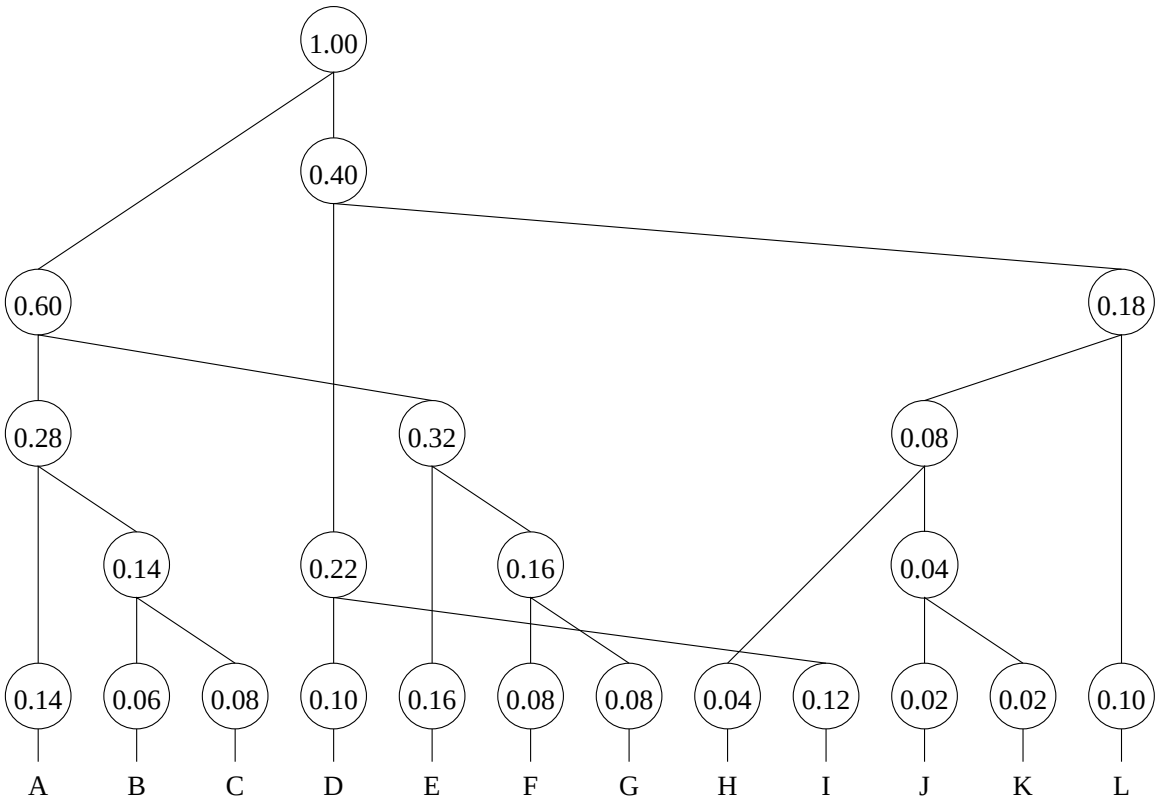


Claude Elwood Shannon (30 avril 1916 à Gaylord, Michigan - 24 février 2001), ingénieur électrique, est l'un des pères, si ce n'est le père fondateur, de la théorie de l'information. Son nom est attaché à un célèbre "schéma de Shannon" très utilisé en sciences humaines, qu'il a constamment désavoué.

Il étudia le génie électrique et les mathématiques à l'Université du Michigan en 1932. Il utilisa notamment l'algèbre booléenne pour sa maîtrise soutenue en 1938 au MIT. Il y expliqua comment construire des machines à relais en utilisant l'algèbre de Boole pour décrire l'état des relais (1 : fermé, 0 : ouvert).
Shannon travailla 20 ans au MIT, de 1958 à 1978. Parallèlement à ses activités académiques, il travailla aussi aux laboratoires Bell de 1941 à 1972.
Claude Shannon était connu non seulement pour ses travaux dans les télécommunications, mais aussi pour l'étendue et l'originalité de ses hobbies, comme la jonglerie, la pratique du monocycle et l'invention de machines farfelues : une souris mécanique sachant trouver son chemin dans un labyrinthe, un robot jongleur, un joueur d'échecs (roi tour contre roi)...
Souffrant de la maladie d'Alzheimer dans les dernières années de sa vie, Claude Shannon est mort à 84 ans le 24 février 2001.

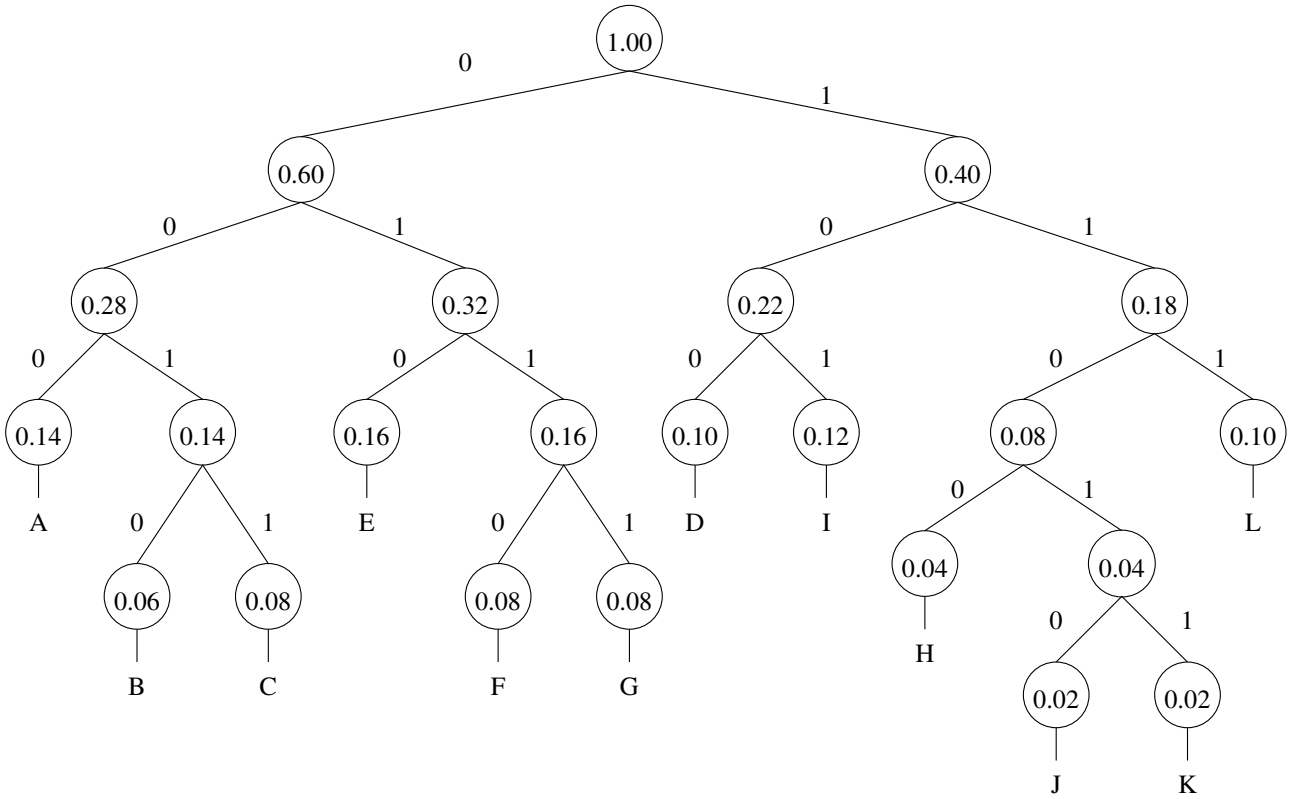
Wikipedia

Huffman's algorithm (1951)



A	0.14
B	0.06
C	0.08
D	0.10
E	0.16
F	0.08
G	0.08
H	0.04
I	0.12
J	0.02
K	0.02
L	0.10

Algorithme de Huffman (1951)



A	0.14	000
B	0.06	0010
C	0.08	0011
D	0.10	100
E	0.16	010
F	0.08	0110
G	0.08	0111
H	0.04	1100
I	0.12	101
J	0.02	11010
K	0.02	11011
L	0.10	111

Optimal code : L-inf = 3.42, Entropy = 3.38

Height = $-\log_2$(probability)

Generalization Lempel-Ziv(1978) and Lempel-Ziv-Welsh(1984)

Huffman's algorithm (1951) : Implantation

HUFFMAN_ALGORITHM (p)
Data: Set of k symbols $\mathcal{S}$ and weight p
Result: Optimal prefix code
Node x,y,z
Priority_Queue F
for $s \in \mathcal{S}$
  z=new_node(p(s),/,/)
  `Insert` (F,z)
for $i = 1$ **to** $K - 1$
  x=`Extract` (F)
  y=`Extract` (F)
  z=new_node(x.p+y.p,x,y)
  `Insert` (F,z)
`Return Extract` (F)

Huffman's algorithm (1951) : Proof

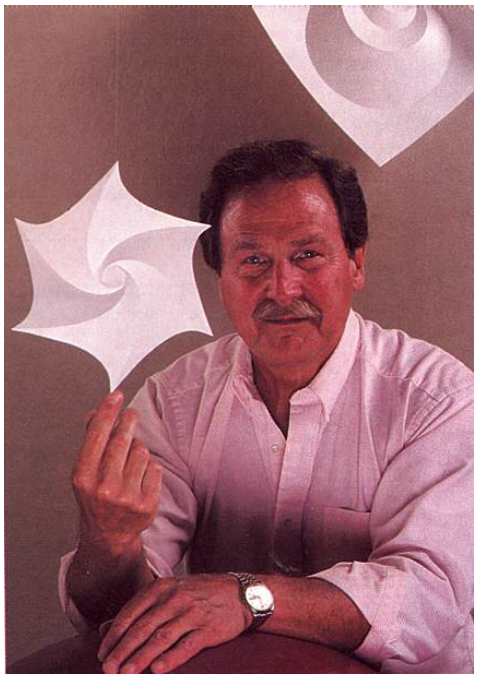
Optimality

Huffman's algorithm provides an optimal prefix code

At any time the priority queue contains a sub-forest of an optimal tree
The complexity of the algorithm is $\mathcal{O}(k \log k)$
Proof by 2 pictures

David A. Huffman (1925-1999)

Avec ses sculptures



Huffman joined the faculty at MIT in 1953. In 1967, he went to University of California, Santa Cruz as the founding faculty member of the Computer Science Department. He played a major role in the development of the department's academic programs and the hiring of its faculty, and served as chair from 1970 to 1973. He retired in 1994, but remained active as an emeritus professor, teaching information theory and signal analysis courses.
Huffman made important contributions in many other areas, including information theory and coding, signal designs for radar and communications applications, and design procedures for asynchronous logical circuits. As an outgrowth of his work on the mathematical properties of "zero curvature Gaussian" surfaces, Huffman developed his own techniques for folding paper into unusual sculptured shapes (which gave rise to the field of computational origami).

http://www.huffmancoding.com/my-family/my-uncle/scientific-american

Synthesis

- Control complexity : a nice dream
- Coherent theory : link between probability theory and computation theory
- Computable approximation of complexity : Entropy
- Coding application (prefix optimal codes)

To go further ...

References

Cover, T. and Thomas, J. (2006), *Elements of Information Theory* 2nd Edition, Wiley-Interscience
Shannon, C. (1948), *A mathematical theory of communications*, Bells Systems Technical Journal.
Li, M. et Vitananyi, P. (1990), *Kolmogorov Complexity and its Applications*, Elsevier Science Publisher, chapter 4, pp. 189-254.
Delahaye, J.-P. (1999), Information, complexité et hasard, Hermes.

Internet links biographies et les photos)

`fr.wikipedia.org`
`http://www-groups.dcs.st-and.ac.uk/~history/`
Chaitin's web page : `http://www.cs.auckland.ac.nz/~chaitin`
Levin's web page : `http://www.cs.bu.edu/~lnd`
Delahaye's web page : `http://www2.lifl.fr/~delahaye/`

I've lost some web references, with great pleasure I'll add them or suppress some part of texts according the author's decision.