COMPLEXITÉ DE KOLMOGOROV

J'AI LANCÉ UNE PIÈCE 20 FOIS ET J'AI OBTENU :

① 10 10 10 10 10 10 10 10 10 10

QU'EN DITES VOUS ? VOUS NE ME CROYEZ PAS.

② 11 10 01 10 11 00 10 00 11 10

QU'EN DITES VOUS ? PEUT-ÊTRE ...

POURQUOI ? PROBABILITÉ IDENTIQUE : 2^{-20}.

LA THÉORIE CLASSIQUE DES PROBABILITÉS NE CAPTURE PAS NOTRE NOTION INTUITIVE DE "ALÉATOIRE".

REMARQUE : ON PEUT GÉNÉRER ① AVEC UN PROGRAMME "RÉPÉTER '10' 10 FOIS" DE LONGUEUR $c + \log_2(10)$, MAIS POUR ② CELA SEMBLE IMPOSSIBLE ... INCOMPRESSIBLE ?

LA COMPLEXITÉ DE KOLMOGOROV EST UNE DES FAÇONS DE DÉFINIR L'INFORMATION, ET LA QUANTIFIER : LA PLUS PETITE DESCRIPTION ALGORITHMIQUE.

DÉFINITION : SOIT $x \in \{0,1\}^*$. LA DESCRIPTION MINIMALE DE x, NOTÉE $d(x)$, EST LE PLUS PETIT MOT $(\langle M \rangle, w)$ AVEC M UNE M.T. QUI SUR L'ENTRÉE w, S'ARRÊTE AVEC x ÉCRIT SUR SON RUBAN. LA COMPLEXITÉ DE KOLMOGOROV DE x EST $K(x) = |d(x)|$.

TAILLE + LEX

ENCODAGE DES COUPLES (u,v), PAR EXEMPLE $\text{double}(|u|)\,01\,uv$ EN $2\log_2(|u|) + 2 + |u| + |v|$ BITS.

EXERCICE : UN ENCODAGE ASYMPTOTIQUEMENT MEILLEUR ?

$\text{double}(\log_2(|u|))\,01\,|u|\,uv$

EN $2\log_2(\log_2(|u|)) + 2 + \log_2(|u|) + |u| + |v|$ BITS.

DÉFINITION ALTERNATIVE : $K'(x) = \min\{|\langle M \rangle| \mid M(\varepsilon) = x\}$ MAIS LA DÉFINITION AVEC $(\langle M \rangle, w)$ EST PRATIQUE POUR HARD-CODER DES VALEURS AVEC M CONSTANTE.

THM : $\exists c, \forall x, K'(x) \leq c\,K(x) + c$ IDÉE : SUR ε ON AJOUTE $|w|$ TRANSITIONS POUR ÉCRIRE w

$\forall x, K(x) \leq K'(x) + 2\log_2(K'(x)) + 2$ IDÉE : $\langle M \rangle \mapsto (\langle M \rangle, \varepsilon)$.

THM : $\exists c : \forall x : K(x) \leq |x| + c$.

PREUVE : AVEC LA M.T. M QUI S'ARRÊTE IMMÉDIATEMENT SUR TOUTE ENTRÉE ET $w = x$. □

THM : $\exists c : \forall x : K(xx) \leq K(x) + c$.

PREUVE : SOIT LA M.T. M QUI, SUR L'ENTRÉE $(\langle N \rangle, w)$:

1. SIMULE N SUR L'ENTRÉE w (S'ARRÊTE AVEC x SUR LE RUBAN)
2. DUPLIQUE LE CONTENU DU RUBAN (S'ARRÊTE AVEC xx SUR LE RUBAN)

UNE DESCRIPTION DE xx EST $(\langle M \rangle, d(x))$. □

COMPLEXITÉ DE LA CONCATÉNATION.

THM : $\exists c, \forall x, y, \; K(xy) \leq 2 \cdot \log_2(K(x)) + K(x) + K(y) + c$.

PREUVE : $(\langle M \rangle, (d(x), d(y)))$ AVEC M LA M.T. QUI SIMULE $d(x)$ PUIS $d(y)$ POUR ÉCRIRE xy. □

THM : $\forall c, \exists x, y : K(xy) > K(x) + K(y) + c$.

PREUVE : ON A BESOIN DU RÉSULTAT SUIVANT :

THM : $\forall k \in \mathbb{N}$, POUR TOUT $x \in \{0,1\}^*$ SUFFISAMMENT LONG IL EXISTE UN PRÉFIXE $y \sqsubseteq x$ TEL QUE $K(y) < |y| - k$.

PREUVE : SOIT f UNE BIJECTION STANDARD (LONGUEUR-LEX) ENTRE $\{0,1\}^*$ ET $\mathbb{N}$.
SOIT $z \sqsubseteq x$ ET $f(z) = n$.
SOIT y L'EXTENSION DE LONGUEUR n DE z SELON x,
C'EST-À-DIRE $y = z\sigma \sqsubseteq x$ AVEC $|\sigma| = n$.
IL EXISTE UNE M.T. M TELLE QUE $M(\sigma) = z\sigma$, EN UTILISANT $f^{-1}(|\sigma|) = z$.
DONC $K(y) \leq |\sigma| + c$ AVEC c INDÉPENDENTE DE y (DE z ET DE σ).
EN PRENANT $|z| > k + c$ ON OBTIENT
$K(y) \leq |\sigma| + c = |y| - |z| + c < |y| - k - c + c = |y| - k$. ◇

SOIT c' TELLE QUE $\forall x : K(x) \leq |x| + c'$.
SOIT z SUFFISAMMENT LONG ET VÉRIFIANT $K(z) \geq |z|$ (INCOMPRESSIBLE, EXISTENCE PAGE SUIVANTE)
SOIT $k = c + c'$. PAR LE THM IL EXISTE $x \sqsubseteq z$ TEL QUE $K(x) < |x| - c - c'$.
ALORS POUR y TEL QUE $z = xy$ ON A :
$K(x) + K(y) + c < \underbrace{|x| - c - c'} + \underbrace{|y| + c'} + c = |x| + |y| = |z| \leq K(z) = K(xy)$. □

OPTIMALITÉ DE LA DESCRIPTION

DÉF : $K_P(x)$ LA COMPLEXITÉ DE KOLMOGOROV RELATIVE AU LANGAGE DE PROGRAMMATION P. (DE DESCRIPTION)

THM : $\forall P : \exists c : \forall x : K(x) \leq K_P(x) + c$.

PREUVE : IL FAUT QUE LA M.T. INTERPRÈTE (COMPILE + EXÉCUTE) LE PROGRAMME EN P. □

K N'EST PAS CALCULABLE

THM : $K : \{0,1\}^* \to \mathbb{N}$ N'EST PAS CALCULABLE.

PREUVE AVEC AUTO-RÉFÉRENCE : $Y(\lambda f.\ \text{PROGRAMME})$

PAR L'ABSURDE, SUPPOSONS QUE K SOIT CALCULABLE.
ALORS ON PEUT CONSTRUIRE M QUI, SUR TOUTE ENTRÉE :

1. OBTIENT SON PROPRE CODE $\langle M \rangle$
2. ÉNUMÈRE LES MOTS BINAIRES JUSQU'À TROUVER x TEL QUE $K(x) > |\langle M \rangle|$
3. ÉCRIT x ET S'ARRÊTE. (ou $K(x) > 2\log_2(|\langle M \rangle|) + 2 + |\langle M \rangle|$)

ALORS M DÉCRIT x MAIS EST PLUS PETITE QUE $K(x)$. $\lightning$ □
($\langle M \rangle, \varepsilon$)

PREUVE SANS AUTO-RÉFÉRENCE :

PAR L'ABSURDE, SUPPOSONS QUE K SOIT CALCULABLE.
ALORS ON PEUT CONSTRUIRE M QUI, SUR L'ENTRÉE w :

1. ÉNUMÈRE LES MOTS BINAIRES JUSQU'À TROUVER x TEL QUE $K(x) > 2.|w|$.
2. ÉCRIT x ET S'ARRÊTE.

ALORS $(\langle M \rangle, w)$ DÉCRIT x MAIS LA TAILLE DE CETTE DESCRIPTION EST $c + |w|$ (MACHINE M)
DONC POUR w SUFFISAMMENT GRAND ON A $|(\langle M \rangle, w)| = c + |w| \leq 2|w| < K(x)$. $\lightning$ □
($|w| \geq c$)

THM : $K : \{0,1\}^* \to \mathbb{N}$ EST APPROXIMABLE "PAR AU-DESSUS" :
$\exists (K_i : \{0,1\}^* \to \mathbb{N})_{i \in \mathbb{N}}$ CALCULABLES TELLES QUE $\forall x : \lim_{i \to +\infty} K_i(x) = K(x)$ ET $\forall i : K_i(x) \geq K(x)$.

PREUVE : SOIT c TELLE QUE $\forall x : K(x) \leq |x| + c$.
K_i EXÉCUTE TOUTES LES DESCRIPTIONS DE TAILLE $< |x| + c$ POUR i ÉTAPES DE TEMPS, ET DONNE EN SORTIE LA TAILLE DE LA PLUS COURTE QUI A DONNÉ x EN SORTIE (SI AUCUNE ALORS $|x| + c$). □

MOTS INCOMPRESSIBLES.

DEF : x EST c-COMPRESSIBLE LORSQUE $K(x) \leqslant |x| - c$.

x EST INCOMPRESSIBLE = 1-INCOMPRESSIBLE = $K(x) \geqslant |x|$

(PAS DE DESCRIPTION PLUS PETITE QUE LUI-MÊME).

THM. IL EXISTE DES MOTS BINAIRES INCOMPRESSIBLES DE TOUTE TAILLE.

PREUVE : COMPTAGE POUR UNE TAILLE n :

NOMBRE DE MOTS BINAIRES DE TAILLE n : $= 2^n$

NOMBRE DE DESCRIPTIONS DE LONGUEUR $< n$: $\leqslant \sum_{i=0}^{n-1} 2^i = 2^n - 1$.

OR CHAQUE DESCRIPTION NE DÉCRIT QU'UN SEUL MOT. □

EXERCICE : COMBIEN DE MOTS BINAIRES DE TAILLE n SONT c-INCOMPRESSIBLES ?

AU MOINS $2^n - 2^{n-c+1} + 1$.

SOIT $L_K = \{ x \mid K(x) \geqslant |x| \}$ LES INCOMPRESSIBLES.

EXERCICE : AUCUN SOUS-ENSEMBLE INFINI DE L_K (DONT L_K LUI-MÊME) N'EST SEMI-DÉCIDABLE.

PREUVE : PAR L'ABSURDE, SOIT M QUI SUR L'ENTRÉE w :

1. ÉNUMÈRE DES INCOMPRESSIBLES JUSQU'À TROUVER UN TEL x AVEC $|x| > 2|w|$,
2. ÉCRIT x ET S'ARRÊTE.

ALORS $(\langle M \rangle, w)$ DÉCRIT x ET LA TAILLE DE CETTE DESCRIPTION EST $c + |w|$.

DONC POUR $|w| \geqslant c$ ON AURA $|(\langle M \rangle, w)| = c + |w| \leqslant 2|w| < |x| \leqslant K(x)$. ↯ □

THM : LES DESCRIPTIONS MINIMALES SONT c-INCOMPRESSIBLES :

$\exists c : \forall x \in \{0,1\}^* : K(d(x)) \geqslant |d(x)| - c$.

PREUVE : SOIT M LA MT QUI, SUR L'ENTRÉE $(\langle R \rangle, y)$, CALCULE $R(y) = (\langle S \rangle, z)$ (REJETTE SI PAS DE LA BONNE FORME) PUIS $S(z)$ ET LAISSE CE RÉSULTAT SUR SON RUBAN.

SOIT $c = 2\log_2(|\langle M \rangle|) + |\langle M \rangle| + 3$.

PAR L'ABSURDE, SI $d(x)$ EST c-COMPRESSIBLE ALORS $|d(d(x))| \leqslant |d(x)| - c$

MAIS $(\langle M \rangle, d(d(x)))$ DÉCRIT x AVEC LONGUEUR

$\underbrace{2\log_2(|\langle M \rangle|) + 2 + |\langle M \rangle|}_{c-1} + |d(d(x))| \leqslant c - 1 + |d(x)| - c = |d(x)| - 1$

CE QUI CONTREDIT LA MINIMALITÉ DE $d(x)$. ↯ □

APPLICATION À LA DENSITÉ DES NOMBRES PREMIERS

IDÉE : LA DÉCOMPOSITION EN FACTEURS PREMIERS PERMET DE REPRÉSENTER SUCCINCTEMENT CERTAINS ENTIERS (PETITS BINAIRES) MAIS LES INCOMPRESSIBLES RESTENT INCOMPRESSIBLES.

DEF : SOIT $\pi(n)$ LE NOMBRE D'ENTIERS PREMIERS $\leq n$. $\pi(11) = 5$.

THM : $\pi(n) \geq \dfrac{\log n}{\log\log n} - o(1)$ AVEC $o(1)$ UNE FONCTION QUI TEND VERS 0 QUAND $n \to +\infty$.

PREUVE : SOIT $n \in \mathbb{N}$ AVEC $n = p_1^{e_1} \cdot p_2^{e_2} \cdots p_m^{e_m}$ SA DÉCOMPOSITION EN FACTEURS PREMIERS.

ALORS ON PEUT REPRÉSENTER n AVEC $(\langle M \rangle, (e_1, \dots, e_m))$ où :

- $e_i \leq \log n$ DONC $\log\log n$ BITS DONC $(e_1, \dots, e_m)$ PEUT ÊTRE ENCODÉ PAR $0^{\log\log\log n} 1\, e_1 e_2 \dots e_m$ (POUR DÉCODER LA SUITE)
- M EST UNE MACHINE CONSTANTE QUI DÉCODE $e_1, \dots, e_m$, CALCULE $p_1, \dots, p_m$, PUIS n.

DONC $K(n) \leq c + m \cdot \log\log n + \log\log\log n$

POUR n INCOMPRESSIBLE ON A $K(n) \geq \log n$ D'OÙ

$$\log n \leq c + m \cdot \log\log n + \log\log\log n$$

$$\Rightarrow \frac{\log n - c - \log\log\log n}{\log\log n} \leq m$$

$$\Rightarrow \frac{\log n}{\log\log n} - o(1) \leq m \leq \pi(n). \quad \square$$

NB : HADAMARD ET DE LA VALLÉE POUSSIN 1896 : $\pi(n) \sim \dfrac{n}{\ln n}$.

THM : $\forall t \in \mathbb{N}, \exists x \in \{0,1\}^* : K(x) > t$.

PREUVE : SOIT $t \in \mathbb{N}$.

IL Y A $\leq 2^t$ PROGRAMMES DE TAILLE $\leq t$.

CHAQUE PROGRAMME DÉCRIT AU PLUS UNE NOUVELLE CHAÎNE,

DONC CES $\leq 2^t$ PROGRAMMES DÉCRIVENT $\leq 2^t$ CHAÎNES.

C'EST UNE QUANTITÉ FINIE DONC $\exists x$ QUI N'EST

PAS PARMI ELLES. CETTE CHAÎNE x N'EST DÉCRITE

PAR AUCUN PROGRAMME DE TAILLE $\leq t$,

C'EST-À-DIRE $K(x) > t$. $\square$