

# Théorie et théories de l'Information

Qq éléments de :

**Information, complexité et hasard** *Jean-Paul Delahaye*

revus par *Pierre Bernhard* et

disponibles aussi sur **)i(nterstices**

Oct08

# Vous avez dit information ?

Dans la suite :

00000 00000 00000 00000 00000 00000 00000 00000

il y a sûrement moins d'information que dans la suite :

26535 89793 23846 26433 83279 50288 41971 69399

qui est plus *complexe* mais . .

**pourquoi ?**

Serait-ce que la 2ème suite  
est (complexité aléatoire) :

*sans ordre, sans régularité, aléatoire, chaotique ?*

ou (complexité organisée)

*organisé, fortement structuré, riche en information ?*

# Vous avez dit information ?

Il se trouve que :

26535 89793 23846 26433 83279 50288 41971 69399  
est la suite des décimales de  $\pi$  à partir de la sixième (après 14159).

Distinguons donc bien :

- Un contenu **brut** en information qui ne dépend pas du *sens*.
- La **valeur** d'une information dépend du *but* fixé.

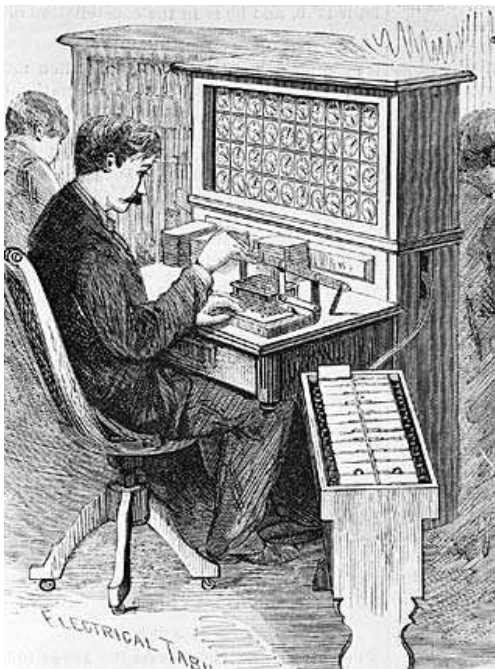
avec deux grandes théories (selon le but) :

- **Théorie probabiliste de l'information** (Shannon / thermodynamique)
- **Théorie algorithmique de l'information** (Kolmogorov / Bennett)

en lien avec :

l'informatique théorique, les neurosciences computationnelles etc..

# Vous avez dit information ?



# Contenu brut en information

-1- Le contenu brut en information d'un booléen

(1/0 vrai/faux oui/non yin/yang) 1.

C'est l' "atome" d'information : le "bit".

-2- Le contenu en information d'une valeur  $v \in \{1, N\}$

(digits ( $N = 10$ ), lettre de l'alphabet ( $N = 26$ ), ..) est  $\log_2(N)$ .

C'est le nombre de bits pour "coder" la valeur, en cohérence avec -1-

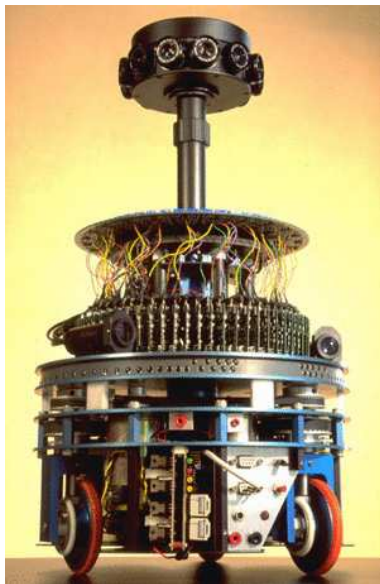
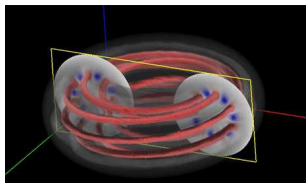
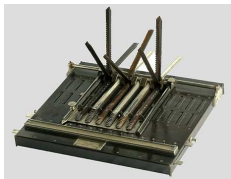
-3- Le contenu brut en information de  $n$  éléments "indépendants" est la *somme* des contenus bruts.

mais inférieure si il y a de la *redondance* (compression)

mais sans *valeur* si il n'y a aucun lien entre les éléments!

- C'est le nombre minimal de questions oui/non à poser pour trouver la valeur donnée!

# Vous avez dit information ?



# Contenu probabiliste en information

Pour le contenu probabiliste en information qui :

1/ N'est fonction que de la probabilité des évènements.

2/ Est additif pour deux sources indépendantes.

3/ Croît linéairement avec le nombre de réponses équiprobables.

On parle d'*entropie* au sens de Shannon et obtient une définition :

$$\mathcal{H}(p : \{1..n\} \rightarrow [0, 1]) = - \sum_n p(n) \log_2 (p(n))$$

*contenu moyen en information d'une distribution de probabilité.*

# Contenu probabiliste en information

- Pour une *distribution binaire*  $p : \{0, 1\} \rightarrow [0, 1]$ 
  - >  $\mathcal{H} = 1$  si  $p(0) = p(1) = 1/2$  : hasard complet, non-redondance
  - >  $\mathcal{H} = 0$  si  $p(0) = 1, p(1) = 0$  : valeur fixe, rien à découvrir
- Pour une *distribution uniforme*  $p : \{0, N\} \rightarrow [0, 1], p(i) = 1/N$ .
  - >  $\mathcal{H}$  est maximale
  - >  $\mathcal{H} = \log_2(N)$  : nb de bits pour coder la valeur
- Au delà :
  - >  $\mathcal{H}$  est une valeur positive, bornée.
  - > Elle se définit dans le cas continu à une *résolution donnée*.



# Utilisation de l'information probabiliste : estimation

Que veut dire . . . estimer une valeur incertaine ?

- Si je sais rien de rien, j'peux qu'agir au hasard . . .
- Si je sais quelque chose, c'est moins aléatoire !

On cherche alors  $p : \{0, N\} \rightarrow [0, 1]$ ,  $\sum_n p(n) = 1$  t.q. :  
 $\sum_n p(n) f_i(n) = v_i, i \in \{1, m\}$  (contraintes),  $\max_p \mathcal{H}(p)$  (entropie)

Une solution vérifie à minima les contraintes (rasoir d'Occam) :

$$p(n) = \frac{1}{Z} e^{-\sum_i \lambda_i f_i(n)}$$

pour des "multiplicateurs"  $\lambda_i$  donnés par les contraintes (Gibbs)

Si on donne moyenne/écart-type d'un réel, c'est la Gaussienne :

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}}$$

Si on donne le rythme moyen d'une série de "bips", c'est le Poisson

$$p(d) = r e^{-r} d, P\{N \text{ tops pendant } T\} = \frac{(rT)^N}{N!} e^{-rT}$$

# Utilisation de l'information probabiliste : transmission

Quelle information véhiculée par une réponse  $\mathbf{r}$  à propos d'un stimulus  $\mathbf{s}$  ?

- Une source  $s$ , grâce à un codage, envoie un message à un récepteur  $r$   
→ qui effectue le décodage dans un contexte perturbé de bruit.

On utilise la probabilité conditionnelle  $p(\mathbf{r}|\mathbf{s})$  :

- > Si  $p(\mathbf{r}|\mathbf{s}) = p(\mathbf{r}) p(\mathbf{s})$  les deux n'ont rien à voir (indépendance).
- > Si  $p(\mathbf{r}|\mathbf{s}) = p(\mathbf{s})$  les deux sont fonctionnellement liées.

On définit alors l'information mutuelle (dépendance statistique)

$$\begin{aligned} \mathcal{I}(\mathbf{r}, \mathbf{s}) &= \mathcal{H}[\text{réponse}] - \mathcal{H}[\text{variabilité non due à la source}] \\ &= \mathcal{H}[\text{réponse}] - \mathcal{H}[\text{réponse}|\text{source}] \\ &\leq \mathcal{H}[\text{réponse}] + \mathcal{H}[\text{source}] \end{aligned}$$

symétrique, positive, etc..

Qui utilise la règle de Bayes :

$$p(\mathbf{r}|\mathbf{s}) = p(\mathbf{r}, \mathbf{s})/p(\mathbf{s}) = p(\mathbf{s}|\mathbf{r}) p(\mathbf{r})/p(\mathbf{s})$$

# Utilisation de l'information probabiliste : autres applications

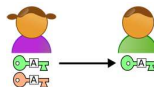


- Codage de l'information

- Compression de données

- Cryptographie

- Corrections d'erreurs



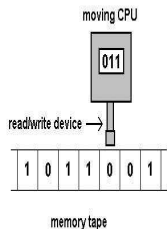
*Une variable d'entropie  $\mathcal{H}(x)$  sera "presque sûrement" codée avec  $\mathcal{H}(x)$  bits*



# Notion de machine universelle

**Thèse de Church-Turing** : Tout problème “algorithmique” peut être résolu par une machine de Turing.

- 0- Un ruban plein de '0' où on peut écrire des '1'
- 1- Une tête qui peut lire/écrire des 0/1
- 2- Et se déplacer vers la gauche ou droite
- 3- Une table d'actions
  - . qui définit en fonction du symbole lu :
  - . – quoi écrire,
  - . – comment se déplacer,
  - . – et quelle nouvelle action.



- Tou(te) *fonction calculable, programme/automate* se réduit à “ça”
- Les actions peuvent être codées sur un autre ruban pour être jouées :  
machine universelle

## Notion de fonctions non calculables

- Les fonctions calculables sont dénombrables (elles sont définies par un ruban) mais les fonctions sur un domaine infini contiennent  $\mathcal{R}$
- Il y a plein d'exemples "utiles" :
  - Pas de programme universel qui en temps fini, renvoie « oui » un programme va finir et « non » s'il va "boucler sans fin".et par déduction (Th. de Rice) :
  - \* « le programme ne termine pas par une erreur d'exécution »
  - \* « le programme calcule le spécifié »
- La complexité de Kolmogorov n'est pas calculable :  
Car si  $K(n)$  est calculable, alors le programme  

```
for(int n = 1; K(n) <= x; n++); print n;
```

qui est de longueur  $L = O(\log_2(x))$  génère un entier  $n$  donc la complexité est au plus  $L$ .  
On peut alors montrer avec  $x$  suffisamment grand on obtient  $n > L$  : contradiction !

# A propos d'échelles de complexité

## - **Récurusif** : *effectif* :

on peut déterminer son contenu par un moyen sans ambiguïté.

Classes L et P : temps de calcul logarithmique ou polynomial

Classe NP : résolu en énumérant les solutions possibles

(classifications selon temps de calcul et espace mémoire).

## - **Récurсивement énumérable** : *constructif* :

il se précise au fur et à mesure du temps, mais à chaque instant impossible de le cerner complètement.

## - **Non récurсивement énumérable** : *prospectif* :

accessible à l'intelligence humaine, mais rien ne l'épuise et rien ne pourra jamais l'épuiser.

*Approximables* : par une suite croissante d'ensembles récurсивement énumérable.

*Immunes* : on ne peut en décrire récurсивement que des parties finies.

*Incompressibles* : on ne peut aller plus loin que les axiomes de base.

# Théorie algorithmique de l'information

La complexité de Kolmogorov  $K(m)$  :

La **longueur** du plus petit programme écrit pour une machine universelle qui génère le message.

- Elle est bornée : au pire avec : `print m` et non additive.
- Elle capture la complexité aléatoire de l'information
- On peut la relier au notions d'  
*incompressibilité = imprévisibilité = absence de structure*
- L'entropie de Shannon correspond à la complexité de Kolmogorov moyenne (Grünwald Vitányi 2004)



# Théorie algorithmique de l'information

La profondeur logique de Bennet  $P(m)$  :

Le **temps de calcul** du plus petit programme écrit pour une machine universelle qui génère le message.

et non le temps de calcul du programme le plus rapide : c'est  $\text{print } m!$

- Capture la notion de complexité organisée :

simple et peu profond :	cristal
simple et profond :	$\pi$ à $10^{-10}$ près
aléatoire et peu profond :	gaz parfait
aléatoire et profond :	être vivant

- Approximativement invariant à la machine (si la longueur est définie à  $k$  près)
- Non additivité (deux moutons ne sont pas deux fois plus complexes qu'un)
- Croissance lente (la profondeur augmente avec du calcul, mais lentement)
- Apparition "spontanée" de complexité organisée par le calcul
- Non calculabilité

Vous avez dit information ?

