Théorie de l'Information

Philippe Duchon

ENSEIRB

2008-09

“Théorie de l’information”...

- Définir proprement la *quantité d’information* apportée par la connaissance d’un fait
- Idée (très) informelle : c’est le *nombre de questions à réponse oui/non que cette connaissance nous évite d’avoir à poser*

Ce dont on va parler. . .

- **Notions mathématiques :** entropie, information mutuelle
- **Codage et décodage :** codes, déchiffrabilité, efficacité d'un code
- **Transmission de l'information :** comment transmettre correctement une information en présence d'erreurs de transmission
- **Compression des données :** essentiellement, compression sans pertes ("conservative")
- Le langage mathématique est celui de la **théorie des probabilités :** variables aléatoires et processus discrets (élémentaires)

Ce dont on ne va *pas* parler...

- **Cryptographie :** cours en deuxième année
- On s'intéresse à *transmettre* de l'information, pas à la *cacher* ou à la *protéger*
- Les erreurs sont le fait d'imperfections du système de transmission, pas d'interférences malveillantes d'un *adversaire*

Historiquement

- **Rudolf Clausius** (1822-1888) : définit l'entropie comme une mesure du *désordre* d'un système
- **Ludwig Boltzmann** (1844-1906) : l'entropie d'un état macroscopique est proportionnelle au *logarithme* du nombre d'états microscopiques correspondants
- **Ronald Fisher** (1890-1962) : utilise le mot *information* dans un contexte mathématique
- **Harry Nyquist** (1889-1976) (bruit, fréquence d'échantillonnage), **Ralph Hartley** (1888-1970)
- **Claude Shannon** (1916-2001), premiers théorèmes sur l'information en théorie de la communication

Quantité d'information

- $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $A \in \mathcal{A}$ un événement (de probabilité non nulle)
- On définit la **quantité d'information** apportée par la réalisation de A (**self-information**), la quantité

$$I(A) = -\log \mathbb{P}(A).$$

- On utilise quasi systématiquement le *logarithme à base* 2, l'unité d'information est alors le **bit**.

Événements indépendants

- Si (et seulement si) les événements A et B sont **indépendants**,

$$I(A \cap B) = I(A) + I(B).$$

- "La quantité d'information apportée par la *réalisation conjointe* de deux événements indépendants, est égale à la *somme* de celles apportées séparément par chacun d'eux."

Retour sur la définition

La propriété d'additivité en cas d'indépendance implique l'usage de la fonction logarithme :

Proposition

Si on souhaite définir la quantité d'information par une fonction positive de la probabilité : $I(A) = f(\mathbb{P}(A))$, avec $f : [0,1] \to \mathbb{R}^+$, avec

- *f continue*
- *$I(A \cap B) = I(A) + I(B)$ dès que A et B sont indépendants*

alors I est forcément de la forme

$$I(A) = -C \log(\mathbb{P}(A))$$

avec $C > 0$ une constante arbitraire (choix de l'unité de mesure).

Système simple

Définition (Système simple)

*On appelle **système simple**, soit une variable aléatoire discrète X, soit la donnée d'une partition $\Omega = A_1 + A_2 + \cdots + A_n$ d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.*

Les deux définitions sont fonctionnellement équivalentes (on ne s'intéresse pas aux valeurs de la variable aléatoire).
Notation : $X = \{A_i\}_{1 \leq i \leq n}$, $p_i = \mathbb{P}(A_i)$

Entropie : un exemple

L'enseignant choisit aléatoirement un élève parmi les présents, et on considère la variable aléatoire X qui désigne le *rang* où se trouve assis l'élève choisi.

- Que vaut l'information "l'élève choisi est au premier rang" ?
- Que vaut l'information "l'élève choisi est au rang x" ? (ça *dépend* de x, sauf si tous les rangs comptent le même nombre d'élèves)
- Que vaut **en moyenne** l'information du rang auquel se trouve l'élève choisi ? **(notion d'entropie)**

Entropie d'un système simple

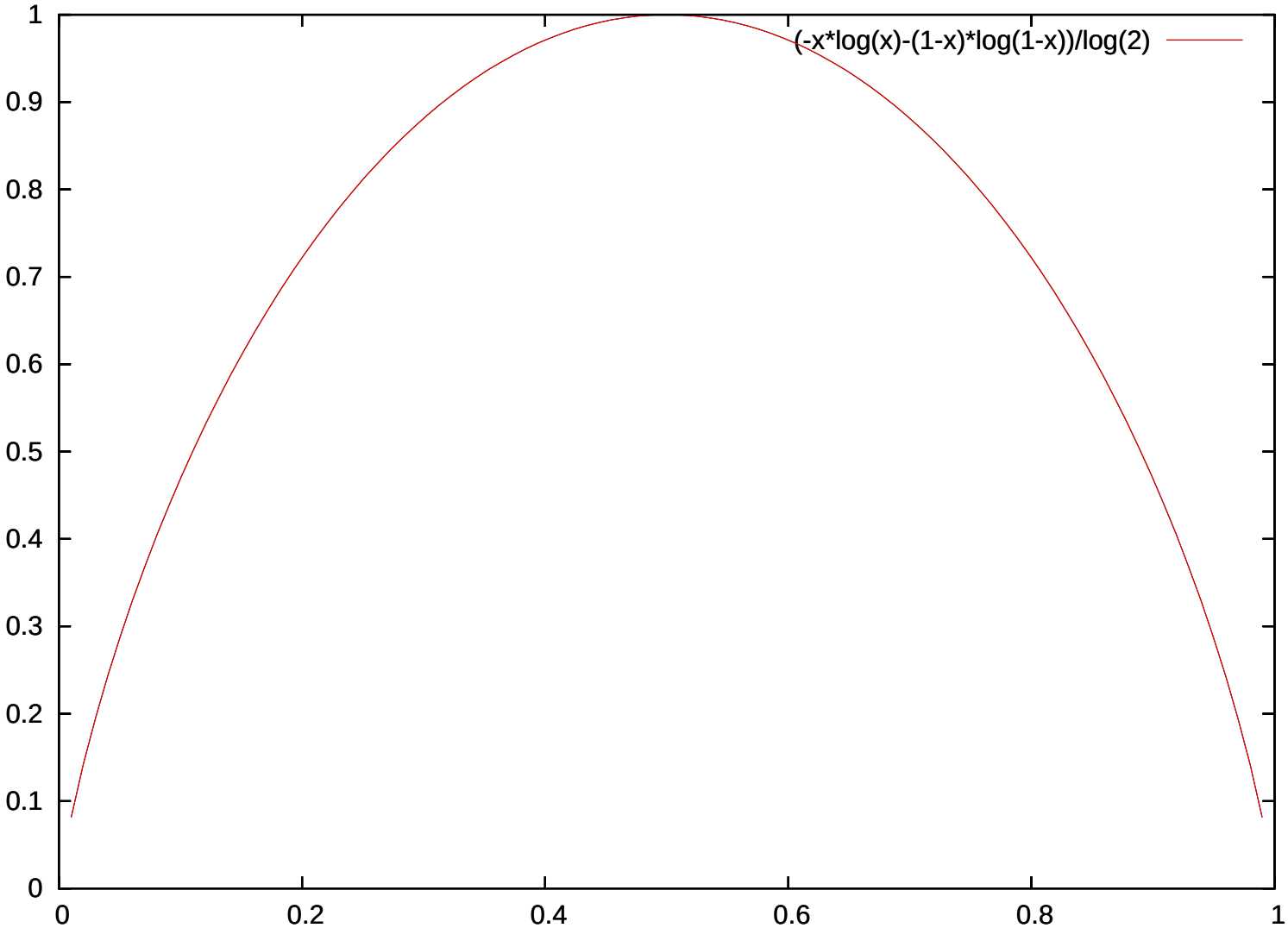
Définition (Entropie d'un système simple)

*On appelle **entropie du système** X, l'espérance de la quantité d'information apportée par la réalisation d'un événement du système.*

$$H(X) = \mathbb{E}(I(X)) = \sum_{A \in X} \mathbb{P}(A) I(A) = -\sum_{i} p_i \log(p_i).$$

(convention : pour $x = 0$, $x \log(x) = 0$)

Entropie d'un système de deux événements



Entropie des lettres du français

- On considère un alphabet de 29 symboles : 26 lettres, l'espace, le point et "les autres signes de ponctuation".
- Si on accorde la même probabilité $1/29$ à chaque symbole : $H(U_{29}) = \log_2(29) \simeq 4.848$ bits.
- Si on prend des probabilités plus proches de la fréquence empirique des lettres dans la langue française :

E, espace	0.14	I, N, T	0.08
A, R, S	0.06	D, L, O, U	0.04
C, M, P	0.02	ponctuation	0.018
point	0.014	12 autres lettres	0.004

on obtient $H(X) \simeq 4.05$ bits, soit un peu plus de 80% de l'entropie du système uniforme (l'entropie est peu sensible à des variations autour de l'équiprobabilité)

Lemme de Gibbs

Le "lemme de Gibbs" permet, entre autres, de prouver que l'entropie est maximale pour une répartition uniforme des probabilités.

Lemme (Gibbs)

Soient $\mathbf{p} = (p_1, \ldots, p_n)$ *et* $\mathbf{q} = (q_1, \ldots, q_n)$ *deux vecteurs de probabilités* ($p_i \geq 0$, $q_i \geq 0$, $\sum_i p_i = \sum_i q_i = 1$).
Alors

$$-\sum_{i=1}^{n} p_i \log(p_i) \leq -\sum_{i=1}^{n} p_i \log(q_i),$$

avec égalité si et seulement si $p_i = q_i$ *pour tout* i.

Entropie maximale d'un système à n symboles

Proposition

Si X est un système simple à n symboles, alors

$$H(X) \leq \log_2(n),$$

avec égalité si et seulement si $p_i = 1/n$ pour tout $i, 1 \leq i \leq n$.

(Preuve par le lemme de Gibbs)