Réseaux, information et communications (INFO-F303)
Partie Théorie de l'Information
3. Entropie et codage efficace

Christophe Petit

Université libre de Bruxelles

Plan du cours

1. Notion de code
2. Source aléatoire et codes efficaces
3. Entropie et codage efficace
4. Compression sans perte
5. Canal bruité
6. Codes correcteurs d'erreurs
7. Codes linéaires
8. Quelques familles de codes linéaires

A. Rappels mathématiques (chapitre 7.1 du syllabus)

Entropie et codage efficace

- But : quantifier l'information d'une source en fonction de la loi de probabilité
- Entropie : définitions et propriétés
- Inégalité de Gibbs
- Extension de la source
- Premier théorème de Shannon :
  - Entropie = mesure du codage le plus efficace
  - On peut se rapprocher asymptotiquement de l'efficacité maximale (par extension de la source)

Contexte

- Source aléatoire, indépendante identiquement distribuée
  - Chaque symbole s_i est associé à une probabilité $p_i > 0$, avec $\sum_{i=1}^{q} p_i = 1$
  - Indépendance :

$$\mathbb{P}[s_{i_1} s_{i_2} \dots s_{i_n}] = p_{i_1} \cdot p_{i_2} \cdot \dots \cdot p_{i_n}$$

- Comment *quantifier* l'information associée à un symbole ou une source ?

Quantité d'information et probabilité d'occurence

- Quelle quantité d'information contenue dans “e”, “z”, “as” et “wo" pour un texte en français ?
- Intuition : quantité d'information plus grande pour les symboles plus rares

Quantité d'information d'un symbole

- Fonction de la probabilité associée au symbole

$$\mathcal{I}^*\left(\{s_i\}\right) = \mathcal{I}(p_i)$$

que l'on veut

- Positive $\mathcal{I}(p_i) \geq 0$
- Additive : $\mathcal{I}^*\left(\{s_i, s_j\}\right) = \mathcal{I}(p_i \cdot p_j) = \mathcal{I}(p_i) + \mathcal{I}(p_j)$
- Continue

$$\mathcal{I}(p) = -\log_b p, \qquad b > 1$$

- Corollaire (Shannon, 1948) :

$$\mathcal{I}(p) = \mathcal{I}_b(p) = -\log_b p$$

avec $b > 1$

Démonstration

- On montre $\mathcal{I}(p^\alpha) = \alpha \cdot \mathcal{I}(p)$ pour tout $\alpha \in \mathbb{R}$
  - Vrai pour $\alpha = n \in \mathbb{Z}$, par l'axiome d'additivité
  - S'étend à $\alpha = 1/n$ avec $n \in \mathbb{Z}$ par manipulations
  - S'étend à $\alpha \in \mathbb{Q}$
  - S'étend à $\alpha \in \mathbb{R}$ par continuité
- On déduit $\mathcal{I}(p) = k \ln p$
- De plus $k < 0$ car $\mathcal{I}(p)$ positif et $p \leq 1$

Entropie

- L'entropie en base b d'une source / variable aléatoire S de distribution $p_1, p_2, \ldots, p_q$ est la **quantité d'information moyenne**

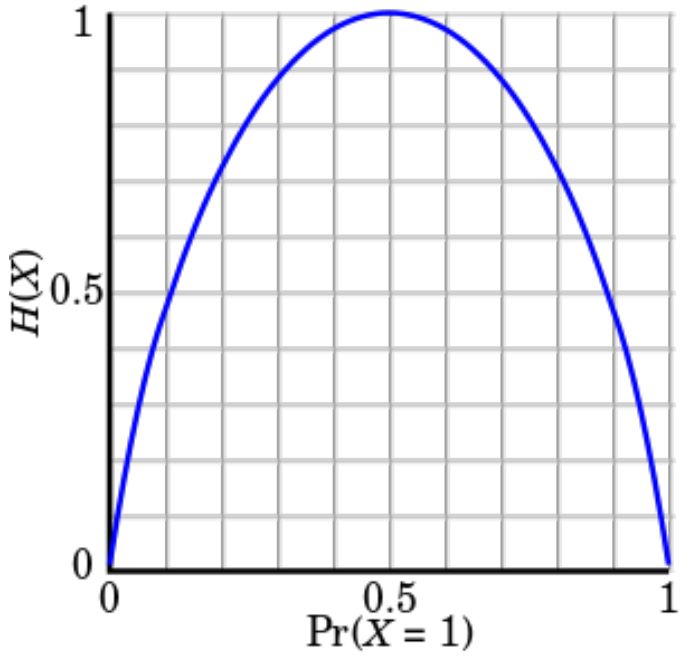
$$H_b(S) = H_b(p_1, p_2, \ldots, p_q) = -\sum_{i=1}^{q} p_i \log_b p_i$$

- Par continuité, on pose $0 \cdot \log 0 = 0$

Entropie d'une source binaire

- Pour une source binaire $p_1 = p$ et $p_2 = 1 - p$, on a

$$\mathcal{H}_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$



Crédit image : Wikipedia

Propriétés de l'entropie

1. H est positive : $H(p_1, \ldots, p_q) \geq 0$
2. $H(p_1, \ldots, p_q) = H(p_1, \ldots, p_q, 0)$
3. $H(p_1, \ldots, p_q)$ est continue et symétrique en ses q variables
4. $H(p_1, \ldots, p_q) \leq H\left(\frac{1}{q}, \ldots, \frac{1}{q}\right)$
5. H est cohérente : $H(p_1, \ldots, p_q) = H((p_1+p_2), p_3, \ldots, p_q) + (p_1+p_2)H\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$
6. On a $0 \leq H_b(S) \leq \log_b q$; entropie nulle ssi $\exists i : p_i = 1$; entropie maximale ssi $\forall i : p_i = \frac{1}{q}$

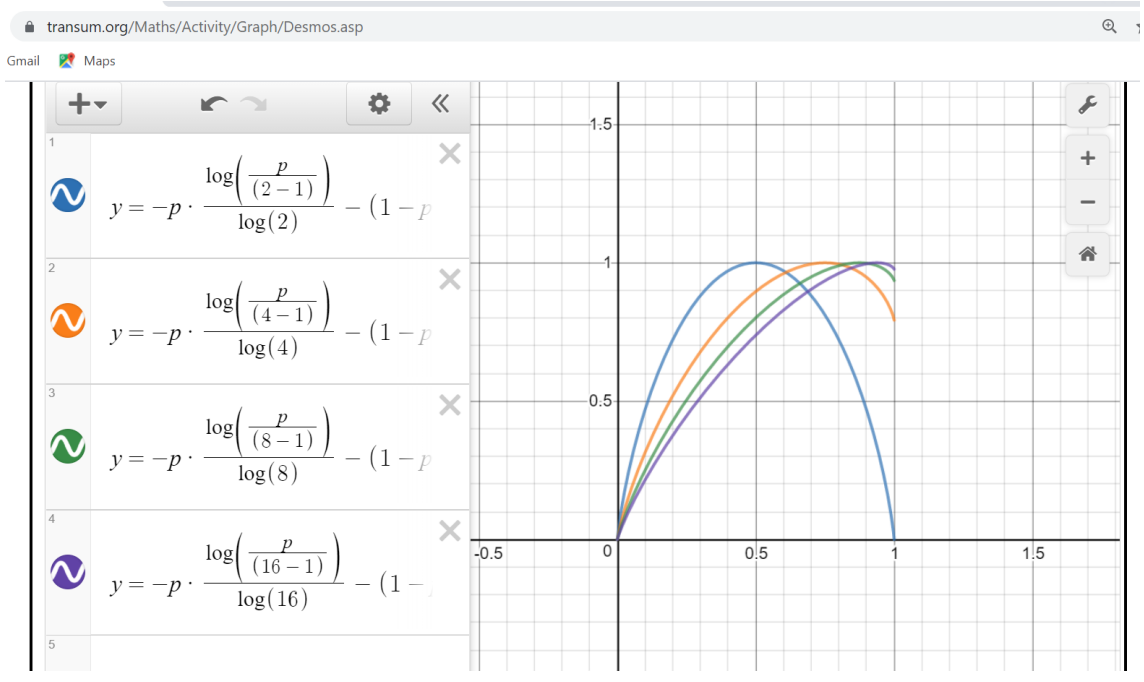
NB : propriétés 1-5 définissent l'entropie

Choix de l'unité d'information

- Entropie $H_b(S) = H_b(p_1, p_2, \ldots, p_q) = -\sum_{i=1}^{q} p_i \log_b p_i$
- Choix de b arbitraire, détermine l'unité d'information
- Choix usuels
  - $b = q$: entropie pour symboles équiprobables vaut 1
  - $b = 2$: unité est appelée le bit
  - $b = e$: unité est appelée le nat

Fonction d'entropie généralisée

- q symboles ; $p_1 = 1 - p$ and $p_i = \frac{p}{q-1}$ for $i \neq 1$



$$\mathcal{H}_q(p) = -p \log_q \frac{p}{q-1} - (1-p) \log_q (1-p)$$

Analogie : entropie en thermodynamique

- En théorie de l'information, l'entropie est une mesure de la **quantité d'information** contenue dans une variable aléatoire
  - Entropie minimale si un seul symbole a une probabilité 1
  - Entropie maximale si tous les symboles équiprobables
- En thermodynamique, l'entropie est une mesure du **désordre** (H=Heat=chaleur)
  - Entropie minimale si les molécule sont immobiles, la matière est dans un seul état stable
  - Entropie maximale si molécules en mouvement

Entropie et codage optimal

- La longueur moyenne minimum $L_{\min}(S)$ d'un code pour une source S de distribution $p_1, p_2, \ldots, p_q$ satisfait

$$H_r(S) \leq L_{\min}(S) \leq H_r(S) + 1.$$

(rappel : r = taille de l'alphabet)

- Preuve : voir ci-dessous
  - Borne supérieure satisfaite par le code de Shannon
  - Borne inférieure satisfaite pour tout code univoque

Entropie et codage optimal : borne supérieure

- Rappel : pour le code de Shannon $\ell_i = \lceil -\log_2 p_i \rceil$.
  La longueur moyenne vaut

$$\begin{aligned} \sum_{i=1}^{q} p_i \cdot \lceil -\log_2 p_i \rceil &\leq \left(-\sum_{i=1}^{q} p_i \log_2 p_i \right) + 1 \\ &= H_2(\{p_1, p_2, \ldots, p_q\}) + 1 \end{aligned}$$

- En particulier, la longueur moyenne minimum d'un code est bornée supérieurement par cette valeur

Théorème de Gibbs

- Pour toute source S de q symboles suivant une loi de probabilité $\{p_i\}$ et pour toute fonction $f(s_i) = f_i$ réelle positive définie sur ces mêmes symboles, on a

$$\sum_{i=1}^{q} f_i \leq 1 \quad \Rightarrow \quad \sum_{i=1}^{q} p_i \log_b \frac{f_i}{p_i} \leq 0$$

- NB : on peut réécrire l'inégalité de droite comme

$$-\sum_{i=1}^{q} p_i \log_b f_i \;\geq\; -\sum_{i=1}^{q} p_i \log_b p_i = H_b(\{p_i\})$$

(entropie est le minimum de $-\sum_{i=1}^{q} p_i \log_b f_i$ sur les fonctions f de L_1 norme bornée par 1)

Démonstration (théorème de Gibbs)

- Il suffit de prouver le résultat pour $b = e$
- Lemme : $\ln x \leq x - 1$
- On développe

$$\begin{aligned}
\sum_{i=1}^{q} p_i \log_b \frac{f_i}{p_i} &\leq \sum_{i=1}^{q} p_i \left(\frac{f_i}{p_i} - 1 \right) \\
&= \sum_{i=1}^{q} f_i - \sum_{i=1}^{q} p_i \\
&\leq 1 - 1 = 0
\end{aligned}$$

Entropie et codage optimal : borne inférieure

- Pour toute source S et pour tout code univoque K, la longueur moyenne du code $L_K(S)$ satisfait

$$H_r(S) \leq L_K(S)$$

Démonstration (borne inférieure)

- Notons ℓ_i les longueurs des mots de K
- De $\ell_i = -\log_r r^{-\ell_i}$, on a

$$\begin{aligned} H_r(S) - L_K(S) &= -\sum_{i=1}^{q} p_i \log_r p_i - \sum_{i=1}^{q} p_i \ell_i \\ &= -\sum_{i=1}^{q} p_i \log_r p_i + \sum_{i=1}^{q} p_i \log_r r^{-\ell_i} \\ &= \sum_{i=1}^{q} p_i \log_r \frac{r^{-\ell_i}}{p_i} \end{aligned}$$

- On a $\sum_{i=1}^{q} r^{-\ell_i} \leq 1$ par l'inégalité de Kraft
- L'inégalité de Gibbs donne le résultat

Entropie et codage optimal

- La longueur moyenne minimum $L_{\min}(S)$ d'un code pour une source S de distribution $p_1, p_2, \ldots, p_q$ satisfait

$$H_r(S) \leq L_{\min}(S) \leq H_r(S) + 1.$$

(rappel : $r =$ taille de l'alphabet)

- Parfois $L_{min}(S)$ est proche de $H_r(S) + 1$
(Exemple : $p_1 = \epsilon$, $p_2 = 1 - \epsilon$ avec ϵ très petit)
Peut-on faire mieux ?

Extension de la source

- Une **extension d'une source** S de longueur n est l'ensemble S^n constitué des n-uplets de S muni de sa loi de probabilité : $\mathbb{P}[s_{i_1} s_{i_2} \dots s_{i_n}] = p_{i_1} \cdot p_{i_2} \cdot \dots \cdot p_{i_n}$
- Plutôt que de coder chaque symbole individuellement, on peut alors définir un mot du code pour chaque n-uplet et calculer l'entropie résultante $H(S^n)$
- On obtient facilement :

$$H(S^n) = nH(S)$$

Premier théorème de Shannon (1948)

- **"Théorème du codage sans bruit de la source" (Noiseless Coding Theorem)**
- La longueur moyenne minimum $L_{\min}(S)$ d'un code pour une source S de distribution $p_1, p_2, \ldots, p_q$ satisfait

$$H_r(S) \leq L_{\min}(S) \leq H_r(S) + 1$$

- Borne supérieure atteinte uniquement pour une source dégénérée ($\exists i : p_i = 1$), pour laquelle $H = 0$ et $L = 1$
- De plus,

$$\lim_{n \to \infty} \frac{L_{\min}(S^n)}{n} = H_r(S) = \frac{H(S)}{\log_q r}$$

Premier théorème de Shannon : signification

- Si l'on choisit un code *efficace* pour une source étendue, la longueur moyenne *par symbole* est *asymptotiquement* celle de l'entropie de la source

$$\lim_{n\to\infty} \frac{L_{\min}(S^n)}{n} = H_r(S)$$

$n=1$		
Symboles :	0	1
Probabiltés :	3/4	1/4
Mots du code :	0	1
Longueur :	$1/1 \cdot 4/4 = 1$	

$n=2$				
Symboles :	00	01	10	11
Probabiltés :	9/16	3/16	3/16	1/16
Mots du code :	0	11	100	101
Longueur :	$1/2 \cdot 27/16 = 0.84375$			

$n=3$								
Symboles :	000	001	010	100	011	101	110	111
Probabiltés :	27/64	9/64	9/64	9/64	3/64	3/64	3/64	1/64
Mots du code :	1	001	010	011	00000	00001	00010	00011
Longueur :	$1/3 \cdot 158/64 = 0.82291666...$							

$n=\infty$	
Entropie :	$2 - \frac{3}{4}\log 3 \approx 0.811278124459...$

Crédit image : syllabus Y Roggeman

Démonstration (Noiseless Coding Theorem)

- $H_r(S) \leq L_{\min}(S) \leq H_r(S) + 1$ prouvé ci-dessus
- Borne supérieure atteinte si $\exists i : p_i = 1$
- Si $\forall i : p_i < 1$ la borne supérieure est stricte pour le code de Shannon
- On construit un code pour S^n

$$\begin{aligned} H(S^n) \leq & \quad L_{\min}(S^n) & \leq H(S^n) + 1 \\ nH(S) \leq & \quad L_{\min}(S^n) & \leq nH(S) + 1 \\ H(S) \leq & \quad \frac{L_{\min}(S^n)}{n} & \leq H(S) + 1/n, \end{aligned}$$

et donc $\lim_{n\to\infty} \frac{L_{\min}(S^n)}{n} = H(S)$, par le théorème du sandwich

Questions ?

?

Crédits et remerciements

- Mes transparents suivent fortement les notes de cours développées par le Professeur Yves Roggeman pour le cours INFO-F303 à l'Université libre de Bruxelles
- Une partie des transparents et des exercices ont été repris ou adaptés des transparents développés par le Professeur Jean Cardinal pour ce même cours
- Je remercie chaleureusement Yves et Jean pour la mise à disposition de ce matériel pédagogique, et de manière plus large pour toute l'aide apportée pour la reprise de ce cours
- Les typos et erreurs sont exclusivement miennes (merci de les signaler !)