Théorie de l'information

Cours 4 - Information élémentaire et entropie d'une source

Laurent Oudre
laurent.oudre@univ-paris13.fr

Université Paris 13, Institut Galilée
Master Ingénierie et Innovations en Images et Réseaux - $1^{\text{ère}}$ année
2017-2018

Sommaire

Sommaire

Information élémentaire

- Comme nous l'avons vu dans l'introduction, l'un des buts de Shannon était d'associer une valeur calculable à la quantité d'information contenue dans un événement ou dans un message.
- Nous allons considérer un événement élémentaire A, ayant une probabilité d'apparition p, et construire une fonction $I(.)$ qui associe à A une quantité d'information élémentaire $I(A)$
- La vision de Shannon de l'information est une vision probabiliste : le sens et la sémantique n'interviennent pas. La quantité d'information contenue dans un événement (ou un message) ne dépendra que de sa probabilité d'apparition.

Information élémentaire

- **Postulat 1** : La quantité d'information d'un événement élémentaire ne dépend que de sa probabilité d'apparition (cf introduction). $I(A)$ est donc une fonction de p. On appelera dans la suite cette fonction Ψ

$$I(A) = \Psi(p)$$

- **Postulat 2** : Un événément certain n'apporte aucune information

$$\Psi(1) = 0$$

- **Postulat 3** : Plus un événement est probable, moins il apporte d'information

$$\Psi(p) \text{ est une fonction décroissante}$$

- **Postulat 4** : Un événement impossible apporte une information infinie

$$\Psi(0) = +\infty$$

Information élémentaire

Il reste encore beaucoup de fonctions possibles... Et que se passe-t-il si on nous transmet non pas une, mais deux informations ?

$$A: \text{ Il va pleuvoir demain}$$

$$B: \text{ Il y a un examen demain}$$

- Les deux informations n'ont aucun rapport entre elles, on reçoit donc la somme des deux informations.
- A et B sont indépendants, il n'y a pas de relations entre eux
- **Postulat 5** : L'information $I(A \cap B)$ apportée par deux événements indépendants correspond à la somme des informations apportées par l'un et l'autre

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

- Quelle conséquence sur la fonction Ψ ?

Information élémentaire

- Notons q la probabilité de l'événement B :

$$\begin{aligned} I(A \cap B) &= \Psi(\mathbb{P}(A \cap B)) \\ &= \Psi(\mathbb{P}(A) \times \mathbb{P}(B)) \text{ car } A \text{ et } B \text{ indépendants} \\ &= \Psi(pq) \end{aligned}$$

- Or, on a également par le postulat 5 :

$$\begin{aligned} I(A \cap B) &= I(A) + I(B) \\ &= \Psi(p) + \Psi(q) \end{aligned}$$

- On a donc :

$$\Psi(pq) = \Psi(p) + \Psi(q)$$

Information élémentaire

- Etant donnés les postulats...

$$\Psi(pq) = \Psi(p) + \Psi(q)$$

$$\Psi(1) = 0,\ \Psi(0) = +\infty \text{ et } \Psi \text{ décroissante}$$

- ... il ne reste qu'une possibilité !

$$\Psi(p) = -\lambda \ln(p) \text{ avec } \lambda > 0$$

- La valeur couramment choisie pour λ est $\frac{1}{\ln(2)}$, ce qui nous ramène à un logarithme en base 2, pratique pour traiter les communications binaires

Information élémentaire

Information élémentaire

Étant donné un événement A ayant une probabilité $\mathbb{P}(A)$, on définit son **information élémentaire** $I(A)$ comme la quantité :

$$I(A) = -\log_2(\mathbb{P}(A))$$

Cette quantité est exprimée en bits.

Sommaire

Entropie d'une source

- Considérons une source (sans mémoire) modélisée par une variable aléatoire X à valeurs dans $\mathcal{X}$.
- Chacun des symboles de l'alphabet de source est associé à une certaine information élémentaire
- Comment caractériser l'information apportée en moyenne par une source ?

Source binaire

On considère une source binaire avec X dans $\mathcal{X} = \{0, 1\}$.
On suppose que $P_X(X = 0) = 0.9$.

- L'information élémentaire du symbole 0 est $I_0 = -\log_2(0.9) = 0.1520$ bits
- L'information élémentaire du symbole 1 est $I_1 = -\log_2(0.1) = 3.3219$ bits
- Information moyenne apportée par la source ?
  - $\frac{I_0+I_1}{2}$? ? NON car en pratique, le symbole 0 sera envoyé beaucoup plus fréquemment !
  - Il faut pondérer l'information élémentaire du symbole par sa probabilité d'apparition si on veut rendre compte de la vraie information moyenne

Entropie d'une source

Entropie d'une source

Étant donnée une source modélisée par une variable aléatoire X à valeurs dans $\mathcal{X}$, on définit **l'entropie de la source** et on note $H(X)$ la quantité :

$$H(X) = -\sum_{x \in \mathcal{X}} P_X(X = x) \log_2(P_X(X = x))$$

Cette quantité est exprimée en bits.

Remarque : On peut remarquer que

$$H(X) = E[-\log_2(P_X(X = x))]$$

c'est à dire que l'entropie correspond à l'espérance (moyenne statistique) de l'information élémentaire, d'où l'interprétation en terme d'information moyenne.

Propriétés de l'entropie

- La fonction

$$\begin{array}{ccc} [0,1] & \longmapsto & \mathbb{R} \\ x & \longmapsto & -x\log_2(x) \end{array}$$

est toujours positive
- L'entropie est donc également toujours positive : $\boxed{H(X) \geq 0}$
- L'entropie est nulle si et seulement si l'un des événements de la source est certain (et les autres impossibles). En effet, si l'un des événements est certain, la source n'émet aucun information.

Entropie d'une source

Comparaison de deux sources

On considère deux sources X et X' à valeurs dans $\mathcal{X} = \{1, 2, 3\}$ dont les lois de probabilités sont données par le tableau suivant :

	$p_X(x)$	$p_{X'}(x)$
1	$\frac{1}{2}$	$\frac{1}{3}$
2	$\frac{1}{4}$	$\frac{1}{3}$
3	$\frac{1}{4}$	$\frac{1}{3}$

- $H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 1.5$ bits
- $H(X') = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 1.6$ bits
- Pourquoi l'entropie de la deuxième source est-elle plus élevée ?

Entropie et incertitude

- On a vu que l'entropie correspond à l'information moyenne apportée par une source
- Mais on a également vu que la quantité d'information d'un événement était liée à la surprise qu'il procurait pour le destinataire
- Ainsi, plus l'entropie d'une source est élevée, plus elle est susceptible de surprendre le destinataire
- L'entropie de la source est donc liée au **degré d'incertitude** pour le destinataire

Entropie et incertitude

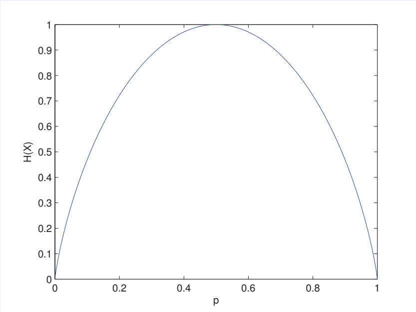
Source binaire

On considère une source binaire avec X dans $\mathcal{X} = \{0, 1\}$.
On suppose que $P_X(X = 0) = p$.

- L'entropie est donc : $H(X) = -p\log_2(p) - (1-p)\log_2(1-p)$
- Si $p = 1$, on envoie que des symboles 0 et on a : $H(X) = 0$ bits
  Normal car le destinataire n'a aucune incertitude sur ce qu'il va recevoir !
- Si $p = 0$, on envoie que des symboles 1 et on a : $H(X) = 0$ bits
  Normal car le destinataire n'a aucune incertitude sur ce qu'il va recevoir !
- En revanche si $p = \frac{1}{4}$, on a : $H(X) = 1$ bit
  En effet, le destinataire a une vraie incertitude : il ne sait pas ce qu'il va recevoir !

Entropie et incertitude

Source binaire (suite)



$$H(X) = -p\log_2(p) - (1-p)\log_2(1-p)$$

- On voit que l'entropie est maximale ici pour $p = \frac{1}{2}$
- Logique, car c'est la configuration où il y a le plus d'incertitude pour le destinataire
- On démontrera dans la suite que la distribution uniforme est celle qui maximise l'entropie

Sommaire

Comparaison de deux sources

Considérons deux sources X et X' à valeurs dans le même alphabet de source $\mathcal{X}$.

- Elles ont le même alphabet de source, mais pas la même loi de probabilité.
- Comment comparer ces deux sources ?
- On va introduire une quantité, appelée **entropie relative**, qui permet d'évaluer la *proximité* de leurs distributions de probabilité

Définition

Entropie relative

Etant données deux variables aléatoires discrètes X et X' à valeurs dans $\mathcal{X}$, dont on notera respectivement $p(x) = P_X(X = x)$ et $q(x) = P_{X'}(X' = x)$ leurs lois de probabilité. On appelle **entropie relative** ou **divergence de Kullback-Leibler** la quantité :

$$D_{KL}(p||q) = \sum_{x\in\mathcal{X}} p(x)\log_2\left(\frac{p(x)}{q(x)}\right)$$

Interprétation

- Comme nous allons le voir, cette quantité peut être interprétée comme une *distance* entre deux distributions de probabilité
- En revanche, attention, elle ne vérifie pas tous les axiomes d'une distance et en particulier on a *a priori*

$$D_{KL}(p||q) \neq D_{KL}(q||p)$$

  C'est pour cette raison qu'on l'appelle divergence et non distance.
- En revanche, tout comme une distance on a :

$$\begin{aligned} D_{KL}(p||p) &= \sum_{x \in \mathcal{X}} p(x) \log_2 \left(\frac{p(x)}{p(x)} \right) \\ &= \sum_{x \in \mathcal{X}} p(x) \log_2 (1) \\ &= 0 \end{aligned}$$

Propriété fondamentale

Étant données deux distributions de probabilités discrètes p et q on a :

$$\boxed{D_{KL}(p||q) \geq 0}$$

Avec une égalité si et seulement si $p = q$.

- Nous allons maintenant démontrer cette propriété...
- ... et nous verrons qu'elle est très importante pour comprendre l'entropie et nous aidera à démontrer de nombreux résultats dans la suite du cours.

Quelques rappels de mathématiques

Pour démontrer cette propriété, nous avons besoin des deux résultats suivants :

- Soit f une fonction $\mathcal{C}^2$ à valeurs dans $\mathbb{R}$, on dit qu'elle est concave si et seulement si

$$\frac{d^2 f(x)}{dx^2} \leq 0$$

  On dit qu'elle est strictement concave si l'inégalité précédente est stricte.
- Soit f une fonction concave, $\{z_i\}_{1 \leq i \leq M}$ M réels strictement positifs, $\{\lambda_i\}_{1 \leq i \leq M}$ M réels strictement positifs tels que $\sum_{i=1}^{M} \lambda_i = 1$, on a :

$$\sum_{i=1}^{M} \lambda_i f(z_i) \leq f\left(\sum_{i=1}^{M} \lambda_i z_i \right)$$

  Il s'agit de l'inégalité de Jensen (dans sa version renversée ici). Si la fonction est strictement concave, l'inégalité devient une égalité si et seulement si les z_i sont tous égaux

Démonstration : $D_{KL}(p||q) \geq 0$

- Étape 1 : La fonction $\log_2$ est (strictement) concave

$$\begin{aligned} \frac{d \log_2(x)}{dx} &= \frac{1}{\ln(2)} \frac{d \ln(x)}{dx} \\ &= \frac{1}{x \ln(2)} \end{aligned}$$

$$\begin{aligned} \frac{d^2 \log_2(x)}{dx^2} &= \frac{1}{\ln(2)} \frac{d \frac{1}{x}}{dx} \\ &= -\frac{1}{x^2 \ln(2)} < 0 \end{aligned}$$

Démonstration : $D_{KL}(p||q) \geq 0$

- Étape 2 : Utilisation de l'inégalité de Jensen renversée. Notons $\mathcal{X} = \{x_1, \dots, x_M\}$ et considérons $\lambda_i = p(x_i)$ et $z_i = \frac{q(x_i)}{p(x_i)}$

$$
\begin{aligned}
& \sum_{i=1}^{M} p(x_i) \log_2 \left(\frac{q(x_i)}{p(x_i)} \right) \leq \log_2 \left(\sum_{i=1}^{M} p(x_i) \frac{q(x_i)}{p(x_i)} \right) \\
\iff \quad & -\sum_{i=1}^{M} p(x_i) \log_2 \left(\frac{p(x_i)}{q(x_i)} \right) \leq \log_2 \left(\sum_{i=1}^{M} q(x_i) \right) \\
\iff \quad & -D_{KL}(p||q) \leq \log_2(1) \\
\iff \quad & D_{KL}(p||q) \geq 0
\end{aligned}
$$

- Étape 3 : Égalité si et seulement si les z_i sont tous égaux. Il existe donc une constante *cst* telle que

$$
\forall i \in [\![1, M]\!], \qquad z_i = cst \quad \iff \quad \forall i \in [\![1, M]\!], \qquad q(x_i) = cst \times p(x_i)
$$

Comme les $\{p(x_i)\}_{1 \leq i \leq M}$ et les $\{q(x_i)\}_{1 \leq i \leq M}$ sont des distributions de probabilités, elles somment à 1 et on a forcément $cst = 1$
Donc l'égalité a lieu si et seulement si $p(x_i) = q(x_i)$

Interprétation

- La divergence de Kullback Leibler peut être interprétée comme une *distance* entre deux lois de probabilité discrète.
- Mathématiquement en revanche, il ne s'agit pas d'une distance car elle ne vérifie ni la propriété de symétrie, ni l'inégalité triangulaire. En revanche on a bien la propriété de séparation :

$$
\boxed{D_{KL}(p||q) = 0 \iff \forall x \in \mathcal{X}, \quad p(x) = q(x)}
$$

- La quantité $D_{KL}(p||q)$ est toujours positive, et plus elle est proche de 0, plus cela signifie que les distributions de probabilité $p(x)$ et $q(x)$ sont *proches*

Sommaire

Entropie maximale ?

- On a vu précédemment que pour une source binaire, l'entropie de la source était maximale pour $p = \frac{1}{2}$, c'est à dire si les deux symboles étaient équiprobables
- Intuitivement, cela fait sens car la distribution uniforme est celle qui maximise l'incertitude du destinataire. Comment prouver cela ?
- On va considérer deux sources X et X' à valeurs dans $\mathcal{X}$ dont on notera respectivement $p(x) = P_X(X = x)$ et $q(x) = P_{X'}(X' = x)$ leurs lois de probabilité. On va suppose que :
  - X suit une distribution de probabilité quelconque $p(x)$
  - X' suit une distribution de probabilité uniforme $q(x) = \dfrac{1}{\text{card}(\mathcal{X})}$

... et les comparer grâce à l'entropie relative.

Entropie maximale ?

$$
\begin{aligned}
D_{KL}(p||q) \geq 0 \quad &\Longleftrightarrow \quad \sum_{x\in\mathcal{X}} p(x)\log_2\left(\frac{p(x)}{q(x)}\right) \geq 0 \\
&\Longleftrightarrow \quad \sum_{x\in\mathcal{X}} p(x)\log_2\left(p(x)\right) - \sum_{x\in\mathcal{X}} p(x)\log_2\left(q(x)\right) \geq 0 \\
&\Longleftrightarrow \quad -H(X) + \log_2\left(\mathrm{card}(\mathcal{X})\right)\sum_{x\in\mathcal{X}} p(x) \geq 0 \\
&\Longleftrightarrow \quad H(X) \leq \log_2\left(\mathrm{card}(\mathcal{X})\right)
\end{aligned}
$$

- L'entropie d'une source ayant un alphabet de source à $\mathrm{card}(\mathcal{X})$ éléments ne peut pas dépasser la valeur de $\log_2\left(\mathrm{card}(\mathcal{X})\right)$
- Cette entropie maximale est atteinte si et seulement si les symboles de la source sont équiprobables.

Bilan

A retenir

Étant donnée une source modélisée par une variable aléatoire X à valeurs dans $\mathcal{X}$, on a :

$$0 \leq H(X) \leq \log_2\left(\mathrm{card}(\mathcal{X})\right)$$

L'entropie est maximale si et seulement si :

$$\forall x \in \mathcal{X}, \qquad P_X(X = x) = \frac{1}{\mathrm{card}(\mathcal{X})}$$