Espace probabilisé discret

L'alphabet est $\mathcal{X}$ ($\mathcal{X}$ est discret).

Variable aléatoire X à valeurs dans $\mathcal{X}$

Loi de probabilité $p_X(x) = \mathbf{Prob}(X = x), x \in \mathcal{X}$. Quand il n'y a pas d'ambigüité, on la notera $p(x)$.

Espérance de la variable aléatoire réelle $V(X)$ (V est une fonction de $\mathcal{X}$ dans $\mathbb{R}$)

$$\mathbb{E}(V) = \sum_{x \in \mathcal{X}} p(x) V(x)$$

Espace probabilisé joint

Alphabet $\mathcal{X} \times \mathcal{Y}$ muni de la loi $p(x, y)$.

Variables aléatoires X et Y à valeurs dans $\mathcal{X}$ et $\mathcal{Y}$ respectivement.

Lois marginales

$$\mathbf{Prob}(X = x) = p_X(x) = \sum_y p(x, y)$$

$$\mathbf{Prob}(Y = y) = p_Y(y) = \sum_x p(x, y)$$

Probabilité conditionnelle

$$\mathbf{Prob}[X = x \mid Y = y] = \frac{p(x, y)}{p_Y(y)}$$

$$\mathbf{Prob}[Y = y \mid X = x] = \frac{p(x, y)}{p_X(x)}$$

En l'absence d'ambiguïté, nous noterons $p(x), p(y), p(x \mid y), p(y \mid x)$, les quantités respectives $\mathbf{Prob}(X = x)$, $\mathbf{Prob}(Y = y)$, $\mathbf{Prob}(X = x|Y = y)$ et $\mathbf{Prob}(Y = y|X = x)$. X et Y sont indépendantes si et ssi

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, p(x, y) = p(x)p(y)$$

Entropie – Propriétés

Définition[Entropie]

$$H(X) \stackrel{\text{def}}{=} -\sum_x p(x) \log_2 p(x)$$

Pour un tirage de Bernouilli ($X = 0$ avec proba p, $X = 1$ avec une proba $1 - p$), on a la fonction d'entropie

$$h(p) = -p \log p - (1 - p) \log(1 - p)$$

Entropie conditionnelle, Information mutuelle

Définition[entropie d'un couple de variables aléatoires]

$$H(X,Y) \stackrel{\text{def}}{=} -\sum_{x,y} p(x,y) \log_2 p(x,y)$$

Définition[entropie conditionnelle]

$$\begin{aligned} H(X|Y=y) &\stackrel{\text{def}}{=} -\sum_{x} p(X=x|Y=y) \log_2 p(X=x|Y=y) \\ H(X|Y) &\stackrel{\text{def}}{=} \sum_{y} H(X|Y=y)p(y) \end{aligned}$$

Définition[Information mutuelle]

$$I(X;Y) = H(X) - H(X|Y)$$

Propriétés

Théorème 1. *1.* $I(X;Y) = \sum p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

2. $I(X;Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y) = I(Y;X)$.

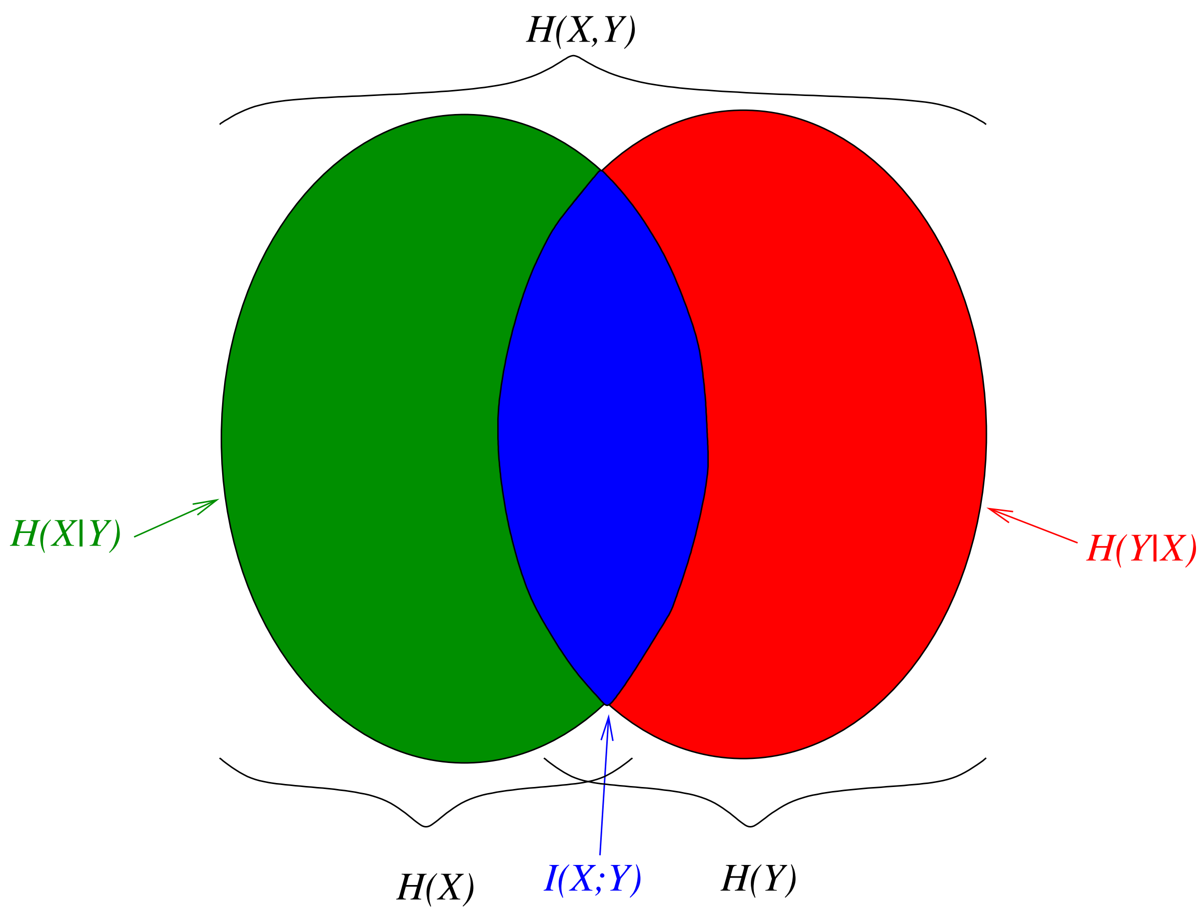
3. $I(X;Y) \geq 0$

4. $I(X;Y) = 0$ *si et ssi* X *et* Y *sont indépendants.*

5. $H(X|Y) \leq H(X)$.

6. $H(X) \leq \log |\mathcal{X}|$ *pour* X *prenant ses valeurs dans* $\mathcal{X}$.

$$I(X;Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y) = I(Y;X)$$



Preuve

1.

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= -\sum_x p(x)\log p(x) + \sum_y p(y) \sum_x p(x|y)\log p(x|y) \\
&= -\sum_{x,y} p(x,y)\log p(x) + \sum_{x,y} p(x,y)\log p(x|y) \\
&= \sum_{x,y} p(x,y)\log \frac{p(x|y)}{p(x)} \\
&= \sum_{x,y} p(x,y)\log \frac{p(x,y)}{p(x)p(y)}.
\end{aligned}
$$

2. Par symétrie de la formule précédente $I(X;Y) = I(Y;X) = H(Y) - H(Y|X)$.

$$\begin{aligned} I(X;Y) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x,y) \log p(x,y) - \sum_{x,y} p(x,y) \log p(x) - \sum_{x,y} p(x,y) \log p(y) \\ &= -H(X,Y) + H(X) + H(Y). \end{aligned}$$

3. $I(X;Y) = D(p(x,y)||p(x)p(y))$.

4. $D(p(x,y)||p(x)p(y)) = 0 \Rightarrow p(x,y) = p(x)p(y)$.

5. $H(X) - H(X|Y) = I(X;Y)$.

6.

$$
\begin{aligned}
-H(X) + \log|\mathcal{X}| &= \sum_x p(x)\log p(x) + \sum p(x)\log|\mathcal{X}| \\
&= \sum_x p(x)\log\frac{p(x)}{\frac{1}{|\mathcal{X}|}} \\
&= D(p||u)
\end{aligned}
$$

La divergence de Kullback

Définition[Divergence de Kullback] Pour deux distributions de probabilité p et q sur un même ensemble discret $\mathcal{X}$:

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

Théorème 2.

$$D(p||q) \geq 0$$

Il y a égalité si et seulement si $p = q$.

Preuve

$$
\begin{aligned}
D(p||q) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \\
&= \sum_{x \in \mathcal{X}} p(x) \left(-\log \frac{q(x)}{p(x)} \right) \\
&\geq -\log \left(\sum_{x} p(x) \frac{q(x)}{p(x)} \right) \\
&= 0.
\end{aligned}
$$