Nicolas Sendrier

Majeure d'informatique

Introduction la théorie de l'information

Cours n°1

Une mesure de l'information

Espace probabilisé discret

L'alphabet est $\mathcal{X}$ (fini en pratique)

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

Loi de probabilité $p_X(x), x \in \mathcal{X}$

Moyenne d'une variable aléatoire réelle

$$\overline{V} = \sum_{x \in \mathcal{X}} p_X(x) V(x) = \sum_x p(x) V(x)$$

Espace probabilisé joint

Alphabet $\mathcal{X} \times \mathcal{Y}$ muni de la loi produit $p_{XY}(x, y), (x, y) \in \mathcal{X} \times \mathcal{Y}$

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

Lois marginales

$$p_X(x) = \sum_y p_{XY}(x, y)$$

$$p_Y(y) = \sum_x p_{XY}(x, y)$$

Probabilité conditionnelle

$$\mathrm{Pr}_{XY}[X = x \mid Y = y] = \frac{p_{XY}(x, y)}{p_Y(y)}$$

$$\mathrm{Pr}_{XY}[Y = y \mid X = x] = \frac{p_{XY}(x, y)}{p_X(x)}$$

En l'absence d'ambiguïté, nous noterons $p(x), p(y), p(x \mid y), p(y \mid x)$, ou parfois $p(x \mid y) = p_{X|Y}(x \mid y)$ et $p(y \mid x) = p_{Y|X}(y \mid x)$.

X et Y sont indépendantes si

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, p_{XY}(x, y) = p_X(x) p_Y(y)$$

Incertitude et information

La quantité d'information obtenue lorsque l'évènement $X = x$ se réalise est liée à l'incertitude sur cet évènement. Nous cherchons

- fonction positive et décroissante de la probabilité : $I(x) = f(p(x))$
- l'évènement certain ne produit aucune information : $f(1) = 0$
- un évènement impossible fournit une quantité infinie d'information : $f(0) = \infty$
- fonction additive : l'information de deux évènements indépendants s'additionne : $f(p(x)p(y)) = f(p(x)) + f(p(y))$

Information propre

Nous utiliserons comme mesure de l'incertitude la quantité suivante :

$$I(x) = \log_b \frac{1}{p(x)}$$

qui sera appelée *information propre* de x. La base du logarithme est arbitraire. L'unité d'information que nous utiliserons est le *bit*, défini par

> *Un bit est égal à la quantité d'information fournie par le choix d'une alternative parmi deux équiprobables.*

Autrement dit nous utiliserons le logarithme en base 2.

Information mutuelle

Si nous voulons quantifier la corrélation entre deux évènements, il faut se demander comment la réalisation de l'un d'entre eux « $Y = y$ » va modifier l'incertitude sur l'autre « $X = x$ ».

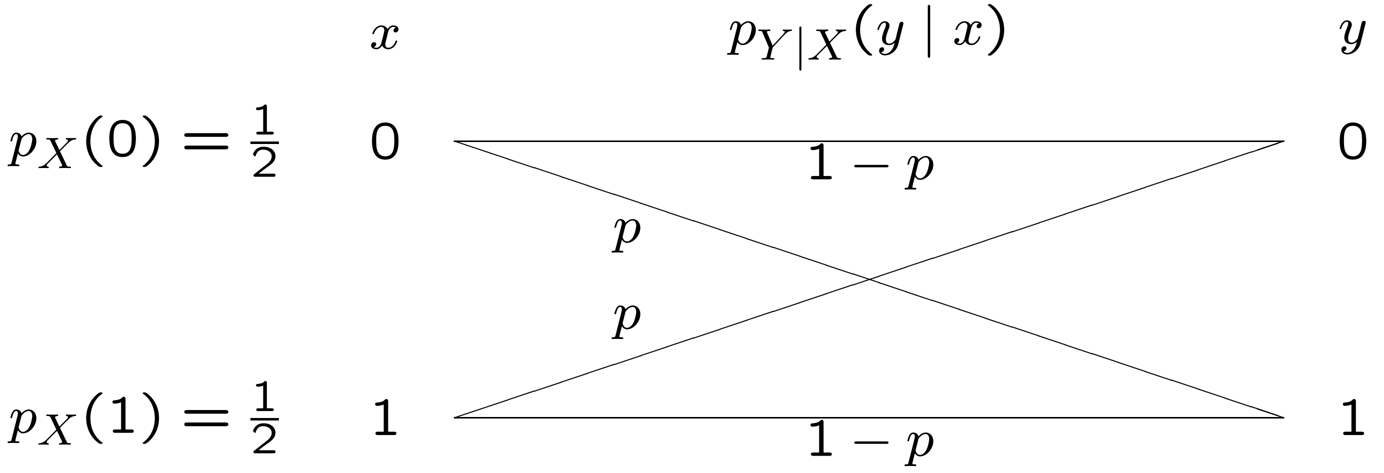
La probabilité *a priori* $p(x) = \Pr[X = x]$ va devenir la probabilité *a posteriori* $p(x \mid y) = \Pr[X = x \mid Y = y]$.

La différence entre les deux « quantités d'incertitude » correspondantes sera l'*information mutuelle* entre x et y :

$$I(x; y) = I(x) - I(x \mid y) = \log_2 \frac{\Pr[X = x \mid Y = y]}{\Pr[X = x]} = \log_2 \frac{p(x \mid y)}{p(x)}$$

Exemple

Soit le canal binaire symétrique de probabilité de transition $p \leq \frac{1}{2}$. Les symboles 0 et 1 sont émis selon une loi uniforme.



$$
\begin{aligned}
I(0;0) = I(1;1) = & \quad \log_2(2-2p) & \geq 0 \\
I(0;1) = I(1;0) = & \quad \log_2(2p) & \leq 0.
\end{aligned}
$$

Interprétation de l'information mutuelle

L'information mutuelle est symétrique

$$I(x;y) = I(y;x) = \log_2 \frac{p(x,y)}{p(x)p(y)}$$

- $I(x;y) \geq 0$ ssi $p(x \mid y) \geq p(x)$; la réalisation de y augmente la probabilité d'occurence de x,
- $I(x;y) \leq 0$ ssi $p(x \mid y) \leq p(x)$; la réalisation de y diminue la probabilité d'occurence de x,
- $I(x;y) = 0$ ssi $p(x \mid y) = p(x)$; les évènements sont indépendants.

Information mutuelle moyenne

Soient X et Y deux variables aléatoires discrètes

Définition (information mutuelle moyenne)

$$I(X;Y) = \sum_{x,y} p(x,y) I(x;y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$

Théorème

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

avec égalité si et seulement si X et Y sont indépendantes

Fonctions convexes

Définition f est convexe sur l'intervalle (a, b) si pour tous $x, y \in (a, b)$

$$\forall \lambda, 0 < \lambda < 1, f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

f est strictement convexe si l'égalité n'est vraie que lorsque $x = y$.

Lemme (inégalité de Jensen) f une fonction strictement convexe sur (a, b). $\forall p_1, \ldots, p_n > 0, \sum_i p_i = 1$, $\forall x_1, \ldots, x_n \in (a, b)$

$$f\left(\sum_i p_i x_i\right) \leq \sum_i p_i f(x_i)$$

Avec égalité si et seulement si les x_i sont tous égaux.

Entropie – Propriétés

Définition (Entropie)

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

Proposition Pour une source de cardinal K

$$0 \leq H(X) \leq \log_2 K$$

Définition (Entropie conditionnelle)

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

Theorème (le conditionnement réduit l'entropie)

$$H(X \mid Y) \leq H(X)$$

Preuve : $I(X;Y) = H(X) - H(X|Y) \geq 0$.