Théorie et codage de l'information

Mesure quantitative de l'information

- Chapitre 2 -

Information propre et mutuelle

Quantité d'information propre d'un événement

Soit A un événement de probabilité $P(A)$ non-nulle.

L'information $h(A)$ apportée par la réalisation de A est d'autant plus grande qu'elle est improbable. Elle peut s'exprimer ainsi :

$$h(A) = f\left(\frac{1}{P(A)}\right).$$

La fonction $f(\cdot)$ vérifie les contraintes suivantes :

- ▷ $f(\cdot)$ est croissante
- ▷ info. apportée par 1 événement sûr est nulle : $\lim_{p\to 1} f(p) = 0$
- ▷ info. apportée par 2 événements indépendants : $f(p_1 \cdot p_2) = f(p_1) + f(p_2)$

Ceci nous conduit à utiliser la fonction logarithmique pour $f(\cdot)$

Information propre et mutuelle

Quantité d'information propre d'un événement

Lemme 1. *La fonction* $f(p) = -\log_b p$ *est la seule qui soit à la fois positive, continue sur* $]0,1]$*, et qui vérifie* $f(p1 \cdot p2) = f(p1) + f(p2)$.

Preuve. La démonstration comporte les étapes suivantes :

1. $f\left(p^n\right) = n\, f(p)$
2. $f\left(p^{1/n}\right) = \frac{1}{n}\, f(p)$ après avoir remplacer p par $p^{1/n}$
3. $f\left(p^{m/n}\right) = \frac{m}{n}\, f(p)$ en combinant le deux égalités précédentes
4. $f(p^q) = q\, f(p)$ où q désigne un nombre rationnel positif quelconque
5. $f\left(p^r\right) = \lim_{n\to+\infty} f\left(p^{q_n}\right) = \lim_{n\to+\infty} q_n\, f(p) = r\, f(p)$

Soient p et q appartenant à $]0,1[$. On peut écrire $p = q^{\log_q p}$, ce qui entraîne

$$f(p) = f\left(q^{\log_q p}\right) = f(q)\, \log_q p.$$

On aboutit finalement au résultat escompté, soit

$$\boldsymbol{f(p) = -\log_b p}$$

Information propre et mutuelle

Quantité d'information propre d'un événement

Définition 1. *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* A *un événement de* $\mathcal{A}$ *de probabilité* $P(A)$ *non-nulle. On associe à la réalisation de* A *la quantité d'information propre :*

$$h(A) = -\log P(A).$$

Unités. L'unité dépend de la base choisie pour le logarithme.

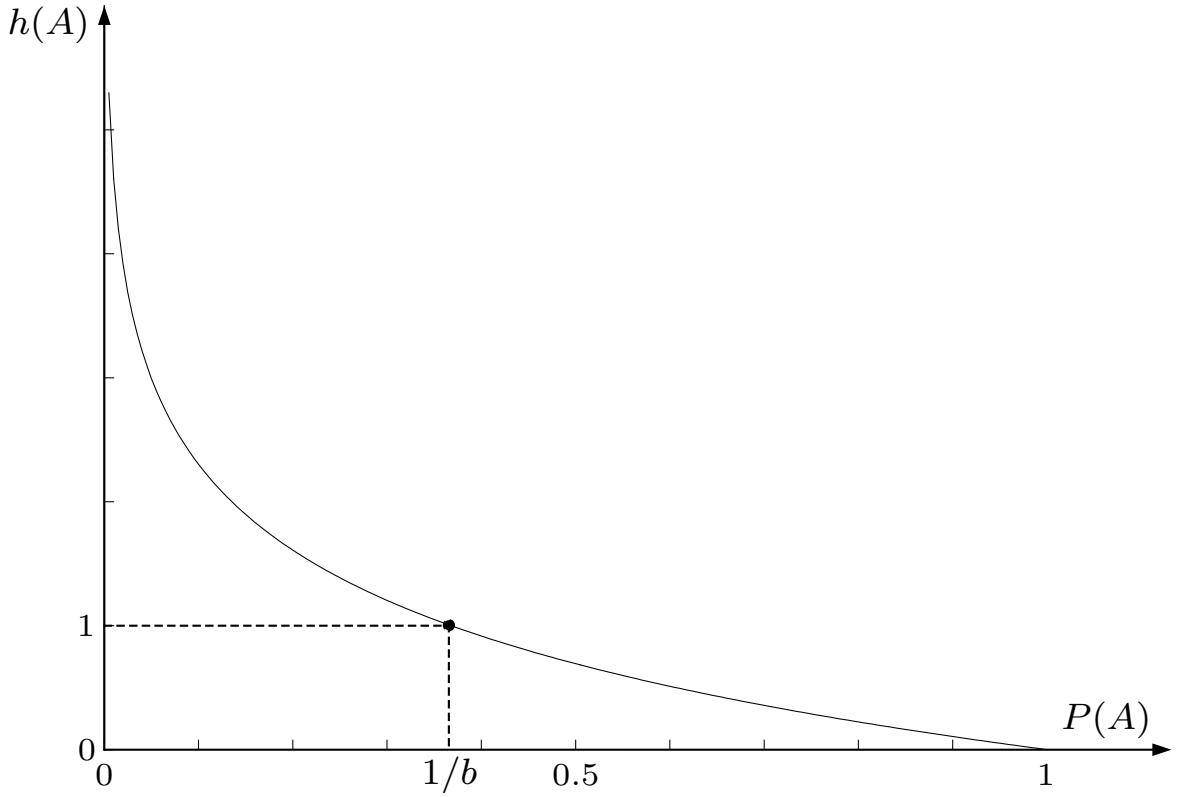
- ▷ $\log_2$: Shannon, bit (binary unit)
- ▷ $\log_e$: logon, nat (natural unit)
- ▷ $\log_{10}$: Hartley, decit (decimal unit)

Vocabulaire. $h(\cdot)$ est désigné par *incertitude* ou encore *quantité d'information*.

Information propre et mutuelle

Quantité d'information propre d'un événement

Quantité d'information propre ou incertitude : $h(A) = -\log_b P(A)$



Information propre et mutuelle

Quantité d'information propre d'un événement

Exemple 1. Dans le cas d'une source binaire $\{0, 1\}$ telle que $P(0) = P(1) = 0.5$, l'information propre associée à chaque symbole binaire, ou bit au sens informatique du terme, vaut $h\left(\frac{1}{2}\right) = \log 2$, soit 1 bit ou Shannon.

Exemple 2. On considère une source S sélectionnant aléatoirement et indépendamment du passé chaque symbole émis parmi les 16 éléments d'un l'alphabet $\{s_0, \ldots, s_{15}\}$, tous équiprobables. L'information propre véhiculée par chacun d'eux est $\log 16$, soit 4 Shannon.

Attention ! Le bit informatique (*binary digit*) et le bit issu de la théorie de l'information (*binary unit*) sont deux notions distinctes.

Information propre et mutuelle

Quantité d'information conditionnelle

La définition de la quantité d'information propre s'applique à la réalisation conjointe de A et B. En remarquant que $P(A \cap B) = P(A)\, P(B|A)$, on obtient :

$$h(A \cap B) = -\log P(A \cap B) = -\log P(A) - \log P(B|A)$$

On note que $-\log P(B|A)$ correspond à la quantité d'information propre de B que ne fournit pas l'observation de A.

Définition 2. *On appelle information conditionnelle de B sachant A la quantité :*

$$h(B|A) = -\log P(B|A),$$

soit, en d'autres termes : $h(B|A) = h(A \cap B) - h(A)$.

Exercice. Étudier et interpréter le cas $A = B$, puis indépendance de A et B

Information propre et mutuelle

Quantité d'information mutuelle

La définition de l'information conditionnelle amène directement une autre définition, celle de l'information mutuelle, qui correspond à la part d'incertitude commune à deux événements.

Définition 3. *On appelle information mutuelle entre A et B la quantité :*

$$i(A,B) = h(A) - h(A|B) = h(B) - h(B|A).$$

Exercice. Étudier les cas $A = B$, $B \subset A$, enfin A et B indépendants.

Entropie d'une variable aléatoire

Définition de l'entropie

Soit une source S d'information sans mémoire sélectionnant aléatoirement un symbole parmi les n éléments d'un alphabet $\{s_1, \ldots, s_n\}$. Soit p_i la probabilité d'apparition de s_i. La quantité d'information moyenne associée à l'apparition de chaque symbole possible est donnée par :

$$H(S) = E\{h(s)\} = -\sum_{i=1}^{n} p_i \log p_i.$$

L'entropie est une quantité d'information moyenne.

Définition 4. *Soit X une variable aléatoire à valeurs dans $\{x_1, \ldots, x_n\}$. L'entropie de X est définie comme suit :*

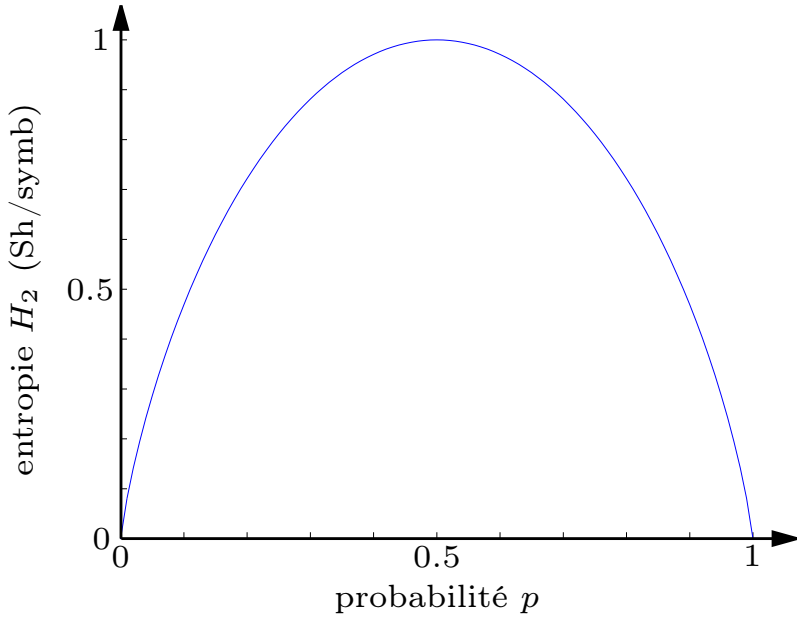
$$H(X) = -\sum_{i=1}^{n} P(X = x_i) \log P(X = x_i).$$

Entropie d'une variable aléatoire

Exemple d'une variable aléatoire binaire

L'entropie d'une variable aléatoire binaire est donnée par :

$$H(X) = -p \log p - (1-p) \log(1-p) \triangleq H_2(p).$$



Entropie d'une variable aléatoire

Notation et propriété préalables

Lemme 2 (Inégalité de Gibbs). *Étant donné 2 distributions de probabilité discrètes* $(p_1, \ldots, p_n)$ *et* $(q_1, \ldots, q_n)$ *sur un même univers fini, l'inégalité suivante est satisfaite :*

$$\sum_{i=1}^{n} p_i \log \frac{q_i}{p_i} \leq 0,$$

l'égalité étant obtenue lorsque $\forall i : p_i = q_i$.

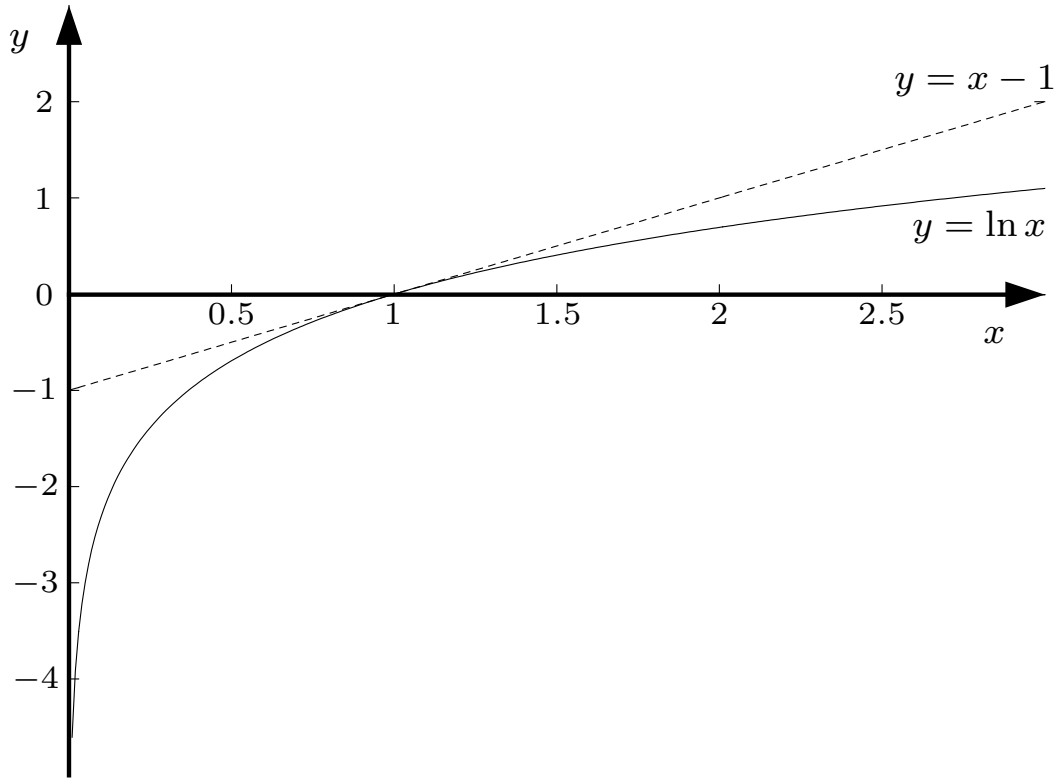
Preuve. On effectue la démonstration dans le cas du logarithme népérien et on note que $\ln x \leq x - 1$, l'égalité étant obtenue pour $x = 1$. On pose $x = \frac{q_i}{p_i}$ et on a

$$\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} \leq \sum_{i=1}^{n} p_i (\frac{q_i}{p_i} - 1) = 1 - 1 = 0.$$

Entropie d'une variable aléatoire

Notation et propriété préalables

Vérification graphique de l'inégalité $\ln x \leq x - 1$



Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

Propriété 1. *L'entropie vérifie l'inégalité suivante*

$$H_n(p_1, \ldots, p_n) \leq \log n,$$

l'égalité étant réalisée dans le cas d'une loi uniforme, c'est-à-dire $\forall i : p_i = \frac{1}{n}$.

Preuve. A partir de l'inégalité de Gibbs, on pose $q_i = \frac{1}{n}$. L'incertitude sur le résultat d'une expérience est d'autant plus grande que tous les résultats possibles sont équiprobables.

Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

Propriété 2. *L'entropie augmente lorsque le nombre d'états du système augmente.*

Preuve. Soit X une variable aléatoire discrète à valeurs dans $\{x_1, \ldots, x_n\}$ avec les probabilités $(p_1, \ldots, p_n)$. On suppose que l'état x_k est scindé en deux sous-états x_{k_1} et x_{k_2}, de probabilités respectives p_{k_1} et p_{k_2} non-nulles telles que $p_k = p_{k_1} + p_{k_2}$.

L'entropie de la variable aléatoire résultante X' s'écrit :

$$\begin{aligned} H(X') &= H(X) + p_k \log p_k - p_{k_1} \log p_{k_1} - p_{k_2} \log p_{k_2} \\ &= H(X) + p_{k_1}(\log p_k - \log p_{k_1}) + p_{k_2}(\log p_k - \log p_{k_2}). \end{aligned}$$

La fonction logarithmique étant strictement croissante, on a : $\log p_k > \log p_{k_i}$. Il en résulte directement que $H(X') > H(X)$.

Interprétation. Second Principe de la Thermodynamique

Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

Propriété 3. *L'entropie H_n est une fonction concave de $p_1, \ldots, p_n$.*

Preuve. Soient 2 distributions de probabilité discrètes $(p_1, \ldots, p_n)$ et $(q_1, \ldots, q_n)$. La concavité de la fonction entropie qu'il s'agit de démontrer se traduit par le fait que pour tout λ de l'intervalle $[0, 1]$, on a :

$$H_n(\lambda p_1 + (1-\lambda)q_1, \ldots, \lambda p_n + (1-\lambda)q_n) \geq \lambda H_n(p_1, \ldots, p_n) + (1-\lambda)H_n(q_1, \ldots, q_n).$$

En posant $f(x) = -x \log x$, on peut écrire que :

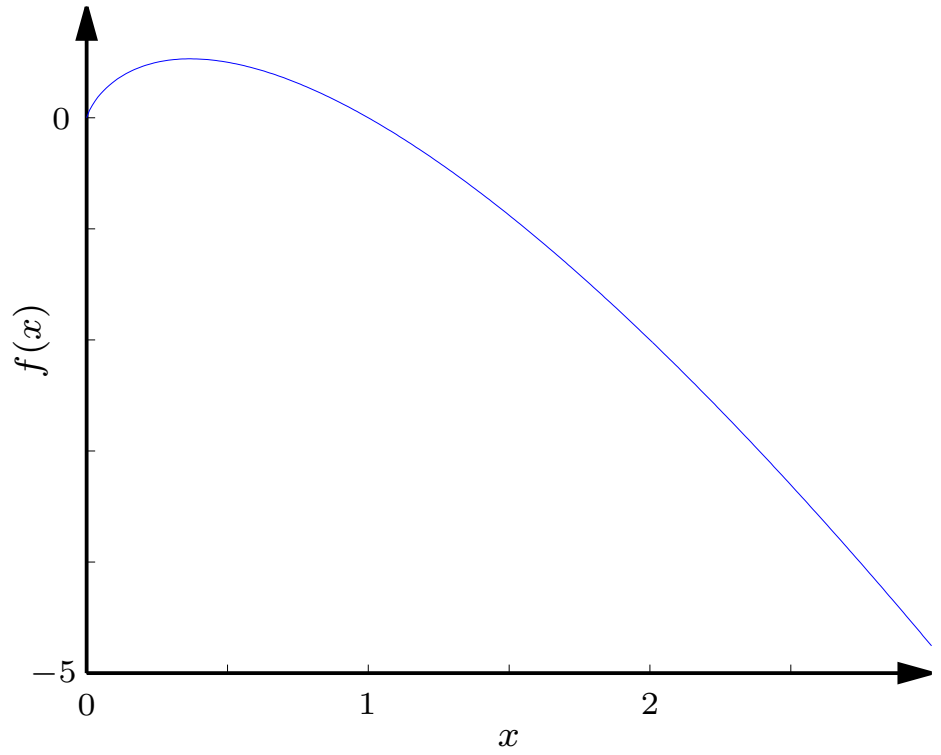
$$H_n(\lambda p_1 + (1-\lambda)q_1, \ldots, \lambda p_n + (1-\lambda)q_n) = \sum_{i=1}^{n} f(\lambda p_i + (1-\lambda)q_i).$$

Le résultat découle directement de la concavité de $f(\cdot)$.

Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

Vérification graphique de la concavité de $f(x) = -x \log x$



Entropie d'une variable aléatoire

Quelques propriétés de l'entropie

La convexité de H_n peut être généralisée à un nombre quelconque de distributions.

Propriété 4. *Étant donné* $\{(q_{1j}, \ldots, q_{nj})\}_{j=1}^{m}$ *un ensemble fini de distributions de probabilité discrètes, l'inégalité suivante est satisfaite :*

$$H_n(\sum_{j=1}^{m} \lambda_j \, q_{1j}, \ldots, \sum_{j=1}^{m} \lambda_j \, q_{mj}) \geq \sum_{j=1}^{m} \lambda_j \, H_n(q_{1j}, \ldots, q_{mj}),$$

avec $\{\lambda_j\}_{j=1}^{m}$ *des éléments de* $[0, 1]$ *tels que* $\sum_{j=1}^{m} \lambda_j = 1$.

Preuve. Comme dans le cas précédent, la démonstration de cette inégalité repose sur la concavité de la fonction $f(x) = -x \log x$. A charge du lecteur de le vérifier.

Couple de variables aléatoires

Entropie conjointe

Définition 5. *Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \ldots, x_n\}$ et $\{y_1, \ldots, y_m\}$, respectivement. L'entropie conjointe de X et Y est donnée :*

$$H(X,Y) \triangleq -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j).$$

▷ l'entropie conjointe est une grandeur symétrique : $H(X,Y) = H(Y,X)$

Exemple. Cas de variables aléatoires indépendantes ou strictement liées.

Couple de variables aléatoires

Entropie conditionnelle

Définition 6. *Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \ldots, x_n\}$ et $\{y_1, \ldots, y_m\}$. L'entropie conditionnelle de X sachant que $Y = y_j$ est donnée :*

$$H(X|Y = y_j) \triangleq -\sum_{i=1}^{n} P(X = x_i|Y = y_j) \log P(X = x_i|Y = y_j).$$

▷ incertitude moyenne sur le résultat de X sachant que celui de Y est y_j

Définition 7. *L'entropie conditionnelle moyenne de X sachant Y est donnée par :*

$$H(X|Y) \triangleq \sum_{j=1}^{m} P(Y = y_j)\, H(X|Y = y_j),$$

Exemple. Cas de variables aléatoires indépendantes ou strictement liées.

Couple de variables aléatoires

Relations entre les entropies

La première relation énoncée lie ainsi les diverses entropies définies précédemment :

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y).$$

Ces égalités se démontrent aisément en écrivant d'abord

$$\log P(X = x, Y = y) = \log P(X = x|Y = y) + \log P(Y = y),$$

puis en prenant l'espérance mathématique de chacun des membres.

Propriété 5 (règle de chaînage). *L'entropie jointe de de n variables aléatoires peut être évaluée selon la règle de chaînage suivante :*

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_1 \ldots X_{i-1}).$$

Couple de variables aléatoires

Relations entre les entropies

Chaque élément de $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$ est positif. On en déduit immédiatement que :

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

Couple de variables aléatoires

Relations entre les entropies

A partir de la *convexité généralisée* de l'entropie, en posant $q_{ij} = P(X = x_i|Y = y_j)$ et $\lambda_j = P(Y = y_j)$, on aboutit à l'inégalité fondamentale suivante :

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

Le conditionnement d'une variable aléatoire diminue son entropie. Sans démonstration, il se généralise ainsi :

Propriété 6 (décroissance par conditionnement). *L'entropie d'une variable aléatoire décroît par conditionnements successifs, soit*

$$H(X_1|X_2, \ldots, X_n) \leq \ldots \leq H(X_1|X_2, X_3) \leq H(X_1|X_2) \leq H(X_1),$$

où $X_1, \ldots, X_n$ désignent n variables aléatoires discrètes.

Couple de variables aléatoires

Relations entre les entropies

Soient X et Y des variables aléatoires à valeurs dans $\{x_1, \ldots, x_n\}$ et $\{y_1, \ldots, y_m\}$, respectivement. Les relations fondamentales vues précédemment se résument ainsi :

$$0 \leq H(X|Y) \leq H(X) \leq H(X,Y) \leq H(X) + H(Y) \leq 2H(X,Y).$$

Couple de variables aléatoires

Information mutuelle moyenne

Définition 8. *L'information mutuelle moyenne de X et Y est définie par*

$$I(X,Y) \triangleq H(X) - H(X|Y)$$

ou encore, de façon équivalente,

$$I(X,Y) \triangleq \sum_{i=1}^{n}\sum_{j=1}^{m} P(X=x_i, Y=y_j)\, \log \frac{P(X=x_i, Y=y_j)}{P(X=x_i)\,P(Y=y_j)}.$$

L'information mutuelle représente l'incertitude moyenne sur X diminuée de celle qui subsiste lorsque Y est connue.

Exercice. Cas de variables aléatoires indépendantes et strictement liées.

Couple de variables aléatoires

Information mutuelle moyenne

Afin de donner une autre interprétation de l'information mutuelle, on rappelle préalablement la définition suivante.

Définition 9. *On appelle distance de Kullback-Leibler entre deux distributions P_1 et P_2, ici supposées discrètes, la quantité*

$$d(P_1, P_2) = \sum_{x \in X(\Omega)} P_1(X = x) \log \frac{P_1(X = x)}{P_2(X = x)}.$$

L'information mutuelle correspond à la distance de Kullback-Leibler entre la loi jointe et le produit des distributions marginales de X et Y.

Couple de variables aléatoires

Diagramme de Venn

Le diagramme de Venn, ici à 2 variables, constitue un moyen mnémotechnique.