Formulaire théorie shannonienne de l'information

H. Glotin, glotin@univ-tln.fr, ref. Cover & Thomas, Wiley Ed.

1 Notations

Soient X et Y deux Variables Aléatoires (VA) prenant respectivement dans leur alphabet les valeurs $x_1, ..., x_i, ..., x_r$ et $y_1, ..., y_j, ..., y_s$, on note $p(x) = p_X(x) = Pr\{X = x\}$, pour x dans l'alphabet de X, p est la distribution de probabilité (ddp ou pdf en anglais), $(x_i|y_j)$ l'évènement x_i conditionné par y_j, (x_i, y_j) leur intersection. On note (X, Y) la VA jointe de X et Y, X^N la séquence $X_1, ..., X_i, ..., X_N$ de N VA. Si le log est en base 2, l'unité est le logon (bit).

2 Indépendance et incompatibilité

$p(x_i \cup y_j) = p(x_i) + p(y_j) - p(x_i, y_j)$
Deux évènements x_i et y_j sont incompatibles (disjoints) ssi $p(x_i \cup y_j) = p(x_i) + p(y_j)$.
$p(x_i, y_j) = p(x_i)p(y_j|x_i) = p(y_j)p(x_i|y_j)$, donc (Bayes) : $\frac{p(x_i|y_j)}{p(x_i)} = \frac{p(y_j|x_i)}{p(y_j)}$
L'évènements x_i est indépendant de y_j ssi $p(x_i, y_j) = p(x_i)p(y_j)$.
La VA X est indépendante de Y ssi $\forall(i, j)$ x_i est indépendant de y_j.
La VA jointe (X, Y) est définie par l'ensemble des couples (x_i, y_j).

3 Incertitude h

$h(x_i) = log\frac{1}{p(x_i)} = -logp(x_i)$
$h(x_i, y_j) = h(x_i) + h(y_j|x_i) = h(y_j) + h(x_i|y_j)$, donc si x_i est indépendant de y_j alors $h(x_i, y_j) = h(x_i) + h(y_j)$.
$0 \leq h(x_i) \leq +\infty$

4 Entropie H (= incertitude moyenne)

$H(X) = \sum_{i=1}^{r} p(x_i).h(x_i) = -\sum_{i=1}^{r} p(x_i)logp(x_i)$, et suivant déf. de (X, Y) on a : $H(X, Y) = -\sum_{i=1}^{r}\sum_{j=1}^{s} p(x_i, y_j)logp(x_i, y_j)$, $H(X, Y) = H(Y, X) = H(X) + H(Y|X) = H(Y) + H(X|Y)$, X indépendant de Y ssi $H(X) = H(X|Y)$ (ssi $H(X, Y) = H(X) + H(Y)$).

L'entropie H(X) est maximale lorsque la ddp de X est uniforme (i.e. système aléatoire), alors $H(X) = log(r)$.

$0 \leq H(\phi(X)) \leq H(X) \leq log(r)$. $H(\phi(X)) = H(X)$ ssi la fonction ϕ est inversible.
$0 \leq H(X|Y) \leq H(X) \leq H(X, Y) \leq H(X) + H(Y) \leq \log(r.s)$
Règle de la chaîne : $H(X^N) = \sum_{n=1}^{N} H(X_n|X^{n-1})$

5 Entropie relative (ou divergence de Kullback-Leibler)

On note $divKL(p||q)$ la divergence (cross-entropy) entre les ddp p et q de même alphabet : $0 \leq divKL(p||q) = \sum_i p(x_i)log\frac{p(x_i)}{q(x_i)} \leq \infty$

$divKL$ est asymétrique et $divKL(p||q) = 0$ ssi $p = q$.

Entropie relative et règle de la chaîne : $divKL(p(x,y)||q(x,y)) = divKL(p(x)||q(x)) + divKL(p(y|x)||q(y|x))$.

$DistKL(p,q) = divKL(p||q) + divKL(q||p) = \sum_x (p-q)log(p/q)$ est symétrique, est dite distance par abus de langage car ne vérifie pas $DistKL(p,r) \leq DistKL(p,q) + DistKL(q,r)$.

6 Pouvoir discriminant et diversité marginale maximale

Le pouvoir discriminant d'une dimension A est défini suivant le critère de 'Maximal Marginal Diversity' sur K classes $C_1, ..., C_K$ par : $J(A) = \sum_k p(C_k).DistKL(p(C_k, A), p(MCK, A)))$, avec p(MCK,A) la ddp du centroïde des K classes dans la dimension A.

7 Information mutuelle évènementielle

$i(x_i, y_j) = i_{x_i \to y_j} = i_{y_j \to x_i} = i(y_j, x_i)$
$i(x_i, y_j) = log\frac{1}{p(x_i)} - log\frac{1}{p(x_i|y_j)} = log\frac{1}{p(y_j)} - log\frac{1}{p(y_j|x_i)}$
x_i et y_j sont indépendants ssi $i(x_i; y_j) = 0$
$-\infty \leq i(x_i, y_j) \leq +\infty$

8 Information mutuelle de VA I (moyenne des MI évènementielles)

$I(X;Y) = I(Y;X) = H(X) + H(Y) - H(X,Y)$
$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
$I(X;Y) = \sum_{i=1}^{r} \sum_{j=1}^{s} p(x_i, y_j) \log \frac{p(x_i,y_j)}{p(x_i)p(y_j)}$.
X et Y sont indépendantes ssi $I(X;Y) = 0$
$0 \leq I(X;Y) \leq \min(H(X), H(Y))$
$0 \leq I(X;Y) \leq I(X;Y,Z)$ où Z est une VA quelconque
Déf. : $0 \leq I(X;Y|Z) = H(X|Z) - H(X|(Y,Z))$
Règle de la chaîne : $I(X^N;Y) = \sum_{i=1}^{N} I(X_i;Y|X^{i-1})$
$I(X;Y) = divKL(p(x,y)||p(x)p(y))$
On peut proposer $I(X;Y;Z) = I(X;Y) - I(X;Y|Z)$, mais peut être négative.

9 Distance entropique

$DH(X,Y) = H(X,Y) - I(X;Y)$ est la distance entropique : DH symétrique, $0 \leq DH$, nulle si $X = Y$, et $D(X,Z) \leq D(X,Y) + D(Y,Z)$.

10 Information mutuelle dirigée

L'information mutuelle directe d'une séquence de variables aléatoires X^N vers Y^N est :
$I(X^N -> Y^N) = \sum_{n=1}^{N} I(X^n; Y_n|Y^{n-1})$

$0 \leq I(X^N -> Y^N) \leq I(X^N; Y^N)$