Formulaire Théorie de l'Information*

1 Notations

Soient X et Y deux Variables Aléatoires (VA) prenant pour valeurs respectives $x_1, ..., x_i, ..., x_r$ et $y_1, ..., y_j, ..., y_s$. On simplifie la notation $P(\text{“}X = x_i\text{”})$ par $P(x_i)$, de distribution de probabilité (ddp) p. On note (X, Y) leur VA jointe. De même (x_i, y_j) représente l'intersection des deux événements. On note $(x_i|y_j)$ l'événement x_i conditionné par y_j. On note X^N la séquence $X_1, ..., X_n, ..., X_N$ de N variables aléatoires. Le log est en base 2, l'unité est le logon ou bit.

2 Probabilités

$P(x_i \cup y_j) = P(x_i) + P(y_j) - P(x_i, y_j)$
Déf. : deux événements x_i et y_j sont incompatibles (disjoints) ssi $P(x_i \cup y_j) = P(x_i) + P(y_j)$.
Déf. : $P(x_i, y_j) = P(X_i).P(y_j|x_i) = P(y_j).P(x_i|y_j)$.
D'où la formule de Bayes : $\frac{P(x_i|y_j)}{P(y_j)} = \frac{P(y_j|x_i)}{P(x_i)}$
Déf. : les événements x_i y_j sont indépendants ssi $P(x_i, y_j) = P(x_i).P(y_j)$.
Déf. : les variables aléatoires X Y sont indépendantes ssi :
$\forall i \in \{1, \ldots, r\}$ $\forall j \in \{1, \ldots, s\}$ x_i et y_j sont indépendants.

3 Incertitude h

$h(x_i) = log\frac{1}{P(x_i)} = -logP(x_i)$
$h(x_i, y_j) = h(x_i) + h(y_j|x_i) = h(y_j) + h(x_i|y_j)$
Cor. : Si x_i et y_j sont indépendants alors $h(x_i, y_j) = h(x_i) + h(y_j)$.
$0 \leq h(x_i) \leq +\infty$

4 Entropie H (incertitude moyenne)

$H(X) = \sum_{i=1}^{r} P(x_i)log\frac{1}{P(x_i)} = -\sum_{i=1}^{r} P(x_i)logP(x_i) = \sum_{i=1}^{r} P(x_i).h(x_i)$
$H(X, Y) = H(Y, X)$
$H(X, Y) = -\sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)logP(x_i, y_j)$
$H(X, Y) = H(X) + H(Y) - I(X; Y)$
$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
X et Y sont indépendantes ssi $H(X) = H(X|Y)$ (alors $H(X, Y) = H(X) + H(Y)$).
L'entropie est maximale lorsque la ddp est uniforme (système aléatoire), elle égale alors log(nombre d'événements possibles=cardinal de l'alphabet).
$0 \leq H(g(X))* \leq H(X) \leq log(r)$, (*) égalité si la fonction g 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 : $H(X^N) = \sum_{n=1}^{N} H(X_n|X^{n-1})$

*Univ. Sud Toulon-Var, H. Glotin

5 Entropie relative (ou divergence de Kullback-Leibler)

La divergence (cross-entropy) des 2 ddp p q est : $0 \leq divKL(p||q) = \sum_x p(x).log\frac{p(x)}{q(x)} \leq \infty$

Prop. : asymétrique. $divKL(p||q) = 0$ ssi $p = q$.

$divKL(p(X,Y)||q(X,Y)) = divKL(p(Y)||q(X)) + divKL(p(X|Y)||q(X|Y))$.

$I(X;Y) = divKL(p(X,Y)||(p(X) * p(Y)))$

$DKL(p,q) = divKL(p||q) + divKL(q||p) = \sum_x (p-q).log(p/q)$ est symétrique mais est nommée à tord distance car $DKL(p,r) \neq DKL(p,q) + DKL(q,r)$.

Le pouvoir discriminant d'une dimension A suivant le critère de 'Maximal Marginal Diversity' peut s'écrire pour K classes : $J(A) = \sum_k p(k).DKL(pdf(k), pdf(moy))$, avec moy le centroïde des K classes dans dim A.

6 Information mutuelle i

$i(X_i, Y_j) = i_{X_i \to Y_j} = i_{Y_j \to X_i} = i(Y_j, X_i)$
$i(X_i, Y_j) = log\frac{1}{P(X_i)} - log\frac{1}{P(X_i|Y_j)} = log\frac{1}{P(Y_j)} - log\frac{1}{P(Y_j|X_i)}$
X_i et Y_j sont indépendants ssi $i(X_i; Y_j) = 0$.
$-\infty \leq i(X,Y) \leq +\infty$.

7 Information mutuelle (moyenne) I

$I(X;Y) = I(Y;X)$.
$I(X;Y) = H(X) + H(Y) - H(X,Y)$.
$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.
$I(X;Y) = \sum_{i=1}^{r} \sum_{j=1}^{s} P(X_i, Y_j) \log \frac{P(X_i,Y_j)}{P(X_i).P(Y_j)}$.
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 : $I(X^N;Y) = \sum_{n=1}^{N} I(X_n;Y|X^{n-1})$
$I(X;Y) = divKL(p(X,Y)||p(X), q(Y))$.

8 Distance entropique

$D_H(X,Y) = H(X,Y) - I(X;Y)$ est nommée distance entropique.
En effet D_H symétrique, $0 \leq D_H$, =0 si $X = Y$, $D(X,Z) \leq D(X,Y) + D(Y,Z)$.

9 Information mutuelle dirigée

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

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