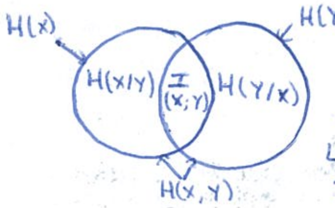
# - Théorie de l'information et Codage

## • Relations

$H(X,Y) = H(X) + H(Y/X)$ → Loi des chaînes
$\quad = H(Y) + H(X/Y)$

$I(X;Y) = H(X) - H(X/Y)$
$\quad = H(Y) - H(Y/X)$
$\quad = H(X) + H(Y) - H(X,Y)$



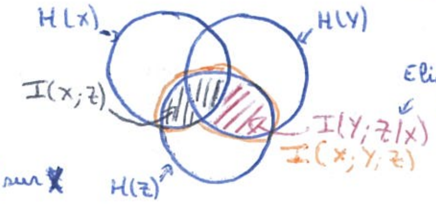

• $I(X;Y/Z) = H(X/Z) - H(X/Y,Z)$
déf de l'information mutuelle conditionnel

• $I(X,Y;Z) = I(X;Z) + I(Y;Z/X)$
Loi des chaînes pour l'information mutuelle
$I(X,Y;Z) = I(Y;Z) + I(X;Z/Y)$

Chaîne de Markov $X \to Y \to Z$

$X \xrightarrow{I(X;Y)} Y \xrightarrow{I(Y;Z)} Z \qquad I(X;Y) \geq I(X;Z)$

Ce qui veut dire que aucun traitement comme ici $Y \to Z$, ne peut augmenter l'information que contient $Y$ sur $X$




Elimine $H(X)$ pour ne pas compter 2 fois la même intersection

## • Formules

$\{x_1, x_2, ...\}$ $M_X$ entrées • $m_i$ = longueur du symbole $x_i$ ; $\{y_1, y_2, ...\}$ $M_Y$ sorties

Quantité d'information : $Q(x_i) = \log_2(1/p(x_i))$ ; Entropie : $H(X) = \sum_i^M p(x_i)\log_2(1/p(x_i))$ ; Taille moyenne : $\bar{m} = \sum_i^M p(x_i) m(x_i)$

$H(X,Y) = \sum_i^{M_X}\sum_j^{M_Y} p(x_i,y_j)\log(1/p(x_i,y_j)) = H(Y,X)$ ; $H(X/Y) = \sum_i^{M_X}\sum_j^{M_Y} p(x_i,y_j)\log(1/p(x_i/y_j))$ ; $P(a/b) = \frac{P(a,b)}{P(b)}$

$H(Y/X) = \sum_i^{M_X}\sum_j^{M_Y} p(x_i,y_j)\log(1/p(y_j/x_i))$ ; $H(X/Y=b) = \sum_i^{M_X} p(x_i/b)\log(1/p(x_i/b))$ ; Relative Entropie : $D(P\|P') = \sum_i^{M_X} p(x_i)\log(P(x_i)/P'(x_i))$

$I(X;Y) = \sum\sum p(x_i,y_j)\log\left(\frac{p(x_i,y_j)}{P(x_i)p(y_j)}\right) = \sum\sum p(x_i,y_j)\log\left(\frac{P(x_i/y_j)}{P(x_i)}\right) = \sum\sum p(x_i,y_j)\log\left(\frac{P(y_j/x_i)}{P(y_j)}\right)$ Théorème de Bayes

Loi marginale : $p(y_1) = p(y_1,x_1) + p(y_1,x_2) + p(y_1,x_3) ...$ ; $p(x_1) = p(x_1,y_1) + p(x_1,y_2) + p(x_1,y_3)$

## • Propriétés

- $Q(x_i) > Q(x_j) \Leftrightarrow P(x_i) < P(x_j)$ - $P(x_i,x_j) = P(x_i)P(x_j) \Leftrightarrow Q(x_i;x_j) = Q(x_i) + Q(x_j)$
- $H(X) = \sum_i^M \frac{1}{M}\log(M) = \log(M)$ si répartition équiprobable - $H(X) \leq \log(M)$ égalité si répartition équipro - $H(X) \leq \bar{m} \leq H(X)+1$ Théorème de Shannon
- $K = \sum_i^n 2^{-m_i} \leq 1$ Inégalité de Kraft : condition d'existence d'un code préfixe - $H(X/Y) \leq H(X)$ conditionnement réduit l'entropie
- $H(Y/X)$ ne dépend pas du canal si celui-ci est symétrique. - $I(X;Y)$ est la quantité moyenne transmise dans le canal
- $H(X/Y)$ est la quantité perdue (appelée aussi ambiguïté) - $I(X;Y) \geq 0$

Si chaîne de Markov $(p(x,y,z) = p(x)p(y/x)p(z/y))$ → $I(X;Y) \geq I(X;Z)$
→ $I(X;Y) \geq I(X;Y/Z)$ } observer $Z$ n'apporte pas plus d'information que d'observer $Y$

Secret Parfait : $H(S/X) = H(S)$ observer $X$ ne donne aucune information sur $S$
$P(X,S) = P(X)P(S)$ Indépendance des variables aléatoires $X$ et $S$
$H(K) \geq H(M)$ Condition de Shannon du secret parfait

- Propriété d'équipartition asymptotique : AEP : $P(X_1, ..., X_m)$ tend vers $2^{-mH(X)}$ quand $m$ tend vers l'infini

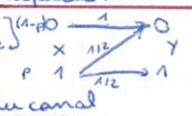
Capacité du canal : $C = \underset{P(X)}{\mathrm{Max}}\, I(X;Y)$ Si canal symétrique la capacité correspond à la répartition équiprobable.

## Terminologie

- Code préfixe : code où le code d'un mot ne peut pas être le début du code d'un autre mot
- Canal symétrique : Pour sa ~~symé~~ matrice de transition, chaque colonne est une permutation d'une autre colonne et chaque ligne est une permutation d'une autre ligne.
- Quantité moyenne transmise dans le signal : $I(X;Y)$ - Quantité perdue (ambiguïté) : $H(X/Y)$
- Théorème de Shannon : $H(X) \leq \bar{m} \leq H(X)+1$ - Conditionnement réduit l'entropie : $H(X/Y) \leq H(X)$
- Espérance : $E[X] = \sum x_i p_i$ - Variance : $Var(X) = E[X^2] - E[X]^2$

## Exercice : Capacité du canal en Z (TD14)

$\Pi = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$




Capacité du canal

$C = \max I(X,Y)$

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

$p(Y=0) = p(Y=0/X=0)p(0) + p(Y=0/X=1)p(1)$
$\quad = 1 - \frac{p}{2}$

$(Y=1) = 1 - p(Y=0) = \frac{p}{2}$

$H(Y) = -\sum p(y)\log_2 p(y)$
$\quad = -\left(\frac{p}{2}\log_2\frac{p}{2} + (1-\frac{p}{2})\log_2(1-\frac{p}{2})\right)$

On calcule $H(Y/X)$ :
$= P(Y=0, X=0) = P(Y=0/X=0)p(0) = 1-p$
$= P(Y=0, X=1) = P(Y=0/X=1)p(1) = P/2$
$= P(Y=1, X=0) = 0$
$= P(Y=1, X=1) = P(Y=1/X=1)p(1) = P/2$

$H(Y/X) = P_1\log(1/P(Y=0/X=0)) + P_2\log(1/P(Y=0/X=1))$
$\quad + P_4\log(1/P(Y=1/X=1))$
$\quad = (1-p)\log_2 1 + \frac{p}{2}\log_2 2 + \frac{p}{2}\log_2 2$
$\quad = 0 + \frac{p}{2} + \frac{p}{2}$
$\quad = P$

$\frac{\partial I(X;Y)}{\partial p} = -\frac{1}{2}\log_2\left(\frac{p}{2}\right) - \frac{p}{2}\times\frac{1}{\ln 2}\times\frac{1}{2}\times\frac{2}{p} - \left[-\frac{1}{2}\log_2(1-\frac{p}{2}) + (1-\frac{p}{2})\frac{1}{\ln 2}\times\frac{-1/2}{(1-\frac{p}{2})}\right] - 1$

$= -\frac{1}{2}\log_2\left(\frac{p}{2}\right) - \frac{1}{2\ln 2} + \frac{1}{2}\log(1-\frac{p}{2}) + \frac{1}{2\ln 2} - 1$

$= -\frac{1}{2}\log_2\left(\frac{p}{2}\right) + \frac{1}{2}\log_2\left(1-\frac{p}{2}\right) - 1$

$= \frac{1}{2}\log_2\left(\frac{1-P/2}{P/2}\right) - 1$

$\mathrm{Max}\, I(X,Y) \Rightarrow \frac{\partial I(X;Y)}{\partial p} = 0$

$\Leftrightarrow \frac{1}{2}\log_2\left(\frac{1-P/2}{P/2}\right) = 1$

$\Leftrightarrow \log\left(\frac{1-P/2}{P/2}\right) = 2$

$\Leftrightarrow \frac{1-P/2}{P/2} = 4$

$\Leftrightarrow P = \frac{2}{5}$

$$\boxed{\begin{array}{l} (uv)' = u'v + uv' \\ (\ln u)' = \frac{u'}{u} \\ \log_2 = \frac{\log x}{\ln 2} \end{array}}$$

$\log a - \log b = \log\left(\frac{a}{b}\right)$

PS* : On trouve $\frac{3}{5}$ si on inverse les probas de base → $p(0) = P$
$p(1) = (1-p)$

---

$\begin{bmatrix} 0,7 & 0,1 \\ 0,1 & 0,7 \end{bmatrix}\begin{bmatrix} 0,2 \\ 0,2 \end{bmatrix}$ → G symétrique
(si On peut sortir deux chiffres identiques et les placer dans une autre matrice)

$\begin{bmatrix} x & y \\ y & x \end{bmatrix}\begin{bmatrix} \alpha \\ \alpha \end{bmatrix}$

xercice 3 TD2 (Démonstration)

$H(X,Y/Z) = H(X/Z) + H(Y/X,Z)$ or $H(Y/X,Z) \geq 0$ donc $\underline{H(X,Y/Z) \geq H(X/Z)}$ VRAI

$H(X/Z) \leq H(Z)$ FAUX Trouver $Z$ tel que $H(Z)=0$ ($Z$ déterministe $p(Z)=1$) ; Trouver $X$ tel que $H(X/Z)>0$ $X, Z$ indépendants $\Rightarrow H(X/Z)=H(X)$ $X$ Bernouilli $(1/2;1/2)$ $H(X)=1$ baud $H(X/Z)=1>H(Z)=0$

Le conditionnement réduit l'entropie $\underline{H(Z/X) \geq H(Z/X,Y)}$

$H(X,Z) = H(X) + H(Z/X) = H(Z) + H(X/Z) \rightarrow H(Z)+H(X/Z) \geq H(X) \rightarrow \underline{H(X/Z) \geq H(X)-H(Z)}$

$X \leftrightarrow Y \leftrightarrow Z$ chaîne de Markov $\rightarrow p(x,y,z) = p(x)p(y/x)p(z/y)$ $\underline{H(X/Y,Z) = H(X/Y)}$ $\rightarrow \underline{H(X/Y,Z) \leq H(X/Z)}$

TD3 exo 2 Nb questions mini pour deviner nbre entre 0 et 63 ? $\log_2 64 = \underline{6}$ (si oui, bit $=1$ ; si non, bit $=0$)

$Q_1$ : $x > 32$ ? $Q_2$ : $x \bmod 32 > 16$ ? $Q_3$ : $x \bmod 16 > 8$ ? $Q_4$ : $x \bmod 8 > 4$ ? $Q_5$ : $x \bmod 4 > 2$ ? $Q_6$ : $x \bmod 2 > 0$ ?

TD6 exo 2 : Conditionnement successif dans une chaîne de Markov

a) $P(X_i/X_1, X_2 \dots X_{i-1}) = p(X_i/X_{i-1})$ car chaîne de Markov

b) $H(X_1, X_2 \dots X_{i-1}) = \sum_i H(X_i/X_{i-1})$ car $\sum p(x_1, x_2 \dots x_{i-1}, x_i) \log_2(1/p(x_i/x_1, \dots x_{i-1}))$ d'après a)

c) $H(X_i/X_{i-1}) = H(X_2/X_1)$ $\forall i \geq 2$ car c'est stationnaire (énoncé) $H(X_1/X_0) = H(X_1) \rightarrow \dfrac{H(X_1, \dots, X_m)}{m} = \dfrac{H(X_1)+(m-1)H(X_2/X_1)}{m} \xrightarrow[m\to\infty]{} H(X_2/X_1)$

TD7 : Inégalité de Fano

1) $H(U,X/Y) = H(U/Y) + H(X/U,Y) = H(X/Y) + H(U/X,Y)$

2) a) $H(U/X,Y) = 0$ car : si on $Y$ on connait $\hat{X}$, puis avec $X$ et $\hat{X}$ on a $U$. On a donc $H(U/U)$ -> ça ne nous apprend rien.

b) $H(U/Y) \leq H(U)$ Conditionnement réduit l'entropie

c) $H(X/Y) = H(U/Y) + H(X/U,Y)$ avec a) et b) on a $H(X/Y) \leq H(U) + H(X/U,Y)$

d) $H(X/U,Y) = -\sum_u \sum_{x,y} p(u,x,y) \log_2 p(x/y,u) = -\sum_u p(u) \sum_{x,y} p(x,y/u) \log_2 p(x/y,u) = -p(u=0) \sum_{x,y} p(x,y/u=0) \log_2 p(x/y,u=0)$

ou $= -P(u=1) \sum_{x,y} p(x,y/u=1) \log_2 p(x/y,u=1)$ d'où $H(X/U,Y) = p(u=0)H(X/Y,u=0) + p(u=1)H(X/Y,u=1)$

e) $P_e = P(u=1)$ f) $H(X/Y,u=0) = 0$ car $Y$ -> on connait $\hat{X}$, $\hat{X}$ et $u$ -> on connait $X$

$H(X/Y,u=1) \leq \log_2(|\mathcal{X}|-1)$ -> $u=1$ donc $X \neq \hat{X}$, on a $|\mathcal{X}|-1$ valeurs possibles -> pire cas : valeurs équiprobables $\left\{ \begin{array}{l} H(X/Y,u=1) \\ \leq \log_2(|\mathcal{X}|-1) \end{array} \right.$

g, h, i, j) $H(X/U,Y) = p(u=0)\underbrace{H(X/Y,u=0)}_{0} + \underbrace{p(u=1)}_{P_e}\underbrace{H(X/Y,u=1)}_{\leq \log_2(|\mathcal{X}|-1)} \leq P_e \log_2(|\mathcal{X}|-1)$

Entropie binaire $H(X/Y) \leq \underbrace{H(u)}_{= H_2(P_e) \leq 1} + H(X/U,Y) \rightarrow H(X/Y) \leq 1 + P_e \log_2(|\mathcal{X}|-1) \rightarrow P_e \geq \dfrac{H(X/Y)-1}{\log_2(|\mathcal{X}|-1)}$

TD9 exercice 1 Information mutuelle

$I(X;Y_1,Y_2) = I(X;Y_1) + I(X;Y_2/Y_1)$ (chain rule) $= I(X;Y_1) + H(Y_2/Y_1) - H(Y_2/X;Y_1)$ ($Y_2$ and $Y_1$ indep conditionnellement à $X$)

$= I(X;Y_1) + H(Y_2/Y_1) - H(Y_2/X) = I(X;Y_1) + H(Y_2/Y_1) - H(Y_2/X) + H(Y_2) - H(Y_2) = I(X;Y_1) + [H(Y_2)-H(Y_2/X)] - [H(Y_2)-H(Y_2/Y_1)]$

$= I(X;Y_1) + I(X;Y_2) - I(Y_1/Y_2)$ CQFD

TD10 L'information mutuelle moyenne comme quantité d'information transmissible dans un canal

1) $P(X_i,Y_j) = P(Y_j/X_i)P(X_i) \rightarrow P(Y=0) = P(Y=0,X=0) + P(Y=0,X=1) = (1-p)P_0 + p(1-P_0) = p + P_0(1-2p)$

$P(Y=1) = 1 - P(Y=0) = 1 - p - P_0(1-2p)$

2) $H(Y/X) = -\sum_x \sum_y P(X_i,Y_j) \log(P(Y_j/X_i)) = -(1-p)P_0 \log(1-p) - p(1-P_0)\log(p) - pP_0\log(p) - (1-p)(1-P_0)\log(1-p)$

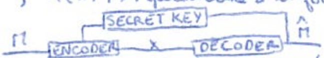
$= -p\log(p) - (1-p)\log(1-p)$ 3) $H(Y/X)$ ne dépend que de $p$, donc du canal car c'est un canal symétrique

$I(X;Y) = H(Y) - H(Y/X) = H(Y) - h(p)$

4) $I(X;Y) = H(Y) - h(p)$ -> maximale si $H(Y)$ maximale -> $H(Y)$ max si $P(Y=0) = P(Y=1) = \frac{1}{2}$ -> $p + P_0(1-2p) = \frac{1}{2} \Leftrightarrow P_0 = \frac{1}{2}$

5) $I(X;Y) = H(X) - H(X/Y)$ -> $H(X)$ : quantité d'information transmise ; $H(X/Y)$ : quantité d'information perdue

6) AN $I(X;Y)$ max -> $I(X;Y) = 1 - h(p)$




$X = f(K,M)$

TD11 Sécurité parfaite ($M$ : message confidentiel, $K$ : clé de cryptage)

$H(K) \geq H(K) - H(K/X,M)$ car $H(\,)$ non négatif -> $H(K) \geq H(K/X) - H(K/X,M)$ car conditionnement réduit entropie -> $H(K) \geq I(K;M/X)$

-> $H(K) \geq H(M/X) - \underbrace{H(M/X,K)}_{0}$ -> $H(K) \geq H(M/X)$ car savoir $X$ et $K$ permet de décoder $M$ -> $H(K) \geq H(M) \Rightarrow$ Sécurité parfaite

TD12 One-time Pad ($M$ et $K$ : variables aléatoires indépendantes) ($X = K \oplus M$ : $\oplus$ addition modulo $|M|$)

Ex : $M = \{0;1;2;3\}$ $|M| = 4$ [ $M=2$, $K=3 \Rightarrow X = 2 \oplus 3 = 1$ Montrer que $I(M,X) = 0 \Leftrightarrow H(M) = H(M/X)$

$P(X/K) = \sum_m P(X,m/K) = \sum_m P(m/K)P(X/K,m)$ or $P(X/K,m) = 1$ ssi $X = K \oplus m$, alors $P(X/K) = P(m/K)$

$P(X) = \sum_k P(X/K)P(K) = \frac{1}{|M|} \sum_k P(X/K) = \frac{1}{|M|} \sum_m P(m) = \frac{1}{|M|}$ -> $H(M) = H(K)$ -> $P(X/m) = P(K/m) = P(K)$ d'où $H(X/M) = H(K)$

## Correction QROC 2015

### Calcul de quantités informatives

$X = \{A, Z, E, R, T, Y\}$ 140 symb

A -> 32, Z -> 8, E -> 64, R -> 16, 16 -> T, 4 -> Y

) $H(X) = \frac{1}{140}\left(32 \times \log_2 \frac{140}{32} + 8 \times \log_2 \frac{140}{8} \dots\right)$

$= \log_2(140) - \frac{1}{140}(32 \times \log_2(32) + \dots)$

$= \log_2(140) - \frac{1}{140}(32 \times 5 + 8 \times 3 \dots)$

$= 7{,}1293 - 5{,}0285 = 2{,}1007$ bits/symb

| | $A_s$ | $Z_s$ | $E_s$ | $R_s$ | $T_s$ | $Y_s$ |
|---|---|---|---|---|---|---|
| A | 16 | 16 | | | | |
| Z | 2 | 4 | 2 | | | |
| E | | 16 | 32 | 16 | | |
| R | | | 4 | 8 | 4 | |
| T | | | | 4 | 8 | 4 |
| Y | | | | | 2 | 2 |

) $H(X,X_s) = \log_2(140) - \frac{1}{140}(1 \times 32 \times 5 + 4 \times 16 \times 4 \dots)$

$= 7{,}1293 - 3{,}6571 = 3{,}4722$ bits/symb

$H(X,X_s) = H(X) + H(X_s/X) \Rightarrow H(X_s/X) = 1{,}3715$

) $I(X;X_s) = H(X_s) - H(X_s/X)$ (Valeurs : énoncé)

$= 2{,}3863 - 1{,}3715 = 1{,}0148$

) Tableau $P(X_s|X) \rightarrow P(X_s|X) = \dfrac{P(X_s,X)}{P(X)}$

) $I(X;X_s) = H(X) - H(X/X_s)$

$\Rightarrow \underbrace{H(X/X_s)}_{\text{pertes du canal (erreurs de frappe)}} = H(X) - \underbrace{I(X;X_s)}_{\text{Q d'info qui passe à travers le canal}} = 1{,}0859$

) Pertes tolérées à 50% $\rightarrow \frac{H(X)}{2} = 1{,}05$

$H(X/X_s) = 1{,}0859 > 50\%$

Ambiguïté > 50%

### 2 Inégalités en théorie de l'information

$X, Y$ v.a $x_1, x_2 \dots x_m$ ; $y_1, y_2 \dots y_n$

$Z = X + Y$

1) $H(Z/X) = H(Y/X)$ ??

$H(Z,Y/X) = H(Z/X) + H(Y/X,Z) = H(Y/X) + H(Z/X,Y)$

Sachant $Z = X+Y$, connaître $Z$ et $X$ détermine $Y$, $Y$ et $X$ détermine $Z$

d'où $H(Y/X,Z) = H(Z/X,Y) = 0$

et donc $H(Z/X) = H(Y/X)$ Vraie

2) Si $X$ et $Y$ indép alors $H(Z) \geq H(Y)$ et $H(Z) \geq H(X)$ ??

Si $X$ et $Y$ indép -> $H(Y/X) = H(Y)$

d'après 1) $H(Z/X) = H(Y)$

or comme le conditionnement réduit l'entropie

$H(Z) \geq H(Z/X) = H(Y)$

d'où $H(Z) \geq H(Y)$ CQFD

Pour $H(Z) \geq H(X)$ :

On reprend à partir de la question 1) en développant cette fois $H(Z,X/Y)$ puis on suit les mêmes étapes (il suffit d'inverser $X$ et $Y$ !)

### 3 Capacité de canaux en cascade

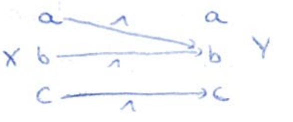
A $P(Y/X)$

| X \ Y | a | b | c |
|---|---|---|---|
| a | 0 | 1 | 0 |
| b | 0 | 1 | 0 |
| c | 0 | 0 | 1 |

B $P(Z/Y)$

| Y \ Z | a | b | c |
|---|---|---|---|
| a | 1 | 0 | 0 |
| b | 0 | 1 | 0 |
| c | 0 | 1 | 0 |

1) Le canal n'est pas symétrique




Capacité $C_A$ du canal A de matrice de transition $P_{Y/X}$ :

Soit $P_X(a), P_X(b), P_X(c) = 1 - (P_X(a) + P_X(b))$

$P_Y(a) = 0$ ; $P_Y(b) = P_X(a) + P_X(b)$ ; $P_Y(c) = P_X(c)$

donc $H(Y) = (P_X(a)+P_X(b))\log\dfrac{1}{P_X(a)+P_X(b)} + P_X(c)\log\dfrac{1}{P_X(c)}$

Il est facile de voir que $H(Y/X) = 0$

$C_A$ est donc obtenue en cherchant la distribution qui maximise $H(Y) - H(Y/X) = H(Y) \rightarrow P_Y(b) = P_Y(c) = \frac{1}{2}$

donc : $P_X(a) + P_X(b) = \frac{1}{2}$ et $P_X(c) = \frac{1}{2}$

avec ces valeurs -> $C_A = 1$ bit/symb

2) Capacité $C_B$ du canal B avec matrice de transition $P_{Z/Y}$

Même raisonnement avec $P_X(a) = \frac{1}{2} = P_X(b) + P_X(c)$

3) Capacité du canal A-B (en cascade A et B)

$P_Z(a) = 0$, $P_Z(b) = 1$, $P_Z(c) = 0$ donc $H(Z) = 0$

et comme $H(Z/X) = 0$ -> capacité $C_{A-B} = 0$ bits/symb

4) Si on peut observer simultanément $Y$ et $Z$ :

comme on observe $Y$, on n'a plus besoin de regarder $Z$ et la capacité $C_{A-B} = C_A = 1$ bit/symb

$$\boxed{\begin{aligned} H(X,Y) &= \sum_i^x \sum_j^y p(Y_j/X_i)P(X_i) \log \frac{1}{P(Y_j;X_i)} \\ &= \sum_i^x \sum_j^y P(X_i/Y_j) P(Y_j) \log \frac{1}{P(X_i,Y_j)} \end{aligned}}$$

d'après le Théorème de Bayes