Théorie de l'Information

Devoir # 2 (Résolution)

SIC-SICOM

Maria-João Rendas

1. Considérez une loi de probabilité d'une variable aléatoire $X \in \mathcal{X}$, où $|\mathcal{X}| = m$: $p = [p_1, p_2, \ldots, p_m]$. Montrez que

$$H(p_1, p_2, \ldots, p_m) = H(p_1+p_2, p_3, \ldots, p_n)+(p_1+p_2)H\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right).$$

Interprétez cette expression.

résolution

$$\begin{aligned}
H(p) &= H(p_1, \ldots, p_m) - (p_1+p_2)\log(p_1+p_2) + (p_1+p_2)\log(p_1+p_2) \\
&= H(p_1+p_2, p_3, \ldots, p_m) \\
&\quad -p_1\log p_1 - p_2 \log p_2 + (p_1+p_2)\log(p_1+p_2) \\
&= H(p_1+p_2, p_3, \ldots, p_m) \\
&\quad +(p_1+p_2)\left[-\frac{p_1}{p_1+p_2}\log p_1 - \frac{p_2}{p_1+p_2}\log p_2 + \frac{p_1+p_2}{p_1+p_2}\log(p_1+p_2)\right] \\
&= H(p_1+p_2, p_3, \ldots, p_m) \\
&\quad +(p_1+p_2)\left[-\frac{p_1}{p_1+p_2}\log\frac{p_1}{p_1+p_2} - \frac{p_2}{p_1+p_2}\log\frac{p_2}{p_1+p_2}\right] \\
&= H(p_1+p_2, p_3, \ldots, p_m) \\
&\quad +(p_1+p_2)H\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right).
\end{aligned}$$

Cette relation peut être interprétée de la façon suivante: la quantité d'information fournie par une expérience qui peut avoir m résultats distincts est égale à la quantité d'information obtenue dans une expérience composée, en deux étapes:

(a) dans la première étape, les deux premiers résulats, que nous nottons x_1, x_2 ne sont pas discernés (ils sont confondus dans un évènement $x' = x_1 \cup x_2$ dont la probabilité est $p_1 + p_2$). La quantit é d'information dans chaque expérience de cette nouvelle variable est $H(p_1 + p_2, p_3, \ldots, p_m)$.

(b) Si le résultat de la première expérience est x', alors une deuxième variable aléatoire nous identifie lequel des deux évènements (x_1 ou x_2) s'est

effectivement produit. La loi de probabilité du résultat de cette deuxième variable est

$$q = p(x|x') = \left[\frac{p_1}{p+1+p_2}, \frac{p_2}{p+1+p_2}\right].$$

La quantité d'information totale est la somme de la quantité d'information associée à la première expérience, $H(p_1+p_2, p_3, \ldots, p_m)$, plus la quantité d'information moyenne de la deuxième expérience, qui est égale à 0 avec probabilité $1-p_1-p_2$, et égale à $H(q)$ avec probabilité p_1+p_2, c'est à dire,

$$H(p) = H(p_1+p_2, p_3, \ldots, p_m) + (p_1+p_2)H(q).$$

2. La définition suivante précise la relation de dispersion plus ou moins élevée des composantes d'un vecteur.

Définition 1 *Majoration*
Soit $x, y \in R^n$ et nottons

$$x_{[1]}, \ldots, x_{[n]}$$

le vecteur obtenu en ordonnant les composantes de x par ordre décroissante. On dira que x est majoré par y si

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \qquad k = 1, \ldots, n-1,$$

et

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}.$$

Montrez que pour toute loi de probabilité $p = [p_1, \ldots, p_n]$,

$$p \text{ majore } \left[\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right].$$

résolution
La deuxième condition est trivialement vérifiée pour des lois de probabilité, qui doivent avoir une somme égale à 1. Nous nous concentrons alors dans la vérification de

$$\sum_{i=1}^{k} \frac{1}{n} = \frac{k}{n} \leq \sum_{i=1}^{k} p_{[i]}, \qquad k = 1, \ldots, n-1 \tag{1}$$

où nous avons considéré déjà la loi uniforme.
Pour simplifier la présentation, nous allons admettre que la loi p est ordonnée, de façon que

$$p_i = p_{[i]}, \qquad i = 1, \ldots, n.$$

Si nous définissons

$$S_k = \sum_{i=1}^{k} p_i - \frac{k}{n},$$

la condition (1) est équivalente à

$$S_k \geq 0, \qquad k = 1, \ldots, n-1.$$

Comme p est une loi de probabilité

$$1 = \sum_{i=1}^{k} p_i + \sum_{i \geq k+1} p_i \overset{(a)}{\leq} \sum_{i=1}^{k} p_i + (n-k)p_{k+1}$$

où en (a) nous avons utilisé le fait que

$$i \geq k+1 \Rightarrow p_i \leq p_{k+1},$$

car la loi p est "ordonnée". Nous pouvons donc écrire

$$\begin{aligned} 1 \quad &= \quad \frac{k}{n} + \frac{n-k}{n} \leq \sum_{i=1}^{k} p_i + (n-k)p_{k+1} \\ &\Rightarrow \quad \frac{n-k}{n} - (n-k)p_{k+1} \leq \sum_{i=1}^{k} p_i - \frac{k}{n} = S_k \\ &\Rightarrow \quad p_{k+1} - \frac{1}{n} \geq -\frac{S_k}{n-k}. \end{aligned}$$

Comme la séquence S_k satisfait la récursion

$$S_{k+1} = S_k + p_{k+1} - \frac{1}{n}, \qquad k = 1, \ldots, n-1,$$

nous pouvons conclure

$$S_{k+1} \geq S_k - \frac{S_k}{n-k} = S_k \frac{n-k-1}{n}.$$

Pour $k \leq n-1$ le facteur qui multiplie S_k dans cette inégalité est toujours positif et donc S_{k+1} doit avoir le même signe que S_k. Si $S_1 \geq 0$, nous pouvons alors garantir que $S_k \geq 0, k = 1, \ldots, k-1$. La preuve que S_1 est positif est obtenue directement du fait que la somme de p est unitaire, et que p_1 est le plus grand élément de p:

$$1 = \sum_{i=1}^{n} p_i \leq np_1 \Rightarrow p_1 \geq \frac{1}{n} \Rightarrow S_1 \geq 0.$$

3. Soit p une loi de probabilité dans le simplex probabiliste de dimension n, telle que

$$p_1 \geq p_2 \geq \cdots \geq p_n.$$

Montrez que

$$H(p_1 - \delta, p_2 + \delta, p_3, \ldots, p_n) \geq H(p).$$

résolution

Considérons que

$$0 \leq \delta \leq \frac{p_1 - p_2}{2} \Leftrightarrow p_1 - \delta \geq p_2 + \delta. \tag{2}$$

Sous cette condition,

$$p_1 - \delta \geq p_2 + \delta \geq p_2,$$

et il est immédiat de vérifier que la loi ordonnée

$$p^{(a)} = [p_1 - \delta, p_2 + \delta, p_3, \cdots, p_n]$$

est majorée par p:

$$\sum_{i=1}^{k} p_i \geq \sum_{i=1}^{k} p_i^{(a)}, k = 1, \ldots, n-1, \qquad \sum_{i=1}^{n} p_i \geq \sum_{i=1}^{n} p_i^{(a)}$$

La deuxième égalité est nécessairement vérifiée, du fait que p et $p^{(a)}$ sont des lois de probabilité. Pour $k = 1$, car $\delta \geq 0$,

$$p_1 \geq p_1 - \delta = p_1^{(a)}$$

Pour $k \geq 2$

$$\begin{aligned} \sum_{i=1}^{k} p_i = p_1 + p_2 + \sum_{i=3}^{k} p_i &= p_1 - \delta + p_2 + \delta + \sum_{i=3}^{k} p_i \\ &= p_1^{(a)} + p_2^{(a)} + \sum_{i=3}^{k} p_i^{(a)} = \sum_{i=1}^{k} p_i^{(a)}. \end{aligned}$$

Nous pouvons utiliser le fait que $H(\cdot)$ est une fonction Schur-concave pour affirmer que

$$p \text{ majore } p^{(a)} \Rightarrow H(p^{(a)}) \geq H(p).$$

Ce résultat est vrai pour δ dans l'intervalle de l'équation (2), de façon que $p^{(a)}$ soit effectivement une loi ordonnée. Il suffit de voir que pour le cas extrême où $\delta = p_1 > (p_1 + p_2)/2$, et donc $p_1^{(a)} = 0$, l'entropie sera nécessairement inférieure :

$$H(p) = H(p^{(a)}) + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) \geq H(p^{(a)}).$$

4. Montrez que

$$H\left(\sum_{i=1}^{k} p_1, 1 - \sum_{i=1}^{k} p_1\right) \leq H(p).$$

Interprétez en termes des partitions de Ω correspondantes aux deux lois de probabilité, p et $\left(\sum_{i=1}^{k} p_1, 1 - \sum_{i=1}^{k} p_1\right)$.
Sous quelle condition peut-on avoir égalité ?

résolution
Soit $X \sim p$ une variable aléatoire qui suit la loi p (et qui a donc entropie $H(X) = H(p)$). Soit $\mathcal{X} = \{a_1, \ldots, a_n\}$ l'ensemble où X prend ses valeurs, avec $n > k$. Considérez la variable aléatoire binaire Y définie à partir de X de la façon suivante:

$$Y = \begin{cases} 1, & X \in \{a_1, \ldots, a_k\} \\ 0, & x \in \{a_{k+1}, \ldots, a_n\} \end{cases}$$

Avec cette définition

$$q_1 = \Pr\{Y = 1\} = \sum_{i=1}^{k} p_k, \qquad q_0 = \Pr\{Y = 0\} = 1 - q_1.$$

Donc, l'entropie indiquée dans le problème est l'entropie de Y:

$$H\left(\sum_{i=1}^{k} p_1, 1 - \sum_{i=1}^{k} p_1\right) = H(Y).$$

Comme $Y = f(X)$

$$H(X, Y) = H(X) = H(p).$$

Mais, pas la régle de la chaîne pour l'entropie,

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

De ces deux inégalités nous concluons directement

$$H(Y) \leq H(p).$$

La résolution fourni directement l'interprétation demandée en termes des partitions de Ω correspondante aux variables aléatoires X et Y : La partition correspondante à X est un raffinement de celle correspondante à Y, et donc elle doit conduire à une variable aléatoire avec un eplus grande entropie.

5. Montrez que

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

Sous quelle condition on observera l'égalité ?
résolution
Par la régle de la chaîne

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z) \leq H(X|Z) + H(Y|Z).$$

où l'inégalité découle du fait que le conditionnement réduit l'entropie, et donc

$$H(Y|X,Z) \leq H(Y|Z).$$

Nous aurons égalité quand, sachant Z, Y est indépendante de X :

$$p(y|x,z) = p(y|z).$$

Par exemple, si $X = S_{n-1}; Z = S_n$ et $Y = S_{n+1}$ et S_n est un processus de Markov d'ordre 1, nous aurons égalité:

$$H(S_{n_1}, S_{n+1}|S_n) = H(S_{n-1}|S_n) + H(S_{n+1}|S_n)$$

qui montre bien le rôle de X_n comme "cut set" entre le futur $S_i, i > n$ et le pass é $S_i, i < n$.

6. Montrez que l'information mutuelle $I(X;Y)$ est

   (a) Une fonction concave de p_X.

   (b) Une fonction convexe de $p_{Y|X}$.

résolution

(a) Nous voulons montrer que pour $p(y|x)$ fixé, I est une fonction concave de $p(x)$:

$$I(\lambda p_1(x)+(1-\lambda p_2(x); p(y|x)) \geq \lambda I(p_1(x); p(y|x))+(1-\lambda)I(p_2(x); p(y|x).$$

Nous savons que

$$I(X;Y) = H(Y) - H(Y|X).$$

Soit

$$p_\lambda(x) = \lambda p_1(x) + (1-\lambda)p_2(x)$$

Alors, pour une loi conditionnelle fixée, $p(y|x)$,

$$\begin{aligned} p_\lambda(y) &= \sum_y p(y|x)p_\lambda(x) \\ &= \lambda \sum_y p(y|x)p_1(x) + (1-\lambda)\sum_y p(y|x)p_2(x) \\ &= \lambda p_1(y) + (1-\lambda)p_2(y) \end{aligned}$$

où les définitions de $p_1(y)$ et $p_2(y)$ sont évidentes. Alors,

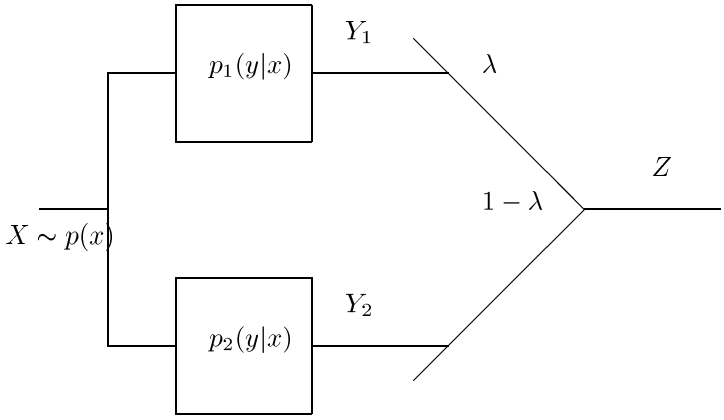
$$H_\lambda(Y) = H(p_\lambda(y)) = H\left(\lambda p_1(y) + (1-\lambda)p_2(y)\right)$$

qui est une fonction concave de λ (Propriété 19 dans la première partie des notes).
Le deuxième terme est, par définition d'entropie conditionnelle :

$$H_\lambda(Y|x) = -\sum_x p_\lambda(x) \sum_y p(y|x) \log p(y|x)$$

et donc il est une fonction linéaire de p_λ pour $p(y|x)$ fixé. Nous avons donc la somme d'une fonction linéaire avec une fonction concave, qui doit être necessairement concave.

(b) Pour démontrer la convexité de I comme fonction de la loi conditionnelle $p(y|x)$ pour $p(x)$ fixé, nous allons considérer le schéma de la figure suivante



Soit Z la variable (indépendante de X) qui indique la position de l'intérrupteur, de façon que

$$Pr\{Z = Y_1\} = \lambda \qquad Pr\{Z = Y_2\} = 1 - \lambda$$

Alors,

$$\begin{aligned} I(X;Y,Z) &= I(X;Y|Z) + I(X;Z) \\ &= I(X;Z|Y) + I(X;Y) \end{aligned}$$

Mais $I(X;Z) = 0$ - car Z est statistiquement indépendant de X – et donc nous pouvons conclure que

$$I(X;Z|Y)+I(X;Y) = I(X;Y|Z) \Rightarrow I(X;Y) = I(X;Y|Z)-I(X;Z|Y) \leq I(X;Y|Z)$$

Mais, par la définition de Z

$$I(X;Y|Z) = \lambda I(X,Y_1) + (1 - \lambda 0 I(I,Y_2)$$

Nous venons donc de montrer que

$$I(p(x); \lambda p_1(y|x)+(1-\lambda)p_2(y|x)) \leq \lambda I(p(x); p_1(y|x))+(1-\lambda)I(p_2(y|x))$$

c'est à dire, que I est une fonction convexe de $p(y|x)$ pour $p(x)$ fixé.

7. Cet exercice généralise l'information mutuelle à des groupes de variables. Il introduit en particulier la notion d'information mutuelle entre deux variables X et Y, conditionnée dans une troisième variable Z, et l'information mutuelle entre une collection de variables (X, Y) et une autre variable Z:

$$\begin{aligned} I(X;Y|Z) &= \mathrm{E}_X \left\{ D\left(p(x,y|z)||p(x|z)p(y|z)\right)\right\} \\ I(X,Y;Z) &= D\left(p(x,y,z)||p(x,y)p(z)\right), \end{aligned}$$

où $D(\cdot||\cdot)$ est l'entropie relative entre deux lois de probabilité, introduite en cours. $I(X;Y|Z)$ est l'information mutuelle entre X et Y sachant Z, et $I(X,Y;Z)$ est l'information mutuelle entre le vecteur (X,Y) et Z.
Montrez que les décompositions suivante sont vraies :

$$I(X;Z) + I(X;Y|Z) = I(X;Y) + I(X;Z|Y) = I(X;Y,Z).$$

résolution
En utilisant les définitions

$$\begin{aligned} I(X;Z) + I(X;Y|Z) &= \sum_{x,z} p(x,z) \log \frac{p(x,z)}{p(x)p(z)} + \sum_z p(z) \sum_{x,y} p(x,y|z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \\ &\overset{(a)}{=} \sum_{x,z} \sum_y p(x,y,z) \log \frac{p(x,z)}{p(x)p(z)} + \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \\ &= \sum_{x,y,z} p(x,y,z) \left[\log \frac{p(x,z)}{p(x)p(z)} + \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \right] \\ &= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,z)p(x,y|z)}{p(x)p(z)p(x|z)p(y|z)} \\ &\overset{(b)}{=} \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x)p(y|z)} \\ &= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)p(z)}{p(x)p(y|z)p(z)} \\ &\overset{(c)}{=} \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)}{p(x)p(y,z)} \\ &= \mathrm{E}_{x,y,z} \left[\log \frac{p(x,y,z)}{p(x)p(y,z)} \right] = I((X;Y,Z)\,. \end{aligned}$$

où les différents pas ont les justifications suivantes:

$(a) : p(x,z) = \sum_y p(x,y,z)$,

$(b) : p(z)p(x|z) = p(x,z)$.

$(c) : p(x,y|z)p(z) = p(x,y,z), p(y|z)p(z) = p(y,z)$.

8. Considérez les définitions de l'exercice précédant. Montrez que si X et Y sont statistiquement indépendantes sachant Z, alors

$$I(X;Z) \geq I(X;Z|Y).$$

Interprétez.

résolution

Le fait que X et Y sont statistiquement indépendants sachant Z implique que

$$p(x,y|z) = p(x|z)p(y|z).$$

De la définition d'information mutuelle :

$$\begin{aligned}
I(X;Z) &= \sum_{x,z} p(x,z) \log \frac{p(x,z)}{p(x)p(z)} \\
&= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,z)p(y|z)}{p(x)p(z)p(y|z)} \\
&= \sum_{x,y,z} p(x,y,z) \log \frac{p(x|z)p(z)p(y|z)}{p(x)p(z)p(y|z)} \\
&= \sum_{x,y,z} p(x,y,z) \log \frac{p(x|z)p(y|z)}{p(x)p(y|z)} \\
&= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x)p(y|z)} \\
&\overset{(a)}{\leq} \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} = I(X;Y|Z).
\end{aligned}$$

où en (a) nous avons utilisé le fait que $\forall z$

$$p(x) \geq p(x|z) \Rightarrow \frac{1}{p(x)} \leq \frac{1}{p(x|z)} \Rightarrow \log \frac{1}{p(x)} \leq \log \frac{1}{p(x|z)}.$$