Inégalités informationnelles et groupes finis

Firas Kraïem, Master 1 CSI, Université Bordeaux 1
Encadré par Christine Bachoc

18 avril 2013

Introduction

Il existe en théorie de l'information des limites aussi fondamentales que le sont la vitesse de la lumière en physique ou le problème de l'arrêt en informatique. Ces limites se présentent sous la forme d'inégalités faisant intervenir les grandeurs informationnelles que l'on peut définir à partir d'un ensemble de variables aléatoires, et notamment leurs entropies respectives. Un problème majeur en théorie de l'information est donc de mettre au jour ces inégalités, afin d'avoir une idée précise de ce qui est possible et de ce qui ne l'est pas, à la fois dans un sens théorique et dans un sens pratique, par exemple lors d'applications au codage de source ou de canal, ou à la cryptographie.

La théorie de l'information étant une discipline jeune (on considère généralement qu'elle fait son apparition avec un article de Shannon publié en 1948 [8]), elle manque encore d'outils puissants permettant de prouver aisément des inégalités de grandeurs informationnelles, et l'on doit souvent avoir recours pour cela à la théorie élémentaire des probabilités et à des calculs fastidieux. Le but de ce TER est de présenter un résultat de Chan et Yeung [1], qui fait apparaître un lien entre ces inégalités et la théorie des groupes finis. Ce résultat permet de ramener l'étude d'une inégalité informationnelle à l'étude d'un problème de théorie des groupes, discipline bien plus ancienne et disposant de nombreux résultats potentiellement exploitables pour prouver de nouvelles inégalités informationnelles.

Afin d'être accessible à des non-spécialistes en théorie de l'information, on commencera au premier chapitre par présenter les grandeurs informationnelles (information, entropie) sur lesquelles portent les résultats présentés, puis par présenter plusieurs inégalités informationnelles dont la signification est aisée à appréhender intuitivement et qui nous serviront d'exemples dans les chapitres suivants. Au deuxième chapitre, on définira les inégalités informationnelles de façon plus rigoureuse qu'au premier, en utilisant les *fonctions d'entropie*. En effet, la recherche d'inégalités informationnelles a pour but ultime de caractériser entièrement l'ensemble des fonctions d'entropie, et c'est donc cet ensemble et les fonctions qui le constituent qui nous intéresseront au plus haut point. Enfin, au troisième chapitre, on montrera comment construire des fonctions d'entropie à partir d'un groupe fini et de ses sous-groupes, et on montrera que, s'il n'est pas possible de construire ainsi

toutes les fonctions d'entropie, il est suffisant, si l'on veut montrer qu'une inégalité informationnelle donnée est vraie pour toute fonction d'entropie, de montrer qu'elle l'est pour toutes les fonctions d'entropie que l'on peut construire ainsi.

Table des matières

Chapitre 1

Théorie de l'information

Dans ce chapitre on donne les résultats élémentaires de théorie de l'information qui permettront au lecteur non-spécialiste d'appréhender les parties suivantes. Le lecteur intéressé pourra ensuite se référer à [2], ou à [11] pour une présentation plus synthétique. On suppose connus les concepts élémentaires de la théorie des probabilités : variable aléatoire, loi de probabilités, etc. Dans toute la suite, les logarithmes sont en base 2.

1.1 Information et entropie

Étant donnée une variable aléatoire X définie sur un ensemble $\mathcal{X}$ et de loi de probabilités p, l'*information* (ou *quantité d'information*) apportée par l'événement $X = x$, pour un certain $x \in \mathcal{X}$, est mesurée en *bits* (ou parfois en *shannons*), et vaut

$$\log \frac{1}{p(X = x)} \text{ bits}\,. \tag{1.1}$$

Intuitivement, l'information apportée par un événement est d'autant plus grande qu'il est difficile de le prévoir *a priori*. L'information est donc nulle si l'événement est certain (*i.e.*, si $p(X = x) = 1$), et on définit de façon analogue comme nulle l'information apportée par un événement impossible (*i.e.*, tel que $p(X = x) = 0$).

Exemple 1.1. Si la variable X représente un lancer d'une pièce équilibrée et si on donne à l'une des deux faces la valeur 0 et à l'autre la valeur 1, l'information apportée par l'événement $X = 0$ (ou $X = 1$) vaut

$$\log \frac{1}{1/2} = \log 2 = 1 \text{ bit}\,.$$

Si maintenant la pièce est déséquilibrée de sorte que la face 0 tombe avec probabilité 3/4, l'information apportée par l'événement $X = 0$ vaut

$$\log \frac{1}{3/4} = \log \frac{4}{3} \approx 0,4 \,\text{bit},$$

alors que l'information apportée par l'événement $X = 1$ vaut

$$\log \frac{1}{1/4} = \log 4 = 2 \,\text{bits}\,.$$

L'événement $X = 1$ apporte plus d'information que l'événement $X = 0$ car il est plus difficle à prévoir, dans le sens où si on prévoit que $X = 1$, on se trompe avec probabilité 3/4, alors que si on prévoit $X = 0$ on se trompe avec probabilité 1/4 seulement.

L'*entropie* de la variable X peut être définie comme l'espérance de la quantité d'information qu'apporte la connaissance de la valeur de X. Elle est notée $H(X)$ ou $H(p)$, et vaut

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} \tag{1.2}$$

$$= -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{1.3}$$

Intuitivement, l'entropie d'une variable aléatoire X est une mesure de la difficulté de deviner la valeur de X. Elle est donc nulle si X prend une valeur donnée de façon certaine (*i.e.*, avec probabilité 1), et maximale si X peut prendre toutes ses valeurs avec la même probabilité (on dit alors que la loi de probabilité de X est *uniforme*). Dans ce cas, l'entropie de X vaut

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} \tag{1.4}$$

$$= \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{1/|\mathcal{X}|} \tag{1.5}$$

$$= \log |\mathcal{X}| \underbrace{\sum_{x \in \mathcal{X}} p(x)}_{=1} \tag{1.6}$$

$$= \log |\mathcal{X}|. \tag{1.7}$$

Exemple 1.2. Comme à l'exemple précédent, X représente le lancer d'une pièce équilibrée. On a alors

$$H(X) = \log 2 = 1 \,\text{bit}\,.$$

Si maintenant on a $X = 0$ avec probabilité $3/4$, on obtient

$$H(X) = \frac{3}{4} \log \frac{4}{3} + \frac{1}{4} \log 4 \approx 0,8 \,\text{bit}\,.$$

Dans le second cas, la valeur de X est plus facile à deviner que dans le premier, dans le sens où la probabilité de se tromper peut être rendue plus faible (elle est de $1/4$ si on devine que $X = 0$, alors que dans le premier cas elle est toujours de $1/2$).

On remarque que l'entropie d'une variable aléatoire dépend uniquement de sa loi de probabilités, et pas des valeurs qu'elle prend. En effet, si dans l'exemple précédent c'est la face 1 et pas la face 0 qui tombe avec probabilité $3/4$, l'entropie de X est inchangée.

Étant données deux variables aléatoires X et Y, définies respectivement sur les ensembles $\mathcal{X}$ et $\mathcal{Y}$, on définit l'*entropie jointe* de X et Y, notée $H(X, Y)$, comme l'entropie de la variable représentant le couple (X, Y). Formellement,

$$H(X, Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(X = x, Y = y) \log \frac{1}{p(X = x, Y = y)}. \tag{1.8}$$

La variable (X, Y) représentant à la fois X et Y, l'entropie jointe $H(X, Y)$ représente la difficulté de deviner les deux valeurs (celle de X et celle de Y). On définit également l'*entropie conditionnelle* $H(X|Y)$ par

$$H(X|Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(X = x, Y = y) \log \frac{1}{P(X = x|Y = y)}. \tag{1.9}$$

De même que la probabilité conditionnelle $p(X = x|Y = y)$ est la probabilité d'avoir $X = x$ en sachant *a priori* que $Y = y$, l'entropie conditionnelle $H(X|Y)$ mesure la difficulté de deviner la valeur de X en connaissant *a priori* la valeur de Y.

Exemple 1.3. On lance un dé équilibré. La variable X représentant la face obtenue est à valeurs dans $\mathcal{X} = \{1, \ldots, 6\}$. La variable Y à valeurs dans $\mathcal{Y} = \{0, 1\}$ vaut 0 si X est paire, et 1 sinon. On a évidemment $H(Y|X) = 0$, car si on connaît la valeur de X, on connaît automatiquement celle de Y. On n'a par contre pas $H(X|Y) = 0$: $H(X|Y)$ vaut $\log 3$, car une fois connue la valeur de Y, on a trois valeurs possibles pour X, de mêmes probabilités.

1.2 Inégalités informationnelles élémentaires

Dans cette section, nous donnons uniquement des justifications intuitives aux inégalités énoncées. Des preuves mathématiquement plus rigoureuses se trouvent dans [2] et [11].

On a d'abord pour toute variable aléatoire X,

$$H(X) \geq 0. \tag{1.10}$$

Cette inégalité est aisée à vérifier mathématiquement : l'entropie est une somme de termes tous positifs. Intuitivement, l'entropie représente la difficulté de deviner la valeur d'une variable aléatoire, et donc avoir une entropie négative n'aurait pas grand sens. On a également

$$H(X) \leq \log |\mathcal{X}|, \tag{1.11}$$

car $\log |\mathcal{X}|$ est la valeur de l'entropie d'une variable aléatoire quand celle-ci est la plus difficile possible à deviner (quand elle prend toutes ses valeurs avec la même probabilité).

Étant données deux variables aléatoires X et Y, on a d'abord $H(X,Y) \geq 0$ et $H(X|Y) \geq 0$, puisque l'entropie est toujours positive. On a également

$$H(X,Y) = H(X) + H(Y|X), \tag{1.12}$$

car deviner la valeur de X puis, connaissant la valeur de X, deviner la valeur de Y revient à deviner la valeur de X et la valeur de Y. De plus,

$$H(X) \geq H(X|Y), \tag{1.13}$$

car il ne peut évidemment pas être plus difficile de deviner la valeur de X quand on connaît celle de Y que quand on ne la connaît pas. Cela peut par contre être plus facile, si Y apporte une quantité non-nulle d'information sur X. La quantité d'information que Y apporte sur X vaut

$$H(X) - H(X|Y). \tag{1.14}$$

Cette quantité est appelée l'*information mutuelle* entre X et Y, et notée $I(X,Y)$. On a alors l'inégalité

$$I(X,Y) \geq 0. \tag{1.15}$$

Par parenthèse, cette inégalité est importante en cryptographie pour évaluer le degré de confidentialité fourni par un système cryptographique : si X désigne le message en clair et Y le message chiffré, on veut que la connaissance de Y apporte à un adversaire le moins d'information possible sur X. Idéalement, on préfère qu'elle n'en apporte aucune, c'est-à-dire avoir

$$H(X) = H(X|Y). \tag{1.16}$$

Un système cryptographique qui possède cette propriété est dit *à confidentialité parfaite*. Pour plus de détails sur les applications de la théorie de l'information en cryptographie, on pourra consulter [10, section 2.4] ou [9, chapitre 2].

Exemple 1.4. Dans la situation de l'exemple 1.3, on a

$$I(X,Y) = \log 6 - \log 3 = \log 2.$$

À partir des inégalités précédentes, on peut déduire plusieurs autres inégalités. Par exemple à partir de (1.12) on obtient

$$H(X|Y) = H(X,Y) - H(Y), \tag{1.17}$$

et en remplaçant $H(X|Y)$ par cette expression dans (1.14) on obtient une autre expression de l'information mutuelle :

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \geq 0, \tag{1.18}$$

d'où l'on peut également déduire

$$H(X,Y) \leq H(X) + H(Y). \tag{1.19}$$

Intuitivement, cette dernière inégalité signifie qu'il ne peut pas être plus difficile de deviner à la fois X et Y (par exemple deviner d'abord X puis, connaissant X, deviner Y) que de les deviner indépendamment. De même que précédemment, cela peut par contre être plus facile, si X fournit une quantité non-nulle d'information sur Y (ou inversement), et alors on a encore une fois $I(X,Y) > 0$. Enfin, à partir de 1.12 on déduit

$$H(X,Y) \geq H(X), \tag{1.20}$$

car on a nécessairement $H(Y|X) \geq 0$. En effet, il ne peut évidemment pas être plus facile de deviner X et Y que de deviner uniquement X.

Chapitre 2

Inégalités informationnelles

Dans ce chapitre nous étudions les inégalités informationnelles de façon plus précise qu'au chapitre précédent, en utilisant les *fonctions d'entropie*, ainsi que le vocabulaire de la topologie des espaces vectoriels normés.

2.1 Fonctions d'entropie

Soient n variables aléatoires discrètes $X_1, \ldots, X_n$, $\mathcal{N} = \{1, \ldots, n\}$, et Ω l'ensemble des parties non-vides de $\mathcal{N}$. Pour tout $\alpha \in \Omega$, on définit la variable aléatoire X_α comme la variable jointe correspondant à la famille de variables $(X_i)_{i\in\alpha}$.

Définition 2.1. Une fonction h définie sur Ω et à valeurs réelles est une *fonction d'entropie* s'il existe n variables aléatoires $X_1, \ldots, X_n$ telles que pour tout $\alpha \in \Omega$, on a $h(\alpha) = H(X_\alpha)$. On note Γ_n^* l'ensemble des fonctions d'entropie (pour un n donné).

Exemple 2.1. Soit $n = 2$. On a alors $\mathcal{N} = \{1, 2\}$ et $\Omega = \{\{1\}, \{2\}, \{1, 2\}\}$. La fonction h définie sur Ω par $h(\{1\}) = h(\{1, 2\}) = \log 6$ et $h(\{2\}) = \log 2$ est une fonction d'entropie. (Elle correspond à la situation de l'exemple 1.3 si on pose $X_1 = X$ et $X_2 = Y$.)

Une fonction d'entropie est donc une fonction à valeurs réelles définie sur l'ensemble Ω à $2^n - 1$ éléments. L'ensemble des fonctions à valeurs réelles définies sur Ω est un $\mathbf{R}$-espace vectoriel de dimension $2^n - 1$, que l'on peut identifier à $\mathbf{R}^{2^n-1}$. Cet espace peut alors être muni de la métrique euclidienne usuelle, et comme Γ_n^* en est une partie, on peut l'étudier par des outils géométriques et topologiques.

La structure de Γ_n^* est complexe. En particulier, Zhang et Yeung [12] montrent que pour $n \geq 3$, il n'est pas fermé. On étudie alors son *adhérence* (*i.e.*, le plus petit fermé qui le contient), notée $\overline{\Gamma}_n^*$. On rappelle les deux définitions suivantes [3, pp. 19,21-22] :

Définition 2.2. Une partie A de $\mathbf{R}^k$ est *convexe* si pour tous $\mathbf{x}, \mathbf{y} \in A$ et tout réel $\lambda \in]0;1[$, on a $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in A$. Géométriquement, cela signifie que tout point du segment $[\mathbf{x};\mathbf{y}]$ est dans A.

Définition 2.3. Une partie A de $\mathbf{R}^k$ est un *cône* si pour tout $\mathbf{x} \in A$ et tout réel $\lambda > 0$, on a $\lambda\mathbf{x} \in A$.

On définit alors naturellement les *cônes convexes*. Nous pouvons maintenant énoncer un premier résultat important, dû à Zhang et Yeung [12] :

Théorème 2.1. *$\overline{\Gamma}_n^*$ est un cône convexe.*

Preuve. Montrons d'abord la convexité. Soit $\lambda \in]0;1[$ et $\lambda' = 1-\lambda$. On commence par montrer que pour tous $g, h \in \Gamma_n^*$, on a $\lambda g + \lambda' h \in \overline{\Gamma}_n^*$, en montrant que $\lambda g + \lambda' h$ est la limite d'une suite de fonctions de Γ_n^*. Soient donc $Y_1, \ldots, Y_n$ et $Z_1, \ldots, Z_n$ deux familles de n variables aléatoires induisant respectivement les fonctions d'entropie $g, h \in \Gamma_n^*$. Pour tout $i \in \{1, \ldots, n\}$ et tout entier positif k, soient $\tilde{Y}_i^k$ et $\tilde{Z}_i^k$ des variables aléatoires définies respectivement comme k copies indépendantes de Y_i et de Z_i. Soit enfin une variable aléatoire U indépendante de toute autre variable aléatoire et définie pour $\delta, \mu \in]0;1[$ par

$$P(U=1) = \delta,\ P(U=2) = \mu,\ P(U=0) = 1-\delta-\mu.$$

On remarque que $H(U)$ peut être rendu arbitrairement petit en prenant δ, μ arbitrairement petits (ce qui a pour effet de rendre $P(U=0)$ arbitrairement proche de 1). On construit les variables aléatoires $X_1, \ldots, X_n$ par

$$X_i = \begin{cases} 0, & \text{si } U = 0 \\ \tilde{Y}_i^k, & \text{si } U = 1 \\ \tilde{Z}_i^k, & \text{si } U = 2. \end{cases}$$

On a alors d'une part

$$\begin{aligned} H(X_\alpha) &\leq H(X_\alpha, U) \\ &\leq H(U) + H(X_\alpha|U) \\ &\leq H(U) + \delta H(\tilde{Y}_\alpha^k) + \mu H(\tilde{Z}_\alpha^k) \\ &\leq H(U) + \delta k H(Y_\alpha) + \mu k H(Z_\alpha), \end{aligned}$$

et d'autre part

$$\begin{aligned} H(X_\alpha) &\geq H(X_\alpha|U) \\ &\geq \delta k H(Y_\alpha) + \mu k H(Z_\alpha). \end{aligned}$$

Il vient alors

$$H(U) \geq H(X_\alpha) - (\delta k H(Y_\alpha) + \mu k H(Z_\alpha)) \geq 0,$$

et en posant $\delta = \lambda/k$ et $\mu = \lambda'/k$:

$$H(U) \geq H(X_\alpha) - (\lambda H(Y_\alpha) + \lambda' H(Z_\alpha)) \geq 0.$$

δ et μ peuvent être rendus arbitrairement petits en prenant k assez grand, et donc $H(U)$ peut également être rendu arbitrairement petit, ce qui signifie que $H(X_\alpha)$ peut être rendu arbitrairement proche de $\lambda g + \lambda' h$. $\lambda g + \lambda' h$ est donc dans $\overline{\Gamma}_n^*$ car c'est la limite de la suite des fonctions d'entropie induites par $X_1, \ldots, X_n$ quand $k \to \infty$ (et c'est donc un *point adhérent* à Γ_n^* [5, p. 153-155]). Enfin, si g et h sont deux éléments de $\overline{\Gamma}_n^*$, ils sont les limites respectivement de deux suites (g_k) et (h_k) d'éléments de Γ_n^*, et $\lambda g + \lambda' h$ est la limite de la suite $\lambda(g_k) + \lambda'(h_k)$. C'est une suite d'élements de $\overline{\Gamma}_n^*$, et donc $\lambda g + \lambda' h$ est dans l'adhérence de $\overline{\Gamma}_n^*$, qui est $\overline{\Gamma}_n^*$ lui-même car $\overline{\Gamma}_n^*$ est fermé.

On a donc montré que $\overline{\Gamma}_n^*$ est convexe. Pour montrer que c'est un cône, on commence par remarquer que la fonction nulle est dans Γ_n^* (et donc dans $\overline{\Gamma}_n^*$) car elle est induite par n variables aléatoires quelconques d'entropies toutes nulles. Ensuite, on remarque que pour toute fonction d'entropie $f \in \Gamma_n^*$, on a $kf \in \Gamma_n^*$ pour tout entier k : en notant $(X_1, \ldots, X_n)$ une famille de n variables aléatoires induisant la fonction f, on construit k familles $(X_{1,1}, \ldots, X_{1,n}), \ldots, (X_{k,1}, \ldots, X_{k,n})$ comme k copies indépendantes de $(X_1, \ldots, X_n)$, et on définit n variables aléatoires $Y_1, \ldots, Y_n$ par $Y_i = (X_{j,i})_{1 \leq j \leq k}$. La famille de variables aléatoires $(Y_1, \ldots, Y_n)$ induit alors la fonction kf, et par passage à la limite on voit que si $f \in \overline{\Gamma}_n^*$, on peut construire une suite de fonctions de Γ_n^* qui converge vers kf, ce qui montre que $kf \in \overline{\Gamma}_n^*$. Enfin, soient $f \in \overline{\Gamma}_n^*$ et un réel $\lambda > 0$. En posant $k = \lceil \lambda \rceil$, λf se trouve sur le segment $[\mathbf{0}; kf]$, et puisque les deux extrémités de ce segment sont dans $\overline{\Gamma}_n^*$ et que $\overline{\Gamma}_n^*$ est convexe, on a $\lambda f \in \overline{\Gamma}_n^*$. □

2.2 Généralités sur les inégalités informationnelles

Une inégalité informationnelle sur n variables aléatoires $X_1, \ldots, X_n$ peut se représenter sous la forme

$$\sum_{\alpha \in \Omega} b_\alpha h(\alpha) \geq 0, \tag{2.1}$$

où la fonction h est la fonction d'entropie induite par les variables $X_1, \ldots, X_n$, et les b_α sont des réels. On peut donc construire les vecteurs $\mathbf{b} = (b_\alpha)_{\alpha \in \Omega}$ et $\mathbf{h} = (h(\alpha))_{\alpha \in \Omega}$, qui sont deux vecteurs de $\mathbf{R}^{2^n - 1}$. Ainsi une inégalité informationnelle peut également se représenter sous la forme

$$\mathbf{b}^{\mathrm{T}} \mathbf{h} \geq 0, \tag{2.2}$$

ou encore en utilisant le produit scalaire usuel sur $\mathbf{R}^k$:

$$\langle \mathbf{b}, \mathbf{h} \rangle \geq 0. \tag{2.3}$$

Exemple 2.2. Soit $n = 2$. On a alors $\Omega = \{\{1\}, \{2\}, \{1, 2\}\}$. Le vecteur $\mathbf{b}$ défini par $\mathbf{b}_{\{1\}} = \mathbf{b}_{\{2\}} = 1$ et $\mathbf{b}_{\{1,2\}} = -1$ définit l'inégalité (1.18).

Une inégalité informationnelle est donc représentée par un vecteur $\mathbf{b} \in \mathbf{R}^{2^n-1}$. On peut alors lui associer ([3, pp. 19-20]) un *hyperplan* de $\mathbf{R}^{2^n-1}$:

$$H_{\mathbf{b}} = \{\mathbf{x} \in \mathbf{R}^{2^n-1} | \langle \mathbf{b}, \mathbf{x} \rangle = 0\}, \tag{2.4}$$

et un *demi-espace* :

$$H_{\mathbf{b}}^+ = \{\mathbf{x} \in \mathbf{R}^{2^n-1} | \langle \mathbf{b}, \mathbf{x} \rangle \geq 0\}. \tag{2.5}$$

On s'intéresse aux inégalités vérifiées par toutes les fonctions d'entropie, c'est-à-dire aux vecteurs $\mathbf{b} \in \mathbf{R}^{2^n-1}$ tels que l'inégalité 2.3 soit vraie pour toute fonction d'entropie h (*i.e.*, pour tout vecteur $\mathbf{h} \in \Gamma_n^*$). Autrement dit, tels que Γ_n^* est inclus dans le demi-espace $H_{\mathbf{b}}^+$. Dans ce cas, comme $H_{\mathbf{b}}^+$ est fermé, il vient par passage à l'adhérence que le cône convexe $\overline{\Gamma}_n^*$ est inclus dans $H_{\mathbf{b}}^+$, comme illustré sur la figure 2.1.

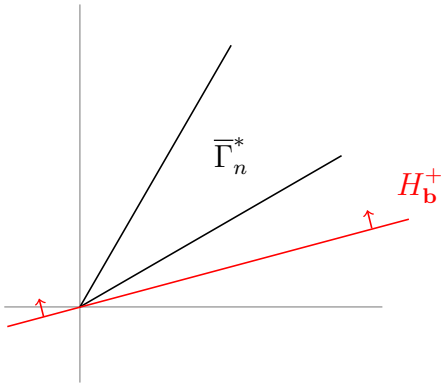

FIGURE 2.1 – Le cône convexe $\overline{\Gamma}_n^*$ est inclus dans le demi-espace $H_{\mathbf{b}}^+$.

Afin de caractériser le plus précisément possible le cône convexe $\overline{\Gamma}_n^*$, on souhaite donc trouver des inégalités telles que l'hyperplan correspondant (la droite rouge sur la figure 2.1) soit le plus proche possible de $\overline{\Gamma}_n^*$.

2.3 Les inégalités de Shannon

Les inégalités informationnelles les plus simples, dont certaines ont été présentées à la section 1.2, sont appelées les *inégalités de Shannon*, et correspondent à

la positivité des grandeurs informationnelles élémentaires : entropie jointe, entropie conditionnelle, information mutuelle et information mutuelle conditionnelle. Nous reprenons ces inégalités, en les exprimant cette fois dans le vocabulaire de la section 2.1.

- **Entropie jointe :** pout tout $\alpha \in \Omega$, on a

$$H(X_\alpha) = h(\alpha) \geq 0.$$

- **Entropie conditionnelle :** pour tous $\alpha, \beta \in \Omega$, on a

$$\begin{aligned} H(X_\alpha | X_\beta) &= H(X_\alpha, X_\beta) - H(X_\beta) \\ &= h(\alpha \cup \beta) - h(\beta) \\ &\geq 0. \end{aligned}$$

- **Information mutuelle :** pour tous $\alpha, \beta \in \Omega$, on a

$$\begin{aligned} I(X_\alpha, X_\beta) &= H(X_\alpha) + H(X_\beta) - H(X_\alpha, X_\beta) \\ &= h(\alpha) + h(\beta) - h(\alpha \cup \beta) \\ &\geq 0. \end{aligned}$$

- **Information mutuelle conditionnelle :** pour tous $\alpha, \beta, \gamma \in \Omega$, on a

$$\begin{aligned} I(X_\alpha; X_\beta | X_\gamma) &= H(X_\alpha, X_\gamma) + H(X_\beta, X_\gamma) - H(X_\alpha, X_\beta, X_\gamma) - H(X_\gamma) \\ &= h(\alpha \cup \gamma) + h(\beta \cup \gamma) - h(\alpha \cup \beta \cup \gamma) - h(\gamma) \\ &\geq 0. \end{aligned}$$

Si l'on définit de plus $X_\varnothing$ comme une variable aléatoire qui prend une valeur donnée de façon certaine (de sorte qu'on a $h(\varnothing) = 0$), la fonction h est définie sur l'ensemble des parties de $\{1, \ldots, n\}$, et on peut alors retrouver les quatre inégalités ci-dessus à partir de la seule information mutuelle conditionnelle :

- **Entropie jointe :** si $\alpha = \beta$ et $\gamma = \varnothing$, l'information mutuelle conditionnelle $I(X_\alpha; X_\beta | X_\gamma)$ devient

$$I(X_\alpha; X_\alpha | X_\varnothing) = I(X_\alpha, X_\alpha) = H(X_\alpha).$$

- **Entropie conditionnelle :** si $\alpha = \beta$, $I(X_\alpha; X_\beta | X_\gamma)$ devient

$$I(X_\alpha; X_\alpha | X_\gamma) = H(X_\alpha | X_\gamma).$$

- **Information mutuelle :** si $\gamma = \varnothing$, $I(X_\alpha; X_\beta | X_\gamma)$ devient

$$I(X_\alpha; X_\beta | X_\varnothing) = I(X_\alpha, X_\beta).$$

Les inégalités de Shannon donnent donc $(2^n)^3 = 8^n$ inégalités informationnelles sur n variables aléatoires, une pour chaque triplet (α, β, γ) de $\mathcal{P}(\{1, \ldots, n\})^3$. (Le nombre d'inégalités « utiles » est en fait inférieur, car des doublons et des inégalités triviales apparaissent.)

On note Γ_n l'ensemble des vecteurs de $\mathbf{R}^{2^n-1}$ qui vérifient toutes les inégalités de Shannon. Γ_n étant défini uniquement par des inégalités informationnelles, c'est l'intersection de tous les demi-espaces correspondants, et c'est donc un cône fermé convexe. C'est même un cône *polyédral*, car c'est l'intersection d'un nombre *fini* de demi-espaces [3, pp. 56-57]. Comme toute fonction d'entropie vérifie les inégalités de Shannon, Γ_n contient Γ_n^* et par passage à l'adhérence il contient également le cône fermé convexe $\overline{\Gamma}_n^*$, comme illustré sur la figure 2.2.

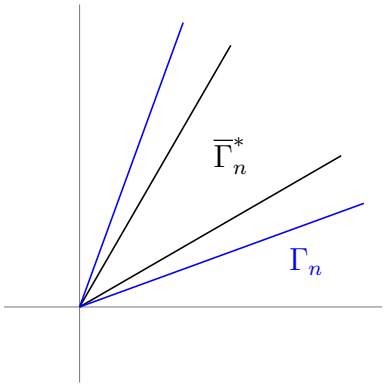

FIGURE 2.2 – Le cône convexe Γ_n contient le cône convexe $\overline{\Gamma}_n^*$.

2.4 Des inégalités plus fortes

On se demande alors si Γ_n n'est pas en fait égal à $\overline{\Gamma}_n^*$, voire à Γ_n^*. Zhang et Yeung [12] montrent que pour $n = 2$, on a $\Gamma_2 = \Gamma_2^*$ (qui est donc également $\overline{\Gamma}_2^*$, puisque Γ_2 est fermé). Pour $n = 3$, ils montrent également que $\Gamma_3 = \overline{\Gamma}_3^* \neq \Gamma_3^*$ (ce qui montre que Γ_3^* n'est pas fermé). Cela implique que des inégalités informationnelles plus fortes que les inégalités de Shannon (*i.e.*, qui sont vraies pour toutes les fonctions de Γ_n^* mais pas pour toutes les fonctions de Γ_n) ne peuvent exister que pour $n \geq 4$. En effet, si une telle inégalité existe, Γ_n^* est inclus dans le demi-espace qu'elle délimite. Comme ce demi-espace est fermé, il contient également $\overline{\Gamma}_n^*$, mais comme il ne contient pas Γ_n, on a finalement $\overline{\Gamma}_n^* \neq \Gamma_n$.

Réciproquement, si $\overline{\Gamma}_n^* \neq \Gamma_n$, il existe une fonction h de Γ_n qui n'est pas dans $\overline{\Gamma}_n^*$. Comme $\overline{\Gamma}_n^*$ est fermé convexe, il existe un hyperplan séparant h et $\overline{\Gamma}_n^*$ [3, pp. 51-52]. Cet hyperplan séparateur définit alors une inégalité informationnelle plus forte que les inégalités de Shannon. La figure 2.3 illustre ce phénomène.

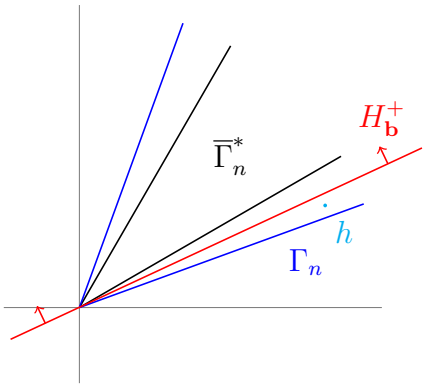

FIGURE 2.3 – L'inégalité représentée par l'hyperplan $H_{\mathbf{b}}$ séparant h et $\overline{\Gamma}_n^*$ est plus forte que les inégalités de Shannon.

On a donc $\overline{\Gamma}_n^* \neq \Gamma_n$ si et seulement si il existe des inégalités plus fortes que les inégalités de Shannon. On se demande donc si de telles inégalités existent. Zhang et Yeung ont commencé par montrer dans [12] que si X, Y, Z, U sont quatre variables aléatoires telles que

$$I(X;Y) = I(X;Y|Z) = 0, \tag{2.6}$$

alors on a l'inégalité

$$I(U;Z|X) + I(U;Z|Y) - I(U;Z) \geq 0. \tag{2.7}$$

Cette inégalité est plus forte que les inégalités de Shannon, car la fonction h définie sur $\mathcal{P}(\{X, Y, Z, U\})$ par

$$\begin{aligned}
h(\varnothing) &= 0, \\
h(X) &= h(Y) = h(Z) = h(U) = 2a > 0, \\
h(X,Y) &= 4a, \\
h(X,U) &= h(X,Z) = h(Y,U) \\
&= h(Y,Z) = h(Z,U) = 3a, \\
h(X,Y,Z) &= h(X,Y,U) = h(X,Z,U) \\
&= h(Y,Z,U) = h(X,Y,Z,U) = 4a
\end{aligned}$$

vérifie les inégalités de Shannon ainsi que l'égalité (2.6), mais pas l'inégalité (2.7) (et n'est donc pas une fonction d'entropie). Cela montre que Γ_4^* est strictement inclus dans Γ_4. Cependant, comme l'inégalité (2.7) n'est vraie que pour certaines fonctions d'entropie, elle ne correspond pas à la situation de la figure 2.3, et on ne peut donc pas en conclure que $\overline{\Gamma}_4^*$ est strictement inclus dans Γ_4.

Un an plus tard, Zhang et Yeung [13] donnent l'inégalité informationnelle suivante sur quatre variables aléatoires X_1, X_2, X_3, X_4 :

$$I(X_1, X_2)+I(X_1; X_3, X_4)+3I(X_3; X_4|X_1)+I(X_3; X_4|X_2)-2I(X_3; X_4) \geq 0. \quad (2.8)$$

Cette inégalité est vraie pour toute fonction d'entropie, mais pas pour toutes les fonctions de Γ_4 (elle n'est pas vérifiée par exemple par la fonction h du paragraphe précédent, ceci sera vérifié numériquement à l'annexe A). C'est donc une inégalité plus forte que les inégalités de Shannon, ce qui implique que $\overline{\Gamma}_4^* \neq \Gamma_4$. Ce résultat se généralise à tout $n \geq 4$. De plus, Matúš [6] a montré que $\overline{\Gamma}_n^*$ n'est *pas* polyédral, ce qui signifie qu'il existe une infinité de telles inégalités.

Chapitre 3

Inégalités informationnelles et groupes finis

Dans ce chapitre, on montre comment construire des fonctions d'entropie à partir de groupes finis, et comme dans la section précédente on étudie certaines propriétés de l'ensemble des fonctions d'entropie que l'on peut construire ainsi. En particulier, on montre qu'une inégalité informationnelle est vraie pour toute fonction d'entropie si et seulement si elle est vraie pour toutes les fonctions d'entropie que l'on peut construire ainsi. Les résultats présentés ici sont dus à Chan et Yeung [1].

3.1 Construction de fonctions d'entropie à partir de groupes finis

On rappelle que pour un entier positif n, on note $\mathcal{N} = \{1, \ldots, n\}$, et Ω l'ensemble des parties non-vides de $\mathcal{N}$. Soit G un groupe fini, $\alpha \in \Omega$, et $(G_i)_{i\in\alpha}$ une famille de sous-groupes de G. L'intersection des G_i est encore un sous-groupe de G, qu'on note G_α. Soit maintenant une famille $(a_iG_i)_{i\in\alpha}$ de classes des G_i dans G. On a le résultat suivant :

Proposition 3.1. *L'intersection des a_iG_i est soit vide, soit de cardinal égal au cardinal de G_α.*

Preuve. On suppose que l'intersection des a_iG_i est non-vide. Soit $b \in \bigcap_{i\in\alpha} a_iG_i$, cela signifie que b est dans la classe a_iG_i de G_i dans G pour tout i, et donc que les

classes a_iG_i et bG_i sont en fait la même classe. Ainsi on a

$$\begin{aligned}\bigcap_{i\in\alpha} a_iG_i &= \bigcap_{i\in\alpha} bG_i \\ &= b\bigcap_{i\in\alpha} G_i \\ &= bG_\alpha,\end{aligned}$$

et donc l'intersection des a_iG_i est une classe de G_α dans G, et est de même cardinal que G_α. □

Nous pouvons maintenant énoncer :

Théorème 3.1. *Soit G un groupe fini, et $G_1, \ldots, G_n$ une famille de n sous-groupes de G. Pour tout $\alpha \in \Omega$, on pose*

$$h(\alpha) = \log \frac{|G|}{|G_\alpha|}. \tag{3.1}$$

Alors h est une fonction d'entropie.

Preuve. Soit Λ une variable aléatoire discrète à valeurs dans G et de loi uniforme. Pour tout $i \in \{1, \ldots, n\}$, on définit la variable X_i comme une variable à valeurs dans l'ensemble des classes de G_i dans G, et on pose $X_i = aG_i$ si $\Lambda = a$. On veut alors montrer que pour tout $\alpha \in \Omega$, on a

$$H(X_\alpha) = \log \frac{|G|}{|G_\alpha|}.$$

On cherche donc à calculer $H(X_\alpha)$, et pour cela on cherche à déterminer la loi de probabilités de X_α. Pour tout $\alpha \in \Omega$, la valeur de X_α est déterminée par les valeurs des X_i pour $i \in \alpha$, c'est-à-dire par la famille des a_iG_i telle que $X_i = aG_i = a_iG_i$ pour tout i. On cherche alors, pour toute famille a_iG_i de classes des G_i dans G, la probabilité d'avoir $aG_i = a_iG_i$ pour tout i. Cela se produit si et seulement si $a \in a_iG_i$ pour tout i, et donc si $a \in \bigcap_{i\in\alpha} a_iG_i$. On a montré (proposition 3.1) que le cardinal de $\bigcap_{i\in\alpha} a_iG_i$ est soit nul, soit égal à $|G_\alpha|$. a étant choisi uniformément dans G, cela signifie que la probabilité d'avoir $aG_i = a_iG_i$ pour tout i vaut $|G_\alpha|/|G|$ si l'intersection des a_iG_i est non-vide, et 0 sinon. La loi de probabilités de X_α est donc uniforme si on se restreint aux valeurs de probabilité non-nulle, et on obtient le résultat voulu. □

Exemple 3.1. Soit $G = \mathbf{Z}/12\mathbf{Z} = \{0, \ldots, 11\}$, $G_1 = \{0, 6\}$, et $G_2 = \{0, 2, \ldots, 10\}$. On a alors $G_1 \cap G_2 = G_1$. La construction par la formule (3.1) donne

$$h(\{1\}) = \log \frac{|G|}{|G_1|} = \log \frac{12}{2} = \log 6,$$

et de façon analogue on obtient $h(\{2\}) = \log 2$ et $h(\{1,2\}) = \log 6$. (Ceci correspond encore à la situation des exemples 1.3 et 2.1.)

3.2 Propriétés

Pour un entier positif n, on note Υ_n l'ensemble des fonctions d'entropie que l'on peut construire à partir d'un groupe fini G et d'une famille de n sous-groupes de G selon (3.1). On a évidemment $\Upsilon_n \subseteq \Gamma_n^*$. Les deux propositions suivantes montrent que l'inclusion réciproque est fausse (*i.e.*, qu'il n'est pas possible de construire ainsi toutes les fonctions d'entropie).

Proposition 3.2. *Υ_n est dénombrable.*

Preuve. Une fonction $h \in \Upsilon_n$ peut être vue comme un vecteur $(x_\alpha)_{\alpha\in\Omega} \in \mathbf{R}^{2^n-1}$. On a alors pour tout $\alpha \in \Omega$

$$x_\alpha = \log \frac{|G|}{|G_\alpha|}.$$

Comme $|G|$ et $|G_\alpha|$ sont entiers, $|G|/|G_\alpha|$ est rationnel, et Υ_n peut donc être mis en bijection avec une partie de $\mathbf{Q}^{2^n-1}$, qui est dénombrable. □

Proposition 3.3. *Γ_n^* est indénombrable.*

Preuve. Soit $h \in \Gamma_n^*$, et $X_1, \ldots, X_n$ des variables aléatoires induisant la fonction d'entropie h. On va montrer que $h(\{1\}) = H(X_1)$ peut prendre un nombre indénombrable de valeurs. Soit $\mathcal{X}_1$ l'ensemble des valeurs de la variable X_1. On se restreint au cas où $\mathcal{X}_1 = \{0,1\}$, et on note $p_0 = P(X_1 = 0)$. On a alors

$$H(X_1) = -(p_0 \log p_0 + (1-p_0)\log(1-p_0)).$$

On sait que $H(X_1)$ est minimal si $p_0 = 0$ et maximal si $p_0 = 1/2$, et qu'on a alors respectivement $H(X_1) = 0$ et $H(X_1) = \log 2 = 1$. Il est de plus aisé de vérifier que la fonction $H(X_1)$ de la variable réelle p_0 est continue, et prend donc toutes les valeurs réelles entre 0 et 1. Comme il y a un nombre indénombrable de réels entre 0 et 1, cela montre que Γ_n^* est indénombrable. □

Nous aurons également besoin du résultat suivant :

Proposition 3.4. *Pour tout entier n, la fonction nulle est dans Υ_n.*

Preuve. Soit n un entier quelconque, G un groupe fini quelconque, et posons $G_1, \ldots, G_n = G$. On a alors $G_\alpha = G$ pour tout α, et donc

$$h(\alpha) = \log \frac{|G|}{|G_\alpha|} = \log \frac{|G|}{|G|} = \log 1 = 0.$$ □

Pour une partie A de $\mathbf{R}^k$, on note $\mathrm{conv}(A)$ l'*enveloppe convexe* de A, qui est la plus petite partie convexe de $\mathbf{R}^k$ qui contient A. On note également $\overline{\mathrm{conv}}(A)$ l'adhérence de $\mathrm{conv}(A)$ (qui est alors la plus petite partie fermée convexe de $\mathbf{R}^k$ qui contient A) [3, p. 31]. Le résultat suivant fait le lien entre la structure de Υ_n et celle de Γ_n^*.

Théorème 3.2.

$$\overline{\mathrm{conv}}(\Upsilon_n) = \overline{\Gamma}_n^*.$$

Pour montrer ce résultat nous avons besoin du lemme suivant, dont la preuve est donnée à l'annexe B.

Lemme 3.1. *Soit X une variable aléatoire à valeurs dans un ensemble $\mathcal{X}$ fini et telle que pour tout $x \in \mathcal{X}$, $p(x) = p(X = x)$ est rationnel de dénominateur q fixé. Alors on a pour tout multiple r de q*

$$H(X) - |\mathcal{X}|\frac{\log(r+1)}{r} \leq \frac{1}{r}\log\frac{r!}{\prod_{x\in\mathcal{X}}(rp(x))!} \leq H(X).$$

Preuve. Voir annexe B. □

Une conséquence de ce résultat est que

$$\lim_{r\to\infty}\frac{1}{r}\log\frac{r!}{\prod_{x\in\mathcal{X}}(rp(x))!} = H(X).$$

Nous pouvons maintenant prouver le théorème.

Preuve du théorème 3.2. L'inclusion $\overline{\mathrm{conv}}(\Upsilon_n) \subseteq \overline{\Gamma}_n^*$ est aisée : on sait que

$$\Upsilon_n \subseteq \Gamma_n^*,$$

et on a donc

$$\overline{\mathrm{conv}}(\Upsilon_n) \subseteq \overline{\mathrm{conv}}(\Gamma_n^*).$$

Or, comme $\overline{\Gamma}_n^*$ est convexe (théorème 2.1), on a

$$\overline{\mathrm{conv}}(\Gamma_n^*) = \overline{\Gamma}_n^*.$$

Pour montrer l'inclusion réciproque, on construit, pour une fonction h quelconque de Γ_n^*, une suite de fonctions f_k de Υ_n telle que

$$\lim_{k\to\infty}\frac{1}{k}f_k = h.$$

En effet, comme la fonction nulle est dans Υ_n (proposition 3.4), la fonction $\frac{1}{k} f_k$ sera dans $\mathrm{conv}(\Upsilon_n)$. La suite ci-dessus sera donc une suite de fonctions de $\mathrm{conv}(\Upsilon_n)$ qui converge vers h. Ainsi on aura

$$\Gamma_n^* \subseteq \overline{\mathrm{conv}}(\Upsilon_n),$$

et donc par passage à l'adhérence :

$$\overline{\Gamma}_n^* \subseteq \overline{\mathrm{conv}}(\Upsilon_n).$$

Soit donc $h \in \Gamma_n^*$ une fonction d'entropie induite par les variables aléatoires $X_1, \ldots, X_n$ à valeurs respectivement dans $\mathcal{X}_1, \ldots, \mathcal{X}_n$. On a alors $h(\alpha) = H(X_\alpha)$ pour toute partie non-vide α de $\{1, \ldots, n\}$, et on note également $\mathcal{X}_\alpha$ l'ensemble des valeurs de la variable X_α (on a donc $\mathcal{X}_\alpha = \prod_{i \in \alpha} \mathcal{X}_i$). On commence par se restreindre au cas où tous les ensembles $\mathcal{X}_1, \ldots, \mathcal{X}_n$ sont finis et où pour tout α, la loi de probabilités de la variable X_α, notée Q_α, est rationnelle de dénominateur q fixé. Pour tout multiple r de q, on construit une matrice à n lignes et r colonnes

$$\mathbf{x} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,r} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,r} \end{pmatrix}$$

telle que pour tout vecteur $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{X}_\mathcal{N}$, le nombre de colonnes de $\mathbf{x}$ égales à $\mathbf{a}^\mathrm{T}$ vaut $rQ_\mathcal{N}(\mathbf{a})$. (Cela est possible car les $Q_\mathcal{N}(\mathbf{a})$ sont rationnels de dénominateur q divisant r et de somme 1, et donc les $rQ_\mathcal{N}(\mathbf{a})$ sont entiers de somme r.) Pour toute partie non-vide α de $\{1, \ldots, n\}$, on note $\mathbf{x}_\alpha$ la sous-matrice de $\mathbf{x}$ constituée des lignes indexées par les éléments de α. Pour tout vecteur $\mathbf{a}$ de $\mathcal{X}_\alpha$, le nombre de colonnes de $\mathbf{x}_\alpha$ égales à $\mathbf{a}^\mathrm{T}$ vaut alors $rQ_\alpha(\mathbf{a})$.

Soit $G = \mathfrak{S}_r$, le groupe des permutations de $\{1, \ldots, r\}$. G agit naturellement sur l'ensemble des matrices à r colonnes par permutation des colonnes. Pour tout $i \in \{1, \ldots, n\}$, on note G_i le stabilisateur de $\mathbf{x}_i$ dans G [4, pp. 55-57]. G_i est donc un sous-groupe de G, et il est clair que $G_\alpha = \bigcap_{i \in \alpha} G_i$ est alors le stabilisateur de $\mathbf{x}_\alpha$. Puisque pour tout $\mathbf{a} \in \mathcal{X}_\alpha$ il y a $rQ_\alpha(\mathbf{a})$ colonnes de $\mathbf{x}_\alpha$ égales à $\mathbf{a}^\mathrm{T}$, on a

$$|G_\alpha| = \prod_{\mathbf{a} \in \mathcal{X}_\alpha} (rQ_\alpha(\mathbf{a}))! \tag{3.2}$$

car il y a alors $(rQ_\alpha(\mathbf{a}))!$ permutations qui envoient une colonne égale à $\mathbf{a}^\mathrm{T}$ sur une autre colonne égale à $\mathbf{a}^\mathrm{T}$, et laissent donc $\mathbf{x}_\alpha$ inchangée.

Pour tout r, on définit la fonction f_r par

$$f_r(\alpha) = \log \frac{|G|}{|G_\alpha|},$$

et on a donc $f_r \in \Upsilon_n$. D'après (3.2) on a

$$f_r(\alpha) = \log \frac{r!}{\prod_{\mathbf{a} \in \mathcal{X}_\alpha} (rQ_\alpha(\mathbf{a}))!},$$

et d'après le lemme 3.1,

$$\lim_{r \to \infty} \frac{1}{r} f_r(\alpha) = H(X_\alpha) = h(\alpha),$$

pour tout α, et donc ([5, pp.172-173])

$$\lim_{r \to \infty} \frac{1}{r} f_r = h.$$

En général, si h est une fonction d'entropie quelconque (*i.e.*, sans faire l'hypothèse que les X_i sont à valeurs dans des ensembles finis ou de lois de probabilités rationnelles), on peut construire une suite h_k de fonctions d'entropie à valeurs rationnelles sur des ensembles finis et qui converge vers h, ce qui prouve le théorème. □

3.3 Relation avec les inégalités informationnelles

On a montré à la section précédente qu'on n'a pas $\Upsilon_n = \Gamma_n^*$, mais on montre dans cette section que si une inégalité informationnelle est vraie pour toutes les fonctions de Υ_n, alors elle est vraie pour toutes les fonctions de Γ_n^* (*i.e.*, pour toute fonction d'entropie).

Soit $\mathbf{b}$ un vecteur représentant une inégalité informationnelle. On rappelle que $H_{\mathbf{b}}^+$ est le demi-espace constitué des vecteurs $\mathbf{x}$ tels que $\langle \mathbf{b}, \mathbf{x} \rangle \geq 0$, qui sont exactement les vecteurs satisfaisant l'inégalité représentée par $\mathbf{b}$. On a donc

$$\Gamma_n^* \subseteq H_{\mathbf{b}}^+, \tag{3.3}$$

et comme $H_{\mathbf{b}}^+$ est fermé, on obtient en passant à l'adhérence

$$\overline{\Gamma}_n^* \subseteq H_{\mathbf{b}}^+. \tag{3.4}$$

L'implication réciproque est trivialement vraie, et on a donc l'équivalence entre (3.3) et (3.4). On veut donc montrer l'équivalence entre (3.4) et

$$\Upsilon_n \subseteq H_{\mathbf{b}}^+. \tag{3.5}$$

Comme $\Upsilon_n \subseteq \Gamma_n^*$, il est clair que (3.4) implique (3.5). La réciproque se montre en passant à l'enveloppe convexe puis à l'adhérence, et en utilisant le théorème 3.2. On peut maintenant énoncer :

Théorème 3.3. *L'inégalité informationnelle représentée par* $\mathbf{b} \in \mathbf{R}^{2^n-1}$ *est vraie pour toute fonction d'entropie si et seulement si l'inclusion a lieu dans* (3.5)*,* i.e.*, si et seulement si pour tout groupe fini* G *et toute famille* $G_1, \ldots, G_n$ *de sous-groupes de* G*, on a*

$$\sum_{\alpha \in \Omega} b_\alpha \log \frac{|G|}{|G_\alpha|} \geq 0.$$

Nous illustrons maintenant le théorème 3.3 par quelques exemples, en l'utilisant pour obtenir certaines des inégalités énoncées à la section 1.2.

Exemple 3.2. Pour tout groupe fini G et tout sous-groupe G_1 de G, on a évidemment $|G| \geq |G_1|$. Il vient que $|G|/|G_1| \geq 1$ et donc $\log(|G|/|G_1|) \geq 0$. Le théorème 3.3 nous permet alors d'obtenir $H(X_1) \geq 0$ (1.10).

Exemple 3.3. Pour tout groupe fini G et tous sous-groupes G_1, G_2 de G, on a $G_1 \cap G_2 \subseteq G_1$, et donc $|G_1 \cap G_2| \leq |G_1|$, d'où l'on déduit

$$\frac{|G|}{|G_1 \cap G_2|} \geq \frac{|G|}{|G_1|}.$$

Par passage aux logarithmes on obtient

$$\log \frac{|G|}{|G_1 \cap G_2|} \geq \log \frac{|G|}{|G_1|},$$

et enfin

$$\log \frac{|G|}{|G_1 \cap G_2|} - \log \frac{|G|}{|G_1|} \geq 0.$$

Le théorème 3.3 nous permet alors d'obtenir

$$H(X_1, X_2) - H(X_1) \geq 0,$$

qui est l'inégalité (1.20).

Exemple 3.4. Pour tout groupe fini G et tous sous-groupes G_1, G_2 de G, on a

$$\frac{|G|}{|G_1 \cap G_2|} \leq \frac{|G|}{|G_1|} \times \frac{|G|}{|G_2|}. \tag{3.6}$$

En effet, puisque $|G|$ est strictement positif, l'inégalité devient

$$\frac{1}{|G_1 \cap G_2|} \leq \frac{|G|}{|G_1||G_2|},$$

et donc

$$|G| \geq \frac{|G_1||G_2|}{|G_1 \cap G_2|}.$$

On utilise alors le résultat bien connu [4, p. 7] :

$$|G_1 G_2| = \frac{|G_1||G_2|}{|G_1 \cap G_2|},$$

et comme $G_1 G_2 \subseteq G$, on a $|G| \geq |G_1 G_2|$, ce qui prouve l'inégalité (3.6). Par passage aux logarithmes, on obtient

$$\log \frac{|G|}{|G_1 \cap G_2|} \leq \log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_2|},$$

et donc

$$\log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_2|} - \log \frac{|G|}{|G_1 \cap G_2|} \geq 0.$$

Le théorème 3.3 nous permet alors d'obtenir

$$H(X_1) + H(X_2) - H(X_1, X_2) \geq 0,$$

et on retrouve l'inégalité (1.18). Réciproquement, l'inégalité (1.18) peut être utilisée pour montrer l'inégalité (3.6) par un raisonnement inverse.

Annexe A

Vérification numérique de l'inégalité 2.8

Afin de vérifier numériquement que l'inégalité (2.8) est bien une inégalité plus forte que les inégalités de Shannon, nous avons utilisé le logiciel Sage [7], plus précisément son module de programmation linéaire [14, chapitre 17]. Nous redonnons ici l'inégalité (2.8) : pour toutes variables aléatoires X_1, X_2, X_3, X_4, on a

$$I(X_1, X_2)+I(X_1; X_3, X_4)+3I(X_3; X_4|X_1)+I(X_3; X_4|X_2)-2I(X_3; X_4) \geq 0. \quad \text{(A.1)}$$

On veut montrer que cette inégalité est plus forte que les inégalités de Shannon, c'est-à-dire qu'il existe un vecteur $h \in \mathbf{R}^{2^n-1}$ qui statisfait les inégalités de Shannon mais qui ne satisfait pas l'inégalité (A.1). Le programme Sage utilisé pour trouver un tel vecteur est le suivant :

```
var = Subsets([1,2,3,4])

p = MixedIntegerLinearProgram()
h = p.new_variable()
p.set_objective(None)
p.add_constraint(h[Set([1])] + h[Set([2])] - h[Set([1,2])] +
                 h[Set([1])] + h[Set([3,4])] - h[Set([1,3,4])] +
                 3*h[Set([1,3])] + 3*h[Set([1,4])] - 3*h[Set([1,3,4])] - 3*h[Set([1])] +
                 h[Set([2,3])] + h[Set([2,4])] - h[Set([2,3,4])] - h[Set([2])] -
                 2*h[Set([3])] - 2*h[Set([4])] + 2*h[Set([3,4])], min=-1/2, max=-1/2)
p.add_constraint(h[Set([])], min=0, max=0)
for i in CartesianProduct(var, var, var):
    p.add_constraint(
            h[Set(union(i[0], i[2]))] +
            h[Set(union(i[1], i[2]))] -
            h[Set(union(i[0], union(i[1], i[2])))] -
            h[i[2]] >= 0)
p.solve()
for i in var:
    print(i, p.get_values(h[i]))
```

On commence par créer le programme linéaire `p`, puis par créer une variable `h`, qui représente le vecteur cherché. Les variables sur lesquelles le programme opérera seront de la forme `h[Set(t)]`, où `t` est une partie de $\{1, 2, 3, 4\}$ (la variable

`h[Set(t)]` représente alors à la composante du vecteur qui correspond à la partie de $\{1, 2, 3, 4\}$ représentée par `t`).

La première contrainte (lignes 6 à 10) est que le vecteur cherché ne doit *pas* vérifier l'inégalité (A.1). C'est-à-dire qu'on doit avoir

$$I(X_1, X_2)+I(X_1; X_3, X_4)+3I(X_3; X_4|X_1)+I(X_3; X_4|X_2)-2I(X_3; X_4) < 0. \quad \text{(A.2)}$$

Afin d'exprimer les termes de l'inégalité en fonction des seules entropies jointes, on utilise les relations

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

et

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

On obtient alors l'expression des lignes 6 à 10, et on cherche un vecteur tel que cette expression vaut $-1/2$.

On ajoute ensuite les contraintes nécessaires pour que le vecteur trouvé satisfasse les inégalités de Shannon (lignes 11 à 17). On a d'abord (ligne 11) $h_\varnothing = 0$. Ensuite, on a vu à la section 2.3 que la positivité de l'information mutuelle conditionnelle impliquait toutes les inégalités de Shannon, et il est donc suffisant d'ajouter uniquement les contraintes liées à l'information mutuelle conditionnelle : pour toutes parties α, β, γ de $\{1, 2, 3, 4\}$, on doit avoir

$$I(X_\alpha; X_\beta|X_\gamma) \geq 0, \quad \text{(A.3)}$$

et en utilisant la troisième relation ci-dessus on obtient

$$h(\alpha \cup \gamma) + h(\beta \cup \gamma) - h(\alpha \cup \beta \cup \gamma) - h(\gamma) \geq 0. \quad \text{(A.4)}$$

Enfin on résout le programme et on affiche les valeurs trouvées. Le programme affiche alors

```
({}, 0.0)
({1}, 1.0)
({2}, 1.0)
({3}, 1.0)
({4}, 1.0)
({1, 2}, 2.0)
({1, 3}, 1.5)
({1, 4}, 1.5)
({2, 3}, 1.5)
({2, 4}, 1.5)
({3, 4}, 1.5)
({1, 2, 3}, 2.0)
({1, 2, 4}, 2.0)
({1, 3, 4}, 2.0)
({2, 3, 4}, 2.0)
({1, 2, 3, 4}, 2.0)
```

ce qui correpond au vecteur donné à la section 2.4 en prenant $a = 1/2$.

Annexe B

Preuve du lemme 3.1

Soit X une variable aléatoire à valeurs dans un ensemble fini $\mathcal{X}$ et telle que pour tout $x \in \mathcal{X}$, $p(X = x) = p(x)$ est rationnel de dénominateur q. On veut montrer que pour tout multiple r de q, on a

$$H(X) - |\mathcal{X}|\frac{\log(r+1)}{r} \leq \frac{1}{r}\log\frac{r!}{\prod_{x\in\mathcal{X}}(rp(x))!} \leq H(X). \tag{B.1}$$

On utilise pour cela la *méthode des types*, présentée dans [2, pp. 347-352].

Définition B.1. Soit $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$. On appelle le *type* de $\mathbf{x}$, et on note $P_\mathbf{x}$, l'application de $\mathcal{X}$ dans $\mathbf{Q}$ définie par

$$P_\mathbf{x}(x) = N(x|\mathbf{x})/n,$$

où $N(x|\mathbf{x})$ est le nombre d'occurrences de x dans $\mathbf{x}$. On note qu'on a

$$\sum_{x\in\mathcal{X}} P_\mathbf{x}(x) = 1,$$

et que $P_\mathbf{x}$ définit donc une loi de probabilités sur $\mathcal{X}$.

Définition B.2. Pour un entier n fixé, on note $\mathcal{P}_n$ l'ensemble des types de dénominateur n. Si $P \in \mathcal{P}_n$, l'ensemble des vecteurs de $\mathcal{X}^n$ de type P est appelé *classe de type* de P et noté $T(P)$. Formellement :

$$T(P) = \{\mathbf{x} \in \mathcal{X}^n | P_\mathbf{x} = P\}.$$

Exemple B.1. Soit $\mathcal{X} = \{1, 2, 3\}$ et $\mathbf{x} = 11321$. On a $P_\mathbf{x}(1) = 3/5$, $P_\mathbf{x}(2) = 1/5$, et $P_\mathbf{x}(3) = 1/5$. La classe de $P_\mathbf{x}$ est l'ensemble des vecteurs avec trois 1, un 2 et

un 3. En particulier, on a

$$\begin{aligned}|T(P_{\mathbf{x}})| &= \binom{5}{3}\binom{2}{1} \\ &= \frac{5!}{3!2!} \times \frac{2!}{1!1!} \\ &= \frac{5!}{3!1!1!} \\ &= 20.\end{aligned}$$

Le résultat suivant est alors clair.

Proposition B.1. *Si $P \in \mathcal{P}_n$, on a*

$$|T(P)| = \frac{n!}{\prod_{a\in\mathcal{X}}(nP(a))!}.$$

Le résultat suivant montre que le nombre de types de dénominateur n est au maximum polynomial en n.

Proposition B.2.

$$|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|}.$$

Preuve. Pour tout $P \in \mathcal{P}_n$ et tout $x \in \mathcal{X}$, $P(x)$ est de dénominateur n par définition, et son numérateur peut prendre au maximum $n+1$ valeurs (de 0 à n). Il y a donc au maximum $n+1$ valeurs possibles pour $P(x)$. □

Cette borne supérieure sur $|\mathcal{P}_n|$ est évidemment très large, mais suffisante pour montrer le résultat voulu. On peut remarquer que comme le nombre de types est polynomial en n mais que le nombre de vecteurs de $\mathcal{X}^n$ est exponentiel, il y a nécessairement un type qui contient un nombre exponentiel de vecteurs.

Proposition B.3. *Si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes à valeurs dans $\mathcal{X}$ et de même loi Q, et $\mathbf{x} = (x_1, \ldots, x_n)$ un vecteur donné de $\mathcal{X}^n$, la probabilité d'avoir $(X_1, \ldots, X_n) = \mathbf{x}$ dépend uniquement de $P_{\mathbf{x}}$. En notant Q^n la loi de probabilités de la variable $(X_1, \ldots, X_n)$, on a*

$$Q^n(\mathbf{x}) = 2^{-n(H(P_{\mathbf{x}})+D(P_{\mathbf{x}}||Q))},$$

où $D(P_{\mathbf{x}}||Q)$ est la distance de Kullback *entre les lois de probabilités $P_{\mathbf{x}}$ et Q [11, p. 1] [2, p. 19].*

Preuve.

$$\begin{aligned} Q^n(\mathbf{x}) &= \prod_{i=1}^{n} Q(x_i) \\ &= \prod_{a\in\mathcal{X}} Q(a)^{N(a|\mathbf{x})} \\ &= \prod_{a\in\mathcal{X}} Q(a)^{nP_{\mathbf{x}}(a)} \\ &= \prod_{a\in\mathcal{X}} 2^{nP_{\mathbf{x}}(a)\log Q(a)} \\ &= 2^{\sum_{a\in\mathcal{X}} nP_{\mathbf{x}}(a)\log Q(a)}. \end{aligned}$$

Or

$$\begin{aligned} &\sum_{a\in\mathcal{X}} nP_{\mathbf{x}}(a)\log Q(a) \\ &= n\sum_{a\in\mathcal{X}} P_{\mathbf{x}}(a)\log Q(a) - P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a) + P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a) \\ &= n(-D(P_{\mathbf{x}}||Q) - H(P_{\mathbf{x}})). \end{aligned}$$

□

Corollaire B.1. *Si* $\mathbf{x} \in T(Q)$ *(*i.e.*, si* $P_{\mathbf{x}} = Q$*), on a*

$$Q^n(\mathbf{x}) = 2^{-nH(Q)}.$$

Exemple B.2. On suppose que n est un multiple de 6. La probabilité que n lancers successifs d'un dé équilibré produisent une suite donnée contenant exactement $n/6$ occurrences de chaque face est évidemment $1/6^n = 2^{-n\log 6}$ (comme pour toute autre suite). Si le dé a comme loi de probabilités $p(1) = p(2) = 1/3$, $p(3) = 1/6$, $p(4) = p(5) = 1/12$, $p(6) = 0$ (et si n est un multiple de 12), la probabilité d'obtenir une suite donnée constituée de $n/3$ occurrences de 1 et 2, $n/6$ occurrences de 3, $n/12$ occurrences de 4 et 5, et aucun 6 est $2^{-nH(p)}$.

Le résultat suivant donne un encadrement de la taille des classes de types. L'inégalité (B.1) en sera un corollaire.

Proposition B.4. *Pour tout type* $P \in \mathcal{P}_n$*, on a*

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{nH(P)} \leq |T(P)| \leq 2^{nH(P)}.$$

Nous aurons besoin du lemme suivant :

Lemme B.1. *Pour tous entiers positifs m, n, on a*

$$\frac{m!}{n!} \geq n^{m-n}.$$

Preuve. Si $m = n$, l'égalité a lieu. Si $m > n$ on a

$$\begin{aligned}\frac{m!}{n!} &= \prod_{i=n+1}^{m} i \\ &= \prod_{i=1}^{m-n} (n+i),\end{aligned}$$

qui est le produit de $m-n$ facteurs tous supérieurs à n, d'où l'inégalité. Si $m < n$ on a d'une part

$$\frac{m!}{n!} = \frac{1}{n!/m!}$$

et d'autre part

$$n^{m-n} = \frac{1}{n^{n-m}}.$$

L'inégalité devient alors

$$\frac{n!}{m!} \leq n^{n-m},$$

et par un raisonnement analogue au précédent, on a

$$\begin{aligned}\frac{n!}{m!} &= \prod_{i=m+1}^{n} i \\ &= \prod_{i=1}^{n-m} (m+i),\end{aligned}$$

qui est le produit de $n-m$ facteurs tous inférieurs ou égaux à n. □

Preuve de la proposition B.4. Pour l'inégalité de droite, avec les notations de la proposition B.3, P^n est la loi de probabilités de la variable jointe correspondant à n copies indépendantes d'une variable à valeurs dans $\mathcal{X}$ et de loi de probabilités P. P^n est donc à valeurs dans $\mathcal{X}^n$, et la probabilité qu'un vecteur de $\mathcal{X}^n$ choisi selon la loi P^n soit dans $T(P)$ (*i.e.*, soit de type P) vaut

$$\begin{aligned}P^n(T(P)) &= \sum_{\mathbf{x} \in T(P)} P^n(\mathbf{x}) \\ &= \sum_{\mathbf{x} \in T(P)} 2^{-nH(P)} && \text{(Corollaire B.1)} \\ &= |T(P)| 2^{-nH(P)},\end{aligned}$$

et comme $P^n(T(P)) \leq 1$ on obtient $|T(P)| \leq 2^{nH(P)}$.

Pour montrer l'inégalité de gauche, on commence par montrer que la classe $T(P)$ est de probabilité maximale pour la loi P^n. En d'autres termes, qu'un vecteur de $\mathcal{X}^n$ tiré selon la loi P^n aura plus de chances d'être dans $T(P)$ que dans n'importe quelle autre classe de $\mathcal{P}_n$. Soit $P' \in \mathcal{P}_n$, on veut donc montrer que

$$P^n(T(P)) \geq P^n(T(P')).$$

On obtient une borne inférieure sur $\frac{P^n(T(P))}{P^n(T(P'))}$ par

$$\begin{aligned}
\frac{P^n(T(P))}{P^n(T(P'))} &= \frac{|T(P)| \prod_{a\in\mathcal{X}} P(a)^{nP(a)}}{|T(P')| \prod_{a\in\mathcal{X}} P(a)^{nP'(a)}} \\
&= \frac{|T(P)|}{|T(P')|} \cdot \prod_{a\in\mathcal{X}} P(a)^{n(P(a)-P'(a))} \\
&= \prod_{a\in\mathcal{X}} \frac{(nP'(a))!}{(nP(a))!} P(a)^{n(P(a)-P'(a))},
\end{aligned}$$

et en utilisant le lemme B.1 :

$$\begin{aligned}
&\geq \prod_{a\in\mathcal{X}} (nP(a))^{nP'(a)-nP(a)} P(a)^{n(P(a)-P'(a))} \\
&\geq \prod_{a\in\mathcal{X}} n^{n(P'(a)-P(a))} \\
&\geq n^{n\sum_{a\in\mathcal{X}}(P'(a)-P(a))} \\
&\geq n^{n(1-1)} \\
&\geq 1.
\end{aligned}$$

Ainsi on a $\frac{P^n(T(P))}{P^n(T(P'))} \geq 1$, et donc $P^n(T(P)) \geq P^n(T(P'))$. On peut maintenant

obtenir l'inégalité de gauche de la proposition B.4 :

$$\begin{aligned}
1 &= \sum_{Q\in\mathcal{P}_n} P^n(T(Q)) \\
&\leq |\mathcal{P}_n| \max_{Q\in\mathcal{P}_n} P^n(T(Q)) \\
&\leq |\mathcal{P}_n| P^n(T(P)) \\
&\leq (n+1)^{|\mathcal{X}|} P^n(T(P)) && \text{(Proposition B.2)} \\
&\leq (n+1)^{|\mathcal{X}|} \sum_{\mathbf{x}\in T(P)} P^n(\mathbf{x}) \\
&\leq (n+1)^{|\mathcal{X}|} \sum_{\mathbf{x}\in T(P)} 2^{-nH(P)} && \text{(Corollaire B.1)} \\
&\leq (n+1)^{|\mathcal{X}|} |T(P)| 2^{-nH(P)}.
\end{aligned}$$

Et donc

$$(n+1)^{|\mathcal{X}|}|T(P)| \geq 2^{nH(P)},$$

et enfin

$$|T(P)| \geq \frac{1}{(n+1)^{|\mathcal{X}|}} 2^{nH(P)}. \qquad \square$$

On peut alors obtenir l'inégalité (B.1) :

Corollaire B.2. *Si X est une variable aléatoire à valeurs dans un ensemble fini $\mathcal{X}$ et telle que pour tout $x \in \mathcal{X}$, $p(X = x)$ est rationnel de dénominateur q, on a pour tout multiple r de q*

$$H(X) - |\mathcal{X}|\frac{\log(r+1)}{r} \leq \frac{1}{r}\log\frac{r!}{\prod_{x\in\mathcal{X}}(rp(x))!} \leq H(X).$$

Preuve. Soit X une telle variable aléatoire. On construit un vecteur $\mathbf{x} \in \mathcal{X}^r$ qui, pour tout $a \in \mathcal{X}$, contient $rp(a)$ occurrences de a. Cela est possible car les $p(a)$ sont rationnels de dénominateur q divisant r et de somme 1, donc les $rp(a)$ sont entiers de somme r. On pose $P = P_{\mathbf{x}}$, et d'après la proposition B.4, on a

$$\frac{1}{(r+1)^{|\mathcal{X}|}} 2^{rH(P)} \leq |T(P)| \leq 2^{rH(P)}.$$

Par passage aux logarithmes, on obtient

$$\log\frac{1}{(r+1)^{|\mathcal{X}|}} + rH(P) \leq \log|T(P)| \leq rH(P),$$

d'où

$$rH(P) - |\mathcal{X}|\log(r+1) \leq \log|T(P)| \leq rH(P),$$

et enfin

$$H(P) - |\mathcal{X}|\frac{\log(r+1)}{r} \leq \frac{1}{r}\log|T(P)| \leq rH(P),$$

et on conclut en remplaçant $|T(P)|$ par son expression de l'énoncé de la proposition B.1. □

Bibliographie

[1] T. H. Chan et R. W. Yeung, *On a Relation Between Information Inequalities and Group Theory. IEEE Transactions on Information Theory*, vol. 48, pp. 1992-1995, 2002.

[2] T. M. Cover et J. A. Thomas, *Elements of Information Theory*, 2e édition. Wiley, Hoboken, New Jersey, 2006.

[3] J.-B. Hiriart-Urruty et C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer, Berlin Heidelberg, 2001.

[4] H. Kurzweil et B. Stellmacher, *The Theory of Finite Groups*. Springer, New York, 2004.

[5] S. Lang, *Undergraduate Analysis*, 2e édition. Springer, New York, 1997.

[6] F. Matúš, *Infnitely Many Information Inequalities*. 2007 IEEE International Symposium on Information Theory, Nice, 24-29 juin 2007.

[7] *Sage : Open Source Mathematics Software*, version 5.8. `http://www.sagemath.org/`, 2013.

[8] C. E. Shannon, *A Mathematical Theory of Communication. Bell Systems Technical Journal*, vol. 27, pp. 379-423, 623-656, 1948.

[9] D. R. Stinson, *Cryptography, Theory and Practice*, 3e édition. Chapman & Hall/CRC, Boca Raton, Floride, 2006.

[10] G. Zémor, *Cours de cryptographie*. Cassini, Paris, 2000.

[11] G. Zémor, *Mémento de théorie de l'information*. `http://www.math.u-bordeaux.fr/~zemor/TI.pdf`, 2011.

[12] Z. Zhang et R. W. Yeung, *A Non-Shannon-Type Conditional Inequality of Information Quantities. IEEE Transactions on Information Theory*, vol. 43, pp. 1982-1986, 1997.

[13] Z. Zhang et R. W. Yeung, *On Characterization of Entropy Functions via Information Inequalities. IEEE Transactions on Information Theory*, vol. 44, pp. 1440-1452, 1998.

[14] P. Zimmermann, *et al.*, *Calcul mathématique avec Sage*, version 1.0.9. `http://sagebook.gforge.inria.fr/`, 2011.