Eva Corot

Conférence sur l'Entropie et la Combinatoire

Sir Timothy Gowers (notes de cours)

13 Octobre 2025
Collège de France

Annonce : Les étudiants qui souhaitent valider ce cours officiellement sont priés de me communiquer leur nom et leur adresse. Un examen final sera organisé à une date à déterminer. Des feuilles d'exercices sont également disponibles sur demande.

Introduction

Ce cours a pour but de montrer des exemples d'utilisation de l'entropie et de présenter un résultat simple à la fin. Des problèmes de combinatoire n'ont été résolus que grâce à l'entropie.

Le concept d'**entropie de Shannon** est un outil très puissant en combinatoire, qui a permis de résoudre des problèmes autrement inaccessibles. Ce cours a pour but d'expliquer pourquoi l'entropie est si utile et de montrer quelques exemples d'application.

Toutes les variables aléatoires considérées seront discrètes, à support fini.

Définition 0.1: Entropie de Shannon

Si X est une variable aléatoire prenant ses valeurs dans un ensemble fini A, alors l'entropie de Shannon de X est donnée par :

$$H[X] = \sum_{a \in A} \mathbb{P}(X = a) \log \left(\frac{1}{\mathbb{P}(X = a)} \right)$$

En notant $p_a = \mathbb{P}(X = a)$, on peut aussi écrire :

$$H[X] = \sum_{a \in A} p_a \log \left(\frac{1}{p_a} \right) = - \sum_{a \in A} p_a \log(p_a)$$

Cette dernière forme, $-\sum p_a \log(p_a)$, est moins préférée.

Remarque : Tous les logarithmes dans ce cours sont en base 2. Par convention, $x \log(x) = 0$ pour $x = 0$.

1 Présentation axiomatique de l'entropie

L'idée est d'utiliser les axiomes et pas la formule.

L'entropie peut être caractérisée de manière unique (à une constante multiplicative près) par un ensemble d'axiomes, dits de **Khinchin**.

Soit H une application définie sur les variables aléatoires à support fini, à valeurs réelles.

Axiome 1 : Normalisation

Si X prend les valeurs 0 et 1 avec probabilité $1/2$ chacune, alors $H[X] = 1$.

Cet axiome sert à avoir la formule pas à une constante multiplicative près devant.

Axiome 2 : Invariance

Si $\phi : A \to B$ est une bijection, X une v.a. sur A et $Y = \phi(X)$, alors $H[Y] = H[X]$.

C'est-à-dire, l'entropie d'une variable aléatoire ne dépend que des probabilités, c'est tout ce qui compte.

Axiome 3 : Extension

Si X est définie sur $A \subset B$ et si Y est définie sur B par $\mathbb{P}(Y = b) = \mathbb{P}(X = b)$ pour $b \in A$ et $\mathbb{P}(Y = b) = 0$ sinon, alors $H[Y] = H[X]$.

Axiome 4 : Continuité

$H[X]$ dépend de façon continue des probabilités $\mathbb{P}(X = a)$.

On utilisera cet axiome pour dire que si l'on prouve un résultat où toutes les probabilités sont rationnelles, alors par ***densité****, le résultat s'applique pour toutes les probabilités réelles.*

Passons à deux axiomes plus importants pour les calculs : la maximalité et l'additivité.

Axiome 5 : Maximalité

Si X et Y prennent leurs valeurs dans le même ensemble fini A et si Y est **uniforme** sur A, alors $H[X] \leq H[Y]$.

C'est-à-dire, parmi toutes les variables aléatoires qui prennent des valeurs dans un ensemble, celle qui a la plus grande entropie est une variable uniformément distribuée.

L'additivité est l'axiome le plus important et la notion d'entropie conditionnelle est centrale.

Axiome 6 : Additivité

Si X et Y sont deux variables aléatoires, alors :

$$H[X, Y] = H[Y] + H[X|Y]$$

ou encore

$$H[X|Y] = H[X, Y] - H[Y]$$

Définition 1.1: Entropie conditionnelle

La notation $H[X|Y]$ représente l'entropie conditionnelle de X sachant Y. Y défini sur B. Elle est définie comme la **moyenne des entropies** de X lorsque la valeur de Y est fixée :

$$H[X|Y] = \sum_{b \in B} \mathbb{P}(Y = b) H[X|Y = b]$$

où $H[X|Y = b]$ est l'entropie de la v.a. X sous la probabilité conditionnelle $\mathbb{P}(\cdot|Y = b)$.

Cela veut dire que si je fixe $Y = b$, alors $H[X|Y = b]$ est bien définie. ***L'entropie conditionnelle est la moyenne*** *de ces entropies.*

1.0 Intuition : l'entropie comme quantité d'information

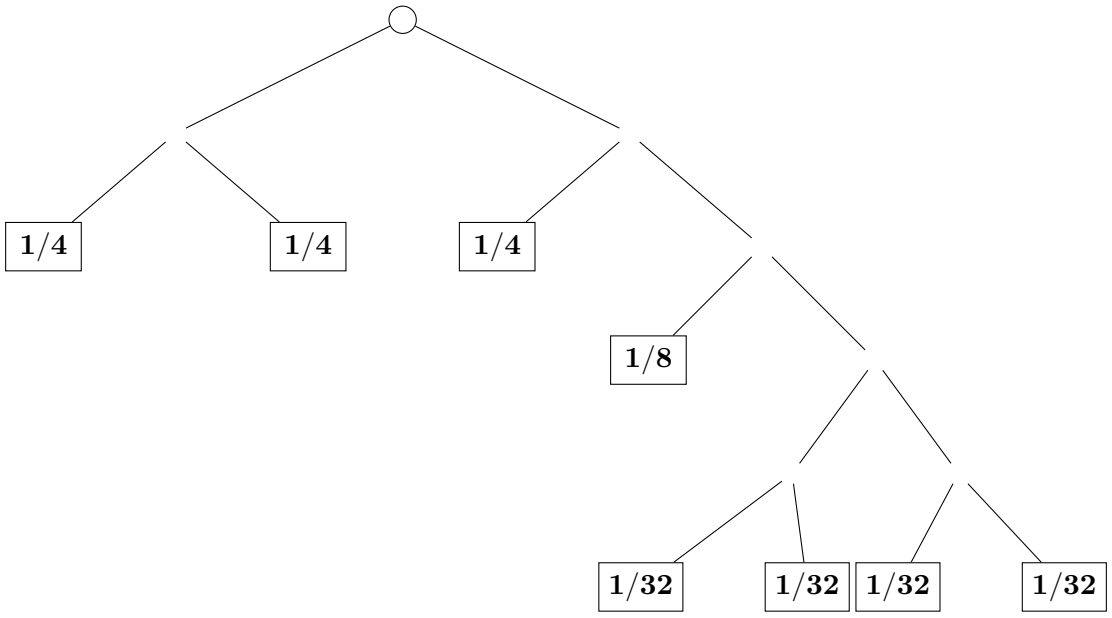
L'entropie est le nombre de bits d'information qu'il faut en moyenne pour déterminer la valeur de la variable aléatoire. Plus la variable est uniforme, plus il sera difficile de déterminer sa valeur.

Pour aller à l'autre extrême : l'entropie est 0 si on n'a pas besoin de poser de question pour trouver la valeur.

L'entropie peut être interprétée comme le **nombre moyen de bits d'information** nécessaires pour déterminer la valeur d'une variable aléatoire.

Exemple (Jeu de 20 questions) :

Imaginons une marche aléatoire sur un arbre binaire. On part de la racine et à chaque nœud, on choisit d'aller à gauche ou à droite avec une probabilité de 1/2, jusqu'à atteindre une feuille. La valeur de la variable aléatoire est la feuille atteinte.



Cette interprétation avec les questions fonctionne bien car c'est un arbre binaire où chaque branche est prise avec probabilité 1/2, mais l'entropie est une généralisation à des cas où ce n'est pas le cas.

L'entropie est le nombre **moyen** de questions "oui/non" (comme "êtes-vous allé à gauche au premier embranchement ?") qu'il faut poser pour identifier la feuille atteinte. Le calcul donne :

$$\begin{aligned} H[X] &= -\left(3 \times \frac{1}{4}\log\left(\frac{1}{4}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right) + 4 \times \frac{1}{32}\log\left(\frac{1}{32}\right)\right) \\ &= 3 \times \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + 4 \times \frac{1}{32} \times 5 \\ &= \frac{6}{4} + \frac{3}{8} + \frac{20}{32} = \frac{3}{2} + \frac{3}{8} + \frac{5}{8} = 2.5 \text{ bits} \end{aligned}$$

2 Propriétés de l'entropie

Lemme 2.1: Indépendance

Si X et Y sont indépendantes, alors $H[X,Y] = H[X] + H[Y]$.

Preuve

Par additivité, $H[X,Y] = H[Y] + H[X|Y]$. Puisque X et Y sont indépendantes, la connaissance de Y ne donne aucune information sur X. Donc, pour tout y, $H[X|Y=y] = H[X]$.

$$H[X|Y] = \sum_y \mathbb{P}(Y=y)H[X|Y=y] = \sum_y \mathbb{P}(Y=y)H[X] = H[X]\sum_y \mathbb{P}(Y=y) = H[X]$$

Ainsi, $H[X,Y] = H[Y] + H[X]$.

Corollaire 2.2

Si $X_1, \dots, X_n$ sont mutuellement indépendantes, alors $H[X_1, \dots, X_n] = \sum_{i=1}^{n} H[X_i]$.

Preuve
Par récurrence évidente sur n. □

Corollaire 2.3

Si X est uniforme sur $\{0,1\}^n$, alors $H[X] = n$.

Preuve
On peut voir X comme un n-uplet $(X_1, \dots, X_n)$ où chaque X_i est une **variable de Bernoulli uniforme** sur $\{0,1\}$ et les X_i sont indépendants. Par normalisation, $H[X_i] = 1$ pour tout i. Par le corollaire précédent, $H[X] = \sum_{i=1}^{n} H[X_i] = n$. □

Lemme 2.4: Formule en chaîne pour l'entropie

L'idée intuitive *de l'entropie d'une variable est le nombre de bits d'information qu'elle contient.* ***L'entropie de X sachant Y est le nombre de bits d'information supplémentaires qu'il faut pour déterminer X quand on connaît déjà Y.*** *Ici, pour déterminer le vecteur $(X_1, \dots, X_n)$, on détermine X_1, puis on regarde combien d'info il faut en plus pour X_2 sachant X_1, et ainsi de suite.*

L'entropie d'une variable jointe se décompose de la manière suivante :

$$H[X_1, \dots, X_n] = H[X_1] + H[X_2|X_1] + \cdots + H[X_n|X_1, \dots, X_{n-1}]$$

Preuve
Par application répétée de l'additivité : $H[X_1, \dots, X_n] = H[X_1, \dots, X_{n-1}] + H[X_n|X_1, \dots, X_{n-1}]$. Le résultat s'ensuit par une récurrence immédiate. □

Proposition 2.5: Entropie de la loi uniforme

Ceci est une généralisation du corollaire précédent. Par ***invariance****, on sait que pour un ensemble de taille 2^n, une variable uniforme a une entropie de n. On va voir ici que l'entropie est bien le logarithme de la taille de l'ensemble, même si cette taille n'est pas une puissance de 2.*

Soit A un ensemble de taille n. Si X est une variable aléatoire uniforme sur A, alors $H[X] = \log(n)$.

Preuve
Soient $X_1, \dots, X_r$ des copies iid indépendantes de X. Le vecteur $(X_1, \dots, X_r)$ est uniforme sur A^r, qui est un ensemble de taille n^r. Par le premier corollaire, $H[X_1, \dots, X_r] = rH[X]$. Pour tout $r \in \mathbb{N}^*$, il existe un entier s tel que $2^s \leq n^r < 2^{s+1}$. Par monotonicité de l'entropie (conséquence des axiomes de **maximalité** et d'**extension**) et invariance, l'entropie d'une variable uniforme sur un ensemble de taille k est une fonction croissante de k. Ainsi :

$$H[\text{Unif}(\{1, \dots, 2^s\})] \leq H[\text{Unif}(A^r)] < H[\text{Unif}(\{1, \dots, 2^{s+1}\})]$$

$$s \leq rH[X] < s + 1$$

En divisant par r, on obtient $\frac{s}{r} \leq H[X] < \frac{s+1}{r}$. De l'encadrement initial $2^s \leq n^r < 2^{s+1}$, en appliquant le log en base 2, on a $s \leq r\log(n) < s+1$, ce qui donne $\frac{s}{r} \leq \log(n) < \frac{s+1}{r}$. Les deux encadrements sont valables pour tout r, et la différence entre la borne supérieure et inférieure est $1/r$. Quand $r \to \infty$, cette différence tend vers 0. On en conclut que $H[X] = \log(n)$.

La preuve générale ne fonctionne pas si $|A| = 1$ car on a utilisé des entiers positifs. Le cas $|A| = 1$ doit être traité séparément. C'est facile mais ***amusant****.*

Cas particulier : Si $|A| = 1$, X est déterministe. X est indépendante d'elle-même (ce qui est un peu bizarre, mais c'est le cas pour une v.a. déterministe). D'une part, $H[X, X] = H[X]$ par invariance

(l'application $x \mapsto (x, x)$ est une bijection sur le support). D'autre part, $H[X, X] = H[X] + H[X] = 2H[X]$ par indépendance. Donc $H[X] = 2H[X]$, ce qui implique $H[X] = 0$. □

Lemme 2.6

Si $Y = f(X)$, alors $H[X, Y] = H[X]$. De plus, $H[Z|X, Y] = H[Z|X]$.

Autrement dit Y n'ajoute aucune info quand on connaît la veleur de X.

Preuve
L'application $\theta : x \mapsto (x, f(x))$ est une bijection du support de X vers celui de (X, Y). Comme $(X, Y) = \theta(X)$, par invariance, $H[X, Y] = H[X]$.
Pour la seconde identité, par additivité :
$H[Z|X, Y] = H[Z, X, Y] - H[X, Y]$. Or $H[Z, X, Y] = H[Z, X]$ (car Y est fonction de X) et $H[X, Y] = H[X]$.
Donc, $H[Z|X, Y] = H[Z, X] - H[X] = H[Z|X]$.□

Proposition 2.7: Formule de Shannon

Si H vérifie les axiomes de Khinchin, alors pour toute v.a. X, on a $H[X] = \sum_x p_x \log(\frac{1}{p_x})$, où $p_x = \mathbb{P}(X = x)$.

Preuve
Par continuité, il suffit de prouver le résultat pour des probabilités rationnelles. Supposons que pour chaque x dans le support A, $p_x = n_x/n$ pour des entiers n_x et n. Soit Y une v.a. uniforme sur $[n] = \{1, \dots, n\}$. On a $H[Y] = \log(n)$. On peut "partitionner" $[n]$ en sous-ensembles $(E_x)_{x \in A}$ tels que $|E_x| = n_x$.

Partitionner "entre guillemets", car E_x peut être vide.

Attention pour d'autres auteurs en combinatoire : $[n] = \{0, \dots, n\}$.

On peut supposer (par invariance) que l'événement $\{X = x\}$ est équivalent à l'événement $\{Y \in E_x\}$. La probabilité que $Y \in E_x$ est $|E_x|/n = n_x/n = p_x$. Par additivité : $H[Y] = H[X, Y] = H[X] + H[Y|X]$. Calculons $H[Y|X]$:

$$H[Y|X] = \sum_{x \in A} \mathbb{P}(X = x) H[Y|X = x] = \sum_{x \in A} p_x H[Y|X = x]$$

Sachant que $X = x$, Y est uniforme sur l'ensemble E_x de taille n_x. Donc $H[Y|X = x] = \log(n_x)$.

$$H[Y|X] = \sum_{x \in A} p_x \log(n_x) = \sum_{x \in A} p_x \log(n \cdot p_x) = \sum_{x \in A} p_x (\log(n) + \log(p_x))$$

$$H[Y|X] = \log(n) \sum_{x \in A} p_x + \sum_{x \in A} p_x \log(p_x) = \log(n) + \sum_{x \in A} p_x \log(p_x)$$

En remplaçant dans l'équation de départ :

$$\log(n) = H[Y] = H[X] + \log(n) + \sum_{x \in A} p_x \log(p_x)$$

En simplifiant par $\log(n)$, on obtient $0 = H[X] + \sum_{x \in A} p_x \log(p_x)$, d'où le résultat.□

Pause

3 Conséquences de la formule et inégalités fondamentales

La formule $H[X] = -\sum p_x \log(p_x)$ montre directement que l'entropie est **non-négative**, car $p_x \in [0,1]$ implique $\log(p_x) \leq 0$.

positif en français.

Proposition 3.1: Sous-additivité

Pour toutes variables aléatoires X, Y, on a $H[X|Y] \leq H[X]$.

Preuve

Les trois inégalités suivantes sont équivalentes par la règle d'additivité : $H[X|Y] \leq H[X]$, $H[Y|X] \leq H[Y]$ et $H[X,Y] \leq H[X] + H[Y]$. (sous additivité)

Si X uniforme sur A, $H[X,Y] = \sum_y \mathbb{P}[Y=y]\, H[X|Y=y] \leq \sum_y \mathbb{P}[Y=y]\, H[X]$ par maximalité. $\leq H[X]$
Soit n tq $p_{ab} = n_{ab}/n$ où n_{ab} non nul $|E_{ab}| = n_{ab}$

Idée de la preuve du cours :
on suppose les probabilités $p_{ab} = \mathbb{P}(X=a, Y=b)$ rationnelles.
Soit W une v.a. uniforme sur $[n]$ et une partition (E_{ab}) de $[n]$ telle que $\mathbb{P}(X=a,Y=b) \Leftrightarrow W \in E_{ab}$ par invariance.
On construit une variable Z sur $[n]$ telle que Z et W soient indépendantes sachant Y.
L'idée est que sachant $Y=b$, on sait que W est dans $E_b = \cup_{a \in A} E_{ab}$.
On choisit alors Z uniformément dans E_b.
Y devient une fonction de Z.
Alors $H[X|Y] = H[X|Y,Z]$ (car X, Z indép. sachant Y (càd quel b on a))
$= H[X|Z]$ (car Y est fonction de Z)
$\leq H[X]$ (car Z est uniforme sur un ensemble plus grand).

La preuve formelle repose sur la concavité de la fonction $f(t) = -t \log t$ et l'inégalité de Jensen.

Corollaire 3.2

Si $X_1, \ldots, X_n$ sont des variables aléatoires, $H[X_1, \ldots, X_n] \leq \sum_{i=1}^n H[X_i]$.

Preuve

Par récurrence, en utilisant la **formule en chaîne** et la **sous-additivité** : $H[X_1, \ldots, X_n] = H[X_1, \ldots, X_{n-1}] + H[X_n | X_1, \ldots, X_{n-1}] \leq H[X_1, \ldots, X_{n-1}] + H[X_n]$.□

Corollaire 3.3: Positivité de l'entropie

$H[X] \geq 0$.

Preuve

Sans la formule

On a $H[X,X] = H[X]$ par invariance. Par sous-additivité, $H[X,X] \leq H[X] + H[X] = 2H[X]$. Donc $H[X] \leq 2H[X]$, ce qui implique $0 \leq H[X]$.□

Proposition 3.4: Inégalité de sous-modularité

Soient X, Y, Z des variables aléatoires. Alors $H[X|Y,Z] \leq H[X|Z]$.

Va être utilisé très souvent pendant le cours.

Preuve

C'est une conséquence directe de la sous-additivité.

$$\begin{aligned} H[X|Y,Z] &= \sum_z \mathbb{P}(Z=z) H[X|Y,Z=z] \\ &\leq \sum_z \mathbb{P}(Z=z) H[X|Z=z] \quad \text{(par sous-additivité sur la distribution conditionnelle)} \\ &= H[X|Z] par definition. \end{aligned}$$

Formes équivalentes de la sous-modularité :

En développant

$$\boxed{H[X,Y,Z] - H[Y,Z] \leq H[X,Z] - H[Z]}$$

Et on obtient en ajoutant $+H[Z]$:

$$\boxed{H[X,Y,Z] + H[Z] \leq H[X,Z] + H[Y,Z]}$$

Ou encore :

$$\boxed{H[X,Y,Z] \leq H[X,Z] + H[Y,Z] - H[Z]}$$

Corollaire 3.5: Sous-modularité fonctionnelle

Si Z est une fonction de Y (i.e. $Z = f(Y)$), alors $H[X|Y] \leq H[X|Z]$.

Intuitivement, si Z est une fonction de Y, elle donne moins d'information que Y, donc l'entropie de ce qu'on sait sachant Y est plus petite que sachant Z.

Preuve
Comme Z est fonction de Y, $H[X|Y,Z] = H[X|Y]$.
Par la sous-modularité, $H[X|Y] = H[X|Y,Z] \leq H[X|Z]$. (forme 1) □

Proposition 3.6

Si Z est à la fois fonction de X et fonction de Y, alors $H[X,Y] + H[Z] \leq H[X] + H[Y]$.

Preuve
Partons de l'inégalité de sous-modularité : $H[X,Y,Z] + H[Z] \leq H[X,Z] + H[Y,Z]$. (forme 3)
Comme Z est fonction de X, $H[X,Z] = H[X]$.
Comme Z est fonction de Y, $H[Y,Z] = H[Y]$.
De plus, Z est fonction de (X,Y), donc $H[X,Y,Z] = H[X,Y]$.
En substituant, on obtient : $H[X,Y] + H[Z] \leq H[X] + H[Y]$.□

4 Information Mutuelle

Définition 4.1: Information mutuelle

Soient X et Y deux variables aléatoires. L'information mutuelle entre X et Y est :

$$I[X:Y] = H[X] + H[Y] - H[X,Y]$$

C'est l'information partagée par la v.a.

I[X :Y] = I[Y :X] par symétrie.

Remarques :

- Par sous-additivité, $H[X,Y] \leq H[X] + H[Y]$, donc $I[X:Y] \geq 0$.
- Si X et Y sont indépendantes, $H[X,Y] = H[X] + H[Y]$, donc $I[X:Y] = 0$.
- Si $X = Y$, alors $H[X,Y] = H[X]$, donc $I[X:X] = H[X] + H[X] - H[X] = H[X]$.

L'identité $H[X] + H[Y] = H[X,Y] + I[X:Y]$ rappelle cette formule.

L'analogie avec la formule du crible pour les ensembles ($|A| + |B| = |A \cup B| + |A \cap B|$) fonctionne bien pour deux variables. Si on en a plus, des problèmes apparaissent, mais l'idée reste utile pour l'intuition.

Définition 4.2: Information mutuelle conditionnelle

L'information mutuelle conditionnelle de X et Y sachant Z est :

$$I[X:Y|Z] = H[X|Z] + H[Y|Z] - H[X,Y|Z]$$

C'est la forme que nous utiliserons le plus souvent. par symétrie on peut aussi écrire :

$$I = I[Y:X|Z] = H[X|Z] + H[Y|Z] - H[X,Y|Z]$$

car

$$I[Y:X|Z] = \sum_z \mathbb{P}(Z=z)\, I[X:Y|Z=z] = \sum_z \mathbb{P}(Z=z)\Big(H[X|Z=z] + H[Y|Z=z] - H[X,Y|Z=z]\Big)$$

Proposition 4.1: Formule en chaîne pour l'information mutuelle

$$I[X:Y,Z] = I[X:Z] + I[X:Y|Z]$$

Preuve

Le professeur a eu une hésitation à ce moment du cours pour retrouver la preuve car il n'utilise pas les diagramme de Venn.

On peut aussi définir $I[A:B] = H[A] - H[A|B]$. Alors :

$$\begin{aligned} I[X:Z] + I[X:Y|Z] &= \big(H[X] - H[X|Z]\big) + \big(H[X|Z] - H[X|Y,Z]\big) \\ &= H[X] - H[X|Y,Z] \\ &= I[X:Y,Z] \end{aligned}$$

(La conférence s'est interrompue ici, pendant la démonstration.)

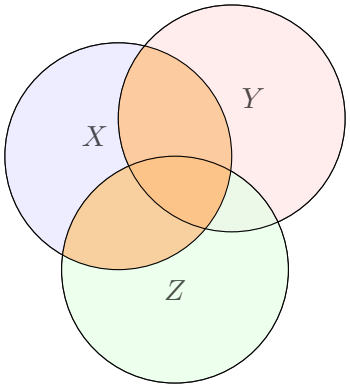
Fin du Cours 1

20 Octobre 2025

Rappels : Information Mutuelle

L'information mutuelle $I[X:Y,Z]$ est **l'information partagée** entre les variables aléatoires.

*Ce qui nous intéresse sur un **diagramme de Venn** c'est l'intersection de X avec l'union de Y et Z.*



Proposition 4.2: Formule en chaîne pour l'information

$$I[X : Y, Z] = I[X : Z] + I[X : Y|Z]$$

Preuve
On développe des 2 côtés en utilisant la définition.
Il faut prouver que $H[X] + H[Y, Z] - H[X, Y, Z] = H[X] + H[Z] - H[X, Z] + H[X|Z] + H[Y|Z] - H[X, Y|Z]$.
On cancel $H[X]$ des deux côtés. On a aussi $H[Y, Z] = H[Z] + H[Y|Z]$ qu'on cancel aussi. On a aussi $H[X, Z] = H[X|Z] + H[Z]$ donc on va ajouter $H[Z] - H[Z]$ à droite pour canceller $-H[X, Z] + H[Y|Z] + H[Z]$ et finalement $H[X, Y, Z] = H[X, Y|Z] + H[Z]$ donc tout va bien c'est fini.

On peut obtenir 2 autres relations par permutation de X Y Z parfois utile en combinaison.

On peut en déduire d'où le nom formule en chaîne. Si on a l'information mutuelle de $X : Y_1, \ldots, Y_n$ ça égale l'info $I[X : Y_1] + I[X : Y_2|Y_1] + \ldots$ comme la formule en chaîne pour l'entropie. On n'utilisera pas cette formule dans ce cours.

Corollaire 4.3: Inégalité de traitement des données (Data Processing Inequality)

Soient X, Y, Z trois variables aléatoires, telles que X et Z sont indépendants sachant Y. Alors **l'info mutuelle de X et Z est au plus l'information mutuelle de X sachant Y** :

$$I[X : Z] \leq I[X : Y]$$

Preuve
En utilisant la formule en chaîne deux fois, on note que : $I[X : Z] + I[X : Y|Z] = I[X : Y, Z] = I[X : Y] + I[X : Z|Y]$. Mais puisque X et Z sont indépendants sachant Y, le dernier terme $I[X : Z|Y]$ est 0. Et l'information mutuelle est toujours non négative $I[X : Y|Z] \geq 0$, le résultat suit. □

Les définitions et propriétés de base sont finies. Maintenant prouvons plusieurs théorèmes et applications de l'entropie.

5 Applications de l'entropie - Chemin dans un graphe

2 applications où on n'a pas besoin de nouvelles idées. Ensuite on verra un lemme important de Shearer pour prouver beaucoup d'autres théorèmes. Voici un théorème comme l'utilisation de l'entropie dans la combinatoire que j'aime beaucoup.

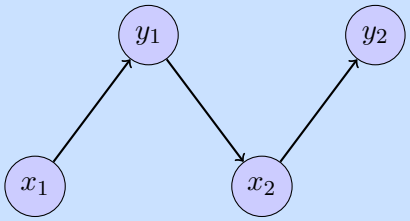
Théorème 5.1: Chemins de longueur 3 dans un graphe biparti

Soit G un graphe biparti fini et soit A et B les ensembles de sommet de G. Soit α la densité de G, c'est-à-dire : $|E(G)| = \alpha|A| \cdot |B|$.

$E(G)$ est l'ensemble des arêtes de G (car en anglais E = Edge).

Alors le nombre de quadruplets $(x_1, y_1, x_2, y_2) \in A \times B \times A \times B$ tel que $(x_1, y_1), (y_1, x_2)$ et (x_2, y_2) sont tous des arêtes de G,

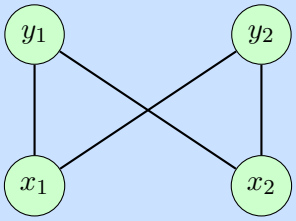
Ça veut dire qu'on a un chemin de longueur 3. C'est à peu près la même chose que le nombre de chemins de longueur 3, sauf que les chemins sont étiquetés (on commence par x_1) et aussi les chemins dégénérés sont permis (x_1 peut être égal à x_2). C'est donc à peu près le nombre de chemins de longueur 3.



est au moins égal à $\alpha^3|A|^2|B|^2$.

C'est-à-dire si on choisit deux sommets indépendants au hasard dans A et deux au hasard dans B, la proba que ces arêtes qui les relient (cf dessin) sont dans le graphe, est au moins α^3. Autrement dit, le nombre de ces quadruplets, est au moins ce qu'on attend dans un graphe aléatoire de densité α. Si on choisit toutes les arêtes, et chaque arête indépendamment de proba α, alors la probabilité que ce quadruplet avec le chemin de 3 (sauf dégénéré) va être de proba α^3. Donc autrement dit : pour minimiser les chemins de longueur 3, il faut prendre un graphe aléatoire, c'est ça l'idée.

C'est possible mais dur de prouver ce théorème sans entropie. Ce théorème ressemble à un théorème où on compte les cycles de longueur 4. Pour ce théorème il y a une preuve en 1 ligne avec l'inégalité de Cauchy-Schwartz, on a α^4. Mais pour notre théorème ici avec seulement 3 arêtes, est très dur sans entropie. La preuve avec l'entropie est facile à comprendre (pas forcément à trouver car il y a une idée).



Preuve

Soit P l'ensemble de quadruplets (x_1, y_1, x_2, y_2) qu'on cherche.

C'est-à-dire qui vérifie la condition de former un chemin de longueur 3 dans le graphe. Alors il suffit de trouver une variable aléatoire X qui prend les valeurs dans P tq $H[X] \geq \log(\alpha^3|A|^2|B|^2)$. Car si on fait ça, on sait que $H[X]$ est au plus le log de la taille de P ($\log|P|$) et donc la taille de P doit être au moins ce qu'on veut.

Pourquoi pas choisir une v.a. distribuée uniformément sur P pour maximiser l'entropie ? Mais si on fait ça, il reste à prouver que le $\log(|P|)$ est au moins $\alpha^3|A||B|$ ce qui revient au point de départ. Donc on va choisir une autre distribution sur P qui est plus facile à calculer avec les axiomes de Khinchin. Il ne faut pas choisir la distrib uniforme sur P mais une autre.

On va choisir la distribution suivante : Soit $X = (X_1, Y_1, X_2, Y_2)$ où (X_1, Y_1) est choisi uniformément de l'ensemble d'arêtes $E(G)$. Après, X_2 est un voisin aléatoire de Y_1. Et Y_2 un voisin aléatoire de X_2.

Aléatoire, ça veut dire choisi uniformément.

Observation expliquant que c'est une très belle distribution choisie : Si on choisit X_1, Y_1 uniformément dans les arêtes $E(G)$, on choisit Y_1 avec une proba proportionnelle au degré du sommet. Si on choisit un voisin aléatoire de Y_1 [...] le même argument dit que (X_2, Y_2) est aussi choisi uniformément parmi les arêtes de G. Fin de l'observation.

On choisit Y_1 avec une probabilité proportionnelle à son degré, et donc les distributions de (X_1, Y_1) et (Y_1, X_2) sont pareilles. Alors en suivant cette idée, (X_1, Y_1), (Y_1, X_2), et (X_2, Y_2) sont tous distribués uniformément sur $E(G)$. Maintenant on est près à faire un calcul : Par la formule en chaîne pour l'entropie, $H[X]$ =H[X_1, Y_1, X_2, Y_2] $= H[X_1] + H[Y_1|X_1] + H[X_2|X_1, Y_1] + H[Y_2|X_1, Y_1, X_2]$.

Observation : une fois qu'on sait ce qu'est Y_1, on n'a plus besoin de X_1 car X_2 est un voisin aléatoire de Y_1, cad X_2 et X_1 sont indépendants sachant Y_1, et on peut donc éliminer X_1 dans $H[X_2|X_1, Y_1]$.

$$H[X] = H[X_1] + H[Y_1|X_1] + H[X_2|Y_1] + H[Y_2|X_2]$$

Par additivité, $H[X_1] + H[Y_1|X_1] = H[X_1, Y_1]$. On applique l'additivité aux autres termes :

$$H[X] = H[X_1, Y_1] + (H[X_2, Y_1] - H[Y_1]) + (H[X_2, Y_2] - H[X_2])$$

Et on sait que ces 3 entropies ($H[X_1, Y_1]$, $H[X_2, Y_1]$ et $H[X_2, Y_2]$) sont égales car ces variables aléatoires sont distribuées uniformément sur $E(G)$, donc

$$H[X] = 3\log(|E(G)|) - H[X_2] - H[Y_1]$$

Où X_2 prend des valeurs dans A, et Y_1 dans B. Donc par maximalité, $H[X_2] \leq \log|A|$ et $H[Y_1] \leq \log|B|$.

$$H[X] \geq 3\log(\alpha|A||B|) - \log|A| - \log|B| = \log(\alpha^3|A|^2|B|^2)$$

L'idée de la preuve était de choisir une distribution qui permet de faire les simplifications dans les calculs.

6 Applications de l'entropie - Permanent d'une matrice

Il s'agit d'un permanent d'une matrice de 0 et 1.

Définition 6.1: Permanent

Soit A une matrice $n \times n$. Le **permanent** de A est la $\sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i,\sigma(i)}$.

Ça rappelle le déterminant mais on n'a pas les signes des permutations. C'est la seule différence. Mais même si la définition ressemble il y a une grosse différence entre le déterminant facile à calculer et le permanent qui est très difficile à calculer pour une matrice. Mais il est intéressant et important quand même.

Définition 6.2: Graphe biparti

Si A est une **matrice de 0s et 1s**, soit G_A le **graphe biparti** où xy est une arête ssi $A_{xy} = 1$. (C'est le graphe biparti qui correspond à la matrice). Donc le produit $\prod_{i=1}^{n} A_{i,\sigma(i)} = 1$ ssi la **permutation** σ **définit un couplage parfait** dans G_A et 0 sinon.

Définition 6.3: Couplage parfait

Un **couplage parfait** ça veut dire si on a un **motgraphe biparti**, c'est pour n sommets d'une partie et n sommets de l'autre partie avec n arêtes qui lient chaque sommet d'une partie avec un sommet de