

Entropie algébrique

Entropie des relations, des partitions et des applications

Hugues Genzrin

Octobre 2026

Résumé

On cherche à porter la notion d'entropie, issue de la physique et de la biologie, au niveau des structures mathématiques élémentaires : relations, relations d'équivalence, ensembles quotients, applications. Au point de départ se trouve des relations *inter-quotient* et *intra-quotient* d'une collection d'éléments affectée d'une information de Hartley . On définira le *nombre de complexions* de Boltzmann, que l'on reformulera comme le nombre d'applications ayant un profil de fibres donné. On décompose alors l'entropie de l'ensemble de départ, comme on décompose une variance, en une entropie *inter-quotient*, qui mesure l'incertitude sur la classe d'un élément, et une entropie *intra-quotient*, qui mesure l'incertitude qui subsiste à l'intérieur de sa classe ; leur somme est constante. On établit, avec démonstrations, la monotonie de l'entropie par raffinement des partitions, une identité exacte de composition, l'entropie de mélange, la conservation de l'entropie par les bijections et sa décroissance par itération d'une application non injective. Enfin, l'évolution d'un état à un autre est modélisée par une chaîne de Markov ; on démontre que la néguentropie, définie comme entropie relative à l'état stationnaire, décroît à chaque pas. Ces résultats rejoignent des théories établies : entropie des partitions, méthode des types, entropie des systèmes dynamiques.

Table des matières

1	Relations de correspondance	2
1.1	Définitions	2
1.2	Relations d'équivalence et ensembles quotients	3
1.3	La décomposition canonique	3
1.4	Patrons canoniques et prédicats	4
2	La composition abaisse l'information de Hartley	4
3	Shannon comme entropie de décomposition	5
4	Shannon sans les axiomes de Shannon	5
5	L'adjonction composition $\dashv$ décomposition, et son asymétrie	6
6	Trois quantités à ne pas confondre	6
7	Lecture en termes de jauge	6
8	Nombres de complexions	7
8.1	Complexions déterministes	7
8.2	Le profil d'une application	7
8.3	Complexions et entropie : la méthode des types	7

9	Entropie inter-quotient et entropie intra-quotient	8
9.1	Définitions	8
9.2	L'analogie avec l'analyse de la variance	9
9.3	Lecture du point de vue de l'application	9
9.4	Composition	10
10	L'entropie des partitions	10
10.1	Définition et raffinement	10
10.2	Deux partitions : information mutuelle	11
11	L'entropie de mélange	11
12	Dynamique : itérer une application	12
13	Chemins entre états : chaînes de Markov	13
13.1	Le cadre	13
13.2	Le second principe discret	13
13.3	Un exemple : le modèle d'Ehrenfest	13
13.4	Réversibilité et récurrence	14
13.4.1	La chaîne est réversible	14
13.4.2	Tout état revient	15
13.4.3	Bilan : trois sens du mot réversible	15
13.4.4	La récurrence comme conversion de néguentropie potentielle	16
14	Probabilités et énoncés	17
15	Synthèse	18

1 Relations de correspondance

1.1 Définitions

Définition 1.1 (Relation de correspondance). Soient E et F deux ensembles. Une *relation de correspondance* de E vers F est un triplet $R = (\Gamma, E, F)$ où $\Gamma \subseteq E \times F$ est son *graphe*. On note $x R y$ pour $(x, y) \in \Gamma$; on dit que y est un *conséquent* de x , et x un *antécédent* de y .

- Le *domaine* de R est $\text{Dom } R = \{x \in E : \exists y, x R y\}$.
- L'*image*, ou co-domaine, de R est $\text{Im } R = \{y \in F : \exists x, x R y\}$.
- La *réciproque* de R est $R^{-1} = (\Gamma^{-1}, F, E)$, avec $\Gamma^{-1} = \{(y, x) : (x, y) \in \Gamma\}$.
- L'*identité* de E est $\text{Id}_E = (\{(x, x) : x \in E\}, E, E)$.

Définition 1.2 (Types de relations). Une relation R de E vers F est :

- *univoque*, ou fonctionnelle, si tout x a au plus un conséquent : $x R y$ et $x R y'$ entraînent $y = y'$;
- *co-univoque*, ou injective, si tout y a au plus un antécédent : $x R y$ et $x' R y$ entraînent $x = x'$;
- *biunivoque* si elle est univoque et co-univoque ;
- *multivoque* si elle n'est pas univoque ;
- *polyvoque* si R et R^{-1} sont toutes deux multivoques.

Définition 1.3 (Composition). Si $R = (\Gamma, E, F)$ et $S = (\Delta, F, G)$, la composée $S \circ R = (\Delta \circ \Gamma, E, G)$ est définie par

$$x (S \circ R) z \iff \exists y \in F, \quad x R y \text{ et } y S z.$$

Remarque. La composée de deux relations est *toujours* définie : aucune condition d'univocité n'est requise.

Proposition 1.4. *Pour toute relation R de E vers F :*

1. $\text{Id}_{\text{Im } R} \subseteq R \circ R^{-1}$, avec égalité si et seulement si R est univoque ;
2. $\text{Id}_{\text{Dom } R} \subseteq R^{-1} \circ R$, avec égalité si et seulement si R est co-univoque.

En particulier, $R \circ R^{-1} = \text{Id}$ n'est vrai, en général, que pour une relation univoque, et sur l'image seulement.

Démonstration. On a $y(R \circ R^{-1})y'$ si et seulement s'il existe x tel que $x R y$ et $x R y'$. Si $y \in \text{Im } R$, on peut prendre $y' = y$: d'où l'inclusion. L'égalité demande que $x R y$ et $x R y'$ entraînent $y = y'$: c'est l'univocité. Le second point s'obtient en appliquant le premier à R^{-1} . $\square$

Une *fonction* est une relation univoque ; une *application* est une fonction dont le domaine est l'ensemble de départ tout entier. Une application co-univoque est injective, une application dont l'image est F est surjective.

1.2 Relations d'équivalence et ensembles quotients

Définition 1.5. Une relation $\sim$ de E vers E est une *relation d'équivalence* si elle est réflexive, symétrique et transitive. La *classe* de x est $\hat{x} = \{x' \in E : x' \sim x\}$; l'ensemble des classes est l'*ensemble quotient* $E/\sim$. On appelle *canon* le couple $(\sim, E/\sim)$.

Les classes d'équivalence forment une *partition* de E : des parties non vides, deux à deux disjointes, de réunion E . Inversement, toute partition est l'ensemble quotient d'une unique relation d'équivalence. On passera librement de l'un à l'autre point de vue.

Définition 1.6 (Noyau d'une application). Le *noyau* d'une application $\alpha : E \rightarrow F$ est la relation d'équivalence

$$x \sim_{\alpha} x' \iff \alpha(x) = \alpha(x').$$

Ses classes sont les *fibres* $\alpha^{-1}(y)$, $y \in \text{Im } \alpha$.

C'est la relation d'équivalence *induite* par α ; plus généralement, toute relation univoque induit ainsi une relation d'équivalence sur son domaine.

1.3 La décomposition canonique

Théorème 1.7 (Décomposition canonique). *Toute application $\alpha : E \rightarrow F$ se décompose de manière unique en*

$$\alpha = \iota \circ \bar{\alpha} \circ \nu,$$

où

- $\nu : E \rightarrow E/\sim_{\alpha}$, $x \mapsto \hat{x}$, est la surjection canonique ;
- $\bar{\alpha} : E/\sim_{\alpha} \rightarrow \text{Im } \alpha$, $\hat{x} \mapsto \alpha(x)$, est une bijection ;
- $\iota : \text{Im } \alpha \rightarrow F$ est l'injection canonique, restriction de l'identité.

$$\begin{array}{ccc} E & \xrightarrow{\alpha} & F \\ \nu \downarrow & & \uparrow \iota \\ E/\sim_{\alpha} & \xrightarrow{\bar{\alpha}} & \text{Im } \alpha \end{array}$$

Démonstration. L'application $\bar{\alpha}$ est bien définie, car $\hat{x} = \hat{x}'$ signifie $\alpha(x) = \alpha(x')$. Elle est injective, car $\alpha(x) = \alpha(x')$ entraîne $\hat{x} = \hat{x}'$, et surjective sur $\text{Im } \alpha$ par construction. L'égalité $\alpha = \iota \circ \bar{\alpha} \circ \nu$ se vérifie élément par élément. L'unicité de $\bar{\alpha}$ vient de la surjectivité de ν . $\square$

Dans le vocabulaire adopté ici, ν est la *complexion* : elle regroupe les éléments en classes, ou complexes. Une application qui choisit un élément dans chaque classe est une *décomplexion* ; il y en a $\prod_y \# \alpha^{-1}(y)$. La bijection $\bar{\alpha}$ est une décomplexion triviale.

1.4 Patrons canoniques et prédicats

Définition 1.8 (Patron canonique). Un *patron canonique* est une famille finie $P = (\sim_1, \dots, \sim_m)$ de canons sur E . Il définit l'application

$$\nu_P : E \rightarrow \prod_{j=1}^m E/\sim_j, \quad x \mapsto (\hat{x}^{(1)}, \dots, \hat{x}^{(m)}).$$

Le cas le plus simple est celui de m prédicats binaires $b_1, \dots, b_m$: chaque prédicat sépare E en deux classes, les éléments qui le satisfont et les autres. On obtient alors $\nu_P : E \rightarrow \{0, 1\}^m$.

Proposition 1.9. 1. ν_P est injective si et seulement si les prédicats séparent les éléments : pour $x \neq x'$, il existe j tel que $b_j(x) \neq b_j(x')$.
2. Si $\#E = n$, il faut au moins $\lceil \log_2 n \rceil$ prédicats binaires pour séparer tous les éléments, et ce nombre suffit.

Démonstration. Le premier point est la définition de l'injectivité. Pour le second, une injection de E dans $\{0, 1\}^m$ exige $2^m \geq n$; inversement, l'écriture en base 2 des numéros $0, \dots, n-1$ fournit $\lceil \log_2 n \rceil$ prédicats séparants. $\square$

Le nombre $\log_2 n$ est la *quantité d'information*, en bits, nécessaire pour identifier un élément de E : chaque prédicat binaire est un bit. C'est le point de départ de la théorie de l'information.

Remarque (Suites infinies de prédicats). Les suites infinies de bits forment l'*espace de Cantor* $\{0, 1\}^{\mathbb{N}}$, objet fondamental de la théorie de l'information. Il n'est cependant pas en bijection naturelle avec un intervalle réel, car le développement binaire n'est pas unique : $0,0111\dots = 0,1000\dots$; en outre, $x \mapsto 1/x$ envoie $]0; 1]$ sur $[1; +\infty[$, et ni les opérations de corps ni la complétude ne sont construites. La construction de $\mathbb{R}$ demande les coupures de Dedekind ou les suites de Cauchy. Les prédicats binaires relèvent donc de la mesure de l'information, non de la construction des nombres.

2 La composition abaisse l'information de Hartley

On reprend le cadre de *Applications du programme du grain* (§1.2). Soit un message de N symboles. *Non structuré*, il offre N alternatives, donc une information de Hartley $\mathbb{E}_{NS} = \log_2 N$. *Structuré*, il est partitionné en n sous-segments de tailles $N_1, \dots, N_n$ ($\sum_i N_i = N$). L'espérance structurée met en évidence une information intra-segmentaire ou intra-quotient qui émerge. En posant $\text{Pr}_i = N_i/N$, l'espérance de l'information de Hartley d'un sous-segment tiré selon Pr est $\mathbb{E}_S = \sum_i \text{Pr}_i \log_2 N_i$.

Théorème 2.1 (Composition et néguentropie ; *Applications*, Thm. 1.4). Avec $\text{Pr}_i = N_i/N$,

$$\Delta\mathbb{E} = \mathbb{E}_S - \mathbb{E}_{NS} = -H(\text{Pr}), \quad H(\text{Pr}) = -\sum_i \text{Pr}_i \log_2 \text{Pr}_i.$$

Démonstration. Puisque $\sum_i \text{Pr}_i = 1$, on a $\log_2 N = \sum_i \text{Pr}_i \log_2 N$, d'où

$$\Delta\mathbb{E} = \sum_i \text{Pr}_i \log_2 N_i - \sum_i \text{Pr}_i \log_2 N = \sum_i \text{Pr}_i \log_2 \frac{N_i}{N} = \sum_i \text{Pr}_i \log_2 \text{Pr}_i = -H(\text{Pr}).$$

Chaque étape n'utilise que $\sum_i \text{Pr}_i = 1$ et la différence des logarithmes ; tout reste en base 2, et $\Delta\mathbb{E} = -H$ en bits. $\square$

Composer — regrouper les symboles en sous-segments — retire donc à l'espérance de Hartley exactement $H(\text{Pr})$ bits. C'est le $\Delta\mathbb{E} = -H$ qui donne son nom à la néguentropie de composition.

3 Shannon comme entropie de décomposition

Lisons ce même fait dans le sens réciproque. Partons de la collection composée (on connaît le sous-segment de chaque élément) et *décomposons*-la jusqu'aux éléments natifs. Deux niveaux se présentent :

- localiser le *sous-segment* d'un élément tiré au hasard coûte $H(\text{Pr})$ bits ;
- localiser ensuite l'élément à l'intérieur de son sous-segment coûte, en moyenne, $\mathbb{E}_S = \sum_i \text{Pr}_i \log_2 N_i$ bits.

La somme restitue l'information de Hartley du message entier :

$$\log_2 N = \underbrace{H(\text{Pr})}_{\text{quel sous-segment}} + \underbrace{\sum_i \text{Pr}_i \log_2 N_i}_{\text{quel élément dans le sous-segment}} .$$

Théorème 3.1 (Shannon = entropie de décomposition = H_{intra}). *La décomposition de Hartley ci-dessus est la décomposition inter/intra de Entropie algébrique (Thm. 3.2), à la base du logarithme près :*

$$\mathbb{E}_{NS} = \log_2 N = H_{\text{intra}} + H_{\text{inter}}, \quad H_{\text{intra}} = H(\text{Pr}), \quad H_{\text{inter}} = \sum_i \text{Pr}_i \log_2 N_i = \mathbb{E}_S.$$

En particulier, l'entropie de Shannon de la partition est l'entropie de décomposition de la collection en ses sous-segments :

$$\boxed{H(\text{Pr}) == -\Delta \mathbb{E}_{\text{composition}}}$$

— la part de l'information de Hartley attribuable à la structure de composition, celle que composer met en réserve et que décomposer restitue.

Démonstration. L'identité $\log_2 N = H(\text{Pr}) + \mathbb{E}_S$ est le théorème précédent réarrangé. Or $H(\text{Pr}) = \sum_i \text{Pr}_i \log_2 (1/\text{Pr}_i) = \sum_i \frac{N_i}{N} \log_2 \frac{N}{N_i}$ est exactement l'entropie inter-quotient H_{intra} de la partition en sous-segments (*Entropie algébrique*, Déf. 3.1), et $\mathbb{E}_S = \sum_i \frac{N_i}{N} \log_2 N_i$ en est l'entropie intra-quotient H_{inter} . L'égalité $\log_2 N = H_{\text{intra}} + H_{\text{inter}}$ est donc le Thm. 3.2, et le dernier membre $H(\text{Pr}) = -\Delta \mathbb{E}$ est le théorème 1. $\square$

Corollaire 3.2 (Loi de conservation). *L'information totale se conserve, répartie entre décomposition et composition :*

$$\underbrace{H(\text{Pr})}_{\text{entropie de décomposition}} + \underbrace{\mathbb{E}_S}_{\text{Hartley résiduel intra}} = \log_2 N \quad (\text{constante, fixée par la résolution native}).$$

Composer déplace de l'information de l'inter vers une réserve ; décomposer la ramène. La somme est invariante.

4 Shannon sans les axiomes de Shannon

L'intérêt fondationnel est que H n'est ici ni postulée ni caractérisée par les axiomes de Shannon (continuité, additivité, monotonie). Elle *émerge* de deux ingrédients antérieurs :

1. l'information de Hartley $\log_2 N$ — un pur comptage d'alternatives ;
2. l'opération de composition en sous-segments.

Le théorème 1 montre alors que l'écart $\mathbb{E}_{NS} - \mathbb{E}_S$ vaut $H(\text{Pr})$. On ne contourne donc pas la valeur H : on la *construit* comme entropie de décomposition.

5 L'adjonction composition $\dashv$ décomposition, et son asymétrie

Composition et décomposition forment un couple réciproque : la néguentropie (déficit de Hartley) packagée par la composition égale l'entropie restituée par la décomposition. Au niveau fini, c'est une *identité* ($-\Delta E_{\text{comp}} = H = H_{\text{intra}}$). Mais la réciprocité n'est pas une involution.

Remarque (Asymétrie sur le continu ; *Applications*, Prop. 1.37). Composer un nombre fini de symboles engage une information finie ; décomposer le *continu* en engage une infinie — toute la finesse infinitésimale. La décomposition peut donc diverger là où la composition reste finie. L'adjonction composition $\dashv$ décomposition est *asymétrique* : il faut énoncer la dualité à résolution fixée (le niveau des « éléments natifs »), et cette asymétrie n'est pas une contradiction mais la raison d'être du continu comme réserve inépuisable de décomposition.

Remarque (« Éléments natifs » = choix de résolution). L'entropie de décomposition dépend du niveau auquel on décompose : décomposer en atomes ou en molécules ne donne pas la même valeur. H_{intra} est relative au niveau natif choisi, c'est-à-dire à la strate de résolution. Cette dépendance est celle de la tour du grain, et elle fait de « Shannon = entropie de décomposition » un énoncé *stratifié* : H se lit à chaque niveau de la tour.

6 Trois quantités à ne pas confondre

La dualité ci-dessus concerne l'entropie *inter-quotient* d'une partition. Elle est distincte des deux néguentropies de *Applications du programme du grain* (§1.3), qui mesurent autre chose.

Définition 6.1 (Déficit absolu contre néguentropie relative ; *Applications*, Rem. 1.6, Déf. 1.7). Le déficit de Hartley $-\Delta E = H(\text{Pr})$ est une néguentropie au sens de Schrödinger (déficit au maximum $E_{NS} = \log_2 N$), mais sa valeur *croît* quand la partition se déséquilibre : ce n'est pas la mesure d'ordre usuelle. La *néguentropie relative de composition* corrige la référence :

$$N_{\text{compo}}^{(\text{sous-seg.})} = \log_2 n - H(\text{Pr}) = D_{\text{KL}}(\text{Pr} \parallel u_n) \geq 0,$$

nulle à la partition équilibrée, croissante quand Pr se concentre.

Remarque (Deux axes orthogonaux ; *Applications*, Rem. 1.14). Il ne faut pas confondre l'entropie de décomposition $H(\text{Pr}) = H_{\text{intra}}$ avec les deux néguentropies *d'ordre* :

- la *néguentropie de biais* $N_{\text{biais}} = \log_2 m - H(p) = D_{\text{KL}}(p \parallel u)$ mesure l'inégalité des fréquences *d'une* position ;
- la *néguentropie de composition* propre (Déf. 1.12) $N_{\text{compo}} = \sum_{\ell} H(X_{\ell}) - H(X_1, \dots, X_L) = \text{TC}(X_1, \dots, X_L)$ est la corrélation totale : la *dépendance entre* positions.

Ces deux-là sont orthogonales (une pièce équilibrée mais collante a $N_{\text{biais}} = 0, N_{\text{compo}} > 0$; un dé truqué indépendant a $N_{\text{biais}} > 0, N_{\text{compo}} = 0$) et ne s'additionnent pas. L'entropie de décomposition $H(\text{Pr}) = H_{\text{intra}}$, elle, est l'entropie *brute* de la partition — l'objet du théorème 2, à distinguer de ces mesures d'ordre.

7 Lecture en termes de jauge

Dans le langage des jauges, la lecture est immédiate. Décomposer la collection en ses éléments natifs *augmente la finesse* (on passe de n classes à N éléments) ; son coût est l'entropie de décomposition $H(\text{Pr}) = H_{\text{intra}}$. Composer *package de l'ordre* : la néguentropie relative $\log_2 n - H(\text{Pr})$ mesure cet ordre. L'entropie de Shannon apparaît ainsi comme la *quantité de finesse qu'une décomposition doit créer* pour atteindre le niveau natif — ce qui la rattache directement à la quantification de la finesse de *Jauges entropiques* : H_{intra} y était la fonction réelle fidèle de la finesse, elle est ici l'entropie de décomposition.

8 Nombres de complexions

8.1 Complexions déterministes

Proposition 8.1. Soit $\alpha : E \rightarrow F$ et $F' \subseteq \text{Im } \alpha$ un complexe. Le nombre de parties $S \subseteq E$ telles que $\alpha(S) = F'$ est

$$N_{\alpha, F'} = \prod_{y \in F'} (2^{\#\alpha^{-1}(y)} - 1).$$

Démonstration. Une telle partie S est contenue dans $\alpha^{-1}(F')$, et se décompose en ses traces $S \cap \alpha^{-1}(y)$, $y \in F'$, qui sont indépendantes. La condition $\alpha(S) = F'$ demande que chaque trace soit non vide : $2^{\#\alpha^{-1}(y)} - 1$ choix pour chaque y . $\square$

Pour des fibres de tailles 3, 1 et 2, on trouve $7 \times 1 \times 3 = 21$ restrictions, ce qu'une énumération confirme. Une application injective a $N_{\alpha, F'} = 1$ pour tout complexe : elle n'admet qu'une restriction par image.

8.2 Le profil d'une application

Définition 8.2 (Profil). Soit $\alpha : E \rightarrow F$ avec $\#E = n$. Son *profil* est la famille $(n_y)_{y \in F}$ des tailles de ses fibres : $n_y = \#\alpha^{-1}(y)$, avec $\sum_y n_y = n$.

Théorème 8.3 (Nombre de complexions d'un profil). Le nombre d'applications $E \rightarrow F$ de profil (n_y) est le coefficient multinomial

$$W(n_y) = \frac{n!}{\prod_{y \in F} n_y!}.$$

Démonstration. On range les n éléments de E dans un ordre quelconque, ce qui donne $n!$ arrangements, et l'on envoie les n_{y_1} premiers sur y_1 , les n_{y_2} suivants sur y_2 , et ainsi de suite. Deux arrangements donnent la même application si et seulement s'ils ne diffèrent que par l'ordre à l'intérieur de chaque bloc, ce qui fait $\prod n_y!$ arrangements par application. $\square$

C'est exactement le nombre de complexions de Boltzmann, les éléments de E jouant le rôle des particules et ceux de F celui des niveaux d'énergie. C'est aussi le nombre de façons de répartir E en classes étiquetées de tailles données.

Proposition 8.4. Avec les effectifs cumulés $N_i = n_1 + \dots + n_i$,

$$W = \prod_{i=1}^{k-1} \binom{N_{i+1}}{N_i}.$$

Démonstration. $\binom{N_{i+1}}{N_i} = \frac{N_{i+1}!}{N_i! n_{i+1}!}$; le produit télescope et vaut $\frac{N_k!}{N_1! n_2! \dots n_k!} = \frac{n!}{n_1! \dots n_k!}$, puisque $N_1 = n_1$. $\square$

8.3 Complexions et entropie : la méthode des types

Définition 8.5. L'entropie de Shannon d'une distribution $p = (p_y)$ est $H(p) = -\sum_y p_y \ln p_y$, avec la convention $0 \ln 0 = 0$.

Théorème 8.6 (Méthode des types). Soit un profil (n_y) sur k valeurs, et $p_y = n_y/n$. Alors

$$\frac{e^{nH(p)}}{(n+1)^k} \leq W(n_y) \leq e^{nH(p)}.$$

Démonstration de la borne supérieure. D'après la formule du multinôme,

$$1 = \left(\sum_y p_y \right)^n = \sum_{\text{profils } (m_y)} W(m_y) \prod_y p_y^{m_y} \geq W(n_y) \prod_y p_y^{n_y}.$$

Or $\prod_y p_y^{n_y} = \exp(n \sum_y p_y \ln p_y) = e^{-nH(p)}$. La borne inférieure se démontre en comparant les termes de la somme ; voir Csiszár et Körner, ou Cover et Thomas. $\square$

Ainsi $\frac{1}{n} \ln W \rightarrow H(p)$: l'entropie « algébrique » $\ln W$ est, à l'échelle près, l'entropie de Shannon des proportions. C'est la justification exacte de l'usage de la formule de Stirling.

Remarque. Même pour des fibres de tailles égales, W n'est pas le nombre $(\#F)!$ de permutations des images : avec $k = 6$ correspondances réparties également sur 3 images, $W = \frac{6!}{(2!)^3} = 90$, et non $3! = 6$.

9 Entropie inter-quotient et entropie intra-quotient

9.1 Définitions

On munit E , de cardinal n , de la loi uniforme : X est un élément tiré au hasard dans E . Une application $\alpha : E \rightarrow F$ de profil (n_y) répartit E en classes : les fibres $\alpha^{-1}(y)$, qui sont les classes du noyau $\sim_\alpha$. L'élément X tombe dans la classe $\alpha^{-1}(y)$ avec probabilité n_y/n .

Définition 9.1. — L'*entropie inter-quotient* de α est l'entropie de la répartition de E entre les classes :

$$H_{\text{inter}}(\alpha) = H(\alpha(X)) = \sum_y \frac{n_y}{n} \ln \frac{n}{n_y}.$$

C'est l'information qu'apporte la connaissance de la classe de X : de combien elle *réduit* l'entropie de E .

— L'*entropie intra-quotient* de α est l'entropie moyenne de E à l'intérieur des classes :

$$H_{\text{intra}}(\alpha) = H(X \mid \alpha(X)) = \sum_y \frac{n_y}{n} \ln n_y.$$

C'est l'incertitude qui subsiste sur X une fois sa classe connue : dans la classe $\alpha^{-1}(y)$, il reste $\ln n_y$ à déterminer, et l'on fait la moyenne sur les classes.

Les deux quantités ne dépendent que de la partition de E en classes, c'est-à-dire du noyau de α , et non des valeurs prises par α . C'est cohérent avec le théorème 1.7 : dans la décomposition $\alpha = \iota \circ \bar{\alpha} \circ \nu$, seule la surjection canonique ν regroupe des éléments ; $\bar{\alpha}$ et ι sont injectives. On parlera donc indifféremment de l'entropie inter-quotient de α , de son noyau, ou du quotient $E/\sim_\alpha$.

Théorème 9.2 (Décomposition).

$$\ln n = H_{\text{inter}}(\alpha) + H_{\text{intra}}(\alpha).$$

De plus :

- $H_{\text{intra}}(\alpha) = 0$ et $H_{\text{inter}}(\alpha) = \ln n$ si et seulement si α est injective : toutes les classes sont des singletons ;
- $H_{\text{inter}}(\alpha) = 0$ et $H_{\text{intra}}(\alpha) = \ln n$ si et seulement si α est constante : il n'y a qu'une classe.

Démonstration. On additionne : $\sum_y \frac{n_y}{n} (\ln \frac{n}{n_y} + \ln n_y) = \ln n \sum_y \frac{n_y}{n} = \ln n$. On a $H_{\text{intra}}(\alpha) = 0$ si et seulement si toutes les classes ont un seul élément, c'est-à-dire si α est injective ; et $H_{\text{inter}}(\alpha) = 0$ si et seulement si il n'y a qu'une classe. $\square$

Exemple 9.3. Soit E à 8 éléments, réparti par α en quatre classes de tailles 4, 2, 1 et 1. En bits (logarithmes en base 2) :

Classe	a	b	c	d	Moyenne pondérée
Taille n_y	4	2	1	1	
Probabilité n_y/n	1/2	1/4	1/8	1/8	
Inter : $\log_2(n/n_y)$	1	2	3	3	$H_{\text{inter}} = 1,75$
Intra : $\log_2 n_y$	2	1	0	0	$H_{\text{intra}} = 1,25$

Pour identifier un élément de E , il faut $\log_2 8 = 3$ bits. Connaître sa classe en apporte 1,75 en moyenne ; il en reste 1,25 à trouver à l'intérieur de la classe. Pour un élément de la classe a, qui en compte quatre, il reste deux questions oui/non à poser ; pour un élément des classes c ou d, aucune.

9.2 L'analogie avec l'analyse de la variance

La décomposition du théorème 9.2 est l'exact analogue de la décomposition de la variance en variance inter-classes et intra-classes.

Proposition 9.4 (Théorème de Huygens). *Pour une variable réelle Z définie sur E et une partition de E en classes,*

$$\text{Var}(Z) = \underbrace{\text{Var}(\mathbb{E}[Z \mid \text{classe}])}_{\text{variance inter-classes}} + \underbrace{\mathbb{E}[\text{Var}(Z \mid \text{classe})]}_{\text{variance intra-classes}}.$$

	Variance	Entropie
Totale	$\text{Var}(Z)$	$H(X) = \ln n$
Inter-classes	dispersion des moyennes de classe	H_{inter} : incertitude sur la classe
Intra-classes	dispersion moyenne dans les classes	H_{intra} : incertitude dans la classe
Partition triviale (une classe)	inter nulle	$H_{\text{inter}} = 0$
Partition discrète (single-tons)	intra nulle	$H_{\text{intra}} = 0$

Remarque (Loi quelconque). La décomposition reste vraie pour toute loi de probabilité sur E , et non seulement pour la loi uniforme : c'est la règle de chaînage de Shannon,

$$H(X) = H(\text{classe de } X) + H(X \mid \text{classe de } X),$$

qui découle de ce que la classe est une fonction de X . Le membre de gauche vaut $\ln n$ seulement pour la loi uniforme.

9.3 Lecture du point de vue de l'application

Vues depuis l'application α , les deux entropies ont une seconde lecture.

- L'entropie inter-quotient est l'information que $\alpha(X)$ *transmet* sur X : c'est l'information mutuelle $I(X; \alpha(X))$, égale ici à $H(\alpha(X))$ puisque $\alpha(X)$ est une fonction de X . En ce sens, c'est l'information *conservée* par α .

- L'entropie intra-quotient est ce qu'il faut *ajouter* à la connaissance de $\alpha(X)$ pour retrouver X : l'information *effacée* par α . Shannon appelait cette quantité l'*équivoque*. Si α est réalisée par un dispositif physique à la température T , le principe de Landauer impose de dissiper au moins $k_B T H_{\text{intra}}(\alpha)$ de chaleur. C'est le pont exact entre l'entropie d'une application et l'entropie physique.

Remarque (Deux lectures d'une bijection). On est tenté d'attribuer à une bijection tantôt une entropie nulle (elle est « déterministe »), tantôt l'entropie maximale $\ln n!$ (toutes les permutations sont possibles). Le théorème 9.2 lève l'ambiguïté : il s'agit de deux quantités différentes et complémentaires. L'entropie intra-quotient est nulle pour une bijection, l'entropie inter-quotient y est maximale. D'après le théorème 8.6, $\frac{1}{n} \ln W \approx H_{\text{inter}}(\alpha)$; et $\ln n! - \ln W = \sum_y \ln n_y! \approx n H_{\text{intra}}(\alpha)$ pour de grandes classes.

9.4 Composition

Théorème 9.5. Pour $\alpha : E \rightarrow F$ et $\beta : F \rightarrow G$,

$$H_{\text{intra}}(\beta \circ \alpha) = H_{\text{intra}}(\alpha) + H(\alpha(X) \mid \beta(\alpha(X))) \geq H_{\text{intra}}(\alpha), \quad H_{\text{inter}}(\beta \circ \alpha) \leq H_{\text{inter}}(\alpha).$$

Composer avec β fusionne des classes de α : l'entropie intra-quotient ne peut que croître, l'entropie inter-quotient ne peut que décroître. L'information effacée ne se retrouve jamais.

Démonstration. Notons $Y = \alpha(X)$ et $Z = \beta(Y)$. Comme Y et Z sont des fonctions de X , la règle de chaînage donne

$$H(X \mid Z) = H(X, Y \mid Z) = H(Y \mid Z) + H(X \mid Y, Z) = H(Y \mid Z) + H(X \mid Y),$$

la dernière égalité venant de ce que Z est une fonction de Y . Le terme $H(Y \mid Z)$ est positif. L'inégalité sur H_{inter} s'en déduit par le théorème 9.2. $\square$

10 L'entropie des partitions

10.1 Définition et raffinement

Définition 10.1. L'*entropie* d'une partition P de E , ou de la relation d'équivalence correspondante, est l'entropie inter-quotient de sa surjection canonique :

$$H(P) = \sum_{A \in P} \frac{\#A}{n} \ln \frac{n}{\#A}.$$

Son *nombre de complexions* est $W(P) = \frac{n!}{\prod_{A \in P} (\#A)!}$.

Une partition Q est *plus fine* que P si chaque classe de Q est contenue dans une classe de P ; on note $Q \preceq P$. Pour chaque $A \in P$, Q induit une partition Q_A de A .

Théorème 10.2 (Composition, forme exacte). Si $Q \preceq P$,

$$W(Q) = W(P) \times \prod_{A \in P} W(Q_A) \quad \text{et} \quad H(Q) = H(P) + \sum_{A \in P} \frac{\#A}{n} H(Q_A).$$

Démonstration. Pour les complexions :

$$W(P) \prod_A W(Q_A) = \frac{n!}{\prod_A (\#A)!} \prod_A \frac{(\#A)!}{\prod_{B \in Q_A} (\#B)!} = \frac{n!}{\prod_{B \in Q} (\#B)!} = W(Q).$$

Pour les entropies, on écrit, pour $B \subseteq A$, $\ln \frac{n}{\#B} = \ln \frac{n}{\#A} + \ln \frac{\#A}{\#B}$, puis on somme avec les poids $\frac{\#B}{n}$. $\square$

Corollaire 10.3 (Croissance par raffinement). *Si $Q \preceq P$, alors $W(Q) \geq W(P)$ et $H(Q) \geq H(P)$, avec égalité si et seulement si $Q = P$.*

Démonstration. Chaque $W(Q_A)$ est un entier au moins égal à 1, et chaque $H(Q_A)$ est positive ; $W(Q_A) = 1$ pour tout A signifie que Q_A n'a qu'une classe, donc $Q = P$. $\square$

Remarque. Deux cas particuliers méritent d'être nommés. Ajouter une classe en scindant une classe existante augmente l'entropie quotient : c'est le corollaire 10.3. Composer deux partitions successives dissipe l'entropie de façon exactement comptabilisée : le théorème 10.2 donne une *égalité*, sans facteur correctif. C'est la règle de chaînage de Shannon.

Remarque (Emboîtement). Le théorème 10.2 est une décomposition inter/intra emboîtée. Si $Q \preceq P$, l'entropie inter-quotient de Q se décompose en l'entropie inter-quotient de P et la moyenne, sur les classes A de P , de l'entropie inter-quotient de Q_A à l'intérieur de A . En raffinant une partition, on fait passer de l'entropie de la part intra à la part inter, exactement comme dans une analyse de la variance hiérarchique, où chaque niveau d'emboîtement transfère de la variance intra-classes en variance inter-classes.

10.2 Deux partitions : information mutuelle

La *partition jointe* $P \vee Q$ a pour classes les intersections non vides $A \cap B$, $A \in P$, $B \in Q$: elle correspond au patron canonique formé des deux canons.

Théorème 10.4.

$$\max(H(P), H(Q)) \leq H(P \vee Q) \leq H(P) + H(Q).$$

La quantité $I(P; Q) = H(P) + H(Q) - H(P \vee Q) \geq 0$ est l'information mutuelle des deux partitions. Elle est nulle si et seulement si les partitions sont indépendantes : $\frac{\#(A \cap B)}{n} = \frac{\#A}{n} \cdot \frac{\#B}{n}$ pour tous A, B .

Démonstration. La borne inférieure vient du corollaire 10.3, puisque $P \vee Q$ est plus fine que P et que Q . Pour la borne supérieure, notons $p_{AB} = \#(A \cap B)/n$, p_A et p_B les marges. Alors

$$H(P) + H(Q) - H(P \vee Q) = \sum_{A,B} p_{AB} \ln \frac{p_{AB}}{p_A p_B} = D((p_{AB}) \parallel (p_A p_B)) \geq 0,$$

l'entropie relative étant positive, et nulle seulement si les deux distributions sont égales. $\square$

Deux relations d'équivalence « en corrélation totale » sont deux partitions dont l'une détermine l'autre : leur information mutuelle est alors maximale, égale à la plus petite des deux entropies.

Remarque (Décomposition canonique et réduction d'entropie).

Si un patron canonique $(\sim_1, \dots, \sim_m)$ détermine parfaitement une application α — la valeur $\alpha(x)$ ne dépend que des classes de x —, alors le noyau de α est moins fin que la partition jointe, et $H_{\text{inter}}(\alpha) \leq H(\sim_1 \vee \dots \vee \sim_m) \leq \sum_j H(\sim_j)$. Décrire α par son patron canonique est d'autant plus économique que les canons sont peu nombreux et peu corrélés.

11 L'entropie de mélange

On réunit deux ensembles disjoints E_1 et E_2 , de cardinaux n_1 et n_2 , munis de partitions P_1 et P_2 , de nombres de complexions W_1 et W_2 .

Théorème 11.1 (Mélange). 1. *Si les classes de P_1 et de P_2 restent distinctes dans la réunion,*

$$W_3 = \binom{n_1 + n_2}{n_1} W_1 W_2.$$

2. Si l'on fusionne des classes de même étiquette — P_1 et P_2 ont des classes i de tailles a_i et b_i , et la réunion a des classes de tailles $a_i + b_i$ —,

$$W_1 W_2 \leq W_3 \leq \binom{n_1 + n_2}{n_1} W_1 W_2.$$

Dans les deux cas, $\ln W_3 \geq \ln W_1 + \ln W_2$: le mélange crée de l'entropie.

Démonstration. 1. On choisit les n_1 éléments qui proviennent de E_1 , de $\binom{n_1+n_2}{n_1}$ façons, puis on les répartit en classes de P_1 , de W_1 façons, et les autres en classes de P_2 , de W_2 façons.

2. On a

$$\frac{W_3}{W_1 W_2} = \frac{(n_1 + n_2)!}{\prod_i (a_i + b_i)!} \cdot \frac{\prod_i a_i! \prod_i b_i!}{n_1! n_2!} = \frac{\binom{n_1+n_2}{n_1}}{\prod_i \binom{a_i+b_i}{a_i}}.$$

Le dénominateur compte les façons de choisir n_1 éléments parmi n_1+n_2 en en prenant exactement a_i dans chaque classe i ; le numérateur compte tous les choix de n_1 éléments. Le quotient est donc compris entre 1 et $\binom{n_1+n_2}{n_1}$. $\square$

Le facteur $\binom{n_1+n_2}{n_1}$ est l'entropie de mélange de Gibbs. Quand on mélange des classes identiques, le gain est plus faible : c'est la version combinatoire du paradoxe de Gibbs.

Remarque. On n'a pas en général $W_3 \geq \max(W_1, W_2)^2$: pour E_1 de 10 éléments en singletons ($W_1 = 10!$) et E_2 d'un seul élément ($W_2 = 1$), on a $W_3 = 11 \times 10! \approx 4 \times 10^7$, alors que $(10!)^2 \approx 1,3 \times 10^{13}$. Seule l'inégalité entropique $H_3 \geq H_1 + H_2$ est générale.

12 Dynamique : itérer une application

Soit $\alpha : E \rightarrow E$ une application d'un ensemble fini dans lui-même, et X_0 uniforme sur E . On pose $X_{t+1} = \alpha(X_t)$ et $H_t = H(X_t)$.

Théorème 12.1. 1. Si α est une bijection, $H_t = \ln n$ pour tout t : l'entropie est conservée.
2. En général, la suite (H_t) est décroissante, et constante à partir d'un rang au plus égal à n . Sa limite est l'entropie de la loi de X_t sur l'image ultime

$$E_\infty = \bigcap_{t \geq 0} \alpha^t(E),$$

sur laquelle α induit une bijection.

Démonstration. 1. Une bijection envoie la loi uniforme sur la loi uniforme.

2. X_{t+1} est une fonction de X_t , donc $H_{t+1} \leq H_t$ (théorème 9.5). La suite d'ensembles $\alpha^t(E)$ est décroissante pour l'inclusion ; elle est donc stationnaire à partir d'un rang $t_0 \leq n$, égale à E_∞ , et $\alpha(E_\infty) = E_\infty$. Une surjection d'un ensemble fini sur lui-même est une bijection. Pour $t \geq t_0$, X_t prend ses valeurs dans E_∞ , où α agit comme une bijection, qui ne fait que permuer les probabilités : l'entropie ne change plus. $\square$

Exemple 12.2. Sur $E = \{0, \dots, 7\}$, l'application $\alpha = (1, 2, 0, 0, 3, 4, 4, 6)$ — c'est-à-dire $\alpha(0) = 1$, $\alpha(1) = 2$, etc. — donne les entropies successives

$$H_0 = \ln 8 \approx 2,079, \quad 1,733, \quad 1,560, \quad 1,321, \quad 1,082, \quad 1,082, \dots$$

L'image ultime est le cycle $\{0, 1, 2\}$, et l'information effacée au total vaut $\ln 8 - 1,082 \approx 1,0$ nat.

Une bijection est l'analogue d'une dynamique *réversible* : elle conserve l'entropie, comme le théorème de Liouville en mécanique. Une application non injective est l'analogue d'une dynamique *irréversible* : elle efface de l'information, et par le principe de Landauer cet effacement a un coût physique.

Remarque (Cycles). Une permutation α est d'ordre r si $\alpha^r = \text{Id}$; r est le plus petit commun multiple des longueurs de ses cycles. Un r -cycle, au sens strict, ne fait tourner que r éléments. La cyclicité ne crée pas de néguentropie : le théorème 12.1 montre qu'une permutation conserve exactement l'entropie. Pour des systèmes infinis, l'entropie de Kolmogorov–Sinaï mesure la vitesse à laquelle une dynamique produit de l'information; elle est nulle pour toute dynamique sur un ensemble fini.

13 Chemins entre états : chaînes de Markov

13.1 Le cadre

On étudie le passage d'un état ST_1 — une répartition des éléments en classes — à un état ST_2 . Le cadre rigoureux de ces transitions aléatoires est celui des chaînes de Markov.

Définition 13.1. Une *chaîne de Markov* sur un ensemble fini d'états est donnée par une matrice de transition $P = (P_{ij})$ à coefficients positifs, dont chaque ligne a pour somme 1 : P_{ij} est la probabilité de passer de l'état i à l'état j . Si p_t est la loi de l'état au temps t , alors $p_{t+1} = p_t P$. Une loi π est *stationnaire* si $\pi P = \pi$.

Lemme 13.2 (Inégalité log-somme). *Pour des réels positifs a_i et b_i ,*

$$\sum_i a_i \ln \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \ln \frac{\sum_i a_i}{\sum_i b_i}.$$

Démonstration. La fonction $f(u) = u \ln u$ est convexe. Avec $B = \sum b_i$ et les poids b_i/B , l'inégalité de Jensen donne $\sum_i \frac{b_i}{B} f\left(\frac{a_i}{b_i}\right) \geq f\left(\sum_i \frac{a_i}{B}\right)$, ce qui est l'inégalité annoncée divisée par B . $\square$

13.2 Le second principe discret

Théorème 13.3. *Si π est stationnaire, l'entropie relative à π décroît à chaque pas :*

$$D(p_{t+1} \parallel \pi) \leq D(p_t \parallel \pi).$$

Démonstration. Pour chaque état j , on applique l'inégalité log-somme avec $a_i = (p_t)_i P_{ij}$ et $b_i = \pi_i P_{ij}$. Comme $\sum_i a_i = (p_{t+1})_j$ et $\sum_i b_i = \pi_j$,

$$(p_{t+1})_j \ln \frac{(p_{t+1})_j}{\pi_j} \leq \sum_i (p_t)_i P_{ij} \ln \frac{(p_t)_i}{\pi_i}.$$

On somme sur j , en utilisant $\sum_j P_{ij} = 1$. $\square$

Corollaire 13.4. *Si la matrice P est bistochastique — ses colonnes ont aussi pour somme 1 —, la loi uniforme est stationnaire et l'entropie $H(p_t)$ croît à chaque pas.*

Démonstration. Avec π uniforme sur N états, $D(p \parallel \pi) = \ln N - H(p)$. $\square$

La néguentropie, définie comme entropie relative à l'état d'équilibre, ne peut que diminuer : c'est le « théorème H » de Boltzmann sous une forme discrète et exacte.

13.3 Un exemple : le modèle d'Ehrenfest

On répartit n boules numérotées entre deux urnes. À chaque pas, on tire une boule au hasard et, avec probabilité $\frac{1}{2}$, on la change d'urne. Un *micro-état* précise l'urne de chaque boule (2^n micro-états); un *macro-état* est le nombre K de boules dans la première urne. Cette chaîne modélise le transfert d'éléments entre deux classes.

La chaîne des micro-états est bistochastique; celle des macro-états a pour loi stationnaire la loi binomiale $\pi_k = \binom{n}{k}/2^n$. Partant de toutes les boules dans la première urne, avec $n = 20$:

Pas t	Nombre moyen dans l'urne 1	Néguentropie $D(p_t \parallel \pi)$	Entropie des micro-états S_t	Somme
0	20,0	13,863	0,000	13,863
1	19,5	11,672	2,191	13,863
5	17,7	6,864	6,999	13,863
10	16,0	3,866	9,997	13,863
20	13,6	1,321	12,542	13,863
40	11,3	0,166	13,697	13,863
80	10,2	0,003	13,860	13,863

La néguentropie décroît vers 0 et l'entropie des micro-états croît vers son maximum $20 \ln 2 \approx 13,863$, leur somme restant constante.

Proposition 13.5. *Pour le modèle d'Ehrenfest, à tout instant,*

$$D(p_t \parallel \pi) + S_t = n \ln 2.$$

Démonstration. Par symétrie, sachant $K = k$, tous les $\binom{n}{k}$ micro-états correspondants sont équiprobables. Donc $S_t = H(K) + \mathbb{E}[\ln \binom{n}{K}]$, et

$$D(p_t \parallel \pi) = \sum_k p_t(k) \ln \frac{p_t(k) 2^n}{\binom{n}{k}} = n \ln 2 - H(K) - \mathbb{E}\left[\ln \binom{n}{K}\right]. \quad \square$$

C'est exactement la relation $\mathcal{J} = S_{\max} - S$: la néguentropie est l'écart entre l'entropie maximale et l'entropie du système.

Remarque (Entropie de chemin). L'irréversibilité ne peut pas se lire sur la probabilité d'un événement conjoint : comme $ST_1 \cap ST_2 \subseteq ST_1$, on a toujours $\Pr(ST_1 \cap ST_2) \leq \Pr(ST_1)$. Dans le cadre markovien, un état ST_2 peut succéder à ST_1 si et seulement si $P_{ST_1, ST_2} > 0$, et la flèche du temps est donnée, au niveau de la loi, par le théorème 13.3 : la néguentropie décroît à chaque pas. La sous-section suivante montre que cette flèche est compatible avec une dynamique réversible.

13.4 Réversibilité et récurrence

Le théorème 13.3 semble introduire une flèche du temps dans une dynamique dont chaque pas peut être défait : une boule qui change d'urne peut revenir au pas suivant. Deux objections ont été opposées au théorème H de Boltzmann : l'objection de *réversibilité* (Loschmidt, 1876) et l'objection de *récurrence* (Zermelo, 1896, d'après le théorème de Poincaré). Le modèle des urnes a été proposé par les Ehrenfest (1907) précisément pour montrer qu'elles ne contredisent pas le théorème.

13.4.1 La chaîne est réversible

Définition 13.6. Une chaîne de Markov de matrice P et de loi stationnaire π est *réversible* si elle vérifie le *bilan détaillé* :

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \text{pour tous les états } i, j.$$

Proposition 13.7. *La chaîne d'Ehrenfest des macro-états est réversible.*

Démonstration. Seules les transitions entre k et $k - 1$ sont à vérifier. On a $P_{k,k-1} = \frac{1}{2} \cdot \frac{k}{n}$ et $P_{k-1,k} = \frac{1}{2} \cdot \frac{n-k+1}{n}$, d'où

$$\pi_k P_{k,k-1} = \frac{1}{2^{n+1}n} \binom{n}{k} k = \frac{1}{2^{n+1}n} \binom{n}{k-1} (n-k+1) = \pi_{k-1} P_{k-1,k},$$

puisque $\binom{n}{k} k = \binom{n}{k-1} (n-k+1) = \frac{n!}{(k-1)!(n-k)!}$. □

Proposition 13.8 (Symétrie temporelle). *Soit $(X_t)_{t \in \mathbb{Z}}$ une chaîne réversible à l'équilibre, c'est-à-dire de loi π à chaque instant. Alors le processus renversé (X_{-t}) a la même loi que (X_t) . En particulier, pour tout état i et tout $t \geq 0$,*

$$\Pr(X_{-t} = j \mid X_0 = i) = \Pr(X_t = j \mid X_0 = i).$$

Démonstration. Par la formule de Bayes et le bilan détaillé, $\Pr(X_{-1} = j \mid X_0 = i) = \pi_j P_{ji} / \pi_i = P_{ij}$: le passé immédiat a la même loi que le futur immédiat. On conclut par récurrence sur t , grâce à la propriété de Markov. □

Pour le modèle d'Ehrenfest, cela signifie qu'à l'équilibre, le film de l'évolution passé à l'envers a exactement les mêmes statistiques qu'à l'endroit. Et si l'on surprend, au cours d'une longue observation à l'équilibre, un état rare — toutes les boules dans la première urne —, cet état *venait* le plus probablement de l'équilibre et *y retourne* : la néguentropie décroît dans les deux sens du temps à partir de ce pic.

13.4.2 Tout état revient

Théorème 13.9 (Lemme de Kac). *Pour une chaîne de Markov irréductible sur un ensemble fini, de loi stationnaire π , partant de l'état k , le temps moyen de premier retour en k vaut*

$$\mathbb{E}_k[T_k] = \frac{1}{\pi_k}.$$

On l'admet ici ; c'est un résultat classique de la théorie des chaînes de Markov. Pour le modèle d'Ehrenfest, $\pi_k = \binom{n}{k} / 2^n$.

	$n = 10$	$n = 20$	$n = 100$
Retour à l'équilibre $k = n/2 : 2^n / \binom{n}{n/2}$	4,1 pas	5,7 pas	12,6 pas
Retour à l'état initial $k = n : 2^n$	1 024 pas	1 048 576 pas	$1,3 \times 10^{30}$ pas
Durée, à un pas par seconde	17 minutes	12 jours	4×10^{22} ans

Une simulation de trois millions de pas avec $n = 10$ donne un temps moyen de retour à l'état initial d'environ 1 060 pas, conforme à la valeur 1 024 du lemme de Kac. Pour $n = 100$, le retour prend des milliers de milliards de fois l'âge de l'univers ; pour un gaz réel, qui compte de l'ordre de 10^{23} molécules, il est inobservable.

13.4.3 Bilan : trois sens du mot réversible

Sens	Ehrenfest	Raison
Chaque trajectoire peut revenir en arrière (bilan détaillé)	oui	proposition 13.7
Aucune entropie n'est produite (sens thermodynamique)	non	hors d'équilibre, $D(p_t \parallel \pi)$ décroît strictement
Tout état finit par revenir (récurrence)	oui	lemme de Kac, en un temps astronomique dès que n est grand

Il n'y a pas de contradiction. Le théorème H porte sur la loi p_t , c'est-à-dire sur une moyenne sur toutes les trajectoires issues d'un même état initial ; une trajectoire individuelle fluctue, s'écarte parfois de l'équilibre et revient en particulier, un jour, à son état de départ. La flèche du temps ne vient pas de la dynamique, qui est symétrique (proposition 13.8), mais de la *condition initiale* : on part d'un état très improbable, préparé à la main. L'irréversibilité est statistique, non absolue.

Remarque (Lien avec les bijections). La dynamique microscopique d'un vrai gaz est déterministe : c'est une bijection de l'espace des états, qui conserve l'entropie fine (théorème 12.1 et théorème de Liouville). Le modèle d'Ehrenfest la remplace par un tirage au hasard : sa matrice de transition n'est pas une permutation, mais une matrice bistochastique qui mélange les probabilités, et c'est ce mélange qui fait croître l'entropie de la loi. Le hasard du modèle représente notre ignorance du détail des collisions : l'irréversibilité apparaît avec la description grossière.

13.4.4 La récurrence comme conversion de néguentropie potentielle

On reprend la décomposition de la néguentropie en une part *calculée* et une part *potentielle* (voir le document *Néguentropie du vivant*). On écrit le micro-état ω sous la forme $\omega = (y, z)$, où y est la variable observée — ici le macro-état K — et z le détail qui reste à préciser : *quelles* boules sont dans la première urne. La règle de chaînage de l'entropie relative donne, pour toute loi p sur les micro-états et la loi uniforme u ,

$$D(p \parallel u) = \underbrace{D(p_K \parallel \pi)}_{J_{\text{calc}}} + \underbrace{\sum_k p_K(k) D(p_{\omega|k} \parallel u_k)}_{J_{\text{pot}}},$$

où u_k est la loi uniforme sur les $\binom{n}{k}$ micro-états de la classe k . La néguentropie calculée est l'ordre lisible sur le macro-état ; la néguentropie potentielle est l'ordre caché dans le détail de la configuration. C'est la décomposition inter/intra appliquée à l'entropie relative.

Deux descriptions. Le partage dépend de ce que l'on sait du système.

- *Description par la loi.* Partant de toutes les boules dans la première urne, la loi des micro-états sachant $K = k$ reste uniforme par symétrie : $J_{\text{pot}}(t) = 0$ à chaque instant, et toute la néguentropie est calculée, $J = D(p_t \parallel \pi)$, qui décroît vers 0. La néguentropie se dissipe ; elle n'est mise en réserve nulle part, et la loi ne revient jamais à son état initial.
- *Description par la trajectoire.* Un observateur qui suit le micro-état ω_t le connaît exactement : sa loi est concentrée en un point, et sa néguentropie totale reste égale à $n \ln 2$ à chaque instant. Elle se partage alors entre les deux parts.

Proposition 13.10. *Le long d'une trajectoire, si $K_t = k$,*

$$J_{\text{calc}}(t) = -\ln \pi_k = \ln \frac{2^n}{\binom{n}{k}}, \quad J_{\text{pot}}(t) = \ln \binom{n}{k}, \quad J_{\text{calc}}(t) + J_{\text{pot}}(t) = n \ln 2.$$

Démonstration. La loi de ω est la masse de Dirac en ω_t . Sa marginale en K est la masse de Dirac en k , d'entropie relative $-\ln \pi_k$ par rapport à π ; la loi conditionnelle est la masse de Dirac en ω_t dans une classe de $\binom{n}{k}$ éléments, d'entropie relative $\ln \binom{n}{k}$ par rapport à u_k . $\square$

Corollaire 13.11. *Le temps moyen de retour en k s'écrit*

$$\mathbb{E}_k[T_k] = e^{J_{\text{calc}}(k)} = e^{n \ln 2 - J_{\text{pot}}(k)}.$$

Démonstration. C'est le lemme de Kac (théorème 13.9), puisque $1/\pi_k = e^{-\ln \pi_k}$. $\square$

Boules dans l'urne 1	J_{calc} (nats)	J_{pot} (nats)	Retour moyen (pas)
10	1,736	12,127	5,7
12	2,119	11,744	8,3
14	3,298	10,565	27
16	5,377	8,486	216
18	8,616	5,247	5 519
20	13,863	0	1 048 576

Interprétation. Pour l'observateur qui suit la trajectoire, la dynamique ne crée ni ne détruit de néguentropie : elle la fait passer d'une forme à l'autre. À l'équilibre ($k = 10$), presque toute la néguentropie est *potentielle* : l'ordre est caché dans le détail de la configuration, illisible sur le macro-état. La récurrence vers $k = 20$ est la conversion spontanée de toute cette néguentropie potentielle en néguentropie *calculée*, lisible sur le seul nombre de boules. Le corollaire 13.11 en donne le prix : le temps d'attente croît exponentiellement avec la part de néguentropie qu'il faut convertir.

Échelle de temps. À l'équilibre, le nombre moyen de visites de la classe k pendant τ pas vaut $\tau \pi_k = \tau e^{-J_{\text{calc}}(k)}$. Sur une durée τ , les classes typiquement atteintes sont donc celles dont la néguentropie calculée ne dépasse pas $\ln \tau$: la néguentropie potentielle que la dynamique convertit spontanément en néguentropie calculée, relativement à l'échelle de temps τ , est de l'ordre de $\ln \tau$. Elle ne croît que logarithmiquement avec la durée : doubler le temps disponible ne permet de convertir qu'un bit de plus.

Remarque (Le rôle de l'observateur). La différence entre les deux descriptions est l'information que l'on garde sur le micro-état. L'observateur qui ne suit que la loi — qui oublie le détail des boules — voit la néguentropie disparaître ; celui qui suit la trajectoire la voit seulement changer de forme. D'après le principe de Landauer, c'est cet oubli qui a un coût physique : la croissance de l'entropie de la loi mesure l'information que l'on a choisi de ne pas conserver.

14 Probabilités et énoncés

Définition 14.1 (Axiomes de Kolmogorov, 1933). Un *espace probabilisé* est un triplet $(\Omega, \mathcal{A}, \text{Pr})$ où Ω est un ensemble, $\mathcal{A}$ une tribu de parties de Ω (les événements), et $\text{Pr} : \mathcal{A} \rightarrow [0; 1]$ vérifie :

1. $\text{Pr}(A) \geq 0$ pour tout événement A ;
2. $\text{Pr}(\Omega) = 1$;
3. pour toute suite d'événements deux à deux disjoints, $\text{Pr}(\bigcup_i A_i) = \sum_i \text{Pr}(A_i)$.

Le troisième axiome, l'additivité *dénombrable*, est plus fort que l'additivité finie ; il est nécessaire dès que Ω est infini.

Un *énoncé événementiel* est un énoncé auquel on attache une probabilité $\text{Pr}(e)$, logique au sens de Carnap, fréquentielle ou subjective. Sa *surprise* est $-\ln \text{Pr}(e)$: l'information qu'apporte l'énoncé quand il se réalise. Pour une famille d'énoncés exclusifs et exhaustifs, l'entropie

$$H = - \sum_i \text{Pr}(e_i) \ln \text{Pr}(e_i)$$

est la surprise moyenne : c'est l'incertitude qui précède l'énoncé. La logique floue, qui attribue des degrés de vérité, est un cadre différent, qui ne suit pas les axiomes de Kolmogorov.

15 Synthèse

Notion	Forme mathématique
Relation induite, décomposition canonique	Noyau, théorème de factorisation
Complexions « déterministes »	Proposition 8.1
Complexions « aléatoires »	Multinomial, méthode des types
Entropie d'une fonction	Entropies inter- et intra-quotient, $H_{\text{inter}} + H_{\text{intra}} = \ln n$
Entropie quotient croissante	Croissance par raffinement
Dissipation de composition	Règle de chaînage (égalité exacte)
Entropie de mélange	Théorème 11.1, paradoxe de Gibbs
Entropie de chemin	Chaînes de Markov, théorème 13.3
Cyclicité	Conservation par les bijections
Suites de prédicats	Espace de Cantor, mesure de l'information

Conclusion. L'entropie a une place naturelle au niveau des relations et des partitions, avant toute structure algébrique. Elle s'y décompose, comme une variance, en une part inter-quotient — l'incertitude sur la classe, ce qu'une application conserve — et une part intra-quotient — l'incertitude dans la classe, ce qu'elle efface —, dont la somme est fixée par la taille de l'ensemble de départ. Le raffinement, la composition et le mélange obéissent à des lois exactes ; les bijections conservent l'entropie, les applications non injectives la dissipent, et les dynamiques aléatoires font décroître la néguentropie. Ces résultats relient la combinatoire de Boltzmann, la théorie de Shannon et le principe de Landauer, et ouvrent vers l'entropie des systèmes dynamiques.

Références

- [1] L. Boltzmann, « Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung », *Wiener Berichte* 76, 1877.
- [2] C. E. Shannon, « A mathematical theory of communication », *Bell System Technical Journal* 27, 1948.
- [3] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, 1933.
- [4] I. Csiszár, J. Körner, *Information Theory : Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2^e éd., 2011.
- [5] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Wiley, 2^e éd., 2006.
- [6] V. A. Rokhlin, « Lectures on the entropy theory of measure-preserving transformations », *Russian Mathematical Surveys* 22, 1967.
- [7] R. L. Adler, A. G. Konheim, M. H. McAndrew, « Topological entropy », *Transactions of the American Mathematical Society* 114, 1965.
- [8] R. Landauer, « Irreversibility and heat generation in the computing process », *IBM Journal of Research and Development* 5, 1961.
- [9] M. Kac, « On the notion of recurrence in discrete stochastic processes », *Bulletin of the American Mathematical Society* 53, 1947.
- [10] P. Ehrenfest, T. Ehrenfest, « Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem », *Physikalische Zeitschrift* 8, 1907.

- [11] R. Balian, *Du microscopique au macroscopique*, Ellipses, 1982.
- [12] R. Carnap, *Logical Foundations of Probability*, University of Chicago Press, 1950.