

Entropie algébrique

Entropie des relations, des partitions et des applications

Hugues Genvrin

Octobre 2026

La question

Relations de correspondance

Théorème de Shannon

Nombres de complexions

Entropie inter-quotient et intra-quotient

L'entropie des partitions

L'entropie de mélange

Dynamique : itérer une application

Chemins entre états : chaînes de Markov

Probabilités et énoncés

La question

Porter l'entropie aux structures fondamentales

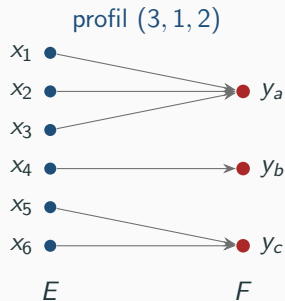
- En physique, l'entropie compte des **micro-états** compatibles avec un macro-état : $S = k_B \ln W$.
- En biologie, la **néguentropie** mesure l'ordre qu'un organisme maintient et consomme.
- Ces deux notions reposent sur un même geste : **regrouper** des éléments en classes, puis **compter**.

Question

Peut-on définir l'entropie dès le niveau le plus élémentaire des mathématiques — relations, relations d'équivalence, applications — avant toute structure algébrique ?

L'idée directrice

- Une application $\alpha : E \rightarrow F$ regroupe les éléments de E selon leur image.
- Les groupes sont ses fibres $\alpha^{-1}(y)$.
- Compter les applications de même profil de fibres, c'est compter des complexions au sens de Boltzmann.
- L'entropie de E se décompose, comme une variance, en une part inter-quotient (entre les classes) et une part intra-quotient (dans les classes).



Relations de correspondance

Définition

Une **relation** 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 $(x, y) \in \Gamma$.

- **Domaine** $\text{Dom } R$: les x qui ont un conséquent ; **image** $\text{Im } R$: les y qui ont un antécédent.
- **Réciproque** : $R^{-1} = (\Gamma^{-1}, F, E)$.
- R est **univoque** (fonctionnelle) si chaque x a au plus un conséquent ;
co-univoque (injective) si chaque y a au plus un antécédent ;
biunivoque si elle est les deux ; **polyvoque** si ni R ni R^{-1} ne sont univoques.
- Une **application** est une relation univoque de domaine E tout entier.

Composition

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

La composée de deux relations est **toujours** définie.

Proposition

Pour toute relation R de E vers F :

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

Idée. $y (R \circ R^{-1}) y'$ signifie qu'un même x a pour conséquents y et y' . L'égalité avec l'identité exige donc $y = y'$: c'est l'univocité.

Relations d'équivalence et noyau

- Une relation d'équivalence $\sim$ sur E découpe E en **classes** : c'est une **partition**. L'ensemble des classes est le quotient $E/\sim$.
- On appelle **canon** le couple $(\sim, E/\sim)$.

Noyau d'une application

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

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

- Toute application induit ainsi une relation d'équivalence sur son ensemble de départ.
- Inversement, toute partition est le noyau de sa surjection canonique.

La décomposition canonique

Théorème

Toute application $\alpha : E \rightarrow F$ s'écrit de manière unique

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

- ν : surjection canonique, $x \mapsto \hat{x}$;
- $\bar{\alpha}$: **bijection** $E/\sim_\alpha \rightarrow \text{Im } \alpha$;
- ι : injection canonique.

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

Vocabulaire

ν est la **complexion** : elle regroupe les éléments en complexes. Choisir un élément dans chaque classe est une **décomplexion** : il y en a $\prod_y \# \alpha^{-1}(y)$.

Toute l'information perdue est dans ν

- $\bar{\alpha}$ est une bijection : elle ne perd rien.
- ι est une injection : elle ne perd rien non plus.
- Seule la surjection ν efface de l'information : elle oublie qui, dans une classe, était l'élément de départ.

Conséquence

L'entropie d'une application ne dépend que de son noyau, c'est-à-dire de la partition de E en fibres. Étudier l'entropie des applications revient à étudier l'entropie des partitions.

Patrons canoniques et prédicats

Patron canonique

Une famille $P = (\sim_1, \dots, \sim_m)$ de canons définit

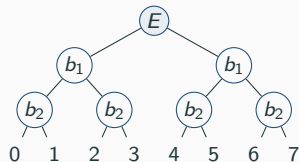
$$\nu_P : x \mapsto (\hat{x}^{(1)}, \dots, \hat{x}^{(m)}).$$

Avec m **prédicats binaires**, $\nu_P : E \rightarrow \{0, 1\}^m$.

Proposition

Pour séparer n éléments, il faut et il suffit de $\lceil \log_2 n \rceil$ prédicats binaires.

Chaque prédicat est **un bit** : $\log_2 n$ est la quantité d'information nécessaire pour identifier un élément.



$n = 8$: trois prédicats suffisent.

Suites infinies de prédicats

- Les suites infinies de bits forment l'espace de Cantor $\{0, 1\}^{\mathbb{N}}$, objet central de la théorie de l'information.
- Il n'est pas en bijection naturelle avec un intervalle réel :

$$0,0111\dots = 0,1000\dots \quad (\text{en base 2}).$$

- Les opérations de corps et la complétude n'y sont pas construites : la construction de $\mathbb{R}$ demande les coupures de Dedekind ou les suites de Cauchy.

À retenir

Les prédicats binaires relèvent de la mesure de l'information, non de la construction des nombres.

Théorème de Shannon

- **Composer** un message en sous-segments abaisse l'information de Hartley d'exactement $H(\text{Pr})$.
- **Décomposer** la collection en ses éléments natifs restitue cette information.
- On en tire une **définition** de l'entropie de Shannon comme *entropie de décomposition* — sans les axiomes de Shannon.

Deux lectures d'un seul fait : $\ln n = H_{\text{inter}} + H_{\text{intra}}$.

Rappel : composition et information de Hartley

Message de N symboles.

- *Non structuré* : N alternatives, Hartley $\mathbb{E}_{NS} = \log_2 N$.
- *Structuré* : n sous-segments de tailles N_i , $\text{Pr}_i = N_i/N$, espérance de Hartley d'un sous-segment $\mathbb{E}_S = \sum_i \text{Pr}_i \log_2 N_i$.

Théorème (Composition et néguentropie ; *Applications*, Thm. 1.4)

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

Idée-clé

Composer retire à l'espérance de Hartley exactement $H(\text{Pr})$ bits : c'est le $\Delta\mathbb{E} = -H$.

Le sens réciproque : décomposer

Partons de la collection composée et **décomposons** jusqu'aux éléments natifs. Deux niveaux :

- localiser le *sous-segment* : $H(\text{Pr})$ bits ;
- localiser l'*élément dans* le sous-segment : $\mathbb{E}_S = \sum_i \text{Pr}_i \log_2 N_i$ bits.

La somme restitue le Hartley du message entier

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

Théorème : Shannon = entropie de décomposition = H_{inter}

Théorème

La décomposition de Hartley est la décomposition inter/intra de Entropie algébrique (Thm. 3.2) :

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

En particulier :

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

L'entropie de Shannon de la partition = l'entropie de **décomposition** de la collection en ses sous-segments : la part de Hartley que composer met en réserve et que décomposer restitue.

Corollaire

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

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

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

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

Idée-clé

On ne contourne pas la valeur H : on la *construit* comme entropie de décomposition, par la voie algébrique, non par la voie axiomatique.

L'adjonction composition $\dashv$ décomposition

La néguentropie packagée par la composition = l'entropie restituée par la décomposition. Au niveau fini : une **identité** ($-\Delta\mathbb{E}_{\text{comp}} = H = H_{\text{inter}}$). Mais pas une involution.

Remarque (Asymétrie sur le continu ; *Applications*, Prop. 1.37)

Composer un nombre fini de symboles : information finie. Décomposer le *continu* : information *infinie* (toute la finesse infinitésimale). L'adjonction est **asymétrique** — à énoncer à résolution fixée.

Remarque (« Éléments natifs » = choix de résolution)

H dépend du niveau natif (atomes ? molécules ?). C'est la dépendance stratale de la tour du grain : « Shannon = entropie de décomposition » est un énoncé *stratifié*.

Trois quantités à ne pas confondre

Entropie de décomposition (l'objet du théorème)

$H(\text{Pr}) = H_{\text{inter}}$: entropie *brute* de la partition. Déficit de Hartley $-\Delta\mathbb{E}$.

Néguentropie de biais (Déf. 1.9)

$\log_2 m - H(p) = D_{\text{KL}}(p \parallel u)$: inégalité des fréquences *d'une* position.

Néguentropie de composition (Déf. 1.12)

$\sum_{\ell} H(X_{\ell}) - H(X_1, \dots, X_L) = \text{TC}$:
dépendance entre positions (corrélacion totale).

Idée-clé

Biais et composition sont *orthogonaux* (Rem. 1.14) ; ni l'un ni l'autre n'est l'entropie de décomposition $H(\text{Pr})$, qui est l'entropie brute de la partition. Attention aussi au déficit absolu (croît au déséquilibre) vs la néguentropie relative $\log_2 n - H$ (vraie mesure d'ordre).

- **Décomposer** en éléments natifs *augmente la finesse* (n classes $\rightarrow N$ éléments) ;
coût = $H(\text{Pr}) = H_{\text{inter}}$.
- **Composer** *package de l'ordre* : néguentropie relative $\log_2 n - H(\text{Pr})$.

L'entropie de Shannon = la **quantité de finesse qu'une décomposition doit créer** pour atteindre le niveau natif.

Idée-clé

On rejoint *Jauges entropiques* : H_{inter} y était la quantification fidèle de la finesse ; elle est ici l'entropie de décomposition. Même objet, deux lectures.

Theorem (Entropie de Shannon)

Shannon = entropie de décomposition.

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

Nombres de complexions

Proposition

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

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

- Dans chaque fibre au-dessus de F' , on choisit une partie **non vide**, indépendamment des autres fibres.
- Fibres de tailles 3, 1, 2 : $N = 7 \times 1 \times 3 = 21$, ce qu'une énumération confirme.
- Une application injective n'admet qu'**une** restriction par image.

Le nombre de complexions d'un profil

Théorème

Le nombre d'applications $E \rightarrow F$ de profil (n_y) , avec $\sum_y n_y = n$, est

$$W(n_y) = \frac{n!}{\prod_y n_y!}.$$

- *Preuve.* On range les n éléments ($n!$ façons), on envoie les n_{y_1} premiers sur y_1 , etc. L'ordre interne aux blocs ne compte pas : on divise par $\prod n_y!$.
- C'est le nombre de complexions de Boltzmann : éléments de E = particules, éléments de F = niveaux d'énergie.
- Forme télescopique, avec $N_i = n_1 + \dots + n_i$: $W = \prod_{i=1}^{k-1} \binom{N_{i+1}}{N_i}$.

Attention

Même pour des fibres de tailles égales, W n'est **pas** le nombre $(\#F)!$ de permutations des images.

Six éléments, trois images, deux éléments par image :

$$W = \frac{6!}{2! 2! 2!} = \frac{720}{8} = \mathbf{90}, \quad \text{et non } 3! = 6.$$

- $3! = 6$ compte les façons de permuter les **étiquettes** des classes.
- 90 compte les façons de répartir les **éléments** dans les classes étiquetées.

Méthode des types : du comptage à Shannon

Théorème

Pour un profil (n_y) sur k valeurs, avec $p_y = n_y/n$ et $H(p) = -\sum_y p_y \ln p_y$:

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

- *Borne supérieure.* Par la formule du multinôme,

$$1 = \left(\sum_y p_y \right)^n \geq W(n_y) \prod_y p_y^{n_y} = W(n_y) e^{-nH(p)}.$$

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

Entropie inter-quotient et intra-quotient

Deux entropies du quotient

On tire X uniformément dans E (n éléments). L'application α répartit E en classes : ses fibres, de tailles n_y .

Entropie inter-quotient

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

Incertitude sur la **classe** de X : ce que la connaissance de la classe **réduit**.

Entropie intra-quotient

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

Incertitude qui **subsiste dans** la classe, une fois celle-ci connue.

H_{inter} : entre les classes

H_{intra} : dans les classes

$\ln n$

Le théorème de décomposition

Théorème

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

- $H_{\text{intra}} = 0$ et $H_{\text{inter}} = \ln n$ si et seulement si α est injective : classes réduites à des singletons ;
- $H_{\text{inter}} = 0$ et $H_{\text{intra}} = \ln n$ si et seulement si α est constante : une seule classe.

Preuve.
$$\sum_y \frac{n_y}{n} \left(\ln \frac{n}{n_y} + \ln n_y \right) = \ln n.$$

- Les deux entropies ne dépendent que du noyau de α , c'est-à-dire de la partition en classes : seule la surjection canonique ν regroupe des éléments.

Exemple : 8 éléments, classes de tailles 4, 2, 1, 1

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	$H_{\text{inter}} = 1,75 \text{ bit}$
Intra : $\log_2 n_y$	2	1	0	0	$H_{\text{intra}} = 1,25 \text{ bit}$

- Identifier un élément de E : $\log_2 8 = 3$ questions oui/non.
- Connaître sa classe en apporte 1,75 en moyenne.
- Il en reste 1,25 à poser à l'intérieur de la classe : deux dans la classe a, aucune dans les classes c et d.

L'analogie avec l'analyse de la variance

Théorème de Huygens

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

	Variance	Entropie
Totale	$\text{Var}(Z)$	$H(X) = \ln n$
Inter	dispersion des moyennes de classe	incertitude sur la classe
Intra	dispersion dans les classes	incertitude dans la classe
Une seule classe	inter nulle	$H_{\text{inter}} = 0$
Singletons	intra nulle	$H_{\text{intra}} = 0$

Pour une loi quelconque : $H(X) = H(\text{classe}) + H(X \mid \text{classe})$, la règle de chaînage de Shannon.

H_{inter} : information transmise

- Ce que $\alpha(X)$ apprend sur X .
- Information mutuelle $I(X; \alpha(X))$.
- En ce sens, ce que α **conserve**.

H_{intra} : équivoque

- Ce qu'il faut **ajouter** à $\alpha(X)$ pour retrouver X .
- L'« équivoque » de Shannon.
- Ce que α **efface**.

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 bilan lève l'ambiguïté

Ce sont deux quantités **différentes et complémentaires** :

- pour une bijection, $H_{\text{intra}} = 0$ et $H_{\text{inter}} = \ln n$ est maximale ;
- en général, $\frac{1}{n} \ln W \approx H_{\text{inter}}(\alpha)$ et $\ln n! - \ln W = \sum_y \ln n_y! \approx n H_{\text{intra}}(\alpha)$.

Le pont avec la physique : Landauer

- $H_{\text{intra}}(\alpha)$ est ce qu'il faut **ajouter** à $\alpha(X)$ pour retrouver X : l'information effacée.
- Le principe de **Landauer** : effacer un bit dans un dispositif à la température T dissipe au moins $k_B T \ln 2$ de chaleur.

Conséquence

Réaliser physiquement α coûte au moins

$$Q_{\min} = k_B T H_{\text{intra}}(\alpha) \quad (\text{par élément traité, } H_{\text{intra}} \text{ en nats}).$$

C'est le lien **exact** entre l'entropie d'une application et l'entropie physique.

Théorème

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

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

Preuve. Avec $Y = \alpha(X)$, $Z = \beta(Y)$, fonctions de X :

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

À retenir

Composer fusionne des classes : H_{intra} ne peut que croître, H_{inter} que décroître. L'information effacée ne se retrouve **jamais**.

L'entropie des partitions

Définition

Pour une partition P de E :

$$H(P) = \sum_{A \in P} \frac{\#A}{n} \ln \frac{n}{\#A}, \quad W(P) = \frac{n!}{\prod_{A \in P} (\#A)!}.$$

- $H(P)$ est l'entropie inter-quotient de la surjection canonique $E \rightarrow E/P$.
- Q est **plus fine** que P ($Q \preceq P$) si chaque classe de Q est contenue dans une classe de P .
- Pour $A \in P$, Q induit une partition Q_A de la classe A .

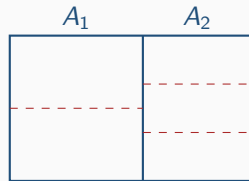
Règle de chaînage : une égalité exacte

Théorème

Si $Q \preceq P$:

$$W(Q) = W(P) \prod_{A \in P} W(Q_A),$$

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



P (traits pleins), Q (tirets)

Preuve. Les factorielles $(\#A)!$ se simplifient ; pour les entropies, $\ln \frac{n}{\#B} = \ln \frac{n}{\#A} + \ln \frac{\#A}{\#B}$.

C'est la règle de chaînage de Shannon, en version **combinatoire** : aucune correction n'est nécessaire.

Corollaire

Si $Q \preceq P$, alors $W(Q) \geq W(P)$ et $H(Q) \geq H(P)$, avec égalité si et seulement si $Q = P$.

- Chaque $W(Q_A)$ est un entier ≥ 1 ; il vaut 1 seulement si Q_A n'a qu'une classe.
- **Scinder une classe** augmente toujours l'entropie quotient.
- **Fusionner des classes** la diminue toujours : c'est la dissipation d'entropie par composition.

Exemple : $n = 4$

$\{1234\} \rightarrow \{12\}\{34\} \rightarrow \{1\}\{2\}\{34\} \rightarrow \{1\}\{2\}\{3\}\{4\}$
 $H = 0 \rightarrow \ln 2 \approx 0,693 \rightarrow 1,040 \rightarrow \ln 4 \approx 1,386.$

Deux partitions : l'information mutuelle

La **partition jointe** $P \vee Q$ a pour classes les intersections non vides $A \cap B$.

Théorème

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

$$I(P; Q) = H(P) + H(Q) - H(P \vee Q) = D((p_{AB}) \parallel (p_A p_B)) \geq 0.$$

- $I = 0$ si et seulement si les partitions sont **indépendantes**.
- I est maximale, égale à $\min(H(P), H(Q))$, si l'une **détermine** l'autre : corrélation totale.
- Un patron canonique qui détermine α donne $H_{\text{inter}}(\alpha) \leq H(\sim_1 \vee \dots \vee \sim_m) \leq \sum_j H(\sim_j)$.

L'entropie de mélange

Réunir deux systèmes

On réunit E_1 et E_2 disjoints (n_1, n_2 éléments), munis de partitions de complexions W_1 et W_2 .

Théorème

1. Classes distinctes : $W_3 = \binom{n_1 + n_2}{n_1} W_1 W_2$.

2. Classes de même étiquette fusionnées (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.

Preuve de 2. $\frac{W_3}{W_1 W_2} = \binom{n_1 + n_2}{n_1} / \prod_i \binom{a_i + b_i}{a_i}$, quotient compris entre 1 et $\binom{n_1 + n_2}{n_1}$.

Le paradoxe de Gibbs, version combinatoire

- Le facteur $\binom{n_1+n_2}{n_1}$ est l'**entropie de mélange** de Gibbs.
- Mélanger des classes **identiques** rapporte moins : la fusion des étiquettes divise par $\prod_i \binom{a_i+b_i}{a_i}$.
- À la limite, mélanger deux échantillons d'un même gaz ne crée rien : c'est le paradoxe de Gibbs.

Pas de borne quadratique

On n'a pas en général $W_3 \geq \max(W_1, W_2)^2$:

E_1 : 10 singletons ($W_1 = 10!$), E_2 : 1 élément ($W_2 = 1$) donnent

$$W_3 = 11 \times 10! \approx 4 \times 10^7 \ll (10!)^2 \approx 1,3 \times 10^{13}.$$

Seule l'inégalité **entropique** $H_3 \geq H_1 + H_2$ est générale.

Dynamique : itérer une application

$\alpha : E \rightarrow E$, X_0 uniforme, $X_{t+1} = \alpha(X_t)$, $H_t = H(X_t)$.

Théorème

1. Si α est une **bijection**, $H_t = \ln n$ pour tout t : l'entropie est conservée.
2. En général, (H_t) **décroît** et devient constante au plus tard au rang n . Sa limite se lit sur l'**image ultime**

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

sur laquelle α induit une bijection.

Idée. X_{t+1} est fonction de X_t , donc $H_{t+1} \leq H_t$. Les images $\alpha^t(E)$ décroissent et se stabilisent ; une surjection d'un ensemble fini sur lui-même est une bijection.

Un exemple : $\alpha = (1, 2, 0, 0, 3, 4, 4, 6)$ sur $\{0, \dots, 7\}$

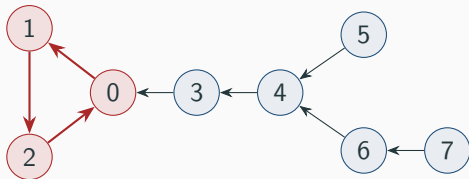
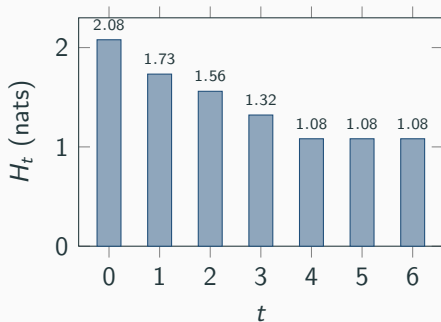


image ultime $\{0, 1, 2\}$



Information effacée au total : $\ln 8 - 1,082 \approx 1,0$ nat. À partir de $t = 4$, toute la masse est sur le cycle, qui ne fait plus que **permuter** les probabilités.

Bijection

- Dynamique **réversible**.
- Conserve exactement l'entropie.
- Analogie du théorème de **Liouville**.

Application non injective

- Dynamique **irréversible**.
- Efface de l'information.
- Coût physique par le principe de **Landauer**.

- Une permutation d'**ordre** r vérifie $\alpha^r = \text{Id}$; r est le ppcm des longueurs de ses cycles.
- La cyclicité **ne crée pas** de néguentropie : une permutation conserve l'entropie.
- Pour les systèmes infinis, l'entropie de **Kolmogorov–Sinaï** mesure l'information produite par unité de temps ; elle est nulle sur un ensemble fini.

Chemins entre états : chaînes de Markov

On étudie le passage d'un état ST_1 — une répartition des éléments en classes — à un état ST_2 , de façon **aléatoire**.

Chaîne de Markov

- Matrice de transition $P = (P_{ij})$, positive, lignes de somme 1.
- Loi au temps t : $p_{t+1} = p_t P$.
- Loi **stationnaire** : $\pi P = \pi$.

L'irréversibilité ne se lit pas sur un événement conjoint

Comme $ST_1 \cap ST_2 \subseteq ST_1$, on a toujours $\Pr(ST_1 \cap ST_2) \leq \Pr(ST_1)$. Un état ST_2 peut succéder à ST_1 si et seulement si $P_{ST_1, ST_2} > 0$.

Lemme

Pour des réels positifs a_i, b_i :

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

- *Preuve.* $f(u) = u \ln u$ est convexe ; on applique Jensen avec les poids $b_i / \sum b_j$.
- C'est la forme « quantitative » de la convexité qui sous-tend toutes les inégalités d'entropie.

Le second principe discret

Théorème (théorème H discret)

Si π est stationnaire,

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

Preuve. Pour chaque j , log-somme avec $a_i = (p_t)_i P_{ij}$, $b_i = \pi_i P_{ij}$:

$$(p_{t+1})_j \ln \frac{(p_{t+1})_j}{\pi_j} \leq \sum_i (p_t)_i P_{ij} \ln \frac{(p_t)_i}{\pi_i} ; \quad \text{on somme sur } j.$$

Corollaire

Si P est **bistochastique**, la loi uniforme est stationnaire et $D(p \parallel \pi) = \ln N - H(p)$: l'entropie $H(p_t)$ **croît** à chaque pas.

La néguentropie comme entropie relative

- La **néguentropie** d'un état est son entropie relative à l'équilibre :

$$\mathcal{J}(p) = D(p \parallel \pi).$$

- Le théorème H dit qu'elle **ne peut que diminuer** sous une dynamique markovienne.
- C'est le théorème H de Boltzmann sous une forme **discrète et exacte**, sans hypothèse de chaos moléculaire.

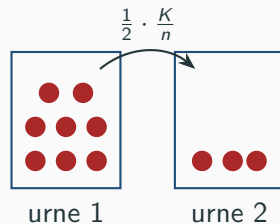
Lien avec $S_{\max} - S$

Quand π est uniforme sur N états,

$$\mathcal{J}(p) = \ln N - H(p) = S_{\max} - S.$$

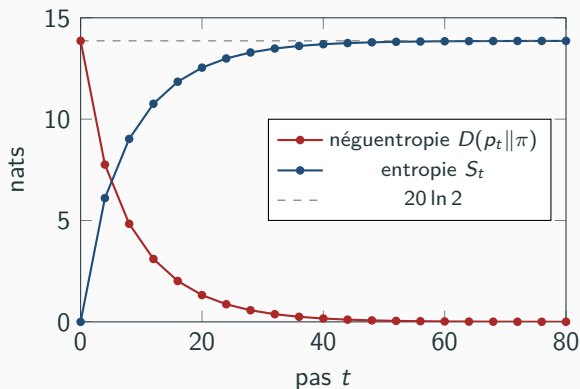
Le modèle d'Ehrenfest

- n boules numérotées, deux urnes.
- À chaque pas : une boule au hasard ; avec probabilité $\frac{1}{2}$, elle change d'urne.
- **Micro-état** : l'urne de chaque boule (2^n états).
- **Macro-état** : le nombre K de boules dans l'urne 1.
- Micro : chaîne bistochastique.
Macro : loi stationnaire $\pi_k = \binom{n}{k} / 2^n$.



Cette chaîne modélise le **transfert d'éléments entre deux classes**.

Ehrenfest, $n = 20$: la néguentropie s'écoule



t	$\mathbb{E}[K]$	D	S_t
0	20,0	13,863	0,000
1	19,5	11,672	2,191
5	17,7	6,864	6,999
10	16,0	3,866	9,997
20	13,6	1,321	12,542
40	11,3	0,166	13,697
80	10,2	0,003	13,860

Départ : toutes les boules dans l'urne 1.

Une loi de conservation

Proposition

Pour le modèle d'Ehrenfest, à tout instant,

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

Preuve. Sachant $K = k$, les $\binom{n}{k}$ micro-états sont équiprobables, donc

$$S_t = H(K) + \mathbb{E} \left[\ln \binom{n}{K} \right], \quad D(p_t \parallel \pi) = n \ln 2 - H(K) - \mathbb{E} \left[\ln \binom{n}{K} \right].$$

Interprétation

C'est exactement $\mathcal{J} = S_{\max} - S$: la négentropie est l'écart entre l'entropie maximale et l'entropie du système. Ce que l'une perd, l'autre le gagne.

Réversible, et pourtant irréversible

Bilan détaillé

$$\pi_k P_{k,k-1} = \pi_{k-1} P_{k-1,k}$$

La chaîne est **réversible** : à l'équilibre, le film passé à l'envers a les mêmes statistiques.

Lemme de Kac

Temps moyen de retour en k :

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

Pas de contradiction

Le théorème H porte sur la **loi** p_t ; une **trajectoire** fluctue et revient. La flèche du temps vient de la **condition initiale**, pas de la dynamique.

	retour à $k = n$	durée (1 pas/s)
$n = 10$	1 024	17 min
$n = 20$	$\approx 10^6$	12 jours
$n = 100$	$\approx 10^{30}$	4×10^{22} ans

La récurrence comme conversion de néguentropie potentielle

Micro-état $\omega = (K, \text{quelles boules})$. Règle de chaînage :

$$D(p \parallel u) = \underbrace{D(p_K \parallel \pi)}_{J_{\text{calc}} : \text{lisible sur } K} + \underbrace{\sum_k p_K(k) D(p_{\omega|k} \parallel u_k)}_{J_{\text{pot}} : \text{caché dans le détail}}$$

Description par la loi

Par symétrie, $J_{\text{pot}} = 0$; $J = J_{\text{calc}}$ décroît vers 0. La néguentropie se **dissipe**.

Description par la trajectoire

On connaît ω_t : $J = n \ln 2$ **constant**.
 $J_{\text{calc}} = -\ln \pi_{K_t}$, $J_{\text{pot}} = \ln \binom{n}{K_t}$.

Temps de retour

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

Ehrenfest, $n = 20$: de la réserve au visible

k	J_{calc}	J_{pot}	retour
10	1,74	12,13	5,7
12	2,12	11,74	8,3
14	3,30	10,57	27
16	5,38	8,49	216
18	8,62	5,25	5 519
20	13,86	0	1 048 576

Néguentropies en nats, retour en pas.

- À l'équilibre, l'ordre est presque tout **potentiel** : caché dans *quelles* boules.
- La récurrence le **convertit** entièrement en néguentropie **calculée**, lisible sur K .
- Le prix est le **temps** : $e^{J_{\text{calc}}}$ pas.
- Sur une durée τ , on ne convertit spontanément que $\approx \ln \tau$: **un bit** de plus à chaque doublement du temps.

Probabilités et énoncés

Axiomes de Kolmogorov (1933)

$(\Omega, \mathcal{A}, \text{Pr})$ avec $\text{Pr}(A) \geq 0$, $\text{Pr}(\Omega) = 1$ et, pour des événements deux à deux disjoints, $\text{Pr}(\bigcup_i A_i) = \sum_i \text{Pr}(A_i)$.

- L'additivité **dénombrable** est nécessaire dès que Ω est infini.
- Un **énoncé événementiel** e porte une probabilité $\text{Pr}(e)$: logique (Carnap), fréquentielle ou subjective.
- Sa **surprise** est $-\ln \text{Pr}(e)$: l'information qu'il apporte en se réalisant.
- Pour des énoncés exclusifs et exhaustifs, l'entropie $H = -\sum_i \text{Pr}(e_i) \ln \text{Pr}(e_i)$ est la **surprise moyenne**.
- La logique floue, à degrés de vérité, est un cadre **différent**.

Synthèse

Notion	Forme mathématique
Relation induite, décomposition canonique	Noyau, théorème de factorisation
Complexions « déterministes »	$N = \prod_y (2^{n_y} - 1)$
Complexions « aléatoires »	Multinomial, méthode des types
Entropie d'une fonction	$H_{\text{inter}} + H_{\text{intra}} = \ln n$ (inter et intra)
Entropie quotient croissante	Croissance par raffinement
Dissipation de composition	Règle de chaînage (égalité exacte)
Entropie de mélange	$W_3 = \binom{n_1+n_2}{n_1} W_1 W_2$, paradoxe de Gibbs
Entropie de chemin	Chaînes de Markov, théorème H discret
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 et une part intra-quotient, de somme $\ln n$.
- Raffinement, composition et mélange obéissent à des lois exactes.
- Les bijections conservent l'entropie ; les applications non injectives la dissipent ; les dynamiques aléatoires font décroître la néguentropie.

Perspectives

Ces résultats relient la combinatoire de Boltzmann, la théorie de Shannon et le principe de Landauer, et ouvrent vers l'entropie des systèmes dynamiques (Kolmogorov–Sinaï, entropie topologique).

Références

- L. Boltzmann, « Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung », *Wiener Berichte* 76, 1877.
- C. E. Shannon, « A mathematical theory of communication », *Bell System Technical Journal* 27, 1948.
- A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, 1933.
- I. Csiszár, J. Körner, *Information Theory : Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2^e éd., 2011.
- T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Wiley, 2^e éd., 2006.
- V. A. Rokhlin, « Lectures on the entropy theory of measure-preserving transformations », *Russian Math. Surveys* 22, 1967.
- R. L. Adler, A. G. Konheim, M. H. McAndrew, « Topological entropy », *Trans. AMS* 114, 1965.
- R. Landauer, « Irreversibility and heat generation in the computing process », *IBM J. Res. Dev.* 5, 1961.
- P. Ehrenfest, T. Ehrenfest, « Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem », *Physikalische Zeitschrift* 8, 1907.
- R. Balian, *Du microscopique au macroscopique*, Ellipses, 1982.
- R. Carnap, *Logical Foundations of Probability*, University of Chicago Press, 1950.