

Que l'entropie de Shannon est une quantité d'information

Notions d'entropie inter et intra-quotient

Hugues Genvrin

octobre 2026

Résumé

On dérive l'entropie de Shannon à partir de la seule information de Hartley et de l'opération de structuration (segmentation en classes), sans recourir à ses axiomes. On montre qu'elle est une *quantité d'information* — une différence d'espérances d'information de Hartley — et qu'elle s'identifie à l'entropie *inter-quotient*, l'incertitude sur la classe d'un élément. Le complément à l'information totale fait alors apparaître, en conclusion, l'entropie *intra-quotient*.

1 Cadre et information de Hartley

Soit un alphabet partitionné en k classes (types) $C_1, \dots, C_k$. La classe C_j compte n_j éléments distinguables, et l'on pose

$$n = \sum_{j=1}^k n_j.$$

On munit l'ensemble des n éléments de la loi uniforme : un élément X est tiré au hasard, et la probabilité qu'il appartienne à la classe C_j est

$$\Pr_j := \Pr(X \in C_j) = \frac{n_j}{n}.$$

Définition 1 (Information de Hartley). *L'information de Hartley d'un ensemble à m alternatives équiprobables est $\log_2 m$: le nombre de bits nécessaires pour en désigner une. Désigner un élément parmi les n de l'alphabet coûte donc $\log_2 n$.*

Remarque 1 (Hypothèse d'équiprobabilité intra-classe). *On suppose les éléments équiprobables à l'intérieur de chaque classe (conséquence de la loi uniforme sur l'ensemble). Dès lors, désigner un élément à l'intérieur de la classe C_j , une fois celle-ci connue, coûte l'information de Hartley $\log_2 n_j$.*

2 L'identité élément par élément

Tout repose sur une décomposition locale de l'information, valable pour un élément quelconque de la classe C_j .

Lemme 1 (Décomposition de l'information d'un élément). *Pour un élément de la classe C_j ,*

$$\underbrace{\log_2 n}_{\text{désigner l'élément}} = \underbrace{\log_2 \frac{n}{n_j}}_{\text{désigner sa classe}} + \underbrace{\log_2 n_j}_{\text{le désigner dans sa classe}}.$$

Démonstration. $\log_2 \frac{n}{n_j} + \log_2 n_j = \log_2 \left(\frac{n}{n_j} \cdot n_j \right) = \log_2 n.$ □

Les deux termes ont un sens distinct, qu'il importe de ne pas confondre. Le premier, $\log_2 \frac{n}{n_j} = -\log_2 \Pr_j$, est l'*information propre* (la surprise) de l'événement « l'élément est de classe C_j ». Le second, $\log_2 n_j$, est l'information de Hartley *résiduelle* : ce qu'il reste à fournir pour localiser l'élément une fois sa classe connue. Ce n'est pas l'information de la classe, mais l'information *intra-classe*.

3 Espérances d'information

On prend l'espérance de ces quantités sur le tirage de X , c'est-à-dire la somme pondérée par les $\Pr_j$. Chaque espérance est une somme ; le produit $\Pr_j \log_2(\cdot)$ en est le terme j , non l'espérance elle-même.

Définition 2 (Information segmentée et non segmentée).

$$E_{NS} = \sum_{j=1}^k \Pr_j \log_2 n = \log_2 n \quad (\text{message non segmenté : information totale}),$$

$$E_S = \sum_{j=1}^k \Pr_j \log_2 n_j \quad (\text{message segmenté : espérance de l'information intra-classe}).$$

La première égalité tient car $\sum_j \Pr_j = 1$: $E_{NS} = \log_2 n \cdot \sum_j \Pr_j = \log_2 n$.

Théorème 1 (Variation de l'espérance d'information).

$$\Delta E = E_S - E_{NS} = \sum_{j=1}^k \Pr_j \log_2 \frac{n_j}{n} = \sum_{j=1}^k \Pr_j \log_2 \Pr_j.$$

Démonstration. $E_S - E_{NS} = \sum_j \Pr_j \log_2 n_j - \sum_j \Pr_j \log_2 n = \sum_j \Pr_j \log_2 \frac{n_j}{n} = \sum_j \Pr_j \log_2 \Pr_j$, puisque $\Pr_j = n_j/n$. Tout reste en base 2. □

4 L'entropie de Shannon est une quantité d'information

Structurer (segmenter) abaisse l'espérance de Hartley. Lisons la variation réciproque — celle qui correspond à la *décomposition* d'un message segmenté en message non segmenté.

Définition 3 (Entropie de Shannon). *La variation d'espérance d'information liée à la décomposition est*

$$H_S := -\Delta E = E_{NS} - E_S = -\sum_{j=1}^k \Pr_j \log_2 \Pr_j.$$

Proposition 1 (H_S est une quantité d'information positive). $H_S \geq 0$, et H_S est une différence de deux espérances d'information de Hartley — donc elle-même une quantité d'information, exprimée en bits.

Démonstration. $H_S = E_{NS} - E_S$ est une différence d'espérances d'information de Hartley. Sa positivité vient de $n_j \leq n$ pour tout j , d'où $E_S = \sum_j \Pr_j \log_2 n_j \leq \sum_j \Pr_j \log_2 n = E_{NS}$. □

Théorème 2 (H_S est l'entropie inter-quotient). *Comme $\Pr_j = n_j/n$ est exactement la distribution d'appartenance aux classes, $H_S = -\sum_j \Pr_j \log_2 \Pr_j$ est l'entropie de la partition en classes : l'incertitude sur la classe d'un élément. C'est l'entropie inter-quotient.*

Démonstration. Par le lemme, l'information « désigner sa classe » d'un élément de C_j est $-\log_2 \Pr_j$. Son espérance sur X est $-\sum_j \Pr_j \log_2 \Pr_j = H_S$: c'est, par définition, l'incertitude moyenne sur la classe, soit l'entropie inter-quotient. $\square$

Remarque 2 (Lecture du côté du récepteur). *Un récepteur reçoit un message segmenté qu'il doit décomposer. Il fait face à l'incertitude sur la classe de chaque symbole reçu — l'hétérogénéité des classes. Cette incertitude est H_S : l'entropie de Shannon est l'information inter-quotient que la décomposition doit lever.*

5 Conclusion : l'entropie intra-quotient

L'information totale d'un message est $E_{NS} = \log_2 n$. Elle se scinde, par le lemme moyenné sur X , en l'entropie inter-quotient H_S (désigner la classe) et le reste (désigner l'élément dans sa classe) :

$$E_{NS} = H_S + E_S.$$

Le complément E_S ne peut donc être autre chose que l'entropie *intra-quotient* — et sa forme directe le confirme :

$$E_S = \sum_{j=1}^k \Pr_j \log_2 n_j = \mathbb{E}[\log_2 n_j] = H(X \mid \text{classe})$$

l'espérance de l'information de Hartley résiduelle, c'est-à-dire l'incertitude qui subsiste sur un élément une fois sa classe connue. On aboutit ainsi à la décomposition de l'information totale en ses deux parts :

$$\underbrace{\log_2 n}_{\text{information totale } E_{NS}} = \underbrace{H_S}_{\text{inter-quotient : quelle classe}} + \underbrace{E_S}_{\text{intra-quotient : quel élément dans la classe}}.$$

Remarque 3. *L'entropie de Shannon n'est donc pas une notion primitive, mais une quantité d'information dérivée : la part inter-quotient de l'information de Hartley totale. L'entropie intra-quotient en est le complément exact. Les deux sont des informations, en bits, et leur somme est l'information de Hartley du message entier.*