

## ENTROPIE D'UNE DISTRIBUTION DE PROBABILITÉS

ENS PARIS/CACHAN MP 2003

Ce problème est à prendre « par le bon bout des calculs » sous risque de s'empêtrer, comme chaque fois qu'il s'agit de manipuler des inégalités de convexité.

Puisqu'on en aura besoin tout du long du problème, rappelons l'allure de la fonction  $h : x \mapsto -x \ln x$  sur  $[0; 1]$  :

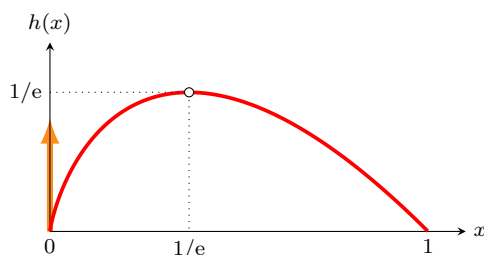


Fig. 1 — La fonction  $h : x \mapsto -x \ln x$ , et sa jolie tangente verticale à l'origine.

## PREMIÈRE PARTIE

1) Soient  $x, y \in I$ . On écrit la formule de Taylor avec reste intégral à l'ordre 2 entre  $x$  et  $y$  :

$$\varphi(y) = \varphi(x) + (y-x)\varphi'(x) + \int_x^y (y-t)\varphi''(t) dt$$

Si  $x \leq y$ , alors  $y-t \geq 0$  dans l'intégrale et l'intégrale est positive. Si  $y \leq x$ , alors  $y-t \leq 0$  et l'intégrale est toujours positive. Dans tous les cas, on a

$$\forall x, y \in I \quad \varphi(y) \geq \varphi(x) + (y-x)\varphi'(x).$$

**Remarque 1** L'interprétation géométrique de cette inégalité est bien connue : une fonction convexe de classe  $\mathcal{C}^1$  est toujours au-dessus de ses tangentes. La démonstration est plus simple que celle du cours car on a l'hypothèse de la classe  $\mathcal{C}^2$ . En réalité, si  $x$  est à l'intérieur de  $I$ , on peut montrer qu'une fonction convexe admet toujours des tangentes à gauche et à droite, et que ces tangentes sont toujours sous la courbe.

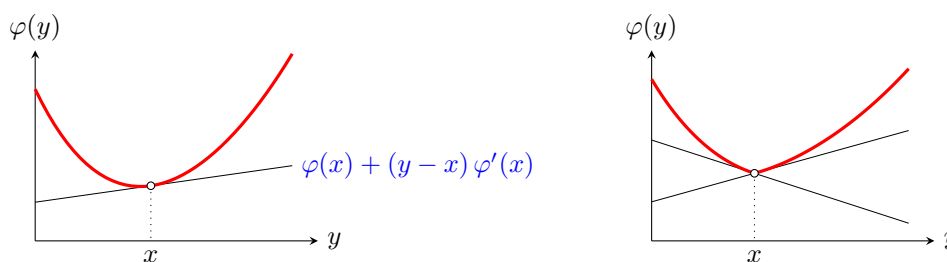


Fig. 2 — À gauche : cas d'une fonction de classe  $\mathcal{C}^2$ . À droite : cas général.

2) On pose  $x = \sum_{i=1}^n a_i x_i$ . En prenant  $y = x_k$  dans la question précédente, on a

$$\varphi(x_k) \geq \varphi\left(\sum_{i=1}^n a_i x_i\right) + \left(x_k - \sum_{i=1}^n a_i x_i\right) \varphi'\left(\sum_{i=1}^n a_i x_i\right) \quad (\text{E}_k)$$

On somme l'équation (E<sub>k</sub>) multipliée par  $a_k$  pour  $k = 1, \dots, n$ , ce qui donne

$$\sum_{k=1}^n a_k \varphi(x_k) \geq \underbrace{\sum_{k=1}^n a_k}_{=1} \varphi\left(\sum_{i=1}^n a_i x_i\right) + \underbrace{\left(\sum_{k=1}^n a_k x_k - \sum_{i=1}^n a_i x_i\right)}_{=0} \varphi'\left(\sum_{i=1}^n a_i x_i\right)$$

et donc l'**inégalité de Jensen** (discrète)

$$\varphi \left( \sum_{i=1}^n a_i x_i \right) \leq \sum_{i=1}^n a_i \varphi(x_i).$$

**3)** La fonction  $g$  est une fonction de poids et joue le même rôle que la famille  $a_1, a_2, \dots, a_n$  de la question précédente, sauf que cette fois-ci on a une somme « continue » et non discrète.

On propose deux méthodes de résolution.

**Première méthode**

À l'image de la question précédente, on applique le résultat de la question **1** en posant  $x = \int_a^b g(t) f(t) dt$  (c'est la valeur moyenne de  $g$  pour la fonction de pondération  $f$ ) et  $y = g(u)$  :

$$\varphi(y) \geq \varphi \left( \int_a^b f(t) g(t) dt \right) + (y - x) \varphi'(x),$$

puis on multiplie par  $f(u)$  et on intègre :

$$\int_a^b \varphi(g(u)) f(u) du \geq \varphi(x) \underbrace{\int_a^b f(u) du}_{=1} + \varphi'(x) \underbrace{\int_a^b [g(u) - x] f(u) du}_{=0} = \varphi(x)$$

ce qui est la formule voulue.

**Deuxième méthode**

On choisit une suite de subdivisions  $(\sigma^{(n)})_{n \in \mathbb{N}}$  régulières. À  $n$  fixé, on note  $\sigma_n = (\xi_0, \xi_1, \dots, \xi_n)$ , on pose

$$\lambda_i^{(n)} = \frac{f(\xi_i)}{\Lambda} \quad \text{où} \quad \Lambda = \sum_{k=0}^n f(\xi_k),$$

ce qui montre que

$$\sum_{k=0}^{n-1} \lambda_i^{(n)} = 1.$$

Alors

$$\frac{b-a}{n} \cdot \frac{\sum_{k=0}^n g(\xi_k) f(\xi_k)}{\sum_{k=0}^n f(\xi_k)} = \frac{1}{b-a} \sum_{k=0}^n \lambda_k g(\xi_k) \xrightarrow{n \rightarrow \infty} \int_a^b g(x) f(x) dx.$$

On compose ensuite par  $\varphi$  et on conclut par continuité.

$$\varphi \left( \int_a^b g(x) f(x) dx \right) \leq \int_a^b \varphi(g(x)) f(x) dx.$$

**DEUXIÈME PARTIE**

Cette partie est consacrée à l'étude de l'entropie de Shannon d'une distribution de probabilité  $(p_1, \dots, p_n)$ .

**4)** On commence par noter que, pour tout  $x \in ]0; 1]$ ,  $-x \ln x \geq 0$  donc  $H_k(\mathbf{p}) \geq 0$ . De plus, l'inégalité de concavité de la fonction  $\ln$  permet d'écrire

$$H_k(\mathbf{p}) = - \sum_{i=1}^k p_i \ln p_i = \sum_{i=1}^k \pi \ln \left( \frac{1}{p_i} \right) \leq \ln \left( \sum_{i=1}^k \frac{p_i}{p_i} \right) = \ln k.$$

Si  $\mathbf{p} \in \mathcal{A}_n$ , on note que si  $p_k = 0$  alors  $p_k \ln p_k = 0$  ce qui donne immédiatement l'inégalité générale :

$$\forall n \geq 2 \quad \forall \mathbf{p} \in \mathcal{A}_n \quad 0 \leq H_n(\mathbf{p}) \leq \ln n.$$

5) La vérification de la convexité de  $\mathcal{A}_n$  est immédiate. La fonction  $\varphi : x \mapsto x \ln x$  est convexe (sa dérivée seconde est  $x \mapsto \frac{1}{x} \geq 0$ ). On a alors

$$H_n(\lambda \mathbf{p} + (1 - \lambda) \mathbf{q}) = - \sum_{i=1}^n \varphi(\lambda p_i + (1 - \lambda) q_i) \geq -\lambda \sum_{i=1}^n \varphi(p_i) + (1 - \lambda) \sum_{i=1}^n \varphi(q_i)$$

donc

$$H_n(\lambda \mathbf{p} + (1 - \lambda) \mathbf{q}) \geq \lambda H_n(\mathbf{p}) + (1 - \lambda) H_n(\mathbf{q}).$$

6.a) Trivialement,  $\mathbf{p} \otimes \mathbf{q} \in \mathcal{A}_{nm}$  (c'est un produit tensoriel, comme le produit de Kronecker de matrices déjà vu en classe). On a alors

$$\begin{aligned} H_{nm}(\mathbf{p} \otimes \mathbf{q}) &= \sum_{i=1}^n \sum_{j=1}^m p_i q_j \ln(p_i q_j) = \sum_{i=1}^n \sum_{j=1}^m p_i q_j (\ln p_i + \ln q_j) \\ &= \sum_{i=1}^n p_i \ln p_i \underbrace{\sum_{j=1}^m q_j}_{=1} + \sum_{j=1}^m q_j \ln q_j \underbrace{\sum_{i=1}^n p_i}_{=1} \end{aligned}$$

donc

$$H_{nm}(\mathbf{p} \otimes \mathbf{q}) = H_n(\mathbf{p}) + H_m(\mathbf{q}).$$

**Remarque 2 (Extensivité en physique)** C'est ce qu'on appelle, en thermodynamique, l'**extensivité** ou l'**additivité** de l'entropie, sachant que l'espace des phases de la réunion de deux systèmes *sans interactions* (indépendants) peut être vue comme le produit tensoriel des espaces des phases de chacun des systèmes.

C'est une notion fondamentale en thermodynamique qui (note historique), a permis l'introduction de la notion d'indiscernabilité des particules classiques, notion dont on a compris par la suite qu'elle était fondamentale dans le traitement quantique des particules.

6.b)

$$H_n(p_1, \dots, p_n) = \sum_{i=1}^n p_i \ln p_i$$

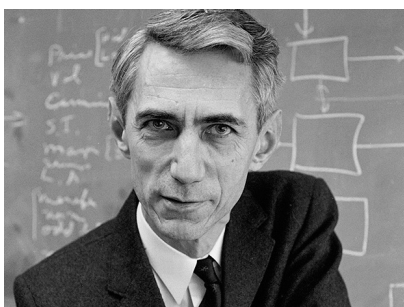
$$\text{et } H_{n-1}(p_1 + p_2, p_3, \dots, p_n) = (p_1 + p_2) \ln(p_1 + p_2) + \sum_{i=3}^n p_i \ln p_i,$$

ce qui, en retranchant la seconde égalité de la première, donne

$$\begin{aligned} H_n(p_1, \dots, p_n) - H_{n-1}(p_1 + p_2, p_3, \dots, p_n) &= p_1 \ln \left( \frac{p_1}{p_1 + p_2} \right) + p_2 \ln \left( \frac{p_2}{p_1 + p_2} \right) \\ &= (p_1 + p_2) \left[ \frac{p_1}{p_1 + p_2} \ln \left( \frac{p_1}{p_1 + p_2} \right) + \frac{p_2}{p_1 + p_2} \ln \left( \frac{p_2}{p_1 + p_2} \right) \right] \end{aligned}$$

c'est-à-dire, puisque  $\frac{p_1}{p_1 + p_2} + \frac{p_2}{p_1 + p_2} = 1$  (propriété indispensable pour pouvoir conclure) :

$$H_n(p_1, \dots, p_n) = H_{n-1}(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2) H_2 \left( \frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2} \right).$$



Le mathématicien, cryptographe, inventeur américain Claude Shannon (1916–2001) est considéré comme le père de la théorie de l'information. S'intéressant à la communication entre machines, il dégage les principes fondamentaux de toute communication entre machine, définit la notion d'entropie d'un signal caractérisé par une distributions probabiliste des lettres de l'alphabet :  $H = - \sum p_i \ln_2 p_i$  et montre l'équivalence avec l'entropie de Boltzmann, et découvre le critère de Shannon sur l'échantillonnage d'un signal : un signal doit être échantillonné à une période au plus égale à la moitié de celle correspondant à la fréquence maximale intervenant dans le signal que l'on veut reconstituer (un CD est échantillonné à 44 100 Hz, soit un peu plus du double de la fréquence maximale entendue par une oreille humaine, qui est  $\approx 20\,000$  Hz).

7)

**7.a)** Si  $\mathcal{I}$  était vide, on aurait  $p_i < q_i$  pour tout  $i \in \llbracket 1; n \rrbracket$ , ce qui met une contradiction avec  $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$ . De même si  $\mathcal{I} = \llbracket 1; n \rrbracket$ . On en déduit que  $\mathcal{I}$  est non vide et de cardinal strictement inférieur à  $n$ .

Soit  $\mathcal{J}$  un sous-ensemble non vide de  $\{1, \dots, n\}$ . Notons  $\alpha := \sum_{i \in \mathcal{J}} p_i$ , alors  $\sum_{i \in \mathcal{J}} \frac{p_i}{\alpha} = 1$ , ce qui permet d'utiliser une inégalité de convexité :

$$\sum_{i \in \mathcal{J}} p_i \ln \frac{q_i}{p_i} = \alpha \sum_{i \in \mathcal{J}} \frac{p_i}{\alpha} \ln \frac{q_i}{p_i} \leq \alpha \ln \left( \sum_{i \in \mathcal{J}} \frac{q_i}{\alpha} \right),$$

donc, en remplaçant  $\alpha$  par sa valeur :

$$\sum_{i \in \mathcal{J}} p_i \ln \frac{q_i}{p_i} \leq \left( \sum_{i \in \mathcal{J}} p_i \right) \ln \left( \frac{\sum_{i \in \mathcal{J}} q_i}{\sum_{i \in \mathcal{J}} p_i} \right).$$

On prend ensuite l'opposé :

$$\boxed{\sum_{i \in \mathcal{J}} p_i \ln \frac{p_i}{q_i} \geq \left( \sum_{i \in \mathcal{J}} p_i \right) \ln \left( \frac{\sum_{i \in \mathcal{J}} p_i}{\sum_{i \in \mathcal{J}} q_i} \right)}.$$

On en déduit, en prenant successivement  $\mathcal{J} = \mathcal{I}$  puis  $\mathcal{J} = \mathcal{I}^c$  :

$$\begin{aligned} D_n(\mathbf{p}, \mathbf{q}) &= \sum_{i \in \mathcal{I}} p_i \ln \left( \frac{p_i}{q_i} \right) + \sum_{i \notin \mathcal{I}} p_i \ln \left( \frac{p_i}{q_i} \right) \\ &\geq \tilde{p} \ln \left( \frac{\tilde{p}}{\tilde{q}} \right) + (1 - \tilde{p}) \ln \left( \frac{1 - \tilde{p}}{1 - \tilde{q}} \right). \end{aligned}$$

$$\boxed{D_n(\mathbf{p}, \mathbf{q}) \geq \tilde{p} \ln \left( \frac{\tilde{p}}{\tilde{q}} \right) + (1 - \tilde{p}) \ln \left( \frac{1 - \tilde{p}}{1 - \tilde{q}} \right)}.$$

**7.b)** On part de

$$0 = 1 - 1 = \sum_{i \in \mathcal{I}} (p_i - q_i) + \sum_{i \in \mathcal{I}^c} (p_i - q_i)$$

pour obtenir

$$\sum_{i=1}^n |p_i - q_i| = \sum_{i \in \mathcal{I}} \underbrace{(p_i - q_i)}_{\geq 0} - \sum_{i \in \mathcal{I}^c} \underbrace{(p_i - q_i)}_{< 0} = 2 \sum_{i \in \mathcal{I}} (p_i - q_i) = 2(\tilde{p} - \tilde{q}).$$

$$\boxed{\sum_{i=1}^n |p_i - q_i| = 2(\tilde{p} - \tilde{q})}.$$

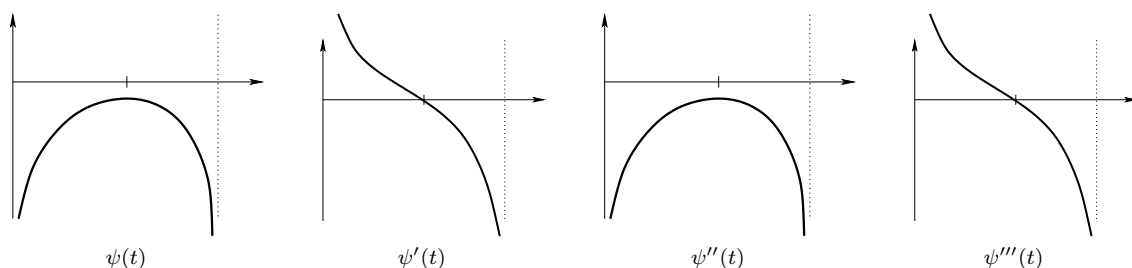
**7.c)** Posons

$$\begin{aligned} \psi: ]0; 1[ &\longrightarrow \mathbf{R} \\ t &\longmapsto \tilde{p} \ln t + (1 - \tilde{p}) \ln(1 - t). \end{aligned}$$

C'est une fonction concave (comme somme de deux fonctions concaves), de dérivées successives

$$\psi'(t) = \frac{\tilde{p} - t}{t(1 - t)} \qquad \psi''(t) = -\frac{\tilde{p} - 2t\tilde{p} + t^2}{t^2(1 - t)^2} \qquad \psi'''(t) = \frac{\tilde{p}^3 - t^3}{t^3(1 - t)^3}.$$

On vérifie que le numérateur de  $\psi''$  ne s'annule jamais. Les graphes de ces fonctions ont donc les allures suivantes, où le maximum de  $\psi$  et  $\psi''$  et le zéro de  $\psi'$  et de  $\psi'''$  sont tous en  $t = \tilde{p}$  :



On utilise alors l'égalité de Taylor avec reste intégral à l'ordre 3 comme dans la question 1 pour dire que

$$\psi(\tilde{q}) - \psi(\tilde{p}) = (\tilde{q} - \tilde{p}) \underbrace{\psi'(\tilde{p})}_{=0} + \frac{1}{2}(\tilde{q} - \tilde{p})^2 \psi''(\tilde{p}) + \int_{\tilde{p}}^{\tilde{q}} \frac{(\tilde{p} - t)^2}{2} \psi'''(t) dt \leq \frac{1}{2}(\tilde{q} - \tilde{p})^2 \psi''(\tilde{p})$$

puisque si  $\tilde{q} \geq \tilde{p}$ , alors  $\psi(t) \geq 0$  et si c'est le contraire, alors  $\psi(t) \leq 0$ .

En prenant l'opposé, on obtient, en notant que  $\psi''(\tilde{p}) = -\frac{1}{\tilde{p}(1-\tilde{p})^2}$ ,

$$D_n(\mathbf{p}, \mathbf{q}) \geq \psi(\tilde{p}) - \psi(\tilde{q}) \geq \frac{1}{2}(\tilde{q} - \tilde{p})^2 \frac{1}{\tilde{p}(1-\tilde{p})^2} = \frac{1}{2}(\tilde{q} - \tilde{p})^2 \underbrace{\frac{1}{\tilde{p}(1-\tilde{p})}}_{\geq \frac{1}{4}} \underbrace{\frac{1}{(1-\tilde{p})}}_{\geq 1},$$

puisque  $x \mapsto x(1-x)$  admet sur  $[0; 1]$  un maximum valant  $1/4$ , soit

$$D_n(\mathbf{p}, \mathbf{q}) \geq 2(\tilde{q} - \tilde{p})^2.$$

En utilisant le résultat de la question 7.c, on conclut donc

$$D_n(\mathbf{p}, \mathbf{q}) \geq \left( \sum_{i=1}^n |p_i - q_i| \right)^2.$$

**8) Une caractérisation de l'entropie.**

**8.a)** Soit  $n \in \mathbf{N}$ . On a alors

$$\psi(n+1) = J_{n+1} \left( \frac{1}{n+1}, \dots, \frac{1}{n+1} \right) \underset{(P3)}{\geq} J_{n+1} \left( \frac{1}{n}, \dots, \frac{1}{n}, 0 \right) \underset{(P4)}{=} J_n \left( \frac{1}{n}, \dots, \frac{1}{n} \right) = \psi(n).$$

De plus, en posant  $\mathbf{p} = \left( \frac{1}{n}, \dots, \frac{1}{n} \right) \in \mathcal{A}_n$  et  $\mathbf{q} = \left( \frac{1}{m}, \dots, \frac{1}{m} \right) \in \mathcal{A}_m$ , on a

$$\psi(nm) = J_{nm}(\mathbf{p} \otimes \mathbf{q}) \underset{(P5)}{=} J_n(\mathbf{p}) + J_m(\mathbf{q}) = \psi(n) + \psi(m),$$

ce qui montre que

$$\psi \text{ est croissante et additive.}$$

**8.b)** • On suppose  $c = 0$ . Alors  $\psi(2) = 0$  donc  $J_2 \left( \frac{1}{2}, \frac{1}{2} \right) = 0$  et donc, par maximalité,  $J_2$  est identiquement nulle. Pour tout  $(\alpha, \beta, \gamma) \in \mathcal{A}_3$ , on a alors

$$J_3(\alpha, \beta, \gamma) \underset{(P6)}{=} J_2(\alpha + \beta, \gamma) + (\alpha + \beta) J_2 \left( \frac{\alpha}{\alpha + \beta}, \frac{\beta}{\alpha + \beta} \right) = 0,$$

ce qui montre que  $J_3$  est identiquement nulle. Une simple récurrence montre que  $J_n$  est nulle pour tout  $n \in \mathbf{N}$  et donc  $\psi$  est la fonction nulle.

• On suppose désormais  $c \neq 0$ . Soient  $n \geq 2$  et  $r \geq 1$  deux entiers. Il existe un entier  $p$  tel que

$$2^p \leq n^r < 2^{p+1},$$

ce qui, en passant au logarithme et en divisant par  $r \ln 2$ , donne

$$\frac{p}{r} \leq \frac{\ln n}{\ln 2} < \frac{p+1}{r}.$$

Utilisons la fonction auxiliaire  $\chi : n \mapsto \frac{\psi(n)}{c} = \frac{\psi(n)}{\psi(2)} \ln 2$ . Alors la fonction  $\chi$  est croissante et additive, ce qui montre par une récurrence immédiate que  $\chi(n^r) = r \chi(n)$  pour tout couple  $(n, r)$  non nuls. Par croissance, on a donc

$$\chi(2^p) \leq \chi(n^r) \leq \chi(2^{p+1})$$

et par additivité

$$p \ln 2 = p \chi(2) \leq r \chi(n) < (p+1) \ln 2,$$

donc

$$\frac{p}{r} \leq \frac{\chi(n)}{\ln 2} < \frac{p+1}{r}.$$

Les deux nombres  $\frac{\chi(n)}{\ln 2}$  et  $\frac{\ln n}{\ln 2}$  sont donc dans le même intervalle de largeur  $1/r$ , donc par inégalité triangulaire

$$\left| \frac{\chi(n)}{\ln 2} - \frac{\ln n}{\ln 2} \right| \leq \frac{1}{r}.$$

Ce résultat étant vrai quelle que soit la valeur de l'entier  $r$ , on en déduit que  $\chi(n) = \ln n$ . Ceci est vrai pour tout  $n \in \mathbf{N}$ ,

$$\boxed{\forall n \in \mathbf{N} \quad \psi(n) = c \ln n.}$$

**Remarque 3** On a donc montré que : toute fonction additive et croissante est un multiple de la fonction logarithme.

**8.c)** Procédons par récurrence sur  $n$ .

- **Au rang  $n = 2$** , si l'on prend  $\mathbf{q}, \mathbf{r} \in \mathcal{A}_2$ , alors

$$\begin{aligned} \text{gauche} &= J_{2n}(p_1 q_1, \dots, p_1 q_n, p_2 r_1, \dots, p_2 r_n) = J_4(p_1 q_1, p_1 q_2, p_2 r_1, p_2 r_2) \\ &= J_3(p_1(q_1 + q_2), p_2 r_1, p_2 r_2) + p_1(q_1 + q_2) J_2\left(\frac{p_1 q_1}{p_1(q_1 + q_2)}, \frac{p_1 q_2}{p_1(q_1 + q_2)}\right) && \text{récursivité (P6)} \\ &= J_3(p_1, p_2 r_1, p_2 r_2) + p_1 J_2(q_1, q_2) && \text{puisque } q_1 + q_2 = 1 \\ &= J_3(p_2 r_1, p_2 r_2, p_1) + p_1 J_2(q_1, q_2) && \text{symétrie (P2)} \\ &= J_2(p_2, p_1) + p_2 J_2(r_1, r_2) + p_1 J_2(q_1, q_2) && \text{récursivité (P6) et } r_1 + r_2 = 1 \\ &= \text{droite} \end{aligned}$$

- **Soit  $n \geq 2$  et supposons la propriété vraie au rang  $n$ .**

Soient  $\mathbf{q}$  et  $\mathbf{r} \in \mathcal{A}_{n+1}$ . Soit  $\mathbf{p} \in \mathcal{A}_2$  avec  $p_1 > 0$  et  $p_2 > 0$ . Notons  $\mathbf{q}' = (q_1, \dots, q_{n-1}, q_n + q_{n+1}) \in \mathcal{A}_n$  et  $\mathbf{r}' = (r_1, \dots, r_{n-1}, r_n + r_{n+1}) \in \mathcal{A}_n$ .

Par symétrie (P2) et récursivité (P6), on a

$$\begin{aligned} J_{2n+2}(p_1 \mathbf{q}, p_2 \mathbf{r}) &= J_{2n}(p_1 \mathbf{q}', p_2 \mathbf{r}') + p_2(r_n + r_{n+1}) J_2\left(\frac{r_n}{r_n + r_{n+1}}, \frac{r_{n+1}}{r_n + r_{n+1}}\right) + \\ & \quad p_2(q_n + q_{n+1}) J_2\left(\frac{q_n}{q_n + q_{n+1}}, \frac{q_{n+1}}{q_n + q_{n+1}}\right). \end{aligned}$$

L'hypothèse de récurrence donne alors

$$\begin{aligned} J_{2n+2}(p_1 \mathbf{q}, p_2 \mathbf{r}) &= J_2(\mathbf{p}) + p_1 J_n(\mathbf{q}') + p_2 J_n(\mathbf{r}') \\ & \quad + p_2(r_n + r_{n+1}) J_2\left(\frac{r_n}{r_n + r_{n+1}}, \frac{r_{n+1}}{r_n + r_{n+1}}\right) + p_1(q_n + q_{n+1}) J_2\left(\frac{q_n}{q_n + q_{n+1}}, \frac{q_{n+1}}{q_n + q_{n+1}}\right). \end{aligned}$$

puis on remarque que

$$p_2 J_n(\mathbf{r}') + p_2(r_n + r_{n+1}) J_2\left(\frac{r_n}{r_n + r_{n+1}}, \frac{r_{n+1}}{r_n + r_{n+1}}\right) = p_2 J_{n+1}(\mathbf{r})$$

en utilisant encore la récursivité et la symétrie, et de même

$$p_1 J_n(\mathbf{q}') + p_1(q_n + q_{n+1}) J_2\left(\frac{q_n}{q_n + q_{n+1}}, \frac{q_{n+1}}{q_n + q_{n+1}}\right) = p_1 J_{n+1}(\mathbf{q}).$$

Au total, on a bien

$$J_{2n+2}(p_1 \mathbf{q}, p_2 \mathbf{r}) = J_2(\mathbf{p}) + p_1 J_{n+1}(\mathbf{q}) + p_2 J_{n+1}(\mathbf{r}),$$

ce qui termine la récurrence.

- **On conclut par récurrence que la propriété est vérifiée pour tout  $n \geq 2$ .**

**8.d)** Soit  $k \in \{2, 3, \dots, n-2\}$ . Posons  $\mathbf{p} = \left(\frac{k}{n}, \frac{n-k}{n}\right)$ . Posons également

$$\mathbf{q} = \left(\underbrace{\frac{1}{k}, \dots, \frac{1}{k}}_{k \text{ fois}}, 0, \dots, 0\right) \quad \text{et} \quad \mathbf{r} = \left(\underbrace{\frac{1}{n-k}, \dots, \frac{1}{n-k}}_{n-k \text{ fois}}, 0, \dots, 0\right).$$

Alors, par extensibilité et symétrie, on note que

$$\psi(n) = J_{2n}(p_1 \mathbf{q}, p_2 \mathbf{r})$$

puis, par le résultat précédent et en supprimant les zéros par extensibilité (P4) :

$$\psi(n) = J_2(\mathbf{p}) + p_1 J_n(\mathbf{q}) + p_2 J_n(\mathbf{r}) = J_2(\mathbf{p}) + p_1 \psi(k) + p_2 \psi(n-k),$$

ce qui nous donne bien la formule voulue en remplaçant  $p_1$  et  $p_2$  par leurs valeurs :

$$\boxed{\psi(n) = J_2\left(\frac{k}{n}, 1 - \frac{k}{n}\right) + \left(1 - \frac{k}{n}\right) \psi(n-k) + \frac{k}{n} \psi(k).}$$

Un argument de densité enfin pour conclure : pour tout rationnel  $p = k/n$ , on a

$$J_2(p, 1-p) = \psi(n) - \frac{k}{n} \psi(k) - \left(1 - \frac{k}{n}\right) \psi(n-k) = c \left\{ \ln n - \frac{k}{n} \ln(k) - \left(1 - \frac{k}{n}\right) \ln(n-k) \right\},$$

soit  $J_2(p, 1-p) = c H_2(p, 1-p)$ . Puisque la fonction  $J_2$  est continue (P1) et que la fonction  $H_2$  l'est aussi, la densité des rationnels dans  $[0; 1]$  montre que

$$\boxed{\forall p \in [0; 1] \quad J_2(p, 1-p) = c H_2(p, 1-p).}$$

**8.e)** La récursivité permet de construire, récursivement, la fonction  $J_n$  à partir des fonctions  $J_2, \dots, J_{n-1}$ .

Puisque  $J_2$  et  $cH_2$  coïncident et vérifient la même propriété de récursivité (question 6.b, le résultat est immédiat.

On a donc montré que : **Les seules fonctions vérifiant les propriétés (P1) à (P6) sont les multiples de l'entropie de Shannon  $H_n$ .**

**Remarque 4 (Sur la constante c)** La valeur de la constante  $c$  est peu importante pour les mathématiciens qui étudient par exemple la théorie du signal. Pour un physicien, elle est au contraire cruciale. On peut montrer en effet que l'entropie de Shannon est reliée à l'entropie de Boltzmann, pour laquelle la constante multiplicative est la **constante de Boltzmann**  $k_B$ , qui n'est pas un nombre sans dimension mais qui s'exprime en Joules par Kelvin ( $k_B = 1,38 \cdot 10^{-23} \text{ J} \cdot \text{K}^{-1}$ ), et qui fixe justement l'échelle des températures par rapport aux autres grandeurs physiques. C'est donc une constante fondamentale pour la conversion entre quantité mécanique et quantité thermodynamique ; on pourrait cependant décider de lui donner une valeur arbitraire en changeant les unités de mesure.

### TROISIÈME PARTIE

La fonction  $f$  est ce que vous avez sans doute vu en mécanique quantique sous le nom de **densité de probabilité**. Elle permet d'effectuer des calculs probabilistes sur des variables aléatoires prenant des valeurs réelles quelconques et caractérisées par une densité :

$$\forall I \subset \mathbf{R} \quad \mathbf{P}\{X \in I\} = \int_I f(x) dx.$$

Bien évidemment, rien sur les variables à densité n'est à savoir !

Notons une bonne fois pour toutes que : *que toute fonction  $f \in \mathcal{F}$  est bornée sur  $\mathbf{R}$  et vérifie  $f(x) = \underset{\pm\infty}{O}\left(\frac{1}{x^2}\right)$ .*

**9)** La fonction  $x \mapsto f(x) \ln f(x)$  est continue sur  $\mathbf{R}$ , bornée puisque composée d'une fonction continue  $u \mapsto u \ln u$  et d'une fonction bornée. De plus

$$f(x) \ln f(x) = \underset{x \rightarrow \pm\infty}{O}\left(\frac{\ln x}{x^2}\right) = \underset{x \rightarrow \pm\infty}{o}\left(\frac{1}{x^{3/2}}\right)$$

ce qui montre que

$$\boxed{x \mapsto f(x) \ln f(x) \text{ est intégrable sur } \mathbf{R}, \text{ et } H(f) \text{ est bien définie.}}$$

**10)**  $g_{\mu, \sigma}$  est une gaussienne d'espérance  $\mu$  et d'écart-type  $\sigma$ , et normée.

On remarque que

$$g_{\mu, \sigma}(x) = \frac{1}{\sigma} g_{0,1}\left(\frac{x - \mu}{\sigma}\right).$$

Voici l'allure de  $g_{0,1}$  et de  $g_{\mu,\sigma}$  :

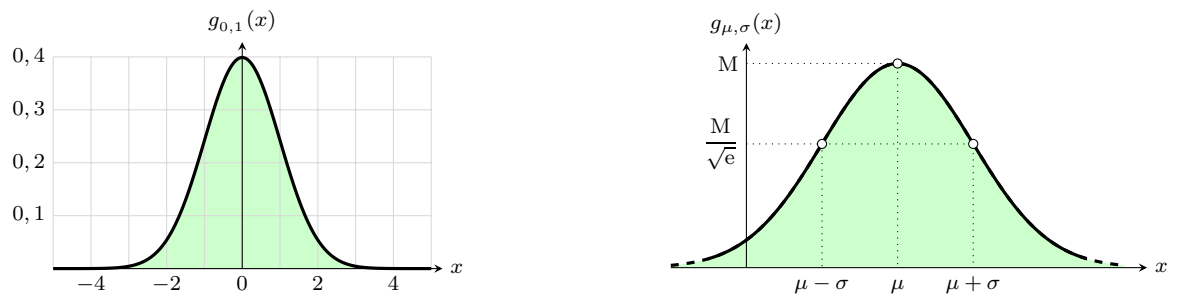


Fig. 3 — Les fonctions  $g_{0,1}$  (loi normale centrée réduite) et  $g_{\mu,\sigma}$  (loi normale d'espérance  $\mu$  et d'écart-type  $\sigma$ ).

On vérifie que c'est un élément de  $\mathcal{F}$  par le changement de variable  $y = (x - \mu)/\sigma\sqrt{2}$  et la valeur de l'intégrale de Gauss rappelée dans le préambule.

Ensuite, un changement de variable évident :

$$\int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \underset{\substack{\uparrow \\ y=x-\mu}}{=} \int_{-\infty}^{+\infty} \frac{y+\mu}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} dy = \int_{-\infty}^{+\infty} \frac{\mu}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} dy = \mu$$

et une intégration par parties :

$$\int_{-\infty}^{+\infty} (x - \mu)^2 g_{\mu,\sigma}(x) dx \underset{\substack{\uparrow \\ y=x-\mu}}{=} \int_{-\infty}^{+\infty} \frac{y^2}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} dy \underset{\substack{\uparrow \\ \text{IPP}}}{=} \sigma^2 \int_{-\infty}^{+\infty} \frac{e^{-y^2/2\sigma^2}}{\sigma\sqrt{2\pi}} dy = \sigma^2$$

On en déduit

$$H(g_{\mu,\sigma}) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \left[ \ln(\sigma\sqrt{2\pi}) + \frac{(x-\mu)^2}{2\sigma^2} \right] dx$$

$$\boxed{H(g_{\mu,\sigma}) = \frac{1}{2} + \ln(\sigma\sqrt{2\pi}).}$$

**11.a)** On reconnaît une forme intégrale de ce que, dans la question 7, on avait appelée *entropie relative*, ici aux distributions  $f$  et  $g_{\mu,\sigma}$ .

On a  $\ln(g_{\mu,\sigma}(x)) = \ln(\sigma\sqrt{2\pi}) + \frac{(x-\mu)^2}{2\sigma^2}$  donc  $x \mapsto f(x) \ln g_{\mu,\sigma}(x)$  est intégrable, ce qui montre que

$$\boxed{K_{\mu,\sigma}(f) \text{ est bien définie.}}$$

On utilise alors l'« inégalité de Jensen intégrale » obtenue à la question 3 dans le cas d'une intégrale sur un segment, mais dont on admet provisoirement qu'elle est vraie sur un intervalle non compact<sup>1</sup>. On obtient, pour toute fonction  $g \in \mathcal{F}$  d'intégrale 1 :

$$\int_{-\infty}^{+\infty} f(x) \ln g(x) dx \leq \ln \left( \int_{-\infty}^{+\infty} f(x) g(x) dx \right).$$

Notamment, pour la fonction  $g_{\sigma,\mu}$  :

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) \ln \left( \frac{f(x)}{g_{\mu,\sigma}(x)} \right) dx &= - \int_{-\infty}^{+\infty} f(x) \ln \left( \frac{g_{\mu,\sigma}(x)}{f(x)} \right) dx \\ &\geq - \ln \left( \int_{-\infty}^{+\infty} \frac{f(x)}{f(x)} g_{\mu,\sigma}(x) dx \right) = - \ln 1 = 0 \end{aligned}$$

donc

$$\boxed{K_{\mu,\sigma}(x) \geq 0.}$$

**11.b)** Les quantités  $m$  et  $s$  sont respectivement l'espérance et l'écart-type de la densité  $f$ .

<sup>1</sup>Essentiellement, c'est la même démonstration, mais en composant avec la fonction Arctan qui permet de ramener  $\mathbf{R}$  à  $]-\frac{\pi}{2}; \frac{\pi}{2}[$  et la décroissance vers 0 de  $f$  qui permet de « fermer » l'intervalle en  $[\frac{\pi}{2}; \frac{\pi}{2}]$  en prolongeant par continuité.

Pour tout choix de  $\mu$  et  $\sigma$ , on a, d'après la question précédente :

$$H(f) \leq - \int_{-\infty}^{+\infty} f(x) \ln g_{\mu,\sigma}(x) dx.$$

En prenant la valeur  $\mu = m$  et  $\sigma = s$ , on obtient

$$H(f) \leq - \int_{-\infty}^{+\infty} f(x) \left[ \ln \left( s\sqrt{2\pi} \right) + \frac{(x-m)^2}{2s^2} \right] dx = \ln \left( s\sqrt{2\pi} \right) + \frac{1}{2} = H(g_{m,s}).$$

On a donc montré que : **l'entropie d'une distribution est majorée par l'entropie de la gaussienne ayant même espérance et même écart-type.**

### QUATRIÈME PARTIE

**12.a)** La fonction  $f$  est continue sur  $\mathbf{R}$ . De plus, la fonction  $\varphi : x \mapsto x^2 f(x)$  est bornée; si l'on note  $M = \max(\|\varphi\|_\infty, 1)$  un majorant de cette fonction, alors

$$\forall x \in \mathbf{R} \quad 0 \leq f^r(x) \leq \frac{M^r}{x^{2r}} \quad 2r > 1,$$

ce qui montre que  $f^r$  est intégrable sur  $\mathbf{R}$ . De plus, la fonction  $f$  n'est pas nulle (elle est d'intégrale 1) donc  $\int_{-\infty}^{+\infty} f^r > 0$  donc

L'entropie de Rényi  $h_r(f)$  est bien définie.

Montrons que l'on peut utiliser le théorème de dérivation sous le signe somme.

On pose, pour tout  $x \in \mathbf{R}$  et tout  $r > 1/2, r \neq 1$  :

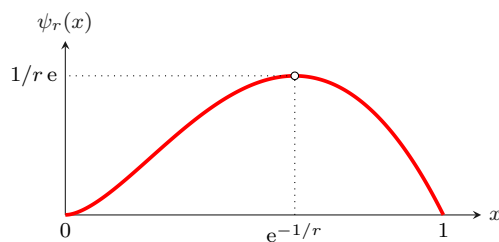
$$h(r, x) = f^r(x).$$

Notons que  $h$  est continue en tant que fonction de deux variables.

C'est un peu délicat. Effectuons d'abord une rapide étude des fonctions

$$\psi_r : x \mapsto -x^r \ln x$$

qui sont croissantes sur  $[0; e^{-1/r}]$ , puis décroissantes :



- Pour tout  $x \in \mathbf{R}$ , la fonction  $h(\cdot, x) : r \mapsto f^r(x)$  est continue sur  $]\frac{1}{2}; +\infty[ \times \mathbf{R}$ ;
- $\frac{\partial h}{\partial r} : (r, x) \mapsto (\ln f(x)) f^r(x)$  est continue par rapport à chacune de ses variables sur  $]\frac{1}{2}; +\infty[ \times \mathbf{R}$  (on a prolongé par continuité en posant  $t^r \ln t = 0$  pour  $t = 0$ );



Alfred « Buba » Rényi (1921–1970), mathématicien hongrois né et mort à Budapest. Il fut détenu dans un camp de travail en 1944 mais réussit à s'échapper, puis, en un acte de courage inouï, délivra ses parents du ghetto de Budapest, après avoir volé un uniforme de soldat, en allant simplement les chercher !

Il travailla avec le célèbre Paul Erdős, étudia des courbes aléatoires remplissant l'espace, et prouva que tout entier pair est la somme d'un nombre premier et d'un nombre « presque premier » (n'ayant que deux facteurs premiers). Puis il consacra sa recherche à la théorie des probabilités. C'est lui qui décréta que : un mathématicien est une machine à transformer le café en théorème. Il fonda l'Institut mathématique de l'Académie des sciences de Hongrie.

« Si je me sens malheureux, je fais des mathématiques pour me rendre heureux. Si je suis heureux, je fais des mathématiques pour rester heureux. »

- Soit  $[a; b] \subset ]\frac{1}{2}; +\infty[$ . La fonction  $f$  tend vers 0 en  $\pm\infty$ . Notons  $A$  un réel (que l'on prendra vérifiant  $A \geq 1$ ) tel que, pour tout  $|x| \geq A$ , on ait  $0 \leq f(x) \leq e^{-1/a} \leq 1$ . Notamment

$$\forall |x| \geq A \quad \forall r \in [a; b] \quad |f(x)^r \ln f(x)| \leq |f(x)^a \ln f(x)|$$

et, par croissance de  $\psi_a$  sur  $[0; e^{-1/a}]$ , la majoration  $f(x) \leq M/x^2$  donne

$$\forall |x| \geq A \quad \forall r \in [a; b] \quad |f(x)^r \ln f(x)| \leq \frac{M^a}{x^{2a}} \left| \ln \left( \frac{M^a}{x^{2a}} \right) \right|$$

La fonction  $(r, x) \mapsto |h(r, x)|$  est continue sur le compact  $[a; b] \times [-A; A]$ ; notons-en  $K$  un majorant. Alors pour tout  $r \in [a; b]$ , on a

$$\forall x \in \mathbf{R} \quad 0 \leq |f^r(x) \ln f(x)| \leq \varphi(x) := \begin{cases} K & \text{si } |x| \leq A \\ \frac{M^a}{x^{2a}} \left| \ln \left( \frac{M^a}{x^{2a}} \right) \right| & \text{sinon.} \end{cases}$$

Enfin, la fonction  $\psi$  est continue par morceaux et intégrable sur  $\mathbf{R}$  car négligeable devant  $1/x^\beta$  où l'on choisit un réel  $\beta$  tel que  $1 < \beta < 2a$ .

En conclusion, on peut appliquer le théorème de dérivation sous le signe somme :

$$\boxed{\text{F est de classe } \mathcal{C}^1}$$

et bien entendu

$$\forall r \in \left] \frac{1}{2}; +\infty \right[ \quad F'(r) = \int_{-\infty}^{+\infty} f^r(x) \ln f(x) dx.$$

En particulier,

$$\boxed{F'(1) = -H(f).}$$

**12.b)** On constate que

$$h_r(f) = -\frac{\ln F(r) - \overbrace{\ln F(1)}^{=0}}{r-1} \xrightarrow{r \rightarrow 1} -[\ln F]'(1) = -\frac{F'(1)}{F(1)}$$

donc, puisque  $F(1) = 1$  :

$$\boxed{\lim_{r \rightarrow 1} h_r(f) = H(f).}$$

**13.a)** Puisque  $f$  et  $g$  sont dans  $\mathcal{F}$ , notamment  $f$  est bornée donc on a la domination

$$\forall (x, y) \in \mathbf{R}^2 \quad |f(x-y)g(y)| \leq \|f\|_\infty |g(y)|$$

donc  $\int_{-\infty}^{+\infty} f(x-y)g(y) dy$  est bien définie.

**13.b)** Montrons maintenant la continuité de  $f * g$  en notant que

$$h : (x, y) \mapsto f(x-y)g(y)$$

est continue par rapport à chacune de ses variables, et uniformément dominée :

$$\forall x \in \mathbf{R} \quad \forall y \in \mathbf{R} \quad |h(xny)| \leq \|f\|_\infty g(y) =: \varphi(y)$$

et bien entendu  $g \in \mathcal{L}^1(\mathbf{R}, \mathbf{R}^+)$ .

$$\boxed{\text{La fonction } f * g \text{ est continue sur } \mathbf{R}.}$$

La même domination permet d'utiliser le théorème de convergence dominée (dans le cas continu) :

$$\lim_{x \rightarrow +\infty} (f * g)(x) = \int_{-\infty}^{+\infty} \underbrace{\lim_{x \rightarrow +\infty} f(x-y)g(y)}_{=0} dy = 0,$$

et de même en  $-\infty$ . Ainsi,  $f * g$  est continu et tend vers 0 en  $\pm\infty$ , donc

$$\boxed{f * g \text{ est uniformément continue.}}$$

**Remarque 5 (Petite preuve)** Prouvons le résultat utilisé : toute fonction continue et tendant vers 0 en  $\pm\infty$  est uniformément continue.

□ Soit  $\varepsilon > 0$ . Il existe  $M > 0$  tel que, pour tout  $x \in \mathbf{R}$  tel que  $|x| \geq M$ , on a  $|f(x)| \leq \varepsilon/2$ . De plus, la fonction  $f$  est uniformément continue sur  $[-M; M]$  : il existe un réel  $\eta > 0$  tel que, si  $x, y \in [-M; M]$  vérifiant  $|x - y| \leq \eta$ , alors  $|f(x) - f(y)| \leq \varepsilon$ . En examinant tous les cas (selon que  $|x|$  et  $|y|$  sont supérieurs ou inférieurs à  $M$ ), on voit que

$$\forall x, y \in \mathbf{R} \quad |x - y| \leq \eta \implies |f(x) - f(y)| \leq 2\varepsilon. \quad \square$$

**14.a)** Supposons  $f, g$  à supports bornés. Il existe donc des compacts  $[-a; a]$  et  $[-b; b]$  en dehors desquels  $f$  (respectivement  $g$ ) sont nuls. Alors, si  $|x| > a + b$ , on a  $f(x - y)g(y) = 0$  pour toute valeur de  $f$ .

$$\boxed{\text{Si } f, g \in \mathcal{F}_0 \text{ alors } fg \in \mathcal{F}_0.}$$

**14.b)** Posons encore  $\varphi : x \mapsto x \ln x$ . On a alors, en utilisant les notations précédentes et en posant  $M = a + b$  :

$$\begin{aligned} \mathbf{H}(f * g) &= - \int_{-M}^M \varphi \left( \int_{-b}^b f(x - y) g(y) dy \right) dx \\ &\stackrel{(*)}{\geq} - \int_{-M}^M \left( \int_{-a}^a \varphi(f(x - y)) g(y) dy \right) dx \\ &\stackrel{(**)}{\geq} - \int_{-a}^a \left( \int_{-M}^M \varphi(f(x - y)) dx \right) g(y) dy \\ &\stackrel{t=x-y}{\geq} - \int_{-a}^a \left( \int_{-\infty}^{+\infty} \varphi(f(t)) dt \right) g(y) dy \\ &\geq \mathbf{H}(f) \int_{-a}^a g(y) dy = \mathbf{H}(f). \end{aligned}$$

où l'on a utilisé : (\*) l'inégalité de Jensen continue et (\*\*) le théorème de Fubini. Les dernières lignes n'utilisent que les définitions et le fait que les fonctions sont toutes à support compact.

$$\boxed{\mathbf{H}(f * g) \geq \mathbf{H}(f).}$$

**Remarque 6** On peut également remarquer que  $f * g = g * f$ , ce qui donne l'inégalité  $\mathbf{H}(f * g) \geq \mathbf{H}(f)$ , et donc  $\mathbf{H}(f * g) \geq \max(\mathbf{H}(f), \mathbf{H}(g))$ .

On peut trouver une interprétation de cette inégalité en terme de traitement du signal (le signal est  $f$ , il est broyé par une fonction de convolution  $g$  sans pertes, son entropie a nécessairement augmenté ; à creuser).

**15.a)** On remarque bien sûr que  $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ , ce qui permet d'utiliser l'inégalité de Young forte :

$$\begin{aligned} h_r(f * g) &= \frac{1}{1-r} \int_{-\infty}^{+\infty} (f * g)^r(x) dx \\ &= \frac{r}{1-r} \ln \left( \int_{-\infty}^{+\infty} (f * g)^r(x) dx \right)^{1/r} \\ &\geq \frac{r}{1-r} \ln \left[ \left( \frac{\mathbf{C}(p) \mathbf{C}(q)}{\mathbf{C}(r)} \right)^{1/2} \left( \int_{-\infty}^{+\infty} f^p \right)^{1/p} \left( \int_{-\infty}^{+\infty} g^q \right)^{1/q} \right] \\ &\geq \frac{r}{1-r} \left\{ \frac{1}{2} \ln \left( \frac{\mathbf{C}(p) \mathbf{C}(q)}{\mathbf{C}(r)} \right) + \underbrace{\frac{1-p}{p}}_{=\lambda \frac{1-r}{r}} h_p(f) + \underbrace{\frac{1-q}{q}}_{=(1-\lambda) \frac{1-r}{r}} h_q(g) \right\} \end{aligned}$$

Quelques calculs relativement fastidieux montrent que

$$\begin{aligned} \alpha &:= \frac{r}{2(1-r)} \ln \left( \frac{\mathbf{C}(p) \mathbf{C}(q)}{\mathbf{C}(r)} \right) \\ &= \frac{r}{2(1-r)} \left[ \frac{1}{p} \ln p + \left(1 - \frac{1}{p}\right) \ln \left(1 - \frac{1}{p}\right) + \frac{1}{q} \ln q + \left(1 - \frac{1}{q}\right) \ln \left(1 - \frac{1}{q}\right) - \frac{1}{r} \ln r - \left(1 - \frac{1}{r}\right) \ln \left(1 - \frac{1}{r}\right) \right] \\ &= \frac{r}{2(1-r)} \left[ -\frac{r-1}{r} \mathbf{H}_2(\lambda, 1-\lambda) + \left(1 - \frac{1}{r}\right) \ln \left(1 - \frac{1}{r}\right) + \frac{1}{p} \ln p + \frac{1}{q} \ln q - \frac{1}{r} \ln r - \left(1 - \frac{1}{r}\right) \ln \left(1 - \frac{1}{r}\right) \right] \\ &= \frac{r}{2(1-r)} \left[ \frac{1}{p} \ln p + \frac{1}{q} \ln q - \frac{1}{r} \ln r \right] + \frac{1}{2} \mathbf{H}_2(\lambda, 1-\lambda), \end{aligned}$$

ce qui mène au résultat voulu :

$$\boxed{h_r(f * g) \geq \lambda h_p(f) + (1-\lambda) h_q(g) + \frac{1}{2} \mathbf{H}_2(\lambda, 1-\lambda) + \frac{r}{2(r-1)} \left( \frac{1}{r} \ln r - \frac{1}{p} \ln p - \frac{1}{q} \ln q \right).}$$

**15.b)** Il ne reste plus qu'à passer à la limite  $r \rightarrow 1$ . Un bête développement limité montre que le dernier terme tend vers 0 tandis que  $p \rightarrow 1$  et  $q \rightarrow 1$ , donc  $h_r(f * g) \rightarrow H(f * g)$ ,  $h_p(f) \rightarrow H(f)$  et  $h_q(g) \rightarrow H(g)$  :

$$\boxed{H(f * g) \geq \lambda H(f) + (1 - \lambda) H(g) + \frac{1}{2} H_2(\lambda, 1 - \lambda).}$$

**15.c)** On étudie la fonction  $M : \lambda \mapsto \lambda H(f) + (1 - \lambda) H(g) + \frac{1}{2} H_2(\lambda, 1 - \lambda)$ . Sa dérivée est

$$M' : \lambda \mapsto H(f) - H(g) - \frac{1}{2} \ln \left( \frac{\lambda}{1 - \lambda} \right),$$

qui s'annule donc lorsque  $\frac{\lambda}{1 - \lambda} = \exp 2(H(f) - H(g))$  soit encore

$$\lambda_0 = \frac{1}{1 + \exp 2(H(g) - H(f))},$$

ce qui donne après deux lignes de calcul

$$M(\lambda_0) = \frac{1}{2} \ln (e^{2H(f)} + e^{2H(g)}).$$

L'inégalité  $H(f * g) \geq M(\lambda_0)$  donne, après multiplication par 2 et passage à l'exponentielle :

$$\boxed{e^{2H(f * g)} \geq e^{2H(f)} + e^{2H(g)}.$$