X-ENS PC 2016 : Entropie de Shannon

Gilbert Primet

24 avril 2016

Partie I

1. φ, produit de deux fonctions $\mathcal{C}^\infty$ est $\mathcal{C}^\infty$ sur $]0,+\infty[$. De plus :
$$\lim_{t\to 0} t\ln(t) = 0$$
donc φ est continue en 0 (donc $\mathcal{C}_0$ sur $[0,+\infty[$.
$$\forall t \in]0,+\infty[\varphi'(t) = -1-\ln(t)$$
et
$$\lim_{t\to 0}\varphi'(t) = +\infty.$$

2. (a) Soit une suite convergente $(p_n) = (p_1^{(n)},\cdots,p_N^{(n)})$ d'éléments de Σ_N de limite $p = (p_1,\cdots,p_N)$. On a donc :
$$\forall i \in [|1,N|] \forall n \in \mathbb{N} p_i^{(n)} \geqslant 0,\ \forall n \in \mathbb{N} \sum_{i=1}^{N} p_i^{(n)} = 1.$$
Par passage à la limite et conservation des inégalités larges, on obtient :
$$\forall i \in [|1,N|] p_i \geqslant 0.\ \sum_{i=1}^{N} p_i = 1.$$
Donc $p \in \Sigma_N$. Ceci prouve par caractérisation séquentielle que Σ_N est fermé.

   (b) Si $p \in \Sigma_N$, alors $\|p\|_1 = \sum_{i=1}^{N} p_i = 1$ donc Σ_N est borné (On peut aussi remarquer que $p \in \Sigma_N \Rightarrow \|p\|_\infty \leqslant 1$.)

   (c) Enfin, si $(p,q) \in \Sigma_N^2, p = (p_1,\cdots,p_N),\ q = (q_1,\cdots,q_N)$, et $\lambda \in [0,1]$, alors, $\lambda p + (1-\lambda)q = (\lambda p_i + (1-\lambda q_i)_{i\in[|1,N|]}$, et :
$$\sum_{i=1}^{N} \lambda p_i + (1-\lambda q_i = \lambda \sum_{i=1}^{N} p_i + (1-\lambda)\sum_{i=1}^{N} q_i = \lambda + (1-\lambda) = 1$$
et
$$\forall i \in [|1,N|] \lambda p_i + (1-\lambda)q_i \geqslant 0.$$
donc
$$\forall (p,q) \in \Sigma_N^2 \forall \lambda \in [0,1] \lambda p + (1-\lambda)q \in \Sigma_N.$$
Σ_N est donc convexe.

3. φ est positive sur $[0,1]$, donc par somme H_N est positive. Les projections $p \mapsto p_i$ sont continues, φ est continue, donc par composée et somme, H_N est continue.
Lorsque $p_i = \frac{1}{N}$, alors
$$H_N(p) = N\varphi\left(\frac{1}{N}\right) = ln(N)$$

4. (a) Posons $g(t)=\varphi(a+t)+\varphi(b-t)-\varphi(a)-\varphi(b)$ pout $t\in[0,b]$. g est définie et continue sur $[0,b]$, dérivable sur $]0,b[$, $g(0)=0$ et $\forall t\in]0,b[\ g'(t)=\varphi'(a+t)-\varphi'(b-t)$. Or, par stricte croissance de ln, φ' décroît strictement sur $]0,+\infty[$. Pour $a+t<b-t$, soit $t<\frac{a+b}{2}$, $g'(t)>0$. g croît donc strictement sur $[0,\frac{a+b}{2}[$, et comme $g(0)=0$, on a donc :

$$\forall t\in\left]0,\frac{a+b}{2}\right[g(t)>0$$

ce qui montre le résultat demandé pour $\varepsilon=\frac{a+b}{2}$

(b) La fonction H_N est continue sur Σ_N . De plus Σ_N est fermé et borné. H_N est donc bornée sur Σ_N et atteint ses bornes. D'après la question précédente, si H_N atteint son maximum en $(p_1,\cdots,p_N)$, tous les p_i sont égaux. Si ce n'était pas le cas, avec $p_i<p_j$, il existerait $t\in]0,p_j[$ tel que

$$\varphi(p_i+t)+\varphi(p_j-t)>\varphi(p_i)+\varphi(p_j),$$

et l'on aurait alors

$$H_N(q_1,\cdots,q_N)>H_N(p_1,\cdots,p_N),$$

où $(q_1,\cdots,q_N)$ s'obtient à partir de $(p_1,\cdots,p_N)$, en posant

$$q_i=p_i+t,\ q_j=p_j-t,\ q_k=p_k \text{ si } k\in[|1,N|]\setminus\{i,j\}.$$

On vérifie aisément que $q\in\Sigma_N$.
Le maximum est donc atteint au seul élément $p=(\frac{1}{N},\cdots,\frac{1}{N})$, qui correspond à la probabilité uniforme sur $\{1,\cdots,N\}$. et ce maximum est $\ln(N)$.

5. (a) $\forall i\in\mathbb{N}^*\varphi(p_i)=-a(1-a)^{i-1}(\ln(a)+(i-1)\ln(1-a))$. On sait que

$$\forall x\in]-1,1[\sum_{n=0}^{\infty}x^n=\frac{1}{1-x},$$

et donc par dérivation d'une série entière à l'intérieur de son intervalle de convergence :

$$\forall x\in]-1,1[\sum_{n=1}^{\infty}nx^{n-1}=\frac{1}{(1-x)^2},$$

d'où :

$$\forall x\in]-1,1[\sum_{n=0}^{\infty}nx^n=\frac{x}{(1-x)^2}.$$

Par le changement d'indice $n=i-1$, on obtient donc :

$$H_\infty(p)=-\ln(a)-\frac{(1-a)\ln(1-a)}{a}=\frac{\varphi(a)+\varphi(1-a)}{a}.$$

Notons $h(a)$ cette quantité. h est dérivable sur $]0,1[$ et :

$$\forall a\in]0,1[h'(a)=\frac{a(\varphi'(a)-\varphi'(1-a))-(\varphi(a)+\varphi(1-a))}{a^2}.$$

$h'(a)$ est donc du signe du numérateur qui vaut :

$$N(a)=(\varphi'(a)-\varphi'(1-a))-(\varphi(a)+\varphi(1-a))=a(-\ln(a)+\ln(1-a)+\ln(a))+(a-1)\ln(1-a),$$

et :

$$N(a)=a\ln(1-a)<0.$$

h décroît donc sur $]0,1[$ de $+\infty$ à 0.

(b) On peut choisir la série de Bertrand convergente de terme général $p_i=\frac{C}{i\ln^2(i)}$.où C est choisi de façon que $\sum_{i=1}^{+\infty}p_i=1$ On a alors, lorsque i tend vers $+\infty$:

$$\ln(p_i)\sim-\ln(i)$$

donc

$$\varphi(p_i)\sim\frac{C}{i\ln(i)}$$

et donc

$$H_n(p)=+\infty.$$

6. Les variables aléatoires$Y_k = \ln(p_{X_k}) = (\ln \circ p)(X_k)$ sont deux à deux indépendantes et de même loi :

$$Y_k(\Omega) = \{\ln(p_1), \cdots, \ln(p_N)\},\ \forall i \in [|1, N|] p([Y_k = \ln(p_i)]) = p_i.$$

. Donc :

$$E(ln(p_{X_k})) = \sum_{i=1}^{n} p_i \ln(p_i) = -H_N(p).$$

(On aurait pu aussi simplement utiliser le théorème de transfert).
Enfin, pour tout entier naturel non nul n :

$$\frac{Y_1 + \cdots + Y_n}{n} = \frac{1}{n} \ln\left(\prod_{k=1}^{n} p_{X_k}\right).$$

Ceci acquis, la propriété demandée traduit donc la loi faible des grands nombres sur la suite de variables aléatoires indépendantes Y_i. (Ces variables, qui sont finies , possèdent toutes une variance, et le théorème peut donc s'appliquer).

Partie II

7. J_f, somme de H_N et d'une fonction polynôme est continue sur Σ_N qui est fermé et borné. J_N est donc bornée et atteint ses bornes. $\Sigma_N(f)$ est donc non vide.
8. (a) D'après la question (4), il existe $\epsilon \in]0, p_2[$ tel que pour tout $t \in]0, \epsilon[$, $\varphi(t) + \varphi(p_2 - t) > \varphi(0) + \varphi(p_2)$. Pour $t \in]0, \epsilon[$, on a :

   $$t f_1 + (p_2 - t) f_2 = t(f_1 - f_2) + p_2 f_2.$$

   Donc , en posant $p' = p + (t, p_2 - t, 0, \cdots, 0)$, alors $p' \in \Sigma_N$ et $J_f(p') - J_f(p) = t(f_1 - f_2) + A(t)$, où :

   $$A(t) = \varphi(t) + \varphi(p_2 - t) - \varphi(0) - \varphi(p_2) = -t \ln(t) + o_0(t \ln(t)).$$

   Donc

   $$J_f(p') - J_f(p) \sim_0 -t \ln(t).$$

   Donc, au voisinage de 0 on a : $J_f(p') - J_f(p) > 0$, cequi contredit le fait que $p \in \Sigma_N(f)$.

   (b) On peut bien sûr répéter le même raisonnement si $p_i = 0$ et $p_j > 0$. Comme les composantes d'un élément de Σ_N ne peuvent pas être toutes nulles (leur somme vaut 1), on obtient le résultat demandé.
9. On munit $\mathbb{R}^N$ du produit scalaire canonique.

   (a) $E_0 = \text{Vect}(1, 1, \cdots, 1)^\perp$, donc E_0 est un sous espace vectoriel de $\mathbb{R}^N$ de dimension $N - 1$ et $E_0^\perp = (\text{Vect}(1, \cdots, 1)^\perp)^\perp = \text{Vect}(1, \cdots, 1)$, puisque $\mathbb{R}^N$ est de dimension finie.

   (b) Si $a = (0, \cdots, 0)$, la question est triviale et tout $\epsilon > 0$ convient.
   Supposons $a \neq (0, \cdots, 0)$. Notons $M = \max_{i \in [|1, N|]} |a_i|$. $M > 0$ puisque les a_i sont non tous nuls par hypothèse.
   $p + ta = (p_i + ta_i)_{i \in [|1, N|]}$. On a :

   $$\sum_{i=1}^{N} p_i + ta_i = \sum_{i=1}^{N} p_i + t \sum_{i=1}^{N} a_i = 1.$$

   D'autre part : Si $|ta_i| < p_i$, alors $p_i + ta_i > 0$. En particulier, si $|t| < \frac{p_i}{M}$, pour tout $i \in [|1, N|]$, alors cette condition est réalisée. En posant $m = \min_{i \in [|1, N|]} p_i > 0$, et $\epsilon = \frac{m}{M}$, on a bien $\tilde{p} \in \Sigma_N$ pour tout $t \in] - \epsilon, +\epsilon[$.
   Enfin $\forall t \in \mathbb{R} \tilde{p}'(t) = a$. Ceci est vrai en particulier pour $t = 0$.

   (c) $\forall t \in] - \epsilon, \epsilon[\ u(t) = J_f(\tilde{p}(t)) = H_N(\tilde{p}(t)) + \sum_{i=1}^{N} (p_i + ta_i) f_i$. u est dérivable sur $] - \epsilon, \epsilon[$, et :

   $$\forall t \in] - \epsilon, \epsilon[u'(t) = \sum_{i=1}^{N} -a_i(1 + \ln(p_i + ta_i) + \sum_{i=1}^{N} a_i f_i.$$

Comme u admet un maximum en 0 (point intérieur à l'intervalle) , $u'(0) = 0$, et compte-tenu du fait que $a \in E_0$, on obtient :

$$\sum_{i=1}^{n} a_i(f_i - \ln(p_i)) = 0.$$

Donc le vecteur $(\ln(p_i) - f_i)_{i \in [|1,N|]}$ est élément de $E_0^\perp = \text{Vect}(1, \cdots, 1)$, et il existe donc $c \in \mathbb{R}$ tel que :$\forall i \in [|1, n|] ln(p_i) = f_i + c$.

10. Comme $p \in \Sigma_N(f)$, on a de plus

$$\sum_{i=1}^{N} p_i = \sum_{i=1}^{N} e^{f_i + c} = 1.$$

Donc $c = -\ln\left(\sum_{i=1}^{N} e^{f_i}\right)$. et $\Sigma_N(f)$ possède un unique élément

$$p = (e^{f_1 + c}, \cdots, e^{f_N + c}).$$

On obtient

$$J_{f,*} = J_f(p) = \sum_{i=1}^{N} p_i(f_i - \ln(p_i)) = -c \sum_{i=1}^{N} p_i = -c = \ln\left(\sum_{i=1}^{N} e^{f_i}\right).$$

11. F est dérivable comme quotient de fonctions dérivables et : $\forall \beta \in]0, +\infty[$ $F'(\beta) = -\frac{N(\beta)}{\beta^2}$ où $N(\beta) = -\ln\left(\sum_{i=1}^{N} e^{\beta f_i}\right) - \frac{\sum_{i=1}^{N} \beta f_i e^{\beta f_i}}{\left(\sum_{i=1}^{N} e^{\beta f_i}\right)}$. Or , en posant $d = -\ln\left(\sum_{i=1}^{N} e^{\beta f_i}\right)$, l'unique élément de $\Sigma_N(f)$ est : $p(\beta) = (e^{f_1 + d}, \cdots, e^{f_N + d})$, et

$$H_N(p(\beta)) = \ln(\sum_{i=1}^{N} e^{\beta f_i}) - \sum_{i=1}^{N} p_i \beta f_i$$

, où

$$p_i = e^{\beta f_i + d} = \frac{e^{\beta f_i}}{\sum_{i=1}^{N} e^{\beta f_i}}.$$

C'est bien le résultat demandé !

12. Comme $N \geqslant 2$, le numérateur de $F(\beta)$ tend vers $\ln(N) > 0$ lorsque β tend vers 0, et donc

$$\lim_{\beta \to 0} F(\beta) = +\infty.$$

Soit $f = \max_{i \in [|1,N|]} f_i$, et $k = \text{card}\{i \in [|1, N|] | f_i = f\}$. Alors $\ln\left(\sum_{i=1}^{N} e^{\beta f_i}\right) = \beta f + \ln\left(\sum_{i=1}^{N} e^{\beta(f_i - f)}\right)$. Le deuxième terme tend vers e^k lorsque β tend vers $+\infty$. On a donc

$$\lim_{\beta \to +\infty} F(\beta) = f = \max_{i \in [|1,N|]} f_i$$

Partie III

13. On a, d'après le théorème de transfert :

$$A_{l,k} = \sum_{i=1}^{N} \mathbf{P}(Y = i)(g_l(i) - m_l)(g_k(i) - m_k) = \mathbf{E}\left((g_l(Y) - m_l)(g_k(Y) - m_k)\right).$$

A est évidemment symétrique. Si $\theta = (\theta_1, \cdots, \theta_d) \in \mathbb{R}^d$, alors

$$\theta^T A \theta = \sum_{(l,k) \in [|1,d|]^2} A_{l,k} \theta_l \theta_k = \mathbf{E}\left(\left(\sum_{j=1}^{d} \theta_j (g_j(Y) - m_j)\right)^2\right) \geqslant 0.$$

(On développe le carré et on utilise la linéarité de l'espérance. On sait de plus que l'espérance d'une variable positive est positive.)

14. (a) D'après la solution de la question précédente, cela signifie que la variable aléatoire $\sum_{j=1}^d \theta_j(g_j(Y)-m_j)$ est nulle. Donc, en reprenant l'indexation de l'énoncé :

$$\sum_{l=1}^d \theta_l g_l(Y) = \sum_{l=1}^d m_l\theta_l = c.$$

Comme pout tout $i \in [|1,N|] p_i \neq 0$, Y peut prendre toute valeur dans $[|1,N|]$ avec une probabilité non nulle. On a donc pour tout $i \in [|1,N|]$:

$$\sum_{l=1}^d \theta_l g_l(i) = c$$

(Pour ceux que cette solution probabiliste rebute, on peut écrire l'espérance précédente comme somme de N termes positifs, pondérés par les $p_i > 0$. Le coefficient de chaque p_i est donc nul.)

(b) on a : $(\theta_1,\cdots,\theta_d,-c) \in \ker \tilde{M}$, donc $\theta = 0$ (et $c = 0$).

15. $f_1,\cdots,f_N$ sont des polynômes, de même que $\theta \mapsto q^T M\theta$ (ce sont même des formes linéaires.) Plus précisément,

$$\forall i \in [|1,N|]\ f_i(\theta) = \sum_{l=1}^d M_{i,l}\theta_l$$

et :

$$q^T M\theta = \sum_{i=1}^N q_i f_i(\theta) = \sum_{(i,l)\in[|1,N|]\times[|1,d|]} q_i M_{i,l}\theta_l$$

L admet donc d'après les théorèmes usuels des dérivées partielles, et :

$$\forall j \in [|1,d|] \frac{\partial L}{\partial \theta_j}(\theta) = \frac{\sum_{i=1}^N M_{i,j}e^{f_i(\theta)}}{\sum_{i=1}^N e^{f_i(\theta)}} - \sum_{i=1}^N q_i M_{i,j}$$

Ces dérivées partielles sont continues, d'après les théorèmes usuels, et en reprenant les notations de l'énoncé :

$$\mathrm{grad}(f) = Mp(\theta)^T - q^T M$$

16. Si θ est point critique de L, alors, on obtient l'égalité demandée par transposition de $Mp(\theta)^T - q^T M = 0$. Or, d'après la définition, $p \in \Sigma_N(\bar{g}, g) \iff p \in \Sigma_N, M^T p = M^T q$. La condition matricielle est réalisée et $p(\theta) \in \Sigma_N$. On a donc bien $p(\theta) \in \Sigma_N(\bar{g}, g)$.

17.

$$\forall k \in [|1,d|] \frac{\partial L}{\partial \theta_k}(\theta) = \frac{\sum_{i=1}^N M_{i,k}e^{f_i(\theta)}}{\sum_{i=1}^N e^{f_i(\theta)}} - \sum_{i=1}^N q_i M_{i,k}$$

On voit que ces fonctions admettent des dérivées partielles, et que :

$$\forall (l,k) \in [|1,d|]^2 \frac{\partial^2 L}{\partial \theta_l \partial \theta_k}(\theta) = \frac{\sum_{i=1}^N M_{i,l}M_{i,k}e^{f_i(\theta)}}{\sum_{i=1}^N e^{f_i(\theta)}} - \frac{\left(\sum_{i=1}^N M_{i,k}e^{f_i(\theta)}\right)\left(\sum_{i=1}^N M_{i,l}e^{f_i(\theta)}\right)}{\left(\sum_{i=1}^N e^{f_i(\theta)}\right)^2}.$$

Donc :

$$\forall (l,k) \in [|1,d|]^2 \frac{\partial^2 L}{\partial \theta_l \partial \theta_k}(\theta) = \sum_{i=1}^N M_{i,l}M_{i,k}p_i(\theta) - \left(\sum_{i=1}^N M_{i,l}p_i(\theta)\right)\left(\sum_{i=1}^N M_{i,k}p_i(\theta)\right)$$

On reconnaît là la covariance des variables aléatoires $g_k(Y)$ et $g_l(Y)$ pour la probabilité de distribution $p_i(\theta)$. Or $m_k(\theta)$ est précisément l'espérance de $g_k(Y)$ pour cette probabilité. On obtient donc , par une autre expression de la covariance :

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = E(X - E(X))(Y - E(Y))$$

la formule demandée.

18. (a) Posons , pour $t \in [0,1]$:

$$u(t) = L(t\theta + (1-t)\theta') = L((t(\theta_i - \theta_i') + \theta')_{i\in[|1,d|]}).$$

Alors u est deux fois dérivable sur $[0,1]$, et, avec des notations pas très heureuses :

$$\forall t \in [0,1]\ u'(t) = \sum_{i=1}^{d}(\theta_i - \theta_i')\frac{\partial L}{\partial \theta_i}(t\theta + (1-t)\theta').$$

En particulier $u'(1) = 0$, puisque toutes les dérivées partielles s'annulent en θ. Puis :

$$\forall t \in [0,1]\ u''(t) = \sum_{i=1}^{d}(\theta_i - \theta_i')\left(\sum_{j=1}^{d}(\theta_j - \theta_j')\frac{\partial^2 L}{\partial \theta_j \partial \theta_i}(t\theta + (1-t)\theta')\right).$$

On reconnaît (?) que cette quantité est : $V(\sum_{k=1}^{d}(\theta_k - \theta_k')g_k(X))$, pour la distribution de probabilités $p(t\theta + (1-t)\theta')$, et cette variance est strictement positive. En effet, si elle s'annulait, d'après la question 14,(qui s'applique ici puisque $p_i(t\theta + (1-t)\theta') > 0$ pour tout i), on aurait $\theta - \theta' = 0$, ce qui n'est pas. u' est donc strictement croissante.

(b) Or $u'(0) = 0$, puisque L admet un point critique en θ'. Ceci contredit donc la stricte croissance de u'. Il existe donc au plus un point critique, donc au plus un point où L atteint sont minimum (Si un tel minimum existe).

19. (a) Supposons d'abord toutes les composantes q_i de q non nulles.
On suppose que L a un minimum global atteint en θ_*. Le gradient de L en θ_* est alors nul. Donc $M^T p(\theta_*)^T = q^T M$, et $p(\theta_*) \in \Sigma_N(\bar{g}, g)$.

D'autre part

$$H_N(p(\theta_*)) = -\frac{1}{Z(\theta_*)}\sum_{i=1}^{N}(f_i(\theta_*) - \ln(Z(\theta_*)))\,e^{f_i(\theta_*)} = \ln(Z(\theta_*)) - E(f(\theta_*)) = \ln(Z(\theta_*)) - \sum_{i=1}^{N} q_i f_i(\theta_*).$$

Donc

$$H_N(p(\theta_*)) - H_N(q) = \ln(Z(\theta_*) - \sum_{i=1}^{N} q_i \ln\left(\frac{e^{f_i(\theta_*)}}{q_i}\right).$$

Or ln est concave (hors programme PC), donc :

$$\sum_{i=1}^{N} q_i \ln\left(\frac{e^{f_i(\theta_*)}}{q_i}\right) \geqslant \ln\left(\sum_{i=1}^{N} e^{f_i(\theta_*)}\right) = \ln(Z(\theta_*)).$$

Donc :

$$H_N(p(\theta^*)) \geqslant H_N(q).$$

Si certaines composantes q_i sont nulles, effectue le même calcul en gardant les composantes non nulles. On obtient alors :

$$H_N(p(\theta_*)) - H_N(q) = \ln(Z(\theta_*) - \sum_{i\in[|1,N|],q_i\neq 0} q_i \ln\left(\frac{e^{f_i(\theta_*)}}{q_i}\right).$$

Et :

$$\sum_{i\in[|1,N|],q_i\neq 0} q_i \ln\left(\frac{e^{f_i(\theta_*)}}{q_i}\right) \leqslant \ln\left(\sum_{i\in[|1,N|],q_i\neq 0} e^{f_i(\theta_*}\right) < \ln(Z(\theta_*).$$

Donc cette fois-ci :

$$H_N(p(\theta_*)) > H_N(q).$$

On obtient cette fois-ci une inégalité stricte. On peut remplacer dans la définition de $\Sigma_N(g, \bar{g})$, l'élément q par n'importe quel élément $q' \in \Sigma_N(g, \bar{g})$. L'inégalité vaut donc pour tout $q' \in \Sigma_N(g, \bar{g})$ et $H_N(p(\theta_*))$ est donc bien la valeur maximale de H_N sur $\Sigma_N(g, \bar{g})$. Remarquons que cette valeur maximale ne peut être atteinte qu'en un $n - uplet$ dont les composantes sont strictement positives.

(b) Réciproquement, supposons H_N atteint son maximum en un point p de $\Sigma_N(g,\bar{g})$ Remarquons d'abord que toutes les composantes de p (où H_N atteint son maximum) sont strictement positives puisque l'on a vu à la question précédente que c'est une condition nécessaire pour que H_N ait un maximum en p.

Remarquons ensuite que pour tout p de Σ_N, on a :

$$p \in \Sigma_N(g,\bar{g}) \iff M^T p = qM^T,$$

$$p \in \Sigma_N(g,\bar{g}) \iff M^T(p-q) = 0 \iff (p-q) \in \ker(M^T).$$

En considérant la somme des composantes de $p-q$, qui est nulle, on voit que les éléments $\Sigma_N(g,\bar{g})$ sont les éléments sont les éléments $q+x$, où $x \in \ker \tilde{M}^T$, qui ont leurs composantes positives ou nulles.

On en déduit aisément que $\Sigma_N(g,\bar{g})$ est convexe.

Si H_N admet un maximum en deux points p et p' distincts de $\Sigma_N(g,\bar{g})$, alors, pour tout $t \in [0,1] tp + (1-t)p' \in \Sigma_N(g,\bar{g})$. Comme p et p' sont à composantes non nulles, cette cette propriété est même encore vraie sur un certain intervalle $[-\epsilon, 1+\epsilon]$, avec $\epsilon > 0$ (On choisit ϵ pour que toutes les composantes des N-uplets restent strictement positives.) Donc $w : t \mapsto H_N(tp + (1-t)p')$ est définie et de classe C^1 sur $[-\epsilon, 1+\epsilon]$ et

$$\forall t \in [0,1]\ w'(t) = -\sum_{i=1}^{N} (p_i - p'_i) \ln(tp_i + (1-t)p'_i).$$

w admet un extrêmum en 0 et 1, donc $w'(0) = w'(1) = 0$. Mais alors :

$$w'(0) - w'(1) = -\sum_{i=1}^{N} (p_i - p'_i)(\ln(p_i) - \ln(p'_i)) = 0$$

Or tous les termes de cette somme sont négatifs ou nuls : ils sont donc tous nuls et donc

$$\forall i \in [|1,N|] p_i = p'_i.$$

On a donc bien montré l'unicité demandée.