UPS **avril 2016**
Corrigé de l'épreuve de mathématiques du concours X-ESPCI-ENS

Question 1. La fonction φ est de classe $\mathcal{C}^\infty$ sur $]0,+\infty[$ car c'est le produit des fonctions $t \mapsto -t$ et $t \mapsto \ln(t)$, qui le sont.

De plus, on sait par les croissances comparées que $t\ln(t)$ tend vers 0 quand t tend vers 0 donc φ est continue en 0. Finalement, la fonction φ est continue sur $[0,+\infty[$.

Pour tout $t>0$, on trouve $\varphi'(t) = -\ln(t)-1$ donc $\varphi'(t)$ tend vers $+\infty$ quand t tend vers 0 dans $]0,+\infty[$.

Question 2. Pour tout i dans $[\![1,\mathrm{N}]\!]$, la fonction coordonnée $p \mapsto p_i$, définie de $\mathbb{R}^\mathrm{N}$ vers $\mathbb{R}$, est continue, donc l'ensemble

$$\left\{p \in \mathbb{R}^\mathrm{N}\ ;\ p_i \geqslant 0\right\}$$

est fermé.

De même, la fonction $p \mapsto p_1 + \cdots + p_n$ est continue donc l'ensemble

$$\left\{p \in \mathbb{R}^\mathrm{N}\ ;\ \sum_{i=1}^{\mathrm{N}} p_i = 1\right\}$$

est fermé.

L'ensemble Σ_N est l'intersection de tous ces ensembles donc c'est un fermé de $\mathbb{R}^\mathrm{N}$.

Soit $p \in \Sigma_\mathrm{N}$. Déjà toutes ses coordonnées sont positives. De plus, pour tout indice i entre 1 et N, on trouve

$$p_i = 1 - \sum_{\substack{1\leqslant j\leqslant \mathrm{N}\\ j\neq i}} p_j \leqslant 1.$$

L'ensemble Σ_N est donc inclus dans $[0,1]^\mathrm{N}$, ce qui prouve qu'il est borné.

Question 3. La fonction φ est continue sur $[0,1]$ donc la fonction

$$(p_1,\ldots,p_\mathrm{N}) \mapsto \sum_{i=1}^{\mathrm{N}} \varphi(p_i)$$

est continue sur $[0,1]^\mathrm{N}$. L'ensemble Σ_N est inclus dans $[0,1]^\mathrm{N}$ et H_N est la restriction de cette fonction à Σ_N donc H_N est continue sur Σ_N.

La fonction ln est négative sur $]0,1]$ donc φ est positive sur $]0,1]$. Elle est nulle en 0 donc elle est positive sur $[0,1]$. On en déduit que la fonction

$$(p_1,\ldots,p_\mathrm{N}) \mapsto \sum_{i=1}^{\mathrm{N}} \varphi(p_i)$$

est positive sur $[0,1]^\mathrm{N}$. Par restriction, la fonction H_N est positive sur Σ_N.

Le N-uplet $p = \dfrac{1}{\mathrm{N}}(1,\ldots,1)$ est bien un élément de Σ_N et on trouve

$$\mathrm{H}_\mathrm{N}(p) = \sum_{i=1}^{\mathrm{N}} \varphi(1/\mathrm{N}) = \mathrm{N} \times \left(-\frac{1}{\mathrm{N}} \ln\left(\frac{1}{\mathrm{N}}\right)\right) = \ln(\mathrm{N}).$$

Question 4.a. On peut définir sur $[0, b]$ la fonction $f : t \mapsto \varphi(a+t) - \varphi(b-t)$. Cette fonction est continue sur $[0, b]$ et dérivable sur $]0, b[$, avec

$$\forall t \in]0, b[, \qquad f'(t) = \varphi'(a+t) - \varphi'(b-t) = -\ln(a+t) + \ln(b-t).$$

Prenons $\varepsilon = \dfrac{b-a}{3}$. On obtient alors pour tout t dans $]0, \varepsilon]$,

$$(b-t) - (a+t) = (b-a) - 2t \geqslant (b-a) - 2\varepsilon > 0 \qquad \text{donc} \qquad \ln(b-t) > \ln(a+t)$$

donc $f'(t) > 0$. La fonction f est donc strictement croissante sur $]0, \varepsilon]$ et même sur $[0, \varepsilon]$ par continuité en 0. En particulier, pour tout t dans $]0, \varepsilon]$, on a $f(t) > f(0)$, c'est-à-dire

$$\varphi(a+t) + \varphi(b-t) > \varphi(a) + \varphi(b).$$

Question 4.b. La fonction $\mathrm{H_N}$ est continue sur Σ_N, qui est fermé et borné, donc elle admet un maximum sur cet ensemble.

Soit p un élément de Σ_N où $\mathrm{H_N}$ atteint son maximum. Raisonnons par l'absurde et supposons que p soit différent de $(1/\mathrm{N}, \ldots, 1/\mathrm{N})$. Il existe alors des indices i et j distincts dans $[\![1, \mathrm{N}]\!]$ tels que $p_i < p_j$.

D'après 4.a, il existe $\varepsilon > 0$ tel que pour tout t dans $]0, \varepsilon]$,

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

Notons q l'élément de $\mathbb{R}^\mathrm{N}$ obtenu à partir de p en remplaçant p_i par $p_i + \varepsilon$ et p_j par $p_j - \varepsilon$. Cet élément est alors de nouveau dans Σ_N et on trouve

$$\mathrm{H_N}(q) - \mathrm{H_N}(p) = (\varphi(p_i + \varepsilon) + \varphi(p_j - \varepsilon)) - (\varphi(p_i) + \varphi(p_j)) > 0,$$

ce qui contredit le fait que $\mathrm{H_N}$ soit maximale en p.

On a prouvé par l'absurde que le maximum de $\mathrm{H_N}$ ne peut être atteint qu'en $(1/\mathrm{N}, \ldots, 1/\mathrm{N})$. Comme ce maximum existe, il est effectivement atteint en ce point.

Question 5.a. La suite p est bien un élément de H_∞ : c'est la loi géométrique de paramètre a. Considérons une variable aléatoire X suivant cette loi. On trouve alors, en appliquant la formule du transfert,

$$\begin{aligned}\mathrm{H}_\infty(p) &= -\sum_{i=1}^{+\infty} p_i \ln(p_i) = -\sum_{i=1}^{+\infty} \mathbb{P}(\mathrm{X} = i)(\ln(a) + (i-1)\ln(1-a)) = -\mathbb{E}(\ln(a) + (\mathrm{X}-1)\ln(1-a)) \\ &= -\ln(a) - \ln(1-a)\left(\frac{1}{a} - 1\right).\end{aligned}$$

Notons $g(a)$ cette valeur. La fonction g est dérivable sur $]0, 1[$ et une dérivation donne

$$\forall a \in]0, 1[, \qquad g'(a) = -\frac{1}{a} + \frac{1}{1-a} \times \frac{1-a}{a} + \frac{\ln(1-a)}{a^2} = \frac{\ln(1-a)}{a^2} < 0.$$

La fonction g est strictement décroissante sur $]0, 1[$. Ses limites en 0 et en 1 sont respectivement $+\infty$ et 0.

Note du rédacteur. Le but de cette question est probablement de montrer que l'entropie n'a pas d'extremum sur Σ_∞. Il est toutefois possible de prouver qu'en restriction à l'ensemble des lois qui ont la même espérance que $\mathcal{G}(a)$, l'entropie est maximale pour la loi géométrique et uniquement pour elle.

Question 5.b. Notons $\alpha = \sum\limits_{n=2}^{+\infty} \frac{1}{n\ln(n)^2}$. On peut alors poser

$$p_1 = 0 \qquad \text{et} \qquad \forall i \geqslant 2, \quad p_i = \frac{1/\alpha}{i\ln(i)^2}.$$

La suite p ainsi définie est un élément de Σ_∞. De plus, pour tout $i \geqslant 2$, on a

$$\ln(p_i) = -\ln(\alpha) - \ln(i) - 2\ln(\ln(i))$$

donc

$$\varphi(p_i) \sim_{i\to+\infty} \frac{1/\alpha}{i\ln(i)}.$$

On en déduit que la série $\sum \varphi(p_i)$ est divergente. C'est une série à termes positifs donc sa somme est infinie [1].

Ce choix de p donne donc $\mathrm{H}_\infty(p) = +\infty$.

Question 6. Remarquons l'égalité

$$\frac{1}{n}\ln\left(\prod_{k=1}^{n} p_{\mathrm{X}_k}\right) = \frac{1}{n}\sum_{k=1}^{n}\ln(p_{\mathrm{X}_k}).$$

Notons $\mathrm{Y}_k = \ln(p_{\mathrm{X}_k})$. Pour tout k dans $[\![1, n]\!]$, la variable aléatoire Y_k a une espérance donnée par la formule du transfert

$$\mathbb{E}(\mathrm{Y}_k) = \sum_{i=1}^{\mathrm{N}}\ln(p_i)\mathbb{P}(\mathrm{X}_k = i) = \sum_{i=1}^{\mathrm{N}}\ln(p_i)p_i = -\mathrm{H}_\mathrm{N}(p).$$

Les variables aléatoires $\mathrm{X}_1, \ldots, \mathrm{X}_n$ sont deux à deux indépendantes donc $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$ le sont aussi. La loi faible des grands nombres donne que pour tout $\varepsilon > 0$, la probabilité

$$\mathbb{P}\left(\left|\frac{\mathrm{Y}_1 + \cdots + \mathrm{Y}_n}{n} - \mathbb{E}(\mathrm{Y}_1)\right| \geqslant \varepsilon\right)$$

tend vers 0 quand n tend vers $+\infty$. Cette probabilité se réécrit

$$\mathbb{P}\left(\left|\frac{1}{n}\ln\left(\prod_{k=1}^{n} p_{\mathrm{X}_k}\right) + \mathrm{H}_\mathrm{N}(p)\right| \geqslant \varepsilon\right).$$

Cette dernière expression tend donc vers 0 quand n tend vers $+\infty$.

Question 7. La fonction H_N est continue et la fonction $p \mapsto \sum\limits_{i=1}^{\mathrm{N}} p_i f_i$ est linéaire donc continue. Ainsi, la fonction J_f est continue. L'ensemble Σ_N est fermé et borné donc la borne supérieure de J_f sur Σ_N est atteinte. Autrement dit, l'ensemble $\Sigma_\mathrm{N}(f)$ est non vide.

Question 8.a. On reprend la méthode de la question 4.a. On définit sur $[0, p_2]$ la fonction

$$h : t \mapsto \mathrm{J}_f(t, p_2 - t, p_3, \ldots, p_\mathrm{N}) = \varphi(t) + \varphi(p_2 - t) + f_2(p_2 - t) + \sum_{i=3}^{\mathrm{N}}(\varphi(p_i) + f_i p_i).$$

La fonction h est continue sur $[0, p_2]$ et dérivable sur $]0, p_2[$, avec

$$\forall t \in]0, p_2[, \qquad h'(t) = \varphi'(t) - \varphi'(p_2 - t) = -\ln(t) + \ln(p_2 - t) - f_2.$$

On observe que $h'(t)$ tend vers $+\infty$ quand t tend vers 0. En particulier, il existe ε dans $]0, p_2[$ tel que

$$\forall t \in]0, \varepsilon], \qquad h'(t) > 0.$$

1. La notion de somme infinie en cas de divergence n'est en fait pas au programme de PC. Il était donc nécessaire pour les candidats d'extrapoler une réponse à cette question.

La fonction h est donc strictement croissante sur $]0, \varepsilon]$ et même sur $[0, \varepsilon]$ par continuité en 0. En particulier, on a l'inégalité $h(\varepsilon) > h(0)$, qui s'écrit aussi

$$\mathrm{J}_f(p') > \mathrm{J}_f(p), \qquad \text{avec} \qquad p' = (\varepsilon, p_2 - \varepsilon, p_3, \ldots, p_\mathrm{N})$$

et p' est bien un élément de Σ_N.

Question 8.b. On peut faire le même raisonnement qu'en 8.a en remplaçant les indices 1 et 2 par n'importe quels indices i et j distincts compris entre 1 et N. On voit alors que si p possède au moins un coefficient nul, alors $\mathrm{J}_f(p)$ n'est pas le maximum de la fonction J_f.

Par contraposée, si p est un élément de $\Sigma_\mathrm{N}(f)$, alors tous ses coefficients sont non nuls, si bien qu'ils sont strictement positifs.

Question 9.a. Notons v le vecteur $(1, \ldots, 1)$ de $\mathbb{R}^\mathrm{N}$. L'ensemble E_0 se réécrit

$$\mathrm{E}_0 = \{a \in \mathbb{R}^\mathrm{N} \ ; \ \langle v, a \rangle = 0\}.$$

Autrement dit, l'ensemble E_0 s'écrit $\mathrm{Vect}(v)^\perp$. En particulier, c'est un sous-espace vectoriel de $\mathbb{R}^\mathrm{N}$ et sa dimension vaut

$$\dim(\mathrm{E}_0) = \dim(\mathbb{R}^\mathrm{N}) - \dim(\mathrm{Vect}(v)) = \mathrm{N} - 1.$$

Enfin, l'orthogonal de E_0 est la droite vectorielle $\mathrm{Vect}(v)$.

Question 9.b. Soit $t \in \mathbb{R}$.

$$\tilde{p}(t) = (p_1 + ta_1, \ldots, p_\mathrm{N} + ta_\mathrm{N}).$$

Déjà, la somme des coefficients de $\tilde{p}(t)$ vaut 1

$$\sum_{i=1}^{\mathrm{N}} (p_i + ta_i) = \sum_{i=1}^{\mathrm{N}} p_i + t \sum_{i=1}^{\mathrm{N}} a_i = 1 + t \times 0 = 1.$$

Il s'agit maintenant de garantir que tous ses coefficients soient positifs. Prenons un indice i dans $[\![1, \mathrm{N}]\!]$. La fonction $t \mapsto p_i + ta_i$ est continue et sa valeur en 0 est strictement positive donc il existe $\varepsilon_i > 0$ tel que pour tout t dans $]-\varepsilon_i, \varepsilon_i[$, le nombre $p_i + ta_i$ soit strictement positif.

En prenant $\varepsilon = \min(\varepsilon_1, \ldots, \varepsilon_\mathrm{N})$, on obtient un nombre $\varepsilon > 0$ tel que pour tout t dans $]-\varepsilon, \varepsilon[$, tous les coefficients de $\tilde{p}(t)$ soient strictement positifs. Ainsi, pour tout t dans cet intervalle, le vecteur $\tilde{p}(t)$ est un élément de Σ_N.

La dérivée de $\tilde{p}$ est la fonction constante, égale au vecteur a. Le vecteur $\tilde{p}'(0)$ est donc le vecteur a.

Remarque. On peut bien sûr être bien plus explicite dans le choix des ε_i. En l'occurrence, si a_i vaut 0, on peut prendre n'importe quelle valeur, par exemple $\varepsilon_i = 1$; si a_i n'est pas nul, on peut choisir $\varepsilon_i = |p_i/a_i|$.

Question 9.c. On peut définir de $]-\varepsilon, \varepsilon[$ dans $\mathbb{R}$ la fonction

$$h : t \mapsto \mathrm{J}_f(\tilde{p}(t)).$$

Cette fonction est donnée par

$$\forall t \in]-\varepsilon, \varepsilon[, \qquad h(t) = -\sum_{i=1}^{\mathrm{N}} (p_i + ta_i) \ln(p_i + ta_i) + \sum_{i=1}^{\mathrm{N}} (p_i + ta_i) f_i.$$

En particulier, la fonction h est dérivable. Cette fonction admet un maximum en 0 et l'intervalle $]-\varepsilon,\varepsilon[$ est ouvert donc $h'(0)$ est nul. Le calcul donne

$$\forall t \in\,]-\varepsilon,\varepsilon[, \qquad h'(t) = -\sum_{i=1}^{\mathrm{N}} a_i \ln(p_i + ta_i) - \underbrace{\sum_{i=1}^{\mathrm{N}} a_i}_{=0} + \sum_{i=1}^{\mathrm{N}} a_i f_i$$

donc, en particulier,

$$h'(0) = -\sum_{i=1}^{\mathrm{N}} a_i \ln(p_i) + \sum_{i=1}^{\mathrm{N}} a_i f_i$$

et on a dit plus haut que cette somme est nulle. On en déduit que le vecteur $(f_i - \ln(p_i))_{1\leqslant i\leqslant \mathrm{N}}$ est orthogonal au vecteur a. Ceci a été obtenu pour un vecteur a quelconque de E_0 donc le vecteur $(f_i - \ln(p_i))_{1\leqslant i\leqslant \mathrm{N}}$ est dans $\mathrm{E}_0^{\perp}$. Il est donc proportionnel au vecteur $v = (1,\ldots,1)$. En notant c le nombre tel que ce vecteur soit égal à $-cv$, il vient

$$\forall i \in [\![1,\mathrm{N}]\!], \qquad \ln(p_i) = f_i + c.$$

Question 10. Soit p un élément de $\Sigma_{\mathrm{N}}(f)$. Comme on l'a vu, il existe une constante c telle que

$$\forall i \in [\![1,\mathrm{N}]\!], \qquad \ln(p_i) = f_i + c.$$

On obtient donc

$$\forall i \in [\![1,\mathrm{N}]\!], \qquad p_i = \mathrm{e}^{f_i + c}$$

puis

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

Cette somme vaut 1 donc c vaut

$$c = -\ln\left(\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{f_i}\right),$$

ce qui donne

$$\forall k \in [\![1,\mathrm{N}]\!], \qquad p_k = \frac{\mathrm{e}^{f_k}}{\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{f_i}}.$$

Il y a donc au plus un élément dans $\Sigma_{\mathrm{N}}(f)$. On a prouvé à la question 7 qu'il y en a au moins un donc la formule ci-dessus donne l'unique élément de cet ensemble.

Un dernier calcul donne

$$\mathrm{J}_{f,*} = -\sum_{i=1}^{\mathrm{N}} p_i \ln(p_i) + \sum_{i=1}^{\mathrm{N}} p_i f_i = -\sum_{i=1}^{\mathrm{N}} p_i c = -c = \ln\left(\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{f_i}\right).$$

Question 11. La fonction $\beta \mapsto \sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i}$ est dérivable, à valeurs strictement positives, donc son logarithme est dérivable et on en déduit que la fonction F est dérivable. On trouve pour tout $\beta > 0$,

$$\begin{aligned}
\mathrm{F}'(\beta) &= -\frac{1}{\beta^2}\ln\left(\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i}\right) + \frac{1}{\beta}\frac{\sum_{i=1}^{\mathrm{N}} f_i \mathrm{e}^{\beta f_i}}{\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i}} \\
&= -\frac{1}{\beta^2}\left(\ln\left(\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i}\right) - \frac{\sum_{i=1}^{\mathrm{N}} \beta f_i \mathrm{e}^{\beta f_i}}{\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i}}\right).
\end{aligned}$$

Considérons le vecteur $p(\beta) = \dfrac{1}{\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i}}(\mathrm{e}^{\beta f_1}, \ldots, \mathrm{e}^{\beta f_{\mathrm{N}}})$. C'est l'unique élément de $\Sigma_{\mathrm{N}}(\beta f)$ comme on l'a vu à la question 10, et on obtient

$$\mathrm{F}'(\beta) = -\frac{1}{\beta^2}\left(\mathrm{J}_{f\beta}(p(\beta)) - \sum_{i=1}^{\mathrm{N}} \beta f_i p(\beta)_i\right) = -\frac{1}{\beta^2}\mathrm{H}_{\mathrm{N}}(p(\beta)).$$

Question 12. Quand β tend vers 0, le facteur $\ln\left(\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i}\right)$ tend vers $\ln(\mathrm{N})$ donc $\mathrm{F}(\beta)$ tend vers $+\infty$.

Pour la limite en $+\infty$, c'est plus délicat. Notons f_+ le plus grand des coefficients $f_1, \ldots, f_n$. On obtient alors l'encadrement

$$\mathrm{e}^{\beta f_+} \leqslant \sum_{i=1}^{\mathrm{N}} \mathrm{e}^{\beta f_i} \leqslant \mathrm{N}\mathrm{e}^{\beta f_+}$$

puis

$$f_+ \leqslant \mathrm{F}(\beta) \leqslant f_+ + \frac{\ln(\mathrm{N})}{\beta}.$$

Par le théorème des gendarmes, on en déduit que F tend vers f_+ en $+\infty$.

Remarque du rédacteur. Je ne comprends pas ce que viennent faire les questions 11 et 12 dans l'histoire.

Question 13. Soit (l, k) un couple d'indices. On remarque l'égalité

$$\mathrm{A}_{l,k} = \sum_{i=1}^{\mathrm{N}} \mathbb{P}(\mathrm{Y} = i)(g_l(i) - m_l)(g_k(i) - m_k),$$

ce qui donne $\mathrm{A}_{l,k} = \mathbb{E}((g_l(\mathrm{Y}) - m_l)(g_k(\mathrm{Y}) - m_k))$ par la formule de transfert.

Cette formule donne[2] en particulier $\mathrm{A}_{l,k} = \mathrm{A}_{k,l}$. C'est vrai pour tout couple d'indices donc la matrice A est symétrique.

Prenons un élément θ de $\mathbb{R}^d$. Un premier calcul matriciel donne

$$\mathrm{A}\theta = \left(\sum_{k=1}^{d} \mathrm{A}_{l,k}\theta_k\right)_{1\leqslant l\leqslant d}.$$

Un deuxième produit matriciel donne ensuite

$$\theta^{\mathrm{T}}\mathrm{A}\theta = \sum_{l=1}^{d}\sum_{k=1}^{d} \mathrm{A}_{l,k}\theta_k\theta_l.$$

Utilisons l'expression des $\mathrm{A}_{l,k}$ obtenue précédemment et exploitons la linéarité de l'espérance

$$\theta^{\mathrm{T}}\mathrm{A}\theta = \mathbb{E}\left(\sum_{l=1}^{d}\sum_{k=1}^{d}(g_l(\mathrm{Y}) - m_l)(g_k(\mathrm{Y}) - m_k)\theta_k\theta_l\right).$$

Introduisons la variable aléatoire $\mathrm{Z} = \sum_{k=1}^{d} \theta_k(g_k(\mathrm{Y}) - m_k)$. La formule précédente se réécrit

$$\theta^{\mathrm{T}}\mathrm{A}\theta = \mathbb{E}(\mathrm{Z}^2).$$

La variable aléatoire Z^2 est positive donc son espérance est positive, c'est-à-dire $\theta^{\mathrm{T}}\mathrm{A}\theta \geqslant 0$.

Question 14.a. On reprend la notation Z introduite à la question précédente.

2. La définition du coefficient $\mathrm{A}_{l,k}$ donne cela aussi.

La variable aléatoire Z^2 est positive et son espérance est nulle donc la variable aléatoire Z^2 a une loi dégénérée en 0, ce qui s'écrit $\mathbb{P}(\mathrm{Z}^2 = 0) = 1$. La variable aléatoire Z a donc également cette loi.

Appliquons la formule des probabilités totales avec le système complet d'événements $([\mathrm{Y} = i])_{1\leqslant i\leqslant \mathrm{N}}$

$$\mathbb{P}(\mathrm{Z} = 0) = \sum_{i=1}^{\mathrm{N}} \mathbb{P}(\mathrm{Y} = i)\mathbb{P}(\mathrm{Z} = 0|\mathrm{Y} = i) = \sum_{i=1}^{\mathrm{N}} p_i \mathbb{P}\left(\sum_{k=1}^{d} \theta_k(g_k(i) - m_k) = 0 \middle| \mathrm{Y} = i\right).$$

L'événement $\left[\sum\limits_{k=1}^{d} \theta_k(g_k(i) - m_k) = 0\right]$ n'est pas aléatoire : il est vrai ou faux. S'il existe un indice j dans $[\![1, \mathrm{N}]\!]$ pour lequel cette égalité est fausse, alors on obtient

$$\mathbb{P}(\mathrm{Z} = 0) \leqslant \sum_{\substack{1\leqslant i\leqslant \mathrm{N}\\ i\neq j}} p_i = 1 - p_j < 1,$$

mais c'est impossible puisque $\mathbb{P}(\mathrm{Z} = 0)$ vaut 1. Ainsi, pour tout i dans $[\![1, \mathrm{N}]\!]$, on obtient

$$\sum_{k=1}^{d} \theta_k g_k(i) = \sum_{k=1}^{d} \theta_k m_k, \qquad \text{c'est-à-dire} \qquad \sum_{k=1}^{d} \theta_k \mathrm{M}_{i,k} = c,$$

en posant $c = \sum\limits_{k=1}^{d} \theta_k m_k$, ce qui est bien un nombre réel indépendant de i.

Autre démonstration. Sans passer par l'espérance, on obtient l'expression

$$\theta^{\mathrm{T}} \mathrm{A} \theta = \sum_{i=1}^{\mathrm{N}} p_i \left(\sum_{l=1}^{d} (\mathrm{M}_{i,l} - m_l)\theta_l\right)^2.$$

On en déduit une autre démonstration de la positivité de la question 13. Dans le cadre de la question 14, nous avons là une somme nulle de termes positifs donc tous ses termes sont nuls. Les p_i étant tous strictement positifs, il reste

$$\forall i \in [\![1, \mathrm{N}]\!], \qquad \sum_{l=1}^{d} (\mathrm{M}_{i,l} - m_l)\theta_l = 0,$$

ce qui mène à la même conclusion que la démonstration précédente. Il semble qu'en ayant orienté ces calculs sur des pistes probabilistes, les auteurs de cet énoncé aient compliqué inutilement cette démonstration.

Question 14.b. La formule précédente donne l'appartenance de $(\theta_1, \ldots, \theta_d, -c)$ à $\mathrm{Ker}(\widetilde{\mathrm{M}})$. Ce noyau est trivial donc les θ_k sont tous nuls. Le vecteur θ est donc nul.

Question 15. Les f_i sont linéaires donc de classe $\mathcal{C}^1$. La fonction z est donc de classe $\mathcal{C}^1$. Elle est à valeurs strictement positives donc la fonction $\theta \mapsto \ln(\mathrm{Z}(\theta))$ est de classe $\mathcal{C}^1$.

La fonction $\theta \mapsto q^{\mathrm{T}}\mathrm{M}\theta$ est linéaire donc de classe $\mathcal{C}^1$.

Par somme, la fonction L est de classe $\mathcal{C}^1$.

Un calcul similaire à celui de la question 13 donne

$$q^{\mathrm{T}}\mathrm{M}\theta = \sum_{l=1}^{d}\sum_{k=1}^{d} \mathrm{M}_{l,k}\theta_k q_l.$$

Sa dérivée par rapport à la variable θ_j s'écrit donc $\sum\limits_{k=1}^{d} \mathrm{M}_{l,j} q_l$.

La fonction f_i admet l'expression $f_i(\theta) = \sum_{k=1}^{d} \mathrm{M}_{i,k}\theta_k$. On peut ainsi dériver la fonction Z

$$\forall \theta \in \mathbb{R}^d, \qquad \frac{\partial \mathrm{Z}}{\partial \theta_j}(\theta) = \sum_{i=1}^{\mathrm{N}} \frac{\partial f_i}{\partial \theta_j}(\theta)\mathrm{e}^{f_i(\theta)} = \sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,j}\mathrm{e}^{f_i(\theta)}.$$

On obtient ensuite

$$\forall \theta \in \mathbb{R}^d, \qquad \frac{\partial \mathrm{L}}{\partial \theta_j}(\theta) = \frac{\sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,j}\mathrm{e}^{f_i(\theta)}}{\sum_{i=1}^{\mathrm{N}} \mathrm{e}^{f_i(\theta)}} - \sum_{l=1}^{d} \mathrm{M}_{l,j} q_j = \sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,j}(p_i(\theta) - q_i).$$

Cette formule est valable pour tout $j \in [\![1, d]\!]$. On reconnaît alors les coefficients d'un produit matriciel. Plus précisément,

$$\nabla \mathrm{L}(\theta) = \mathrm{M}^{\mathrm{T}}(p(\theta) - q).$$

Question 16. On suppose que θ est un point critique de L. On obtient donc l'égalité $\mathrm{M}^{\mathrm{T}}(p(\theta) - q) = 0$, c'est-à-dire $\mathrm{M}^{\mathrm{T}}p(\theta) = \mathrm{M}^{\mathrm{T}}q$.

Par ailleurs, on sait déjà que le vecteur $p(\theta)$ appartient à Σ_{N}. Il reste à vérifier qu'il est dans $\Sigma_{\mathrm{N}}(\overline{g}, g)$. Prenons un indice j dans $[\![1, d]\!]$. On a vu à la question précédente l'égalité

$$\sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,j} p_i(\theta) = \sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,j} q_i.$$

Celle-ci se réécrit

$$\sum_{i=1}^{\mathrm{N}} g_j(i) p_i(\theta) = \sum_{i=1}^{\mathrm{N}} g_j(i) q_i = \overline{g_j}$$

en utilisant le fait que q est dans $\Sigma_{\mathrm{N}}(\overline{g}, g)$. C'est vrai pour tout indice j donc le vecteur $p(\theta)$ est dans $\Sigma_{\mathrm{N}}(\overline{g}, g)$.

Question 17. Les dérivées partielles de L sont de classe $\mathcal{C}^1$ par les théorèmes généraux de stabilité. Reprenons l'expression déjà obtenue

$$\forall \theta \in \mathbb{R}^d, \qquad \frac{\partial \mathrm{L}}{\partial \theta_k}(\theta) = \sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,k}(p_i(\theta) - q_i).$$

Il s'agit maintenant de calculer les dérivées partielles de p_i. Pour cela, on part de l'expression $p_i(\theta) = \mathrm{e}^{f_i(\theta)}/\mathrm{Z}(\theta)$. On obtient pour tout θ dans $\mathbb{R}^d$

$$\begin{aligned} \frac{\partial p_i}{\partial \theta_l}(\theta) &= \frac{\partial f_i}{\partial \theta_l}(\theta)\frac{\mathrm{e}^{f_i(\theta)}}{\mathrm{Z}(\theta)} - \frac{\partial \mathrm{Z}}{\partial \theta_l}(\theta)\frac{\mathrm{e}^{f_i(\theta)}}{\mathrm{Z}(\theta)^2} \\ &= \mathrm{M}_{i,l} p_i(\theta) - \left(\sum_{j=1}^{\mathrm{N}} \mathrm{M}_{j,l}\mathrm{e}^{f_i(\theta)}\right)\frac{p_i(\theta)}{\mathrm{Z}(\theta)} \\ &= \left(\mathrm{M}_{i,l} - \sum_{j=1}^{\mathrm{N}} \mathrm{M}_{j,l} p_j(\theta)\right) p_i(\theta). \end{aligned}$$

La somme $\sum_{j=1}^{\mathrm{N}} \mathrm{M}_{j,l} p_j(\theta)$ est le l-ième coefficient de $\mathrm{M}^{\mathrm{T}}p(\theta)$, c'est-à-dire $m_l(\theta)$. On obtient donc

$$\frac{\partial p_i}{\partial \theta_l}(\theta) = (\mathrm{M}_{i,l} - m_l(\theta))p_i(\theta).$$

Pour tout θ de $\mathbb{R}^d$, on en déduit l'égalité

$$\frac{\partial^2 \mathrm{L}}{\partial\theta_l\partial\theta_k}(\theta) = \sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,k}\frac{\partial p_i}{\partial\theta_l}(\theta) = \sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,j}(\mathrm{M}_{i,l} - m_l(\theta))p_i(\theta).$$

Pour obtenir la formule de l'énoncé, il reste à prouver que les termes manquants ont une somme nulle.

$$\sum_{i=1}^{\mathrm{N}} m_k(\theta)(\mathrm{M}_{i,l} - m_l(\theta))p_i(\theta) = m_k(\theta)\left(\underbrace{\sum_{i=1}^{\mathrm{N}} \mathrm{M}_{i,l}p_i(\theta)}_{=m_l(\theta)} - m_l(\theta)\underbrace{\sum_{i=1}^{\mathrm{N}} p_i(\theta)}_{=1}\right) = 0.$$

On peut donc enfin conclure à la formule attendue

$$\frac{\partial^2 \mathrm{L}}{\partial\theta_l\theta_k}(\theta) = \sum_{i=1}^{\mathrm{N}}(\mathrm{M}_{i,k} - m_k(\theta))(\mathrm{M}_{i,l} - m_l(\theta))p_i(\theta).$$

Question 18.a. On définit de $[0,1]$ dans $\mathbb{R}$ la fonction

$$\alpha : t \mapsto \mathrm{L}(t\theta + (1-t)\theta').$$

La règle de la chaîne donne

$$\forall t \in [0,1], \qquad \alpha'(t) = \sum_{k=1}^{d}(\theta_k - \theta'_k)\frac{\partial \mathrm{L}}{\partial\theta_k}(t\theta + (1-t)\theta')$$

puis, pour tout $t \in [0,1]$,

$$\begin{aligned}\alpha''(t) &= \sum_{k=1}^{d}\sum_{l=1}^{d}(\theta_k - \theta'_k)(\theta_l - \theta'_l)\frac{\partial^2 \mathrm{L}}{\partial\theta_l\partial\theta_k}(t\theta + (1-t)\theta') \\ &= \sum_{k=1}^{d}\sum_{l=1}^{d}(\theta_k - \theta'_k)(\theta_l - \theta'_l)\sum_{i=1}^{\mathrm{N}} p_i(t\theta + (1-t)\theta')(\mathrm{M}_{i,l} - m_l(t\theta) + (1-t)\theta')(\mathrm{M}_{i,k} - m_k(t\theta + (1-t)\theta')).\end{aligned}$$

On reconnaît alors l'expression $\alpha''(t) = (\theta-\theta')^{\mathrm{T}}\mathrm{A}(t)(\theta-\theta')$, où $\mathrm{A}(t)$ désigne la matrice A du préambule de la partie III, dans laquelle le vecteur m a été remplacé par $m(t\theta + (1-t)\theta')$.

Le vecteur $\theta - \theta'$ est non nul par hypothèse donc, d'après 13 et 14.b, on obtient $\alpha''(t) > 0$. C'est vrai pour tout t dans $[0,1]$ donc la fonction α' est strictement croissante sur $[0,1]$.

La première dérivation donne en particulier $\alpha'(1) = \sum\limits_{k=1}^{d}(\theta_k - \theta'_k)\dfrac{\partial \mathrm{L}}{\partial\theta_k}(\theta)$.

On a suppose que L admet un point critique en θ donc les dérivées partielles d'ordre 1 sont nulles en θ, donc $\alpha'(1)$ est nul.

Question 18.b. On suppose que L possède deux points critiques θ et θ' distincts. Reprenons la fonction α de la question précédente. On trouve alors

$$\alpha'(0) = \sum_{k=1}^{d}(\theta_k - \theta'_k)\frac{\partial \mathrm{L}}{\partial\theta_k}(\theta') = 0$$

par le même argument. L'égalité $\alpha'(0) = \alpha'(1)$ entre alors en contradiction avec la stricte croissance de α'.

Cette contradiction prouve que L possède au plus un point critique.

La fonction L est de classe $\mathcal{C}^1$ sur l'ouvert $\mathbb{R}^d$ donc les éventuels points où L admet un extremum sont des points critiques de L. Le nombre de points où la fonction L atteint son minimum vaut donc 1 ou 0.

Question 19.a. Le calcul de la question 10 donne $\mathrm{J}_{f(\theta_*),*} = \ln(\mathrm{Z}(\theta_*)$ puis

$$\mathrm{H_N}(p(\theta_*)) = \ln(\mathrm{Z}(\theta_*)) - \sum_{i=1}^{\mathrm{N}} f_i(\theta_*)p_i(\theta_*) = \ln(\mathrm{Z}(\theta_*)) - p(\theta_*)^{\mathrm{T}}\mathrm{M}p(\theta_*).$$

La question 16 donne $\mathrm{M}^{\mathrm{T}}p(\theta_*) = \mathrm{M}^{\mathrm{T}}q$ donc $p(\theta_*)^{\mathrm{T}}\mathrm{M} = q^{\mathrm{T}}\mathrm{M}$ donc

$$\mathrm{H_N}(p(\theta_*)) = \ln(\mathrm{Z}(\theta_*)) - q^{\mathrm{T}}\mathrm{M}p(\theta_*) = \mathrm{L}(\theta_*).$$

Par ailleurs, la question 10 donne $\mathrm{J}_{f(\theta_*)}(q) \leqslant \mathrm{J}_{f(\theta_*)}(p(\theta_*))$ or on connaît l'égalité $\mathrm{J}_{f(\theta_*)}(p(\theta_*)) = \ln(\mathrm{Z}(\theta_*))$ et on trouve

$$\mathrm{J}_{f(\theta_*)}(q) = \mathrm{H_N}(q) + \sum_{i=1}^{\mathrm{N}} q_i f_i(\theta) = \mathrm{H_N}(q) + q^{\mathrm{T}}\mathrm{M}f(\theta_*).$$

On en déduit l'inégalité $\ln(\mathrm{Z}(\theta_*)) \geqslant \mathrm{H_N}(q) + q^{\mathrm{T}}\mathrm{M}f(\theta_*)$ puis $\mathrm{L}(\theta_*) \geqslant \mathrm{H_N}(q)$, c'est-à-dire, en exploitant le calcul précédent,

$$\mathrm{H_N}(p(\theta_*)) \geqslant \mathrm{H_N}(q).$$

Il s'agit maintenant de comprendre que malgré ce que peut laisser penser la rédaction du préambule de la partie III, l'ensemble $\Sigma_{\mathrm{N}}(\overline{g}, g)$ ne dépend pas de q mais seulement des fonctions g_k et des nombres $\overline{g_k}$. Ainsi, l'inégalité ci-dessus est valable pour n'importe quel élément q de $\Sigma_{\mathrm{N}}(\overline{g}, g)$, si bien que $\mathrm{H_N}(p(\theta_*))$ est la valeur maximale de la fonction $\mathrm{H_N}$ en restriction à $\Sigma_{\mathrm{N}}(\overline{g}, g)$.

Question 19.b. Soit q un élément de $\Sigma_{\mathrm{N}}(\overline{g}, g)$ où $\mathrm{H_N}$ atteint son maximum. En reprenant les calculs de la question précédente, on arrive à l'égalité

$$\mathrm{J}_{f(\theta_*)}(q) = \mathrm{J}_{f(\theta_*)}(p(\theta_*)).$$

Le vecteur q est donc un élément de $\Sigma_{\mathrm{N}}(f(\theta_*))$. D'après l'unicité de la question 10, on en déduit que q est égal au vecteur $p(\theta_*)$.
