1 Maximum d'entropie

1.1 Exercice

Soit $n \in \mathbb{N}$. On pose $U = \left(\mathbb{R}_+^*\right)^n$. On munit $\Omega = [\![1, n]\!]$ de sa tribu discrète. On identifie une probabilité P sur Ω à $\mathbf{p} = \begin{pmatrix} P(1) \\ \vdots \\ P(n) \end{pmatrix}$. On note $\mathcal{P}$ l'ensemble $\{\mathbf{p}, \mathbf{p} \in U\}$. Soit $\varphi : [\![1, n]\!] \to \mathbb{R}$ non constante et $a \in]\min(\varphi), \max(\varphi)[$. Enfin on définit l'entropie :

$$\begin{aligned} H : \mathcal{P} &\to \mathbb{R} \\ H(\mathbf{p}) &= -\sum_{k=1}^{n} \mathbf{p}(k) \log(\mathbf{p}(k)) \end{aligned}$$

1 Montrer que la fonction f définie sur $\mathbb{R}$ par $f(x) = \sum\limits_{k=1}^{n} (\varphi(k) - a) \exp(x(\varphi(k) - a))$ est bijective.

2 Montrer que H atteint son maximum sur $M = \left\{\mathbf{p} \in \mathcal{P}, \sum\limits_{k=1}^{n} \varphi(k)\mathbf{p}(k) = a\right\}$ en un unique point $\mathbf{p_0}$.

Indication : on pourra considérer $\mathcal{X} = \left\{\mathbf{x} \in \mathbb{R}_+^n, \sum\limits_{k=1}^{n} \mathbf{x}(k) = 1, \sum\limits_{k=1}^{n} \varphi(k)\mathbf{x}(k) = a\right\}$.

3 Donner l'expression de $\mathbf{p_0}$ en fonction de ϕ, $f^{-1}(0)$.

1.2 Correction

1 On calcule la dérivée f et on obtient $\sum\limits_{k=1}^{n} (\varphi(k) - a)^2 \exp(x(\varphi(k) - a))$. Puisque la fonction φ n'est pas constante on obtient que la dérivée est strictement positive et donc la fonction est strictement croissante. Soit k_1 tel que $\varphi(k_1) = \min(\varphi)$ et k_2 tel que $\varphi(k_2) = \max(\varphi)$. On a $f_1(x) = (\varphi(k_1) - a) \exp(x(\varphi(k_1) - a))$ qui tend vers $-\infty$ en $-\infty$. De même f_2 tend vers $+\infty$ en $+\infty$ donc la fonction est bijective.

2 $\mathcal{X}$ est compacte. Donc en prolongeant $x \mapsto -x \log(x)$ par 0 en 0 on obtient un prolongement continu. On peut donc définir un prolongement continu de H sur $\mathcal{X}$. Ce prolongement de H admet un maximum. Montrons que ce maximum est un élément de M.
Supposons que $\phi(1) \geq a$ et $\phi(2) \leq a$. Soit $\mathbf{p_1} \in \mathcal{X}$ telle que $\mathbf{p_1}_1(1) = 0$. Il existe $t \in]0, 1]$ tel que $t\phi(1) + (1 - t)\phi(2) = 0$. On pose $\mathbf{p_2} \in \mathcal{X}$ telle que $\mathbf{p_2}(1) = t$, $\mathbf{p_2}(2) = 1 - t$ et zéro ailleurs. Toute combinaison convexe de $\mathbf{p_1}$ et $\mathbf{p_2}$ est dans $\mathcal{X}$. On note

$$\mathbf{p_u} = (1 - u)\mathbf{p_1} + u\mathbf{p_2}$$

En calculant la dérivée de $H(\mathbf{p_u})$ par rapport à u on obtient que H est croissante sur $[0, \epsilon]$ avec ϵ assez petit. Ainsi la distribution de probabilité d'entropie maximale n'est pas atteinte en $\mathbf{p}$ tel que $\mathbf{p}(1) = 0$. En reprenant ce raisonnement pour chacune des composantes on obtient que le maximum est atteint pour un élément de M. Montrons maintenant son unicité.
Soit deux probabilités maximales $\mathbf{p_1}$ et $\mathbf{p_2}$. La stricte concavité de $x \mapsto -x \log(x)$ implique que $\frac{\mathbf{p_1}(k)+\mathbf{p_2}(k)}{2} \log\left(\frac{\mathbf{p_1}(k)+\mathbf{p_2}(k)}{2}\right) > \mathbf{p_1}(k) \log(\mathbf{p_1}(k)) + \mathbf{p_2}(k) \log(\mathbf{p_2}(k))$ et donc $H(\frac{\mathbf{p_1}+\mathbf{p_2}}{2}) > \frac{1}{2}\left(H(\mathbf{p_1}) + H(\mathbf{p_2})\right)$.

3 Le théorème des extrema liés assure

$$\forall k \in \Omega,\ \log(\mathbf{p_0}(k)) + 1 = \alpha + \beta(\varphi(k) - a).$$

Donc $\forall k \in \Omega, \mathbf{p_0}(k) = \frac{\exp(\beta(\varphi(k)-a))}{\sum\limits_{k=1}^{n} \exp(\beta(\varphi(k)-a))}$. On a de plus,

$$\sum_{k=1}^{n} (\phi(k) - a) \exp(\beta(\phi(k) - a)) = 0.$$

Donc $\beta = f^{-1}(0)$ et on obtient l'expression de $\mathbf{p_0}$.

2 Inégalité d'Hadamard

2.1 Exercice

Soit $n \in \mathbb{N}$.

1 Montrer que $\det : (\mathbb{R}^n)^n \to \mathbb{R}$ définie par $\det((u_1,\dots,u_n))$ atteint son maximum sur $\mathcal{M} = \{(u_1,\dots,u_n), \forall k \in [\![1,n]\!], \|u_k\|_2 = 1\}$.

2 Soit $(u_1,\dots,u_n) \in \mathcal{M}$ tel que $\det((u_1,\dots,u_n))$ est maximal. Montrer que $\forall (i,j) \in [\![1,n]\!], i \neq j \Rightarrow \langle u_i, u_j\rangle = 0$.

3 En déduire l'inégalité d'Hadamard,

$$\forall M \in \mathcal{M}_n(\mathbb{R}),\ |\det(M)| \leq \|u_1\| \dots \|u_n\|.$$

Donner le cas d'égalité.

2.2 Correction

1 M est compacte et le déterminant est continu donc il atteint son maximum sur M.

2 Le maximum du déterminant est positif (en effet $\mathrm{Id} \in M$. Donc on peut supposer que $\det(u_1,\dots,u_n)) > 0$. On utilise le théorème des extrema liés avec $g_k(u) = \|u_k\|^2$ (gradients indépendants) et on obtient que

$$\forall k \in [\![1,n]\!],\ \frac{\partial \det}{\partial u_k}((u_1,\dots,u_n))(h) = \det(u_1,\dots,u_{k-1},h,u_{k+1},\dots,u_n) = \alpha\langle u_k, h\rangle.$$

En posant $h = u_k$ on obtient que $\alpha > 0$ puis en posant $h = u_l$ avec $l \neq k$ on a obtient les conditions d'orthogonalité. Dans ce cas $\det((u_1,\dots,u_n)) = 1$.

3 Soit M une matrice. Si une de ses colonnes est nulle le théorème est triviale. Sinon on divise chacune des colonnes de M par la norme de cette colonne. On obtient une matrice $M' \in \mathcal{M}$ qui vérifie $\det(M') \leq 1$ avec égalité si et seulement si les colonnes de M' et donc les colonnes de M sont orthogonales, c'est-à-dire, $M \in \mathcal{O}_n(\mathbb{R})$. Pour pouvoir passer à la valeur absolue, il faut reprendre l'étude précédente pour le minimum du déterminant. On obtient les mêmes conditions et donc $\det(M') \geq -1$.

3 Semi-continuité inférieure et topologie de Sorgenfrey

3.1 Exercice

Attention ! A part les deux premières questions, les questions de cet exercice sont indépendantes. Soit (X,τ) un espace topologique. On dit qu'une fonction f sur X à valeurs réelles est semi-continue inférieurement en $x_0 \in X$ si pour tout $\epsilon \in \mathbb{R}_+^*$ il existe $V \in \mathcal{V}(x_0)$ tel que $\forall x \in V,\ f(x) \geq f(x_0) - \epsilon$.

1 On définit l'épigraphe d'une fonction f, $\mathrm{Epi}(f) = \{(x,\lambda) \in X \times \mathbb{R},\ f(x) \leq \lambda\}$. Montrer que f est semi-continue inférieurement si et seulement si son épigraphe est fermé.

2 En déduire que si $(f_i)_{i\in I}$ est une famille de fonctions semi-continue inférieurement alors si $\sup\limits_{i\in I} f_i$ est bien définie c'est une fonction semi-continue inférieurement.

Sur $\mathbb{R}$ on définit la topologie de Sorgenfrey comme étant la topologie engendrée par les ensembles $]a,b]$ avec $a < b$, On note τ_S cette topologie. On définit la topologie stricte à droite comme étant la topologie engendrée par les ensembles $]a,+\infty[$, on note τ_{sd} cette topologie.

3 Montrer qu'une fonction f sur X à valeurs réelles est semi-continue inférieurement si et seulement si elle est continue pour la topologie τ_{sd}.

4 Soit $(O_n)_{n\in\mathbb{N}}$ une suite d'ouverts denses pour la topologie τ_S. Montrer que $\bigcap_{n\in\mathbb{N}} O_n$ est également dense pour la topologie τ_S.

3.2 Correction

1 On suppose que f est semi-continue inférieurement. On va montrer que $O = \{(x,\lambda) \in X \times \mathbb{R},\ f(x) > \lambda\}$ est ouvert. Soit $(x_0,\lambda_0) \in O$. On a $f(x_0) = \lambda_0 + 2\epsilon$ avec $\epsilon \in \mathbb{R}_+^*$. Il existe un ouvert $V \in \mathcal{V}(x_0)$ et tel que $\forall x \in V, f(x) \geq f(x_0) - \epsilon > \lambda_0 + \epsilon$. Donc $V \times B(\lambda,\epsilon)$ est un voisinage de (x_0,λ_0). Ainsi O est voisinage de chacun de ses points donc ouvert.
Réciproquement, on suppose que O est ouvert. Soit $x_0 \in X$ et $\epsilon \in \mathbb{R}_+^*$ alors $(x_0, f(x_0) - \frac{\epsilon}{2}) \in O$. Donc il existe un voisinage de $(x_0, f(x_0) - \epsilon)$ inclus dans O. Par définition de la topologie produit on peut choisir ce voisinage de la forme $V \times B(f(x_0) - \frac{\epsilon}{2}, \eta)$ avec $\eta \leq \frac{\epsilon}{2}$. Donc pour tout $x \in V$ on a $f(x) > f(x_0) - \frac{\epsilon}{2} - \eta \geq f(x) - \epsilon$ et donc la fonction est semi-continue inférieurement.

2 Il suffit de remarquer que l'épigraphe du sup est l'intersection des épigraphes.

3 Supposons que f est continue pour τ_S alors $]f(x_0) - \epsilon, +\infty]$ est un ouvert de τ_S et il existe un voisinage $V \in \mathcal{V}(x_0)$ tel que $f(V) \subset]f(x_0) - \epsilon, +\infty[$. Maintenant supposons que f est semi-continue inférieurement en x_0. Tout voisinage de $f(x_0)$ contient un ensemble de la forme $]f(x_0) - \epsilon, +\infty[$. La semi-continuité inférieure assure l'existence d'un voisinage de x_0 tel que $f(X)$ est inclus dans le voisinage de $f(x_0)$.

4 Soit $x \in \mathbb{R}$. Il existe $\epsilon \in \mathbb{R}_+^*$ tel que $]x-\epsilon, x+\epsilon] \cap O_1 \neq \emptyset$. Soit $]x_1 - 2\epsilon_1, x_1 + 2\epsilon_1] \subset]x - \epsilon, x+\epsilon] \cap O_1$ avec $\epsilon_1 < 2^{-1}$. On recommence en remplaçant O_1 par O_2, x par x_1 et 2^{-1} par 2^{-2}. Bref, par récurrence on construit une suite de Cauchy $(x_i)_{i\in\mathbb{N}}$. Elle converge et on a que la limite $\overline{x} \in O_i$ pour tous les O_i. De plus $\overline{x}$ est à distance ϵ ou moins de $\overline{x}$.

4 Opérateur proximal

4.1 Exercice

Soit f une fonction convexe définie sur $\mathbb{R}^n$. On suppose que la quantité

$$\inf_{x\in\mathbb{R}^n} \left(f(x) + \frac{1}{2\alpha}\|x-z\|^2 \right),$$

est bien définie pour tout $z \in \mathbb{R}^n$ et $\alpha \in \mathbb{R}_+^*$.

1 Montrer que si

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} \left(f(x) + \frac{1}{2\alpha}\|x-z\|^2 \right),$$

existe alors il est unique. On note $\mathrm{prox}_{\alpha,f}(z)$ cet élément.
La fonction $\mathrm{prox}_{\alpha,f}$ est appelée opérateur proximal de f.

2 Expliciter l'opérateur proximal si f est différentiable.

3 Soit $\alpha \in \mathbb{R}_+^*$. Montrer que z est point fixe de $\mathrm{prox}_{\alpha,f}$ est équivalent à $f(z) = \inf_{x\in\mathbb{R}^n} f(x)$.

4 Montrer que l'opérateur proximal est lipschitzien.

4.2 Correction

1 L'unicité est directe car la fonction norme au carrée est strictement convexe et que la somme d'une fonction convexe et d'une strictement convexe est strictement convexe.

2 $\mathrm{prox}_{\alpha,f}(z) = z + \nabla f(z)$.

3 Si z est tel que pour tout $x \in \mathbb{R}^n, f(z) \leq f(x)$ alors $f(z)+\frac{1}{2\alpha}\|z-z\|^2+f(z) \leq f(x)+\frac{1}{2\alpha}\|x-z\|^2$. Donc $\operatorname{prox}_{\alpha,f}(z)=z$.
Réciproquement, si on suppose que z est point fixe

$$\begin{aligned}
&\forall x \in \mathbb{R}^n, f(z) \leq \frac{1}{2\alpha}\|z-x\|^2+f(x)\\
&\forall (u,t) \in \mathbb{R}^n \times [0,1], f(z) \leq \frac{t^2}{2\alpha}\|u\|^2+f(z+tu)\\
&\forall (u,t) \in \mathbb{R}^n \times [0,1], f(z) \leq \frac{t^2}{2\alpha}\|u\|^2+tf(u)+(1-t)f(z)\\
&\forall (u,t) \in \mathbb{R}^n \times [0,1], t(f(z)-f(u)) \leq \frac{t^2}{2\alpha}
\end{aligned}$$

On divise par t et on fait tendre t vers 0 pour obtenir l'inégalité voulue.

4 On note $Pu = \operatorname{prox}_{\alpha,f}(u)$ et $Pv = \operatorname{prox}_{\alpha,f}(v)$. On a

$$\begin{aligned}
&\forall (x,y) \in \mathbb{R}^n,\ \begin{cases} f(Pu)+\frac{1}{2\alpha}\|u-Pu\|^2 \leq \frac{1}{2\alpha}\|u-x\|^2+f(x)\\ f(Pv)+\frac{1}{2\alpha}\|v-Pv\|^2 \leq \frac{1}{2\alpha}\|v-y\|^2+f(y)\end{cases}\\
&\begin{cases} f(Pu)+\frac{1}{2\alpha}\|u-Pu\|^2 \leq \frac{1}{2\alpha}\|u-tPu-(1-t)Pv\|^2+f(tPu+(1-t)Pv)\\ f(Pv)+\frac{1}{2\alpha}\|v-Pv\|^2 \leq \frac{1}{2\alpha}\|v-tPu-(1-t)Pv\|^2+f(tPu+(1-t)Pv)\end{cases}\\
&\begin{cases} t\left(f(Pu)+\frac{1}{2\alpha}\|u-Pu\|^2\right) \leq \frac{t}{2\alpha}\|u-tPu-(1-t)Pv\|^2+tf(tPu+(1-t)Pv)\\ (1-t)\left(f(Pv)++\frac{1}{2\alpha}\|v-Pv\|^2\right) \leq \frac{1-t}{2\alpha}\|v-tPu-(1-t)Pv\|^2+(1-t)f(tPu+(1-t)Pv)\end{cases}.\\
&0 \leq t\|u-tPu-(1-t)Pv\|^2+(1-t)\|v-tPu-(1-t)Pv\|^2-t\|u-Pu\|^2-(1-t)\|v-Pv\|^2\\
&0 \leq t\left(\|u-Pu+(1-t)(Pu-Pv)\|^2-\|u-Pu\|^2\right)+(1-t)\left(\|v-Pv+t(Pv-Pu)\|^2-\|v-Pv\|^2\right)\\
&0 \leq 2\langle u-Pu, Pu-Pv\rangle+(1-t)\|Pu-Pv\|^2+2\langle Pv-v, Pu-Pv\rangle+t\|Pu-Pv\|^2\\
&0 \leq \|Pu-Pv\|^2-2\|Pu-Pv\|^2+2\langle u-v, Pu-Pv\rangle\\
&\|Pu-Pv\|^2 \leq 2\langle u-v, Pu-Pv\rangle
\end{aligned}$$

En utilisant l'inégalité de Cauchy-Schwarz on peut conclure. On peut donner une constante de Lipschitz égale à 1 en poussant l'analyse avec les sous-gradients et la règle de Fermat on peut montrer que la constante de Lipschitz est au plus égale à 1.