Préparation aux Concours (X)

Dualité

Notations et rappels

- On notera $\mathbb{R}_+$ l'ensemble des réels positifs ou nuls et $\mathbb{R}_+^*$ l'ensemble des réels strictement positifs.
- On rappelle que si $F \subset \mathbb{R}^n$ est un fermé borné et $f : F \to \mathbb{R}$ est continue, alors le minimum de f sur F est atteint, c'est-à-dire qu'il existe $x \in F$ tel que $f(x) \leq f(y)$ pour tout $y \in F$.
- Soit E un espace vectoriel sur $\mathbb{R}$. On dit que $C \subset E$ est un ensemble convexe si pour tous $x, y \in C$ et tout $t \in [0,1]$ on a $(1-t)x + ty \in C$.

Pour $C \subset E$ convexe, une fonction $f : C \to \mathbb{R}$ est dite convexe, si pour tous x, y éléments de C et tout $t \in [0,1]$, on a $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$. On dit que f est strictement convexe si cette inégalité est stricte pour $t \in]0,1[$ et $x \neq y$.

- Soient A et B deux ensembles et $f : A \times B \to \mathbb{R}$. On dit que f admet *un point selle* en $(a_*, b_*) \in A \times B$ si pour tout $(a,b) \in A \times B$ on a

$$f(a_*, b) \leq f(a_*, b_*) \leq f(a, b_*).$$

- Toutes les variables aléatoires seront supposées définies sur un espace probabilisé commun $(\Omega, \mathcal{F}, P)$

Les parties I et II sont indépendantes.

I. Convexité et points selles

Soient m et n deux entiers positifs non nuls et E un espace vectoriel sur $\mathbb{R}$.

(1) Soient $C \subset E$ un ensemble convexe. Soient f et g deux fonctions convexes de C dans $\mathbb{R}$.

(a) Montrer que $f+g$ est convexe, et strictement convexe s'il l'une des deux fonctions f ou g est strictement convexe.

(b) On suppose f strictement convexe. Vérifier que le minimum de f est atteint sur C en au plus un point de C.

(2) Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$ une matrice de m lignes et n colonnes. On note $\langle u, v\rangle_{\mathbb{R}^n}$ le produit scalaire entre deux vecteurs u et v de $\mathbb{R}^n$ et $\langle \mu, \nu\rangle_{\mathbb{R}^m}$ celui entre deux vecteurs μ et ν de $\mathbb{R}^m$.

(a) Montrer que pour tout $(x, \nu) \in \mathbb{R}^n \times \mathbb{R}^m$, on a

$$\langle Ax, \nu\rangle_{\mathbb{R}^m} = \left\langle x, A^\top \nu\right\rangle_{\mathbb{R}^n}$$

où $A^\top$ désigne la matrice transposée de A.

(b) En déduire que $\ker A \subset \left(\operatorname{Im} A^\top\right)^\perp$ où $E^\perp$ désigne l'orthogonal de E pour le produit scalaire sur $\mathbb{R}^n$ pour tout sous-espace vectoriel E de $\mathbb{R}^n$.

(c) Montrer que $\ker A = \left(\operatorname{Im} A^\top\right)^\perp$

(3) On considère un ouvert $U \subset \mathbb{R}^n$, $h : U \to \mathbb{R}$ une application $\mathcal{C}^1$ et $b \in \mathbb{R}^m$. On suppose qu'il existe $x_* \in U$ un minimum de h sur l'ensemble $V_b = \{x \in U \mid Ax + b = 0\}$.

(a) Montrer que pour tout $u \in \mathbb{R}^n$ tel que $Au = 0$ on a $\langle \nabla h(x_*), u\rangle_{\mathbb{R}^n} = 0$ où $\nabla h(x)$ désigne le gradient de h en x.

(b) Montrer l'existence de $\nu_* \in \mathbb{R}^m$ tel que $\nabla h(x_*) - A^\top \nu_* = 0$.

(c) En déduire que l'application $L : U \times \mathbb{R}^m \to \mathbb{R}$ telle que $L(x,\nu) = h(x) - \langle \nu, Ax + b\rangle_{\mathbb{R}^m}$ vérifie $\frac{\partial L}{\partial x_k}(x_*, \nu_*) = 0$ pour tout $1 \leq k \leq n$ où $\frac{\partial L}{\partial x_k}(x,\nu)$ désigne la dérivée partielle de L par rapport à la k-ième coordonnées de $x \in \mathbb{R}^n$.

(d) Conclure que si U est convexe, et h convexe sur U, alors L admet un point selle en (x_*, ν_*), c'est-à-dire que l'on a

$$L(x_*, \nu) \leq L(x_*, \nu_*) \leq L(x, \nu_*)$$

pour tout $(x, \nu) \in U \times \mathbb{R}^m$.

II. Entropie et codage

Soient $\mathscr{X}$ un ensemble fini et $\boldsymbol{p} = (\boldsymbol{p}_x)_{x \in X}$ une loi de probabilité sur $\mathscr{X}$. On suppose que $\boldsymbol{p}$ charge tous les points de $\mathscr{X}$: $\boldsymbol{p}_x > 0$ pour tout $x \in \mathscr{X}$. On appelle *entropie* de $\boldsymbol{p}$ la quantité

$$H(\boldsymbol{p}) = -\sum_{x \in \mathscr{X}} \boldsymbol{p}_x \ln(\boldsymbol{p}_x)$$

On considère l'ensemble $Q_{\mathscr{X}} = \left\{\boldsymbol{q} = (\boldsymbol{q}_x)_{x \in x} \in \mathbb{R}^{\mathscr{X}} \mid \forall x \in \mathscr{X}, \boldsymbol{q}_x \geq 0\right\}$. Pour tous $\boldsymbol{q}, \boldsymbol{q}' \in Q_{\mathscr{X}}$ tels que $\boldsymbol{q}'_x > 0$ pour tout $x \in \mathscr{X}$, on définit :

$$\mathrm{KL}(\boldsymbol{q}, \boldsymbol{q}') = \sum_{x \in \mathscr{X}} \varphi(\boldsymbol{q}_x / \boldsymbol{q}'_x)\, \boldsymbol{q}'_x$$

avec $\varphi : \mathbb{R}_+ \to \mathbb{R}$ définie par $\varphi(x) = x \ln(x) - x + 1$ pour $x > 0$ et prolongée en 0 par continuité.

(4) (a) Préciser $\varphi(0)$.

(b) Vérifier que φ est continue strictement convexe positive et que $\varphi(x) = 0$ si et seulement si $x = 1$.

(c) Montrer que $Q_{\mathscr{X}}$ est convexe et que $\boldsymbol{q} \mapsto \mathrm{KL}(\boldsymbol{q}, \boldsymbol{q}')$ est strictement convexe positive et s'annule ssi $\boldsymbol{q} = \boldsymbol{q}'$.

Soit $\mathcal{A}$ un ensemble fini. On appelle *mot* sur $\mathcal{A}$ une suite finie d'éléments de $\mathcal{A}$, on le note $u = u_1 \dots u_n$ et n est la *longueur* du mot u, notée $|u|$. Le mot vide est noté $\boldsymbol{\epsilon}$, il est de longueur nulle. On note $\mathcal{A}^*$ l'ensemble des mots sur $\mathcal{A}$ et $\mathcal{A}^+ = \mathcal{A}^* \backslash \{\boldsymbol{\epsilon}\}$ l'ensemble des mots privé du mot vide.
On définit la *concaténation* $u \cdot v$ de deux mots $u, v \in \mathcal{A}^*$ par $u \cdot \boldsymbol{\epsilon} = \boldsymbol{\epsilon} \cdot u = u$ et $u \cdot v = u_1 \dots u_{|u|} v_1 \dots v_{|v|}$ si $u, v \in \mathcal{A}^+$. On dit que u est un *préfixe* de v si $v = u \cdot w$ pour $w \in \mathcal{A}^*$.
Soient $\mathscr{X}$ un ensemble fini non vide et $c : \mathscr{X} \to \{0,1\}^+$ une application *injective*. On dira que c est un *code* binaire sur $\mathscr{X}$. On suppose de plus que c est un code préfixe, c'est-à-dire que pour tous $x \neq y$ dans $\mathscr{X}$, $c(x)$ n'est pas un préfixe de $c(y)$.

(5) On définit $\bar{c} : \mathscr{X} \to \{0,1\}^*$ tel que pour tout $x \in \mathscr{X}$, $c(x) = c(x)_1 \cdot \bar{c}(x)$ où $c(x)_1$ est le premier élément du mot $c(x)$.

(a) Vérifier que pour tout $x \neq y \in \mathscr{X}$, si $c(x)_1 = c(y)_1$ alors $\bar{c}(x) \neq \bar{c}(y)$ et $\bar{c}(x)$ n'est pas un préfixe de $\bar{c}(y)$.

(b) Pour $a \in \{0,1\}$ on note $\mathscr{X}_a = \{x \in \mathscr{X} \mid c(x)_1 = a\}$. Montrer que si $\mathscr{X}_a$ contient au moins deux éléments, alors la restriction de $\bar{c}$ à $\mathscr{X}_a$ est un code préfixe sur $\mathscr{X}_a$.

(c) En déduire que $\sum_{x \in \mathscr{X}} 2^{-|c(x)|} \leq 1$. *(Ind. : On pourra décomposer la somme en une somme sur $\mathscr{X}_0$ et $\mathscr{X}_1$ et raisonner par récurrence sur $L(c) = \max\{\, |c(x)| \mid x \in \mathscr{X}\}$.)*

Soient $\boldsymbol{q} = \left(2^{-|c(x)|}\right)_{x \in \mathscr{X}}$ et X une variable aléatoire à valeurs dans $\mathscr{X}$ de loi $\boldsymbol{p}$.

(6) (a) Vérifier que $\ln(2) E(|c(X)|) = -\sum_{x \in \mathscr{X}} \boldsymbol{p}_x \ln(\boldsymbol{q}_x)$.

(b) En déduire que $E(|c(X)|) \geq \frac{H(\boldsymbol{p})}{\ln(2)}$.
(Ind. : On pourra chercher à exprimer $\ln(2) E(|c(X)|)$ en fonction de $H(\boldsymbol{p})$ et $\mathrm{KL}(\boldsymbol{p}, \boldsymbol{q})$.)

III. Transport régularisé

Dans toute la suite I et J désignent deux ensembles finis.

- On considère $\alpha = (\alpha_i)_{i\in I} \in \left(\mathbb{R}_+^*\right)^I$ et $\beta = (\beta_j)_{j\in J} \in \left(\mathbb{R}_+^*\right)^J$ tels que $\sum_{i\in I} \alpha_i = \sum_{j\in J} \beta_j = 1$ si bien que α et β peuvent être considérés comme définissant deux lois de probabilités sur I et J.
- Dans la suite on notera

$$Q = \{\left(\boldsymbol{q}_{ij}\right)_{(i,j)\in I\times J} \in \mathbb{R}^{I\times J} \mid \boldsymbol{q}_{ij} \geq 0 \text{ pour tout } (i,j) \in I \times J\}$$

et

$$F(\alpha, \beta) = \{\boldsymbol{q} \in Q \mid \sum_{j'\in J} \boldsymbol{q}_{ij'} = \alpha_i \text{ et } \sum_{i'\in I} \boldsymbol{q}_{i'j} = \beta_j \text{ pour tout } (i,j) \in I \times J\}.$$

On notera $\boldsymbol{p}$ l'élément de $F(\alpha, \beta)$ défini par $p_{ij} = \alpha_i \beta_j > 0$ pour tout $(i,j) \in I \times J$.

(7) Vérifier que $F(\alpha, \beta)$ est un ensemble convexe de l'espace vectoriel $E = \mathbb{R}^{I\times J}$.

(8) Soient X_1 et X_2 deux variables aleatoires telles que X_1 est à valeurs dans I et X_2 à valeurs dans J.

(a) Vérifier que si $\boldsymbol{q} \in F(\alpha, \beta)$, alors $\sum_{i\in I} \sum_{j\in J} \boldsymbol{q}_{ij} = 1$.

(b) On suppose que $P\left(X_1 = i, X_2 = j\right) = \boldsymbol{q}_{ij}$ avec $\boldsymbol{q} \in F(\alpha, \beta)$. Calculer la loi de X_1 et celle de X_2 en fonction de α et β.

(c) Que dire de X_1 et X_2 lorsque $\boldsymbol{q} = \boldsymbol{p}$?

Soient $C = (C_{ij})_{(i,j)\in I\times J} \in \mathbb{R}_+^{I\times J}$ et $\epsilon > 0$. On considère $J_\epsilon : Q \to \mathbb{R}$ définie par

$$J_\epsilon(\boldsymbol{q}) = \sum_{ij} \boldsymbol{q}_{ij} C_{ij} + \epsilon \operatorname{KL}(\boldsymbol{q}, \boldsymbol{p})$$

où $\operatorname{KL}(\boldsymbol{q}, \boldsymbol{p})$ est défini dans la partie précédente en prenant $\mathscr{X} = I \times J$.

(9) Montrer que J_ϵ est strictement convexe sur Q.

(10) (a) Vérifier que $F(\alpha, \beta)$ est un fermé borné de $\mathbb{R}^{I\times J}$.

(b) Montrer qu'il existe un unique $\boldsymbol{q}(\epsilon) \in Q$ minimisant J_ϵ sur $F(\alpha, \beta)$.

(c) En considérant un contre-exemple simple, montrer que l'unicité n'est plus vraie si on suppose que $\epsilon = 0$.

(11) (a) Vérifier que $\boldsymbol{q}(\epsilon)_{ij} > 0$ pour tout $(i,j) \in I \times J$ *(Ind: On pourra raisonner par l'absurde et considérer pour tout $t \in]0,1[$ $\boldsymbol{q}(\epsilon, t) = (1-t)\boldsymbol{q}(\epsilon) + t\boldsymbol{p}$ puis observer le comportement de $\varphi(x)$ au voisinage de $x = 0$).*

(b) Montrer que ceci n'est plus vrai si on suppose que $\epsilon = 0$.

IV. Dualité

On définit $Q_{>0} = \left(\mathbb{R}_+^*\right)^{I\times J}$ et $\mathscr{L} : Q_{>0} \times \left(\mathbb{R}^I \times \mathbb{R}^J\right) \to \mathbb{R}$ défini par

$$\mathscr{L}(\boldsymbol{q}, (f,g)) = J_\epsilon(\boldsymbol{q}) + \sum_{i\in I} f_i(\alpha_i - \sum_{j\in J} \boldsymbol{q}_{ij}) + \sum_{j\in J} g_j(\beta_j - \sum_{i\in I} \boldsymbol{q}_{ij})$$

(12) (a) Vérifier que $Q_{>0}$ est un ouvert convexe $\mathbb{R}^{I\times J}$.

(b) Montrer qu'il existe $(f(\epsilon), g(\epsilon)) \in \mathbb{R}^I \times \mathbb{R}^J$ tel que $\mathscr{L}(\boldsymbol{q}(\epsilon), (f(\epsilon), g(\epsilon))$ est un point selle de $\mathscr{L}$. *(Indication: On pourra identifier $\mathbb{R}^{I\times J}$ avec $\mathbb{R}^n$ et $\mathbb{R}^I \times \mathbb{R}^J$ avec $\mathbb{R}^m$ pour n cardinal de $I \times J$ et m somme des cardinaux de I et J puis utiliser la question 3 de la partie I.)*

(13) (a) Montrer que pour tout $(f, g) \in \mathbb{R}^I \times \mathbb{R}^J$, le minimum de $\boldsymbol{q} \mapsto \mathscr{L}(\boldsymbol{q}, (f, g))$ sur $Q_{>0}$ est atteint en $\boldsymbol{q}(f, g)_{ij} = \mathrm{e}^{(f_i + g_j - C_{ij})/\epsilon} \boldsymbol{p}_{ij}$.

(b) Calculer la valeur de $G(f, g) = \mathscr{L}(\boldsymbol{q}(f, g), (f, g))$.

(c) Vérifier que G est concave sur $\mathbb{R}^I \times \mathbb{R}^J$.

(14) Vérifier que si $f_* : \mathbb{R}^J \to \mathbb{R}^I$ et $g_* : \mathbb{R}^I \to \mathbb{R}^J$ sont définies par

$$f_*(g)_i = -\epsilon \ln \left(\sum_{j \in J} e^{(g_j - C_{ij})/\epsilon} \beta_j \right) \text{ et } g_*(f)_j = -\epsilon \ln \left(\sum_{i \in I} e^{(f_i - C_{ij})/\epsilon} \alpha_i \right)$$

alors pour tout $(f, g) \in \mathbb{R}^I \times \mathbb{R}^J$, on a $\frac{\partial G}{\partial f_i}(f_*(g), g) = \frac{\partial G}{\partial g_j}(f, g_*(f)) = 0$ pour tout $(i, j) \in I \times J$.

Soit $\left(f^0, g^0\right) \in \mathbb{R}^{I \times J}$. Pour tout $k \geq 0$, on considère

$$g^{k+1} = g_*\left(f^k\right) \text{ et } f^{k+1} = f_*\left(g^{k+1}\right)$$

(15) Montrer que la suite $\left(G\left(f^k, g^k\right)\right)_{k \geq 0}$ est croissante.

(16) On suppose qu'il existe $f^\infty = (f_i^\infty)_{i \in I}$ et $g^\infty = \left(g_j^\infty\right)_{j \in J}$ tel que $\left|f_i^k - f_i^\infty\right| \to 0$ et $\left|g_j^k - g_j^\infty\right| \to 0$ pour tous $i \in I$ et $j \in J$. On note $G_* = \sup\{G(f, g) \mid (f, g) \in \mathbb{R}^I \times \mathbb{R}^J\}$.

(a) Montrer que $G\left(f^\infty, g^\infty\right) = G_*$.

(b) Montrer que $G(f(\epsilon), g(\epsilon)) = G_*$.

(c) Montrer qu'il existe une constante $a \in \mathbb{R}$ telle $f(\epsilon)_i = f_i^\infty + a$ et $g(\epsilon)_j = g_j^\infty - a$ pour tout $(i, j) \in I \times J$.

(d) En déduire que $\boldsymbol{q}\left(f^k, g^k\right) \to \boldsymbol{q}(\epsilon)$.