Inégalité de Muirhead

Sommaire

1. Introduction

L'objectif de ce document est d'énoncer et de démontrer l'inégalité de Muirhead. Celle-ci permet d'affirmer sans trop réfléchir que

$$\forall x, y, z \in \mathbb{R}_+^*, 6xyz \leqslant x^2y + x^2z + xy^2 + y^2z + xz^2 + yz^2$$

ou encore que

$$\forall x, y \in \mathbb{R}_+^*, x + y \leqslant \frac{x^2}{y} + \frac{y^2}{x}$$

On voit que ces inégalités font intervenir des sommes de la forme

$$\sum_{\sigma \in S_n} x_{\sigma(1)}^{p_1} \dots x_{\sigma(1)}^{p_n}$$

où $x = (x_1, ..., x_n) \in {\mathbb{R}_+^*}^n$ et $p = (p_1, ..., p_n) \in \mathbb{R}^n$. Une telle somme sera notée $[p]_x$. Par exemple, les deux inégalités ci-dessus se reformulent en $[(1, 1, 1)]_{(x,y,z)} \leqslant [(2, 1, 0)]_{(x,y,z)}$ et $[(1, 0)]_{(x,y)} \leqslant [(2, -1)]_{(x,y)}$.

Pour pouvoir énoncer l'inégalité de Muirhead, il faut d'abord définir la majorisation :

Définition 1.1: Soient $p = (p_1, ..., p_n), q = (q_1, ..., q_n) \in \mathbb{R}_n$ des vecteurs décroissants (i.e. $p_1 \geqslant ... \geqslant p_n$ et $q_1 \geqslant ... \geqslant q_n$). On dit que p majorise q et on note $p \succeq q$ si et seulement si les deux propriétés suivantes sont vérifiées :
1. $\forall k \in [\![1, n-1]\!], \sum_{i=1}^k p_i \geqslant \sum_{i=1}^k q_i$;
2. $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$.

Si p, q ne sont pas forcément décroissants, on dit que p majorise q et on note $p \succeq q$ si et seulement si $\tilde{p} \succeq \tilde{q}$ où $\tilde{p}$ et $\tilde{q}$ sont les vecteurs obtenus en triant les coefficients de p et q respectivement par ordre décroissant.

On vérifie facilement que la majorisation est un préordre (i.e. une relation binaire réflexive et transitive) sur $\mathbb{R}^n$.

L'inégalité de Muirhead dit alors la chose suivante :

Théorème 1.1 (inégalité de Muirhead): Soient $x \in {\mathbb{R}_+^*}^n$ et $p, q \in \mathbb{R}^n$ tels que $p \succeq q$. Alors $[p]_x \geqslant [q]_x$.

À titre d'exemple, vous pouvez vérifier que l'inégalité de Muirhead permet bien de justifier les deux inégalités données au début de cette section.

Si $p = (1, 0, ..., 0)$ et $q = \left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right)$, on retrouve l'inégalité arithmético-géométrique (mais on aura besoin de cette dernière dans la preuve de l'inégalité de Muirhead, donc ça n'a pas vraiment d'intérêt).

Voici un autre exemple d'application :

Théorème 1.2 (inégalité de Nesbitt): Soient $a, b, c \in \mathbb{R}_+^*$. Alors

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} \geqslant \frac{3}{2}$$

Preuve:

$$\begin{aligned} & \frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} \geqslant \frac{3}{2} \\ \iff & 2(a(a+c)(a+b) + b(b+c)(a+b) + c(b+c)(a+c)) \geqslant 3(b+c)(a+c)(a+b) \\ \iff & 2(a^3 + b^3 + c^3) \geqslant a^2(b+c) + b^2(a+b) + c^2(a+c) \\ \iff & [(3,0,0)]_{(a,b,c)} \geqslant [(2,1,0)]_{(a,b,c)} \end{aligned}$$

ce qui est vrai par l'inégalité de Muirhead puisque $(3, 0, 0) \succeq (2, 1, 0)$. (La page Wikipédia de l'inégalité de Nesbitt recense 8 autres preuves.) ■

Plus généralement, l'inégalité de Muirhead est très puissante pour démontrer des inégalités homogènes et symétriques, mais elle fait souvent apparaître des calculs très bourrins.

2. Idée de la démonstration

La démonstration de l'inégalité de Muirhead est plutôt inattendue puisqu'elle fait intervenir des matrices et des graphes bipartis.

La clé est le théorème suivant :

Théorème 2.1: Soient $p, q \in \mathbb{R}^n$. Alors $p \succeq q$ si et seulement si q est dans l'enveloppe convexe des vecteurs obtenus en permutant les coordonnées de p.

Autrement dit, $(p_1, ..., p_n) \succeq (q_1, ..., q_n)$ si et seulement si il existe des réels positifs de somme 1 $(c_\sigma)_{\sigma \in S_n}$ tels que $q = \sum_{\sigma \in S_n} c_\sigma \left(p_{\sigma(1)}, ..., p_{\sigma(n)}\right)$.

On peut se convaincre du théorème géométriquement si $n = 2$ ou 3, mais la démonstration dans le cas général n'est pas évidente. Nous allons montrer uniquement le sens direct puisque c'est celui qui nous intéresse (en fait, c'est le sens le plus difficile).

Pour $\sigma \in S_n$, on a $\begin{pmatrix} p_{\sigma(1)} \\ \dots \\ p_{\sigma(n)} \end{pmatrix} = P_\sigma \begin{pmatrix} p_1 \\ \dots \\ p_n \end{pmatrix}$ où P_σ est la matrice de permutation définie par $(P_\sigma)_{i,j} = \mathbb{1}_{\sigma(i)=j}$.

Nous allons donc démontrer ceci :

Théorème 2.2: Soient $p, q \in \mathbb{R}^n$ d tels que $p \succeq q$. Alors il existe des réels positifs de somme 1 $(c_\sigma)_{\sigma \in S_n}$ tels que $q = \sum_{\sigma \in S_n} c_\sigma P_\sigma p$.

Commençons par montrer l'inégalité de Muirhead en admettant ce théorème.

Preuve de l'inégalité de Muirhead: Soient $x \in {\mathbb{R}_+^*}^n$ et $p, q \in \mathbb{R}^n$ tels que $p \succeq q$. D'après le théorème précédent, il existe des réels positifs de somme 1 $(c_\sigma)_{\sigma \in S_n}$ tels que $q = \sum_{\sigma \in S_n} c_\sigma P_\sigma p$.

On a alors

$$\begin{aligned}
[p]_x &= \sum_{\sigma \in S_n} c_\sigma [p]_x = \sum_{\sigma \in S_n} c_\sigma [P_\sigma p]_x \\
&= \sum_{\sigma \in S_n} c_\sigma \sum_{\tau \in S_n} x_{\tau(1)}^{p_{\sigma(1)}} \dots x_{\tau(n)}^{p_{\sigma(n)}} = \sum_{\tau \in S_n} \sum_{\sigma \in S_n} c_\sigma x_{\tau(1)}^{p_{\sigma(1)}} \dots x_{\tau(n)}^{p_{\sigma(n)}} \\
&\geqslant \sum_{\tau \in S_n} \prod_{\sigma \in S_n} x_{\tau(1)}^{c_\sigma p_{\sigma(1)}} \dots x_{\tau(n)}^{c_\sigma p_{\sigma(n)}} = \sum_{\tau \in S_n} x_{\tau(1)}^{\sum_{\sigma \in S_n} c_\sigma p_{\sigma(1)}} \dots x_{\tau(n)}^{\sum_{\sigma \in S_n} c_\sigma p_{\sigma(n)}} \\
&= \left[\sum_{\sigma \in S_n} c_\sigma P_\sigma p \right]_x = [q]_x
\end{aligned}$$

où l'on a utilisé l'inégalité arithmético-géométrique pondérée, qui stipule que si $\lambda_1, \dots, \lambda_n$ sont des réels positifs de somme 1 et $a_1, \dots, a_n \in \mathbb{R}^*$ alors $\sum_{i=1}^n \lambda_i a_i \geqslant \prod_{i=1}^n a_i^{\lambda_i}$. Cette inégalité se démontre facilement en appliquant l'inégalité de Jensen à la fonction concave ln (il faut juste faire attention au cas où l'un des λ_i ou l'un des a_i est nul). ■

Il reste à démontrer le Théorème 2.2. Pour cela, on va procéder en deux temps :

1. si $p \succeq q$ alors il existe une matrice bistochastique B telle que $q = Bp$ (une matrice bistochastique étant une matrice à coefficients positifs dont la somme des coefficients de chaque ligne et chaque colonne vaut 1) ;
2. toute matrice bistochastique s'écrit comme combinaison convexe de matrices de permutation.

3. Lien avec les matrices bistochastiques

Théorème 3.1: Soient $p, q \in \mathbb{R}^n$ tels que $p \succeq q$. Alors il existe une matrice bistochastique B telle que $q = Bp$.

Preuve: On suppose sans perte de généralité p et q décroissants : on aura alors montré que $\tilde{q} = B\tilde{p}$ où $\tilde{p}$ et $\tilde{q}$ sont les vecteurs p et q triés par ordre décroissant, mais $\tilde{q} = Pq$ et $\tilde{p} = P'p$ où P et P' sont des matrices de permutation, donc $q = P^{-1}P'Bp$ où $P^{-1}P'B$ est bistochastique. (En effet, on vérifie facilement que l'inverse d'une matrice de permutation est une matrice de permutation et qu'un produit de matrices bistochastiques est une matrice bistochastique.)

L'idée est de construire des vecteurs décroissants $r_0, \dots, r_k \in \mathbb{R}^n$ tels que $p = r_0 \succeq r_1 \succeq \dots \succeq r_k = q$ et pour tout i, $r_i = T_i r_{i-1}$ avec T_i une matrice bistochastique. On aura alors $q = T_k T_{k-1} \dots T_1 p$, d'où le résultat car un produit de matrices bistochastiques est encore bistochastique.

Plus précisément, chaque r_i aura au moins un coefficient de plus que r_{i-1} égal au coefficient correspondant de q. Ceci garantira que la construction termine. On va construire T_1 et r_1, le reste suivra donc par une récurrence triviale.

L'idée est de prendre $T_1 = P_1 T_1'$ où P_1 est une matrice de permutation permettant d'assurer que r_1 soit décroissant, et T_1' est de la forme

$$\begin{pmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & (0) & & & \\ & & 1 & & & & & & & & \\ & & & \alpha & & & 1-\alpha & & & & \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & 1-\alpha & & & & \alpha & & & \\ & & & & & & & & 1 & & \\ & (0) & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{pmatrix}$$

Cette matrice est clairement bistochastique si $\alpha \in [0,1]$. On note j l'indice de la colonne $\begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}$ (qui est aussi l'indice de la ligne $\begin{pmatrix} \alpha & 1-\alpha \end{pmatrix}$) et k l'indice de la colonne $\begin{pmatrix} 1-\alpha \\ \alpha \end{pmatrix}$ (qui est aussi l'indice de la ligne $\begin{pmatrix} 1-\alpha & \alpha \end{pmatrix}$).

On choisit $j = \min\{i; p_i > q_i\}$ et $k = \max\{i; p_i < q_i\}$. Ces deux entiers existent sauf si $p = q$, mais dans ce cas $B = I_n$ convient. On voit alors facilement qu'on a $j < k$ et $p_j > q_j \geqslant q_k > p_k$.

On pose

$$r_1' = T_1' p = \begin{pmatrix} p_1 \\ \vdots \\ p_{j-1} \\ \alpha p_j + (1-\alpha) p_k \\ p_{j+1} \\ \vdots \\ p_{k-1} \\ (1-\alpha) p_j + \alpha p_k \\ p_{k+1} \\ \vdots \\ p_n \end{pmatrix}$$

On choisit $\alpha = \frac{q_j - p_k}{p_j - p_k} \in [0,1]$, de sorte que $\alpha p_j + (1-\alpha) p_k = q_j$.

Soit $r_1 = P_1 r_1'$ où P_1 est une matrice de permutation telle que r_1 soit décroissant. Il reste à montrer que $p \succeq r_1$. Pour tout $i \in [\![1,n]\!]$, on note $S_i = \sum_{j=1}^{i} r_{1,j}'$ et $T_i = \sum_{j=1}^{i} p_j$. On voit facilement que si $i \in [\![1, j-1]\!] \cup [\![k+1, n]\!]$ alors $S_i = T_i$. Soit $i \in [\![j,k]\!]$, montrons que $S_i \leqslant T_i$. On voit que S_i peut être de quatre sortes différentes :

- Si $S_i = p_1 + \ldots + p_{j-1} + p_{j+1} + \ldots + p_{i+1}$ alors $T_i - S_i = p_j - p_{i+1} \geqslant 0$.

- Si $S_i = p_1 + \ldots + p_{j-1} + p_{j+1} + \ldots + p_i + (\alpha p_j + (1-\alpha) p_k)$ alors $T_i - S_i = (1-\alpha)(p_j - p_k) \geqslant 0$.
- Si $S_i = p_1 + \ldots + p_{j-1} + p_{j+1} + \ldots + p_i + ((1-\alpha) p_j + \alpha p_k)$ alors $T_i - S_i = \alpha(p_j - p_k) \geqslant 0$.
- Si $S_i = p_1 + \ldots + p_{j-1} + p_{j+1} + \ldots + p_{i-1} + (\alpha p_j + (1-\alpha) p_k) + (\alpha p_j + (1-\alpha) p_k)$ alors $T_i - S_i = p_i - p_k \geqslant 0$.

Ainsi on a construit un vecteur décroissant r_1 tel que $p \succeq r_1, r_1 = T_1 p$ où T_1 est bistochastique, et qui contient au moins une composante de plus que p égale à la composante correspondante de q. Cela permet de conclure par une récurrence immédiate. ■

4. Théorème de Birkhoff - von Neumann

Il reste à montrer que toute matrice bistochastique s'écrit comme combinaison convexe de matrices de permutation. Ce résultat constitue le théorème de Birhoff - von Neumann. Pour cela, nous allons commencer par démontrer un résultat de théorie des graphes qui n'a a priori rien à voir : le théorème de Hall.

Définition 4.1 : Un graphe (simple et non orienté) est dit biparti si et seulement si son ensemble de sommets peut être partitionné en deux ensembles disjoints X et Y tels que chaque arrête du graphe ait une extrémité dans X et une extrémité dans Y. On notera les graphes bipartis sous la forme (X, Y, E) où E est l'ensemble des arêtes, représentées sous forme d'ensemble de deux sommets.

Définition 4.2 : Soit $G = (X, Y, E)$ un graphe biparti.
- On appelle couplage tout ensemble d'arêtes $C \subseteq E$ deux à deux disjointes.
- Si de plus tout sommet de X est l'extrémité d'une arrête de C, C est dit couplage X-parfait.

Théorème 4.1 (de Hall) : Soit $G = (X, Y, E)$ un graphe biparti. Alors G possède un couplage X-parfait si et seulement si $\forall W \subseteq X, |W| \leqslant |V(W)|$ où $V(W) \subseteq Y$ est l'ensemble des sommets adjacents à au moins un sommet de W.

Preuve:
- Supposons que G possède un couplage X-parfait C. Soit $W \subseteq X$. Pour tout sommet $w \in W$, on dispose d'une arête $\{w, y_w\} \in C$. Par définition d'un couplage, les y_w sont distincts, et les y_w sont dans $V(W)$. Du coup, $|W| \leqslant |V_G(W)|$. (En fait, ce sens nous servira pas dans la suite.)
- Réciproquement, raisonnons par contraposée et supposons que G ne possède pas de couplage X-parfait. Soit C un couplage de G de cardinal maximal. Ce couplage n'est pas parfait, donc il existe un sommet $u \in X$ qui n'est extrémité d'aucune arête de C. Considérons l'ensemble des chemins alternants (i.e. les chemins qui empruntent alternativement les arêtes de C et les arêtes de $E \setminus C$) partant de u. Soient W l'ensemble des sommets de ces chemins qui sont dans X, et Z l'ensemble des sommets de ces chemins qui sont dans Y.

Alors par maximalité de C, chaque sommet de Z est apparié à un sommet de W par le couplage C. En effet, si ce n'était pas le cas, on aurait un chemin alternant $uz_1w_1...w_kz_{k+1}$ où $\{u, z_1\} \notin C$ et $\{w_k, z_{k+1}\} \notin C$. Le couplage

$$C' = (C \cup \{\{u, z_1\}, \{w_1, z_2\}, ..., \{w_k, z_{k+1}\}\}) \setminus \{\{z_1, w_1\}, \{z_2, w_2\}, ..., \{z_k, w_k\}\}$$

serait alors strictement plus grand que C.

Du coup, $|W| - 1 \geqslant |Z|$: à chaque sommet de Z correspond un sommet de $W \setminus \{u\}$ via le couplage C.

Enfin, $Z = V(W)$. En effet, on a clairement $Z \subseteq V(W)$ par construction. Réciproquement, soit $v \in V(W)$, soit donc $w \in W$ tel que $\{v, w\} \in E$. Alors on peut obtenir un chemin alternant de u à v soit en enlevant l'arête $\{v, w\}$ du chemin alternant de u à w si $\{v, w\} \in C$, soit en ajoutant l'arête $\{v, w\}$ au chemin alternant de u à w si $\{v, w\} \notin C$. Ainsi, $|W| - 1 \geqslant |V(W)|$, CQFD.

■

Nous sommes maintenant en mesure de démontrer le théorème de Birkhoff - von Neumann.

Définition 4.3:
- On note $\mathcal{P}_n$ l'ensemble des matrices de permutation de taille n, i.e. l'ensemble des matrices $P_\sigma := \left(\mathbb{1}_{\sigma(i)=j}\right)_{i,j \in [\![1,n]\!]}$ où $\sigma \in S_n$.
- On note $\mathcal{B}_n$ l'ensemble des matrices bistochastiques de taille n, i.e. l'ensemble des matrices à coefficients positifs dont la somme des coefficients de chaque ligne et chaque colonne vaut 1.

Théorème 4.2 (de Birkhoff - von Neumann): $\mathcal{B}_n$ est l'enveloppe convexe de $\mathcal{P}_n$.

Preuve:
- On vérifie facilement que $\mathcal{B}_n$ est un convexe contenant $\mathcal{P}_n$.
- Soit $X = (x_{i,j}) \in \mathcal{B}_n$. On va utiliser le théorème de Hall pour justifier que l'on peut choisir n coefficients non nuls de la matrice X, de sorte qu'il y en ait exactement un par ligne et un par colonne. Considérons le graphe biparti $G = (L, C, E)$ où L est l'ensemble des lignes de X, C est l'ensemble des colonnes de X, et il y a une arête entre la ligne i et la colonne j si et seulement si $x_{i,j} \neq 0$.

  Soit $A \subseteq L$. Comme X est bistochastique, on a pour $\forall i \in A$, $\sum_{j \in V(A)} x_{i,j} = 1$. Du coup

  $$|A| = \sum_{i \in A} 1 = \sum_{i \in A} \sum_{j \in V(A)} x_{i,j}$$

  et on en déduit

  $$|V(A)| = \sum_{j \in V(A)} 1 = \sum_{j \in V(A)} \sum_{i=1}^{n} x_{i,j} \geqslant \sum_{j \in V(A)} \sum_{i \in A} 1 = |A|$$

D'après le théorème de Hall, le graphe G possède un couplage L-parfait, i.e. il existe un ensemble d'arêtes réalisant une bijection σ entre l'ensemble des colonnes et l'ensemble des lignes de X.

Soit $P = (p_{i,j}) = P_\sigma$. Soit $\lambda = \min\{x_{i,j}; p_{i,j} = 1\} \in [0,1]$. Si $\lambda = 1$ alors $p_{i,j} = 0 \Longrightarrow x_{i,j} = 1$ et comme x est bistochastique, on en déduit $X = P$. Dans ce cas, il n'y rien à démontrer. On suppose maintenant $\lambda < 1$ et on considère $X' = (x'_{i,j}) = \frac{X - \lambda P}{1 - \lambda}$.

Vérifions que X' est bistochastique. Déjà, il est clair que la somme des coefficients de chaque ligne et de chaque colonne vaut 1. Il reste à montrer que X' est à coefficients positifs. Soient $i, j \in [\![1, n]\!]$.

- Si $x_{i,j} = 0$ alors $\{i, j\}$ n'est pas une arête du graphe G, donc $i \neq \sigma(j)$ et donc $p_{i,j} = 0$. Dans ce cas, on a donc $x'_{i,j} = 0$.
- Si $x_{i,j} > 0$ et $p_{i,j} > 0$ alors $x'_{i,j} \geqslant 0$ par définition de λ.
- Si $x_{i,j} > 0$ et $p_{i,j} = 0$ alors $x'_{i,j} = 0$.

Ainsi, X' est bien bistochastique. De plus, on a vu que si $x_{i,j} = 0$ alors $x'_{i,j} = 0$, et si i_0, j_0 sont tels que $x_{i_0,j_0} = \lambda$ et $p_{i_0,j_0} = 1$ alors $x'_{i_0,j_0} = 0$. Ainsi X' a au moins un coefficient nul de plus que X, et X s'écrit comme une combinaison convexe de P et X'. En itérant le procédé, on obtient que X s'écrit comme combinaison convexe de matrices de permutations.

■

Récapitulons. Soient $x \in {\mathbb{R}_+^*}^n$ et $p, q \in \mathbb{R}^n$ des vecteurs décroissants tels que $p \succeq q$. D'après le Théorème 3.1, il existe une matrice bistochastique B telle que $q = Bp$. D'après le Théorème 4.2, on peut écrire $B = \sum_{\sigma \in S_n} c_\sigma P_\sigma$ où les $(c_\sigma)_{\sigma \in S_n}$ sont des réels positifs de somme 1, ce qui démontre le Théorème 2.2. Avec la fin de la Chapitre 2, on a bien démontré l'inégalité de Muirhead !