T. D. n° 1 Introduction et questions d'optimalité

Exercice 1. Plans pour pesée

1. **(Inégalité de Hadamard.)**
Soit $\boldsymbol{A}$ une matrice semi-définie positive et a_{11}, ..., a_{nn} les éléments de sa diagonale principale. Montrez l'inégalité suivante :
$$|\boldsymbol{A}| \leqslant a_{11} \cdots a_{nn}.$$
où $|\boldsymbol{A}|$ désigne le déterminant de la matrice carrée $\boldsymbol{A}$.
Montrez que si $\boldsymbol{A}$ est définie positive alors l'égalité est réalisée dans l'inégalité ci-dessus si et seulement si $\boldsymbol{A}$ est une matrice diagonale.
2. Soit $\boldsymbol{X}$ une matrice carrée d'ordre n dont les coefficients x_{ij} vérifient $|x_{ij}| \leqslant 1$, pour tout $1 \leqslant i, j \leqslant n$. Montrez que :
$$|\boldsymbol{X'X}| \leqslant n^n.$$

Une matrice $\boldsymbol{X}$ est appelée matrice de Hadamard si ses coefficients appartiennent tous à $\{-1; 1\}$ et si $\boldsymbol{X'X} = n\boldsymbol{I_n}$.

3. Montrer qu'une matrice réelle $\boldsymbol{X}$ carrée d'ordre n dont les coefficients x_{ij} vérifient $|x_{ij}| \leqslant 1$, pour tout $1 \leqslant i, j \leqslant n$ atteint l'inégalité de Hadamard $|\boldsymbol{M}| \leqslant n^{n/2}$ si et seulement si $\boldsymbol{M}$ est une matrice de Hadamard.

Supposons que nous devions peser n objets avec une balance ayant deux plateaux en réalisant au maximum n pesées. À chacune des pesées, nous pouvons répartir tous ou une partie des objets sur l'un ou l'autre des deux plateaux de la balance. Cette répartition des objets est appelée un *plan pour pesée*. Soit β_1, β_2, ..., et β_n les poids, inconnus, des objets. Nous posons $x_{ij} = 1$ si le $j-$ème objet est dans le plateau de droite au cours de la $i-$ème pesée, $x_{ij} = -1$ si le $j-$ème objet est dans le plateau de gauche au cours de la $i-$ème pesée et $x_{ij} = 0$ si le $j-$ème objet n'est pas utilisé lors de la $i-$ème pesée. Afin d'équilibrer les deux plateaux de la balance lors de la $i-$ème pesée il est possible de rajouter une masse à l'un des deux plateaux de la balance. Le signe de cette masse dépend du plateau sur lequel elle a été posée : positif si elle se trouve sur le plateau de gauche et négatif si elle se trouve sur le plateau de droite. Nous désignons par y_i cette masse signée utilisée lors de la $i-$ème pesée.
Nous aboutissons ainsi naturellement à un modèle linéaire :
$$\mathbb{E}\left[\boldsymbol{y}\right] = \boldsymbol{X\beta}$$

où $\boldsymbol{X} = (x_{ij})_{1\leqslant i,j\leqslant n}$ est fixe, $\boldsymbol{y} = (y_1, \ldots, y_n)$ aléatoire, à cause des erreurs de mesure, et $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$.
Nous supposons que les composantes y_i ne sont pas corrélées et de même variance σ^2.

4. En faisant l'hypothèse supplémentaire que le modèle est régulier, donnez $\widehat{\boldsymbol{\beta}}$ le MELSB (BLUE) de $\boldsymbol{\beta}$ et montrez que sa dispersion $\mathbb{D}\left[\widehat{\boldsymbol{\beta}}\right]$ est égale à $\sigma^2 \left(\boldsymbol{X'X}\right)^{-\mathbf{1}}$.

Il apparaît donc que pour rendre le plan pour pesée le plus précis possible, c'est-à-dire rendre « petite » la variance de l'estimateur $\widehat{\boldsymbol{\beta}}$, il est judicieux de rendre le produit $\boldsymbol{X'X}$ « grand ». Une manière naturelle de mesurer la taille d'une matrice semi-définie positive est de calculer son déterminant. Il faut ainsi construire un plan pour pesée, c'est-à-dire une matrice $\boldsymbol{X}$, telle que la valeur du déterminant de $\boldsymbol{X'X}$ soit maximale. Une telle matrice, si elle existe, est qualifiée de $D-$optimale. Dans le cas d'un modèle linéaire gaussien, cette mesure sera directement liée au volume de l'ellipsoïde de confiance construit autour des paramètres du modèle.

5. Considérons quatre objets et quatre pesées à réaliser. Montrer que, dans ce contexte, la matrice $\boldsymbol{X_1} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$ définit un plan pour pesée $D-$optimal.

6. En utilisant la matrice $\boldsymbol{X_1}$, montrez qu'il existe des plans pour pesée $D-$optimaux dès que $n = 2^k$ avec $k \in \mathbb{N}^\star$.

L'utilisation des corps finis a permis à R. Paley de trouver des procéder de construction de matrices de Hadamard dans les deux cas suivants :
- Si q est une puissance d'un nombre premier congru à 3 modulo 4, il existe une matrice d'Hadamard d'ordre $q + 1$.
- Si q est une puissance d'un nombre premier congru à 1 modulo 4, il existe une matrice d'Hadamard d'ordre $2(q + 1)$.

7. En déduire qu'il existe des plans pour pesée $D-$optimaux dès que $n = 2^k$, $k \in \mathbb{N}$, $n = 1 + q^k$ avec q un nombre premier tel que $q = 3 + 4l$ et $k \in \mathbb{N}$ ou dès que $n = 2(q^k + 1)$ avec q un nombre premier tel que $q = 1 + 4l$ et $k \in \mathbb{N}$.

Conjecture de Hadamard
La question ouverte la plus importante à propos des matrices de Hadamard est celle de leur existence. D'après la conjecture de Hadamard, une matrice de Hadamard d'ordre $4k$ existe pour tout entier positif k. Le plus petit ordre multiple de 4 pour lequel aucune matrice de Hadamard n'est connue est 668. Ces matrices sont répertoriées dans des tables qu'il suffit de consulter pour obtenir un plan pour pesée $D-$optimal.

Exercice 2. Cas de la régression linéaire simple

1. Soit A une matrice carrée symétrique d'ordre 2, $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$. Montrer que A est semi-définie positive si, et seulement si, $a \geqslant 0$, $d \geqslant 0$ et $\det(A) \geqslant 0$.

2. Dans le contexte de l'exemple 1 sur le modèle de régression linéaire simple et gaussienne et en utilisant le résultat précédent, montrer qu'il n'existe pas de plan $\boldsymbol{u}_\star$ qui soit le « meilleur » au sens suivant, ordre de Loewner : $M(\boldsymbol{u}_\star) - M(\boldsymbol{u})$ est semi-définie positive pour tout plan $\boldsymbol{u}$.

Exercice 3. Régression quadratique à une variable

1. Calculer la matrice des moments du plan η_0 introduit dans le cours.

2. Vérifier que le plan η_0 est $D-$optimal.

Exercice 4. Algorithme de Fedorov

Il existe plusieurs algorithme permettant de rechercher numériquement des plans optimaux. Les plus connus sont ceux de Wynn et Fedorov. La bibliothèque `AlgDesign` du langage `R` contient, avec la fonction `optFederov`, une implémentation de l'algorithme de Fedorov pour la recherche de plans D-, A- ou I-optimaux.

1. Installer la bibliothèque.

2. Afficher la vignette avec la commande `vignette("AlgDesign")`.

3. Étudier, en les reproduisant avec `R`, les exemples proposés à la section 4, page 15.

4. Lire l'annexe A, page 47, qui expose le fonctionnement de l'algorithme de Fedorov.

Exercice 5. Proc Optex

La Proc Optex du système SAS permet de construire des plans optimaux pour les critères D et A à l'aide de plusieurs algorithmes dont celui de Fedorov.

1. Consulter l'aide la Proc Optex dans SAS.

2. Étudier, en les reproduisant avec SAS, les exemples proposés.

Exercice 6. $D-$optimalité des plans factoriels complets 2^k et fractions 2^{k-p}

1. Montrer que, pour tout $x \in [-1; 1]^k$, $x'x \leqslant k$.

2. En utilisant la caractérisation des plans $D-$optimaux

$$x' \left[M(\eta_\star)\right]^{-1} x \leqslant 2^k, \quad \forall x \in \mathcal{X}$$

montrer que les plans factoriels complets sont $D-$optimaux pour l'estimation du modèle saturé à k facteurs ayant 2 modalités.

3. Déduire de l'inégalité de Hadamard, voir Exercice **1.**, que, pour un nombre d'essai égal à $m2^k$, les plans 2^k complètement répliqués maximisent le critère de $D-$optimalité d'un modèle saturé à k facteurs à 2 modalités.

4. En utilisant la caractérisation des plans $D-$optimaux

$$x' \left[M(\eta_\star)\right]^{-1} x \leqslant 2^{k-p}, \quad \forall x \in \mathcal{X}$$

montrer que les fractions régulières 2^{k-p} de plans factoriels complets sont $D-$optimaux pour l'estimation de n'importe quel modèle saturé à k facteurs ayant 2 modalités.