Une introduction aux plans d'expériences

1 Modèle linéaire

Pour n un entier naturel non nul, on considère le modèle linéaire :

$$Y_i = a^* + b^* f(x_i) + \varepsilon_i,\ i = 1, \cdots, n,$$

où, pour $i = 1, \cdots, n$, $x_i \in [0, 1]$ et les variables (ε_i) sont i.i.d. de loi $\mathcal{N}(0, \sigma^{*2})$. On suppose que f est une fonction continue non constante sur $[0, 1]$.

1) Que faut-il supposer sur $(x_i)_{i=1,\ldots,n}$ pour que le modèle soit identifiable? A partir de maintenant, on supposera que cette condition est vérifiée et on notera Σ l'ensemble des plans d'expériences qui vérifient cette condition. Soit $\theta^* = (a^*, b^*)^T$. Donner l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ^*. Que vaut la matrice de variance covariance de cet estimateur? On note Γ_{θ^*} cette matrice.

2) Afin de fixer les points $x = (x_i)_{i=1,\ldots,n}$ du plan d'expériences, on considère le critère :

$$D(x) = \frac{1}{\det \Gamma_{\theta^*}}.$$

Calculer D en fonction de x.

3) Montrer qu'il existe $m < M$ tels que la fonction f prenne toutes les valeurs de $[m, M]$. On pose

$$\tilde{D}(y) = D(x) \text{ où } y = (y_i)_{i=1,\ldots,n},\ y_i = f(x_i),\ i = 1, \ldots, n.$$

Montrer que

$$\max_{x \in \Sigma} D(x) = \max_{y \in \tilde{\Sigma}} \tilde{D}(y),$$

où

$$\tilde{\Sigma} = \left\{ y :\ y_i \in [m, M]\ i = 1, \ldots, n,\ \exists i, i' \in \{1, \ldots, n\}, i \neq i', y_i \neq y_{i'} \right\}$$

4) Montrer que

$$\tilde{D}(y) = C_n \sum_{i=1}^{n} (y_i - \overline{y})^2,$$

où $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ et C_n ne dépend que de n. En déduire que

$$\max_{y \in \tilde{\Sigma}} \tilde{D}(y) = \max_{y: y_i \in [m,M],\ i=1\ldots n} \tilde{D}(y)$$

Soit y^* avec $\max_{y: y_i \in [m,M],\ i=1\ldots n} \tilde{D}(y) = \tilde{D}(y^*)$. Supposons qu'il existe un indice $i^* \in \{1, \cdots, n\}$ avec $m < y^*_{i^*} < M$ et considérons la fonction φ sur $]m, M[$ définie, pour $y \in]m, M[$, par

$$\varphi(y) = \tilde{D}(\tilde{y}^*),\ \text{avec, pour } i \neq i^*\ \tilde{y}^*_i = y^*_i \text{ et } \tilde{y}^*_{i^*} = y.$$

Calculer la fonction φ et étudier ses variations sur $]m, M[$. En déduire que l'hypothèse faite sur $y^*_{i^*}$ est incorrecte et que, pour $i = 1, \ldots, n$, $y^*_i \in \{m, M\}$.

5) On pose

$$r^* = \mathrm{Card}\left\{i \in \{1,\dots,n\} : \ y_i^* = m\right\}.$$

On suppose que $n = 2p$ ($p \in \mathbb{N}^*$), que vaut r^*? Dans le cas $n = 2p+1$, quelles sont les valeurs possibles pour r^*.

6) Décrire tous les plans d'expériences D-optimaux.

7) On se place dans le cas où $f(x) = 2\pi x \sin(2\pi x)$. Déterminer précisement les plans D-optimaux.

2 Carré Latin

On se propose de construire un générateur aléatoire de carré latin (de taille $n > 1$). Pour cela, on va utiliser les remarques qui suivent.

- Le carré suivant est latin :

$$\begin{array}{cccc} 1 & 2 & \cdots & n \\ 2 & 3 & \cdots & 1 \\ \vdots & & & \vdots \\ n & 1 & \cdots & n-1 \end{array}$$

- Si l'on dispose d'un carré latin, on construit un nouveau carré latin en permuttant ses lignes et/ou ses colonnes.

En utilisant la procédure `SCILAB grand`, construire une fonction qui génère un carré latin aléatoire.

3 D et E optimalités

On considère, pour $x \in [-1,1]$, le modèle de régression :

$$Y(x) = \sum_{i=1}^{k} a_i^* f_i(x) + \varepsilon(x).$$

Les fonctions de régression $(f_i)_{i=1,\dots,k}$ sont données.

1. On se place dans le cas o $k = 3$ et $f_i(x) = x^{i-1}, i = 1,2,3$. On s'intéresse aux plans d'expériences à $n = 15$ points. En utilisant l'optimiseur de `SCILAB`, construire les plans D et E optimaux.

2. En utilisant les polynômes de Legendre sous `SCILAB`, tracer la courbe représentative de la fonction $(x^2-1)P_2'(x)$ o P_i désigne le polynme de Legendre de degré $i \geq 0$. Vérifier alors le plan D-optimal trouvé en 1).

3. On suppose encore que $n = 6$, mais on s'intéresse ici à $k = 3$ avec pour $x \in [-1,1]$, $f_1(x) = 1$, $f_2(x) = \exp(x)$, $f_3(x) = x^2$. Ecrire un code qui calcule le plan D-optimal. Comparer la variance de ce plan à celle obtenue en utilisant les points $-1, 1, 1/2, -1, 1, 1/2$.

4 D-optimalité en dimension 2

Sur $[-1,1]^2$, on considère le modèle de régression :

$$Z(x,y) = a^* + b^* x + c^* y + d^* x^2 + e^* y^2 + f^* xy + \varepsilon(x,y).$$

Utiliser l'optimiseur de `SCILAB` pour déterminer un plan D-optimal à 20 points. Représenter ce plan dans l'espace à deux dimensions.