Examen de Plans d'expérience du 13 Novembre 2014
Durée : 2h, 13h30-15h30

Notes de cours autorisées.
Les résultats devront être justifiés. Les calculatrices sont autorisées

Problème

On considère le modèle de régression :

$$Y(x_1, x_2) = a_1^* x_1 + a_2^* x_2 + \varepsilon, \quad x_1, x_2 \in \{-1, 0, 1\}.$$

On observe, pour $n \geq 2$ et $j = 1, \cdots, n$,

$$Y_j = a_1^* x_{1,j} + a_2^* x_{2,j} + \varepsilon_j.$$

Où, les variables $\varepsilon_j, \ j = 1, \cdots, n$ sont i.i.d. de loi $\mathcal{N}(0, \sigma_*^2)$ avec $\sigma_*^2 > 0$ et $x_{1,j}, x_{2,j} \in \{-1, 0, 1\}$.

1. À quelles conditions sur le plan d'expériences $(x_{i,j})_{i=1,2, j=1,\cdots,n}$ le modèle linéaire précédent satisfait la propriété classique d'identifiablilé de ses paramètres ? On appelle $\mathcal{P}$ l'ensemble des plans d'expériences satisfaisant ces propriétés. Mettre le modèle linéaire sous la forme habituelle $Y = X\theta^* + \varepsilon$.
2. On pose $A := X^T X$ et l'on note $0 \leq \lambda_1 \leq \lambda_2$ les valeurs propres de A.
   (a) On propose les critères d'optimalité d'un plan d'expériences suivants :

   $$\begin{aligned} F(A) &:= \lambda_1^2 + \lambda_2^2, \\ G(A) &:= \lambda_1 \lambda_2, \\ H(A) &= \lambda_1. \end{aligned}$$

   Expliquer pourquoi la maximisation de ces critères sur $\mathcal{P}$ est une stratégie raisonnable pour bâtir un plan d'expériences.

   (b) Montrer que

   $$\begin{aligned} F(A) &= \sum_{j=1}^{n} (x_{1,j}^2 + x_{2,j}^2), \\ G(A) &= \sum_{j=1}^{n} x_{1,j}^2 \sum_{j=1}^{n} x_{2,j}^2 - \left(\sum_{j=1}^{n} x_{1,j} x_{2,j} \right)^2 \end{aligned}$$

   i. Soit $y = (y_j) \in \mathbb{R}^n$, on pose $\|y\|_0 := \sum_{j=1}^{n} \mathbf{1}_{(y_j \neq 0)}$. Exprimer F à l'aide de la norme $\|\cdot\|_0$. Décrire tous les plans de $\mathcal{P}$ qui maximisent F.

   ii. Montrer quer $G(A) \leq n^2$. En déduire tous les plans de $\mathcal{P}$ qui maximisent G.

   (c) On suppose maintenant que pour $j = 1, \cdots, n$ les vecteurs $\begin{pmatrix} x_{1,j} \\ x_{2,j} \end{pmatrix}$ sont aléatoires et i.i.d.. On pose, pour $\alpha, \beta \in \{-1, 0, 1\}$,

   $$\mathbb{P}(x_{1,j} = \alpha, x_{2,j} = \beta) = p_{\alpha\beta}.$$

i. Calculer $\mathbb{E}(x_{i,j}^2),\ i=1,2,\ j=1,\cdots,n$.

ii. Montrer que

$$\mathbb{E}(F(A)) = n\left(2 - [p_{0-1} + 2p_{00} + p_{01} + p_{-10} + p_{10}]\right).$$

iii. Quelles sont les lois sur $\{-1,0,1\}^2$ qui maximisent $\mathbb{E}(F(A))$? Commenter ce résultat.

iv. On suppose que la loi du couple $(x_{1,j}, x_{2,j})^T$ maximise $\mathbb{E}(F(A))$. Montrer que

$$\left(\sum_{j=1}^{n} x_{1,j}x_{2,j}\right)^2 = \sum_{j=1}^{n} x_{1,j}^2 x_{2,j}^2 + 2 \sum_{1\leq j_1<j_2\leq n} x_{1,j_1}x_{2,j_1}x_{1,j_2}x_{2,j_2}.$$

En déduire que

$$\mathbb{E}(G(A)) = n(n-1)\left(1 - [\mathbb{E}(x_{1,1}x_{2,1})]^2\right).$$

v. Montrer que parmi les lois du couple $(x_{1,j}, x_{2,j})^T$ qui maximisent $\mathbb{E}(F(A))$, celles qui maximisent aussi $\mathbb{E}(G(A))$ satisfont

$$p_{11} + p_{-1-1} - p_{-11} - p_{1-1} = 0.$$

En déduire que l'on peut écrire toutes ces lois sous la forme

$$\begin{aligned}
p_{11} &= \theta_1,\ 0 \leq \theta_1 \leq \frac{1}{2},\\
p_{-1-1} &= \frac{1}{2} - \theta_1,\\
p_{10} &= p_{-10} = p_{01} = p_{0-1} = 0\\
p_{-11} &= \theta_2,\ 0 \leq \theta_2 \leq \frac{1}{2},\\
p_{1-1} &= \frac{1}{2} - \theta_2.
\end{aligned}$$

On suppose de plus que $x_{1,1}$ est indépendant de $x_{2,1}$. Quelles sont alors les lois du couple $(x_{1,j}, x_{2,j})^T$ qui maximisent conjointement $\mathbb{E}(F(A))$ et $\mathbb{E}(G(A))$?