TD5 : Modèles hiérarchiques

# 1 Un cas pratique

On considère $N$ élèves du lycée d'une ville appartenant à $M$ lycées différents.
Le niveau d'un élève $y_{i,j}$, $j$ème élève du lycée $i$ peut être modélisé par une variable réelle $y_{i,j} = \theta_i + \varepsilon_{i,j}$ où $\theta_i$ est le niveau moyen du lycée indicé par $i$ et $\varepsilon_{i,j}$ est une variable gaussienne $\mathcal{N}(0,1)$ et où on suppose que les $\theta_i$ sont des variables i.i.d. suivant une loi $\mathcal{N}(\mu, \tau^2)$, où $\tau^2$ est connu mais où $\mu$ est un hyperparamètre sur lequel on peut mettre un *a priori* .
Si on met un *a priori* uniforme sur $\mu$, (la mesure de Lebesgue sur $\mathbb{R}$ est impropre mais peut être un *a priori* valide si la loi de $\theta$ est gaussienne), on peut estimer la loi *a posteriori* de la moyenne $\theta_j$ du lycée $j$ à partir des élèves de ce lycée. Nous allons voir que proposer un modèle hiérarchique c'est à dire, supposer que les différentes moyennes des lycées sont des variables aléatoires suivant une même loi, induit une corrélation sur les $(\theta_j)_{j\leqslant M}$ et que les estimations des différents $\theta_j$ vont faire intervenir, le niveau des élèves des autres lycées. Les deux cas extrêmes sont

- Le cas limite où $\tau = 0$ c'est à dire où on suppose que les niveaux des différents lycées sont tous identiques. L'estimation des différents $\theta_i$ utilisera de la même manière le niveaux de tous les élèves de tous les lycées.
- Le cas limite où $\tau$ tend vers $+\infty$, le niveau des différents lycées est très hétérogènes et dans ce cas, l'estimation de $\theta_j$ prend essentiellement en compte le niveau des élèves du lycée $j$.

## 1.1 Questions préliminaires

1. Soit $f$ la fonction définie de $\mathbb{R}^2$ dans $\mathbb{R}$ par $f(x_1, x_2) = ax_1^2 + 2cx_1x_2 + bx_2^2$, où $(a, b, c) \in \mathbb{R}^3$. On pose $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Déterminer $A$, matrice symétrique telle que $f(x_1, x_2) = X^tAX$.

   À quelle condition sur $a$, $b$ et $c$ la matrice $A$ est-elle inversible ?

2. Soit $A$ une matrice symétrique et inversible, $(\alpha, \beta) \in \mathbb{R}^2$ et $g(x_1, x_2) = \alpha x_1 + \beta x_2$. Montrer qu'il existe un vecteur $P \in \mathbb{R}^2$ tel que $g(x_1, x_2) = P^tX + X^tP$. En déduire qu'il existe un vecteur $M$ tel que $g(x_1, x_2) = M^tAX + X^tAM$

3. Soit $f$ la fonction définie de $\mathbb{R}^2$ dans $\mathbb{R}$ par $f(x_1, x_2) = ax_1^2 + 2cx_1x_2 + bx_2^2 + \alpha x_1 + \beta x_2 + \gamma$, où $(a, b, c, \alpha, \beta, \gamma) \in \mathbb{R}^6$ tels que $ab - c^2 > 0$. Montrer qu'il existe $A$ matrice symétrique inversible, $M \in \mathbb{R}^2$ et $\delta \in \mathbb{R}$ tels que

$$f(x_1, x_2) = (X - M)^tA(X - M) + \delta.$$

4. Pour quelles valeurs de $(x_1, x_2)$ cette fonction est-elle minimale ? En déduire un moyen de déterminer $M$ simplement.

5. Soit $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ un vecteur aléatoire dont la densité est de la forme

$$h(x_1, x_2) = Ke^{ax_1^2+2cx_1x_2+bx_2^2+\alpha x_1+\beta x_2+\gamma}.$$

Montrer que $X$ est un vecteur gaussien sous certaines conditions sur $a, b, c$ que l'on donnera.

6. À quelle condition les variables $x_1$ et $x_2$ sont-elles indépendantes ?

## 1.2 Retour sur le cas pratique

On va se placer dans le cas simple où $M = 2$, c'est à dire que l'on a deux lycées et on ne considère qu'un seul élève par lycée. On dispose ainsi de deux observations $y_{1,1}$ et $y_{2,1}$, la première $y_{1,1}$ tirée selon une loi $\mathcal{N}(\theta_1, 1)$, la seconde $y_{2,1}$ tirée selon une loi $\mathcal{N}(\theta_2, 1)$ et telles que $y_{1,1}$ et $y_{2,1}$ sont indépendantes conditionnellement à $(\theta_1, \theta_2)$. On suppose que les $\theta_i$ sont i.i.d. et tirés selon une loi $\mathcal{N}(\mu, \tau^2)$ où $\tau^2$ est supposé connu et on considère un *a priori* uniforme (mesure de Lebesgue sur $\mathbb{R}$) pour $\mu$.

1. Exprimer la loi de $\theta_1$ sachant $y_{1,1}$ et celle de $\theta_2$ sachant $y_{2,1}$.

2. Proposer un estimateur de $\theta_1$ et de $\theta_2$ à partir de ces lois *a posteriori* .

3. Justifier que
$$\pi(\theta_1, \theta_2 | y_{1,1}, y_{2,1}) \propto f(y_{1,1}, y_{2,1} | \theta_1, \theta_2) \pi(\theta_1, \theta_2).$$
où $\pi(\theta_1, \theta_2)$ est la loi *a priori* (éventuellement impropre du couple $(\theta_1, \theta_2)$).

4. Montrer que
$$\pi(\theta_1, \theta_2 | \mu) = \frac{1}{2\pi\tau^2} e^{-\frac{(\mu - \frac{\theta_1+\theta_2}{2})^2}{\tau^2} - \frac{(\theta_1-\theta_2)^2}{4\tau^2}}$$

5. Donner une expression de la *loi*
$$\pi(\theta_1, \theta_2) = \int_{\mu \in \mathbb{R}} \pi(\theta_1, \theta_2 | \mu) \pi(\mu) d\mu.$$

6. Pourquoi cette loi est-elle impropre ?

7. Justifier que
$$f(y_{1,1}, y_{2,1} | \theta_1, \theta_2) \pi(\theta_1, \theta_2) \propto e^{-\frac{(y_{1,1}-\theta_1)^2}{2}} e^{-\frac{(y_{2,1}-\theta_2)^2}{2}} e^{-\frac{(\theta_1-\theta_2)^2}{4\tau^2}}$$
et justifier que la loi de $(\theta_1, \theta_2)$ sachant $(y_{1,1}, y_{2,1})$ est bien une loi propre et la reconnaître.

8. Déterminer l'espérance et la matrice de covariance de cette loi. On pourra montrer que les composantes du vecteur $\mathbb{E}\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix} = \begin{pmatrix}m_1 \\ m_2\end{pmatrix}$ vérifient le système
$$\begin{cases} \left(1 + \frac{1}{2\tau^2}\right) m_1 & - & \frac{1}{2\tau^2} m_2 & = & y_{1,1} \\ -\frac{1}{2\tau^2} m_1 & + & \left(1 + \frac{1}{2\tau^2}\right) m_2 & = & y_{2,1} \end{cases}$$

9. Les variables $\theta_1$ et $\theta_2$ sont-elles indépendantes conditionnellement à $y_{1,1}$ et $y_{2,1}$ ?

10. Quelle est la limite de l'espérance de $\theta_1$ sachant $y_{1,1}$ et $y_{2,1}$ quand $\tau^2$ tend vers 0 ?

11. Même question quand $\tau^2$ tend vers $+\infty$.