HMMA307

Modèles linéaires (un peu) mixtes

Cours: Joseph Salmon *Scribes: Mohamed Akli Rabia, Samuel Kaci, Ali Abed Alsater*

1 Rappel sur le modèle

Le modèle est le suivant :

$$y_{i,j} = \mu + A_j + \varepsilon_{i,j} \quad , \tag{1}$$

- μ est déterministe,
- $A_j \sim \mathcal{N}(0, \sigma_A^2)$ est un effet aléatoire,
- $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$,
- $n = \sum_{j=1}^J n_j$, où n_j est le nombre d'observations ayant la modalité j.

Formulation Matricielle du modèle : On observe le vecteur :

$$y = \mu \mathbb{1} + ZA + \varepsilon \quad , \tag{2}$$

avec

$$Z = [\mathbb{1}_{C_1}, \dots, \mathbb{1}_{C_J}], A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_J \end{pmatrix} \in \mathbb{R}^J, \varepsilon \in \mathbb{R}^n \quad . \tag{3}$$

Comme dans les cours précédents on note :

$$\overline{y}_{:,j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{i,j} \quad . \tag{4}$$

Espérance de $\overline{y}_{:,j}$:

$$\mathbb{E}[\overline{y}_{:,j}] = \mu + 0 + 0 = \mu \quad . \tag{5}$$

Variance de $\overline{y}_{:,j}$:

$$\mathrm{Var}[\overline{y}_{:,j}] = \sigma_A^2 + \frac{\sigma_\varepsilon^2}{n_j} := \tau_j^2 \quad . \tag{6}$$

Estimateur de μ :

$$\hat{\mu} = \sum_{j=1}^{J} w_j \overline{y_{:j}} \ \text{ est un estimateur possible de } \mu\text{, avec } \ w_j = \frac{\frac{1}{\tau_j^2}}{\sum_{j'=1}^{J} \frac{1}{\tau_{j'}^2}} \quad . \tag{7}$$

Remarque 1.1. *Cas équilibré, $\forall j \in [\![1, J]\!], n_j = I$ et donc $n = IJ$.*

$$\tau_j^2 = \sigma_A^2 + \frac{\sigma_\varepsilon^2}{I} \implies w_j = \frac{1}{n}, \forall j \in [\![1, J]\!] \ . \tag{8}$$

Notons cependant que σ_A^2 et σ_ε^2 sont en général inconnues.

Théorème 1.1. *Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes telles que $\mu = \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n]$, parmi les estimateurs linéaires (en $X_1, \dots, X_n$), sans biais de μ, celui de variance minimale est donné par*

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \frac{1}{\mathrm{Var}[X_i]} X_i}{\sum_{i'=1}^{n} \frac{1}{\mathrm{Var}[X_{i'}]}} \quad . \tag{9}$$

Démonstration.

$$\begin{cases} \min_{\alpha_1,\dots,\alpha_n} \mathrm{Var}[\sum_{i=1}^{n} \alpha_i X_i] \\ \text{s.c} : \displaystyle\sum_{i=1}^{n} \alpha_i = 1 \end{cases} \iff \begin{cases} \min_{\alpha_1,\dots,\alpha_n} \sum_{i=1}^{n} \alpha_i^2 \underbrace{\mathrm{Var}\left[X_i\right]}_{\sigma_i^2} \\ \text{s.c} : \sum_{i=1}^{n} \alpha_i = 1 \end{cases} \tag{10}$$

On obtient le Lagrangien :

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 + \lambda \left(\sum_{i=1}^{n} \alpha_i - 1 \right) \quad . \tag{11}$$

On résout alors :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \alpha_i} & = 0 \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} & = 0 \end{cases} \iff \begin{cases} \alpha_i \sigma_i^2 + \lambda & = 0 \\ \sum_{i=1}^{n} \alpha_i & = 1 \end{cases}$$

Enfin le précédent système est équivalent à :

$$\begin{cases} \alpha_i = \dfrac{-\lambda}{\sigma_i^2} & \text{(12a)} \\ \displaystyle\sum_{i=1}^{n} \alpha_i = 1 & \text{(12b)} \end{cases}$$

En remplaçant α_i dans (12b) on obtient : $-\sum_{i=1}^{n} \frac{\lambda}{\sigma_i^2} = 1$ d'où $\lambda = -\frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$

Au final :

$$\boxed{\alpha_i = \frac{\frac{1}{\sigma_i^2}}{\displaystyle\sum_{i'=1}^{n} \frac{1}{\sigma_{i'}^2}}} \quad . \tag{13}$$

□

Il faut maintenant proposer des estimateurs de σ_A^2, σ_ε^2.

1.1 "Type 1" (Anova)

C'est une approche de type *pluggin / moments*.

$$\mathbb{E}\left[\frac{1}{n-J}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j}-\overline{y}_{:,j})^2\right]=\sigma_\varepsilon^2 \quad . \quad \text{(admis)} \tag{14}$$

(XXX il faut harmoniser les notations avec Anova.tex, somme i=1.. I , puis de $j = 1...n_1$)

On peut donc proposer :

$$\widehat{\sigma_\varepsilon^2}=\frac{1}{n-J}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j}-\overline{y}_{:j})^2 \quad . \tag{15}$$

$$\mathbb{E}\left[\frac{1}{J-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(\overline{y}_{:,j}-\overline{y}_n)^2\right]=\sigma_\varepsilon^2+\frac{1}{n(J-1)}\left[n^2-\sum_{i=1}^{J}n_j^2\right]\sigma_A^2 \quad . \quad \text{(admis)} \tag{16}$$

Posons :

$$\widehat{S}^2=\frac{1}{J-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(\overline{y}_{:,j}-\overline{y}_n)^2 \quad . \tag{17}$$

On peut prendre comme estimateur (sans biais) de σ_A^2 :

$$\widehat{\sigma}_A^2=\frac{(\widehat{S}^2-\widehat{\sigma_\varepsilon}^2)}{n^2-\sum_{j=1}^{J}n_j^2}n(J-1) \tag{18}$$

Est ce que cet estimateur est bien défini ?

$$n^2=\left(\sum_{j=1}^{J}(n_j)\right)^2=\sum_{j=1}^{J}(n_j^2)+a \tag{19}$$

(XXX détailler la valeur explicite de a ci-dessus) donc $n^2-\left(\sum n_j\right)^2=a>0$ (Sauf si $n_j=0$).

Mais il est possible que $\widehat{\sigma_A^2}$ soit négatif, en général on le remplace par 0 dans ce cas. On a donc : $\widehat{\sigma_A^2}\leftarrow max(0,\widehat{\sigma_A^2})$.

b) Maximum de vraisemblance

$$\begin{aligned}\operatorname{Var}(y)&=\operatorname{Var}(\mu+\mathbb{1}_n+ZA+\varepsilon)\\&=\operatorname{Var}(ZA)+\sigma_\varepsilon^2+\operatorname{Id}_n\\&=ZZ^\top\sigma_A^2+\sigma_\varepsilon^2\operatorname{Id}_n:=V\end{aligned}$$

Ainsi, le vecteur y suit donc une loi gaussienne $y\sim\mathcal{N}(\mu I_n,V)$: (XXX améliorer le texte ici)

$$\begin{aligned}L(y,\mu,\sigma_\varepsilon^2,\sigma_A^2)&=\frac{1}{(2\pi)^{\frac{n}{2}}}\frac{1}{\det(V)^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\left[(y-\mathbb{1}_n\mu)^\top V^{-1}(y-\mathbb{1}_n\mu)\right]\right]\\-\log(L)&=cte+\frac{1}{2}\log\left(\det(V)\right)+\frac{1}{2}\left(y-\mathbb{1}_n\mu\right)^\top V^{-1}\left(y-\mathbb{1}_n\mu\right) \quad .\end{aligned}$$

Pour trouver l'estimateur du maximum de vraisemblance il faut résoudre :

$$\operatorname*{arg\,min}_{\mu,\sigma_A^2,\sigma_\varepsilon^2\in\mathbb{R}} -\log\left(L(y,\mu,\sigma_\varepsilon^2,\sigma_A^2)\right) \quad . \tag{20}$$

Optimisation en μ

Remarque 1.2. *Calculer $\frac{1}{2}(y-(\mu+\delta)\mathbb{1}_n)^\top\ V^{-1}\ (y-(\mu+\delta)\mathbb{1}_n)$ avec $\delta\to 0$.*

$$\begin{aligned}\frac{\partial(-\log L)}{\partial\mu}=0 &\iff \mathbb{1}_n^\top V^{-1}(\hat\mu\mathbb{1}_n - y)=0\\ &\iff \hat\mu(\mathbb{1}_n^\top V^{-1}\mathbb{1}_n)=\mathbb{1}_n^\top V^{-1}y\\ &\iff \hat\mu=\frac{\mathbb{1}_n^\top V^{-1}y}{\mathbb{1}_n V^{-1}\mathbb{1}_n}\quad .\end{aligned}$$

C'est l'estimateur $\widehat{\mu}$ qu'on avait proposé avant.

Remarque 1.3. *L'optimisation en σ_A^2 et en σ_ε^2 n'admet pas de formule explicite.*

1.2 Modélisation linéaire mixte

On va maintenant mélanger effets fixes/effets aléatoires.

$$\begin{aligned}y &= X\beta+ZA+\varepsilon,\quad \text{avec}\quad A\perp\!\!\!\perp\varepsilon\\ &= X\beta+\sum_{j=1}^J Z_jA_j+\varepsilon\quad .\end{aligned}$$

Avec $y\in\mathbb{R}^n, X\in\mathbb{R}^{n\times q}$ et $Z=[Z_1,\dots,Z_J]\in\mathbb{R}^{n\times q}, A=\begin{pmatrix}A_1\\A_2\\\vdots\\A_J\end{pmatrix}$, puis pour tout$j\in[\![1,J]\!]$, $A_j\sim\mathcal{N}(0,\sigma_j^2\operatorname{Id}_{q_j})$, $\varepsilon\sim\mathcal{N}(0,\sigma_\varepsilon^2\operatorname{Id}_n), q=q_1+\cdots+q_J, y\sim\mathcal{N}(X\beta,V)$ avec $V\in\mathbb{R}^{n\times n}$ donnée par

$$\begin{aligned}V=\operatorname{Var}(y) &= \operatorname{Var}(\varepsilon)+\operatorname{Var}(ZA)\\ &= \sigma_\varepsilon^2\operatorname{Id}_n+\sum_{j=1}^J\operatorname{Var}(z_jA_j)\\ &= \sigma_\varepsilon^2\operatorname{Id}_n+\sum_{j=1}^J\sigma_j^2Z_jZ_j^\top\quad .\end{aligned}$$

De manière similaire :

$$-\log L(y,\beta,\sigma_\varepsilon^2,\sigma_1^2,\dots,\sigma_J^2)=cst+\log\det(V)+(y-X\beta)^\top V^{-1}(y-X\beta)\quad . \tag{21}$$

(XXX ajoute du texte ici.)

$$\hat\beta=\frac{\partial L}{\partial\beta}=0\iff \left(X^\top V^{-1}X\right)^{-1}X^\top V^{-1}y \tag{22}$$

Remarque 1.4. *On peut utiliser un solveur de type descente "par coordonnées" (ou par bloc) pour obtenir les quantités $\hat\sigma_\varepsilon^2,\hat\sigma_1^2,\dots,\hat\sigma_J^2$ et $\hat\beta$ (XXX ajoute du texte ici.)*