Table des matières

1 Le modèle linéaire mixte **2**
- 1.1 Du modèle linéaire au modèle linéaire mixte . . . 2
- 1.2 Définition de Rao et Kleffé : approche structurale . . . 6
  - 1.2.1 Exemples . . . 7
- 1.3 Définition de Lindley et Smith. Approche hiérarchique . . . 9
  - 1.3.1 Exemple des données des enfants . . . 10
- 1.4 Exemple de données de Transcriptome : données de nutrition chez la souris . . . 11

2 Estimation - Prédiction **13**
- 2.1 Estimation des effets fixes . . . 13
- 2.2 Prédiction des effets aléatoires . . . 14
  - 2.2.1 Le meilleur prédicteur (Best Predictor) . . . 14
  - 2.2.2 Meilleur prédicteur linéaire (Best Linear Predictor) . . . 15
  - 2.2.3 Meilleur prédicteur linéaire sans biais : le BLUP (Best Linear Unbiased Predictor) . . . 15
- 2.3 Estimation des composantes de la variance . . . 17

3 Tests d'hypothèses - Choix de modèles **19**
- 3.1 Tests sur effets fixes . . . 19
  - 3.1.1 V est connue . . . 20
  - 3.1.2 V est connue à un coefficient près . . . 20
  - 3.1.3 V est quelconque et inconnue . . . 20
  - 3.1.4 Test robuste . . . 21
- 3.2 Choix de modèles . . . 22
  - 3.2.1 Modèles linéaires classiques (modèles à effets fixes) . . . 22
  - 3.2.2 Modèles linéaires mixtes . . . 23
- 3.3 Tests sur composantes de la variance . . . 24
- 3.4 Discussion . . . 24

4 Modèle linéaire mixte et approche bayésienne **26**
- 4.1 Introduction à la statistique bayésienne . . . 26
- 4.2 Traitement bayésien du modèle linéaire . . . 27
- 4.3 Estimation des composantes de la variance . . . 29

5 Quelques Références **30**

Chapitre 1

Le modèle linéaire mixte

1.1 Du modèle linéaire au modèle linéaire mixte

L'analyse de variance, la régression linéaire et les modèles de covariance sont désormais d'usage courant. Ces trois méthodes découlent de la même théorie : la théorie du modèle linéaire. Dans ce modèle, les variables explicatives, qualitatives ou quantitatives, apparaissent sous la forme d'effets fixes. Ainsi lorsque cette variable est qualitative, on s'intéresse à l'effet particulier de chacun de ses niveaux sur la variable à expliquer. Cette façon de procéder suppose que l'on introduise dans le modèle tous les niveaux du facteur susceptibles d'avoir un intérêt. Mais cela n'est pas toujours possible. Par exemple, si on s'intéresse aux performances au champ d'une variété de blé, ou aux performances de croissance (ou production laitière) des animaux d'une race particulière, il est impossible de tester ces performances sur tous les champs ou animaux possibles (qui représenteraient tous les niveaux de la variables explicative). On peut également vouloir s'intéresser à l'effet d'un régime alimentaire sur la croissance des porcs, on ne pourra pas le tester sur tous les porcs. A chaque fois, pour réaliser l'expérience, il faudra prendre quelques individus (ici, des champs ou des porcs) et chercher à étendre les résultats obtenus à la population entière. Si on suppose que les individus ont été tirés au hasard dans la population, on ne s'intéresse plus à l'effet particulier associé à tel individu particulier, mais à la distribution de l'ensemble des effets possibles. L'effet associé à l'individu n'est plus un effet fixe mais devient un effet aléatoire et il faut en tenir compte dans l'analyse. Le modèle linéaire étudié contient un mélange d'effets fixes et d'effets aléatoires, on parle alors de modèle linéaire mixte. Le modèle linéaire mixte constitue une extension du modèle linéaire classique. D'une manière générale, on pourra y faire appel chaque fois que l'on désirera *étendre à une population toute entière des résultats obtenus sur quelques individus pris au hasard dans cette population.*

Définition : Un modèle linéaire mixte se définit comme un modèle linéaire dans lequel toute ou partie des paramètres associés à certaines unités expérimentales sont traités comme des variables aléatoires du fait de l'échantillonnage de ces unités dans une population plus large.

Dans le cadre du modèle linéaire mixte, les différents niveaux des effets fixes étant fixés une fois pour toutes, les effets associés sont des paramètres à estimer qui interviennent dans la moyenne du modèle. Les facteurs à effets aléatoires vont avoir, a priori, une grande quantité de niveaux, les observations réalisées correspondant à un nombre restreint de ces niveaux, pris aléatoirement. On va ainsi modéliser ces niveaux en tant qu'observations d'une variable aléatoire normale, de moyenne nulle (la moyenne du modèle sera définie par les effets fixes) et de variance inconnue à estimer. Chaque facteur à effets aléatoires sera donc caractérisé par un paramètre de variance qu'il faudra estimer en plus de la variance des erreurs du modèle. D'où le nom de composantes de la variance qu'on rencontre également pour de tels modèles. La nécessité d'estimer simultanément plusieurs paramètres de moyenne et plusieurs paramètres de variance dans

les modèles mixtes va compliquer la procédure d'estimation.

Exemple 1 : Supposons que l'on cherche à comparer 2 traitements A et B ; 4 élevages ont été sélectionnés pour participer à cet essai. Dans chaque élevage un échantillon d'animaux a été tiré au hasard, une moitié des animaux de l'échantillon ont reçu le traitement A et l'autre moitié le traitement B. Les données brutes ont été analysées et les analyses ont montré que le traitement B a une plus grande efficacité que le traitement A. Que peut-on conclure ? Pour répondre convenablement à cette question, il est nécessaire de préciser la nature du facteur élevage :

- si les élevages ont été choisis, le facteur élevage est un facteur fixe et les résultats de l'analyse ne peuvent pas être extrapolés à d'autres élevages,

- si les élevages ont été tirés au hasard parmi tous les élevages susceptibles d'utiliser ces produits, le facteur élevage est alors un facteur aléatoire et les résultats de cette analyse peuvent être extrapolés aux autres élevages.

Exemple 2 : En génétique quantitative, il n'est pas raisonnable de supposer que les effets de certains facteurs, comme l'effet de la mère sur les performances des descendants, est fixe. Il est supposé aléatoire, car cet effet peut être assimilé à un tirage aléatoire dans le génome de la mère. De façon plus générale, on considère que les effets d'un facteur sont aléatoires si les niveaux de ce facteur ne sont pas choisis mais tirés aléatoirement dans un ensemble de niveaux.

Exemple 3 : Le modèle linéaire mixte est largement utilisé dans le cadre de l'analyse de données longitudinales (plusieurs observations ont été mesurés sur le même individu au cours du temps) car il permet de prendre en compte des corrélations entre les observations. Si le nombre d'instants de mesures sur un individu est faible, on utilise le modèle mixte pour prendre en compte la dépendance entre observations issues du même individu. En revanche si le nombre d'instants de mesure n'est pas faible et si l'évolution dans le temps est le sujet d'intérêt, alors il faudra utiliser une autre catégorie de modèles, les modèles de séries chronologiques.

Exemple 4 : On a relevé les durées de gestation de 16 filles de 30 taureaux qui avaient été tirés au sort dans la population devant être étudiée. On voudrait savoir dans quelle mesure la durée de gestation est un caractère héréditaire - caractère se transmettant par les pères. Il s'agit de répondre, grâce à un échantillon comportant peu de taureaux, à une question concernant toute la population. Pour pouvoir étendre les résultats obtenus sur l'échantillon, il faut que celui-ci soit représentatif de toute la population et donc qu'il ait été obtenu par tirage au sort (indépendants et équiprobables). Il en découle que les taureaux de l'échantillon sont aléatoires et leurs effets sur leurs descendants sont a fortiori aléatoires.

Le modèle s'écrira

$$y_{ij} = \mu + a_i + e_{ij} \quad j = 1, ...16 \quad i = 1, ...30$$

où y_{ij} représente la durée de gestation de la fille j du père i,
μ est la moyenne générale,
a_i est l'effet du père i, supposé aléatoire puisque le père est un individu tiré aléatoirement,
e_{ij} est le résidu.

On suppose les distributions suivantes :

$$\begin{aligned} a_i &\sim N(0, \sigma_a^2) \\ e_{ij} &\sim N(0, \sigma_e^2) \end{aligned}$$

Les a_i et les e_{ij} sont supposés mutuellement indépendants.

On appelle σ_a^2 et σ_e^2 les composantes de la variance. La quantité $\frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$ est la part de variance "génétique" de la variance totale.

Il faut bien comprendre que $\mu + a_i$ n'est pas l'espérance de y, mais son espérance conditionnelle : $E(y_{ij}|a_i) = \mu + a_i$. De la même manière, la variance conditionnelle de y vaut $Var(y_{ij}|a_i) = \sigma_e^2$.

En appliquant les formules de conditionnement - déconditionnement, on peut calculer l'espérance et la variance de y :

$$\begin{aligned} E(y_{ij}) &= E_a(E_y(y_{ij}|a_i)) \\ &= E_a(\mu + a_i) \\ &= \mu \\ Var(y_{ij}) &= Var_a(E_y(y_{ij}|a_i)) + E_a(Var_y(y_{ij}|a_i)) \\ &= Var_a(\mu + a_i) + E_a(\sigma_e^2) \\ &= \sigma_a^2 + \sigma_e^2 \end{aligned}$$

On calcule aussi les covariances :
La covariance $Cov(y_{ij}, y_{ij'})$ représente la covariance entre 2 observations (durée de gestation) de 2 filles j et j' du même père i, alors que la covariance $Cov(y_{ij}, y_{i'j'})$ représente la covariance entre 2 observations (durée de gestation) de 2 filles j et j' issues de 2 pères différents i et i'.

$$\begin{aligned} Cov(y_{ij}, y_{ij'}) &= Cov_a(E_y(y_{ij}|a_i), E_y(y_{ij'}|a_i)) + E_a(Cov_{y|a}(y_{ij}, y_{ij'})) \\ &= Cov_a(\mu + a_i, \mu + a_i) + E_a(0) \\ &= \sigma_a^2 \\ Cov(y_{ij}, y_{i'j'}) &= Cov_a(E_y(y_{ij}|a_i), E_y(y_{i'j'}|a_{i'})) + E_a(Cov_{y|a}(y_{ij}, y_{i'j'})) \\ &= Cov_a(\mu + a_i, \mu + a_{i'}) + E_a(0) \\ &= 0 \end{aligned}$$

Deux observations de deux pères différents ne sont pas corrélées, alors que deux observations du même père sont corrélés. Le coefficient de corrélation vaut alors :

$$Cor(y_{ij}, y_{ij'}) = \frac{Cov(y_{ij}, y_{ij'})}{\sqrt{(var(y_{ij})var(y_{ij'}))}} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

Exemple 5 : Une compagnie pharmaceutique veut tester les performances d'une méthode de spectroscopie (NIR = Near Infrared Reflectance) permettant de déterminer le contenu en substance active de comprimés. La méthode utilisée en routine (HPLC) est plus coûteuse et laborieuse. Pour cela, 10 comprimés ont été tirés au hasard et les 2 méthodes ont été utilisées sur chacun des comprimés. La question posée est "'Existe-t-il une différence de performance entre les 2 méthodes testées HPLC et NIR ?"'. Le fichier de données est le suivant :

Comprimé	HPLC	NIR	Différence HPLC - NIR
1	10.4	10.1	0.3
2	10.6	10.8	-0.2
3	10.2	10.2	0.0
4	10.1	9.9	0.2
5	10.3	11.0	-0.7
6	10.7	10.2	0.2
7	10.3	10.2	0.1
8	10.9	10.9	0.0
9	10.1	10.4	-0.3
10	9.8	9.9	-0.1
	$\bar{y}_{HPLC} = 10.34$	$\bar{y}_{NIR} = 10.36$	$\bar{diff} = -0.05$
	$s_{HPLC} = 0.3239$	$s_{NIR} = 0.4033$	$s_{diff} = 0.2953$

1. L'analyse la plus simple consiste à considérer les données comme un échantillon apparié et d'utiliser le test de Student correspondant. La moyenne et l'écart-type des différences $diff$ sont égaux à $\bar{diff} = -0.05$ et $s_{diff} = 0.2953$ respectivement. La statistique du test de Student vaut $t = \frac{\bar{diff}}{SE_{\bar{diff}}} = \frac{-0.05}{0.2953/\sqrt{(10)}} = -0.535$ qui donne une p-value égale à 0.61. On en conclut qu'il n'y a pas différence significative entre les 2 méthodes de mesure. L'intervalle de confiance à 95% du biais de la méthode vaut $\bar{diff} \pm t_{(0.975;9)} * SE_{\bar{diff}} = -0.05 \pm 2.262 * 0.2953/\sqrt{(10)} = -0.05 \pm 0.21$; soit $[-0.26; 0.16]$.

Avec le logiciel R, on obtient

```
> t.test(d)

One-sample t-Test

data : d
t = -0.5354, df = 9, p-value = 0.6054
alternative hypothesis : true mean is not equal to 0
95 percent confidence interval :
-0.2612693 0.1612693
sample estimates :
mean of x
-0.05
```

2. ANOVA.

Le modèle d'analyse de variance pour cette situation s'écrit :

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \qquad e_{ij} \sim N(0, \sigma^2)$$

où y_{ij} représente le contenue en substance active du comprimé i avec la méthode j, μ représente la moyenne générale, α_i l'effet du $i$ème comprimé, et β_j l'effet de la méthode j.

La statistique du test F de Fisher est égal au carré de la statistique de Student

$$F = t^2 = (-0.535)^2 = 0.29$$

L'estimation de l'écart-type résiduel est donnée par $\hat{\sigma} = 0.209 = \sqrt{2} s_{diff}$. L'incertitude sur la moyenne des différences est donnée par $SE(\bar{y}_{HPLC} - \bar{y}_{NIR}) = \sqrt{\sigma^2(1/10 + 1/10)} = 0.0934$, exactement comme dans l'approche simple (1).

Si maintenant on s'intéresse à la précision de la valeur moyenne pour la méthode NIR, l'ANOVA donne $SE(\bar{y}_{NIR}) = \hat{\sigma}/\sqrt{10} = 0.066$. D'un autre côté, si on considère l'échantillon des 10 mesures de NIR, on obtient $s_{NIR} = 0.4012$, et donc $SE(\bar{y}_{NIR}) = s_1/\sqrt{10} = 0.127$, valeur très différente de 0.066.

L'ANOVA sous estime beaucoup l'incertitude sur l'estimation de l'effet moyen de NIR. C'est ainsi parce que la variance σ^2 mesure la variabilité résiduelle après que les effets des comprimés aient été corrigés. La différence conceptuelle entre les 2 approches est que l'ANOVA considère que les 10 comprimés n'ont pas été tirés au hasard, alors que la seconde (échantillon des 10 mesures NIR) si. L'ANOVA n'est valide que si l'on s'intéresse aux effets spécifiques des 10 comprimés.

L'idée du modèle mixte est de combiner les 2 approches, c'est à dire utiliser un modèle linéaire et y considérer certains facteurs comme aléatoires.

3. On considère maintenant le modèle linéaire mixte

$$y_{ij} = \mu + a_i + \beta_j + \varepsilon_{ij}$$

où a_i est l'effet aléatoire du $i$ème comprimé. Ses effets sont supposés indépendants et identiquement distribués : $a_i \sim N(0, \sigma_a^2)$.

On peut montrer que

	ANOVA	Modèle mixte :
$E(y_{ij})$	$\mu + \alpha_i + \beta_j$	$\mu + \beta_j$
$Var(y_{ij})$	σ^2	$\sigma^2 + \sigma_a^2$
$Cov(y_{ij}, y_{i'j'}), j \neq j'$	0	σ_a^2 si $i = i'$et 0 sinon

L'écart-type attendu de la moyenne des valeurs de NIR vaut dans le cadre du modèle linéaire mixte : $SE(\bar{y}_{NIR}) = \sqrt{\hat{\sigma}_a^2 + \hat{\sigma}^2}/\sqrt{10} = 0.115$, ce qui est conforme à ce que l'on attendait.

Ce modèle peut s'écrire sous forme matricielle :

$$y = X\beta + Za + \varepsilon$$

avec $y = (y_{1,HPLC}, ..., y_{10,HPLC}, y_{1,NIR}, ...y_{10,NIR})'$ le vecteur de dimension (20×1) des données,
$\beta = (\beta_{HPLC}, \beta_{NIR})'$ le vecteur de dimension (2×1) des effets fixes,
$a = (a_1, ...a_{10})'$ le vecteur de dimension (10×1) des effets aléatoires,
$\varepsilon = (\varepsilon_{1,HPLC}, ..., \varepsilon_{10,HPLC}, \varepsilon_{1,NIR}, ...\varepsilon_{10,NIR})'$ le vecteur de domension (20×1) des résidus,
$X = (x_{HPLC}, x_{NIR})$ la matrice d'incidence associée à β et formée des vecteurs colonnes $x_{HPLC} = (1, ..., 1, 0, ...0)'$ et $x_{NIR} = (0, ..., 0, 1, ...1)'$,
et $Z = (I_{10}, I_{10})'$ la matrice d'incidence associée à a.

Les distributions des effets aléatoires sont :

$$a \sim N_{10}(0, \sigma_a^2 I)$$

$$\varepsilon \sim N_{20}(0, \sigma_e^2 I)$$

et a et ε indépendants.

1.2 Définition de Rao et Kleffé : approche structurale

Un modèle linéaire mixte est un modèle linéaire $y = X\beta + \varepsilon$ avec $\varepsilon \sim N(0, \Sigma)$ dans lequel la variable aléatoire ε est décomposée comme une combinaison linéaire de variables aléatoires

structurales u_k, pour $k = 0, 1, ...K$, non observables :

$$\varepsilon = \sum_{k=0}^{K} Z_k u_k = Zu$$

où $Z_{n\times q} = (Z_0, Z_1, ...Z_k, ...Z_K)$ est une concaténation de matrices Z_k connues de dimension $n \times q_k$, et $u_{q\times 1} = (u'_0, u'_1, ..., u'_k, ...u'_K)$ est le vecteur correspondant des variables structurales $u_k = \{u_{kl}\}$ avec $l = 1, ...q_k$ tel que

$$u \sim N(0, \Sigma_u)$$

Σ_u est une fonction linéaire de paramètres θ_m avec $m = 1, ...M$

$$\Sigma_u = \sum_{m=1}^{M} \theta_m F_m$$

les matrices F_m étant des matrices carrées d'ordre $q = \sum_{k=0}^{K} q_k$.

Remarque : on ne posera pas de contraintes spécifiques sur θ_m et F_m dans le cas général, seuls les paramètres et ces matrices doivent assurer la positivité de Σ_u.

Considérant ce modèle, la matrice Σ de variance - covariance des variables aléatoires observables y est une fonction linéaire des paramètres θ_m. En effet, $\Sigma = Z\Sigma_u Z' = Z(\sum_m \theta_m F_m)Z' = \sum_m \theta_m (ZF_m Z') = \sum_m \theta_m V_m$ avec $V_m = ZF_m Z'$, d'où

$$\Sigma = \sum_{m=1}^{M} \theta_m V_m \tag{1.1}$$

Cette équation est une caractéristique du modèle linéaire mixte, qui est tel qu'à la fois l'espérance des observations vaut $E(y) = X\beta$ et leur matrice de variance - covariance $Var(y) = \Sigma = \sum_m \theta_m V_m$ sont des fonctions linéaires de paramètres (β et θ_m ici).

Remarque : dans cette définition, les effets aléatoires apparaissent comme un moyen de structurer la matrice de variance - covariance des observations.

1.2.1 Exemples

Un exemple classique réside dans le modèle linéaire mixte à K facteurs aléatoires indépendants :

$$y = X\beta + \sum_{k=1}^{K} Z_k u_k + e$$

et si on pose $u_0 = e$ et $Z_0 = I_n$, le modèle ci-dessus s'écrit

$$y = X\beta + \sum_{k=0}^{K} Z_k u_k$$

avec

$$\begin{aligned} E(y) &= X\beta \\ u_k &\sim N(0, \sigma_k^2 I_k) \\ E(u_k u_l) &= 0 \quad \forall k \neq l \\ \Sigma_u &= \bigoplus_{k=0}^{K} \sigma_k^2 I_k \end{aligned}$$

$$\begin{aligned}\Sigma &= \sum_{k=0}^{K} Z_k(\sigma_k^2 I_k)Z_k' \\ &= \sum_{k=0}^{K} \sigma_k^2 Z_k Z_k'\end{aligned}$$

Cas particulier : Modèle linéaire mixte gaussien à 2 effets aléatoires indépendants autres que la résiduelle. Ce modèle peut être considéré pour décrire des donnée longitudinales. On souhaite modéliser la croissance de 11 filles et 16 garçons qui ont été mesurés à 4 âges différents (8, 10, 12 et 14 ans). Chaque individu possède 4 mesures pour chaque âge considéré.

Le modèle peut s'écrire sous la forme suivant :

$$y = X\beta + Z_1 u_1 + Z_2 u_2 + e$$

où y représente le vecteur des observations des 27 * 4 mesures des individus ;
β est le vecteur des effets fixes comprenant une moyenne générale, l'effet du sexe (2 modalités) correspondant à l'ordonnée à l'origine différente selon le sexe des indivius et le coefficient de régression (pente) en fonction de l'âge. On peut éventuellement dans la partie fixe du modèle considérer une régression propre à chacun des sexes (supposant que la croissance des garçons diffère de celle des filles non seulement d'un point de vue de l'ordonnée à l'origine mais aussi de la pente de régression) ;
$u_1 = \{u_{1i}\}$ est un vecteur aléatoire correspondant à des « intercepts » ou « ordonnées à l'origine » des individus (indicés par i) mesurés de façon répétée. Cet effet aléatoire reflète l'existence d'une variabilité des ordonnées à l'origine entre individus ;
$u_2 = \{u_{2i}\}$ est un vecteur alétoire correspondant à des « pentes » des individus (indicés par i). Cet effet aléatoire reflète l'existence d'une variabilité des pentes entre individus.

A travers cette modélisation, on cherche à savoir s'il existe une variabilité entre individus aux niveaux des ordonnées à l'origine et des pentes.

Les matrices de variance-covariance de ce modèle sont :

$$Var\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 I_q & 0_{q\times q} \\ 0 & \sigma_2^2 I_q \end{pmatrix} = \Sigma_u \otimes I_q$$

avec

$$\Sigma_u = Var\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

formée par les variances de l'intercept (σ_1^2), de la pente (σ_2^2) et de leur covariance (ici égale à 0 car les 2 effets alétoires sont supposés indépendants).

On peut alors écrire :

$$Var\begin{pmatrix} e \\ u_1 \\ u_2 \end{pmatrix} = \sigma_0^2 \begin{pmatrix} I_n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \sigma_1^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & I_q & 0 \\ 0 & 0 & 0 \end{pmatrix} + \sigma_2^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_q \end{pmatrix}$$

ainsi que

$$Var(y) = \sigma_0^2 I + \sigma_1^2 Z_1 Z_1' + \sigma_2^2 Z_2 Z_2'$$

qui suit bien la forme linéaire $\Sigma = \sum_m \theta_m Vm$.

On peut noter que les mêmes propriétés de linéarité de Σ en les paramètres subsistent dans le cas où les facteurs u_k sont corrélés entre eux (par exemple pour les modèles à coefficients de

régression aléatoires).

Autre exemple : Modèle à 2 effets aléatoires autres que la résiduelle

$$y_{ijk} = \mu + h_i + s_j + (hs)_{ij} + e_{ijk}$$

avec y_{ijk} représente la $k$ème observation du père j dans le $i$ème « troupeau x année x saison » ;
h_i représente l'effet fixe du $i$ème « troupeau x année x saison » ;
s_j représente l'effet aléatoire du père j tel que $s_j \sim N(0, \sigma_s^2)$ iid (indépendants et identiquement distribués) ;
$(hs)_{ij}$ représente l'effet aléatoire de l'interaction « troupeau i x père j » tel que $(hs)_{ij} \sim N(0, \sigma_{hs}^2)$ iid.

On a les expressions des covariances suivantes :
$Cov(y_{ijk}, y_{i'jk'}) = \sigma_s^2$ (même père, milieu différent)
$Cov(y_{ijk}, y_{ij'k'}) = 0$ (père différent, même milieu)
$Cov(y_{ijk}, y_{ijk'}) = \sigma_s^2 + \sigma_{hs}^2$ (même père, même milieu)
$Cov(y_{ijk}, y_{i'j'k'}) = 0$ (père différent, milieu différent)
$Var(y_{ijk}) = \sigma_s^2 + \sigma_{hs}^2 + \sigma_e^2$ (variance totale)

1.3 Définition de Lindley et Smith. Approche hiérarchique

On considère un processus d'échantillonnage gaussien en 2 étapes relatives aux données et aux paramètres de position :
1)

$$y|\theta_1, C_1 \sim N(A_1\theta_1, C_1)$$

c'est à dire

$$y = A_1\theta_1 + e_1 \qquad e_1 \sim N(0, C_1)$$

2)

$$\theta_1|\theta_2, C_2 \sim N(A_2\theta_2, C_2)$$

c'est à dire

$$\theta_1 = A_2\theta_2 + e_2 \qquad e_2 \sim N(0, C_2)$$

La résultante de ces 2 étapes conduit à la distribution marginale des données

$$y|\theta_2, C_2 \sim N(A_1A_2\theta_2, C_1 + A_1C_2A_1').$$

En effet,

$$y = A_1\theta_1 + e_1 = A_1(A_2\theta_2 + e_2) + e_1 = A_1A_2\theta_2 + A_1e_2 + e_1$$

Ce modèle peut ainsi s'identifier à l'écriture d'un modèle mixte :

$$y = X\beta + Zu + e,$$

$$u \sim N(0, \sigma_u^2 I) \qquad e \sim N(0, \sigma_e^2 I)$$

avec $e = e_1$, $u = e_2$, $Z = A_1$, $X = A_1A_2$, $\beta = \theta_2$.

Pour résumer, on échantillonne d'abord des variables « réponse » sachant les paramètres, puis on échantillonne les paramètres.

Cette approche est celle présentée dans le cadre de la *statistique bayésienne*.

1.3.1 Exemple des données des enfants

Si nous reprenons l'exemple des données de croissance faciale à 4 âges de 11 filles et 16 garçons. Cet exemple est traité en détail dans le livre de Verbeke and Molenberghs (1997).

Un modèle simple d'analyse de ces données consiste en l'ajustement d'une droite de régression propre à chaque individu. Soit i l'indice du sexe ($i = 1$ pour les filles et $i = 2$ pour les garçons),
j l'indice de la période de mesure ($j = 1, ...4$),
t_j le temps correspondant (8, 10, 12, 14),
k l'indice de l'individu intra-sexe ($k = 1, ...11$ pour $i = 1$) et $k = 1, ...16$ pour $i = 2$).

Le modèle s'écrit :

$$y_{ijk} = A_{ik} + B_{ik}t_j + e_{ijk}$$

où A_{ik} est l'intercept (ordonnée à l'origine) propre à l'individu ik, et B_{ik} est la pente. Conditionnellement aux valeurs des coefficients de régression A_{ik}, B_{ik} des individus, le modèle ci-dessus est un modèle linéaire classique ($y = X\beta + e$) à variables aléatoires résiduelles indépendantes.

Si, dans une deuxième phase de raisonnement, on considère que les individus représentent un échantillon aléatoire d'enfants de chaque sexe, les A_{ik} et les B_{ik} sont alors des variables aléatoires. On peut les caractériser par leurs 2 premiers moments (Espérance et Variance) :

$$\begin{pmatrix} A_{ik} \\ B_{ik} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}\right)$$

Cela revient à décomposer l'ordonnée à l'origine (intercept) et la pente en la somme de 2 parties.

$$A_{ik} = \alpha_i + a_{ik}$$

$$B_{ik} = \beta_i + b_{ik}$$

avec une composante systématique α_i et β_i propre à chaque sexe et un écart centré a_{ik} et b_{ik} propre à l'individu k du sexe i.

Le modèle final peut alors s'écrire

$$y_{ijk} = \alpha_i + \beta_i t_j + a_{ik} + b_{ik}t_j + e_{ijk}$$

qui sépare la partie fixe ($\alpha_i + \beta_i t_j$) qui ne dépend de l'individu (non indicée par k) de la partie aléatoire ($a_{ik} + b_{ik}t_j$) propre à chaque individu.

Si on pose $y_{ik} = \{y_{ijk}\}$, $e_{ik} = \{e_{ijk}\}$, $\beta_{4x1} = (\alpha_2, \alpha_1 - \alpha_2, \beta_2, \beta_1 - \beta_2)'$, $u_{ik} = (a_{ik}, b_{ik})'$, $X_{ik} = (\mathbb{1}_4, 0_4, t, 0_4)$ si $i = 2$, $X_{ik} = (\mathbb{1}_4, \mathbb{1}_4, t, t)$ si $i = 1$, $Z_{ik} = (\mathbb{1}_4, t)$ avec $t_{4\times1} = \{t_j\}$, on écrit le modèle sous la forme matricielle

$$y_{ik} = X_{ik}\beta + Z_{ik}u_{ik} + e_{ik}$$

avec

$$u_{ik} \sim N(0, G) \qquad e_{ik} \sim N(0, R)$$

$$G = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$$

$$R = \sigma_e^2 I_4$$

1.4 Exemple de données de Transcriptome : données de nutrition chez la souris

Les données ont été fournies par Thierry Pineau et Pascal Martin de l'unité TOXALIM (INRA, Toulouse - site de Saint-Martin). Elles proviennent d'une étude de nutrition chez la souris. Pour 40 souris, nous disposons :

- des données d'expression de 120 gènes recueillies sur membrane nylon avec marquage radioactif,
- des mesures de 20 acides gras hépatiques.

Par ailleurs, les 40 souris sont réparties selon deux facteurs :

- Génotype (2 modalités) : les souris sont soit de type sauvage (wt) soit génétiquement modifiées (PPAR) ; 20 souris dans chaque cas.
- Régime (5 modalités) : les 5 régimes alimentaires sont notés `ref`, `efad`, `dha`, `lin`, `tournesol` ; 4 souris de chaque génotype sont soumises à chaque régime alimentaire.

Le modèle linéaire va nous permettre de répondre à quelques questions :

- Quels sont les gènes différentiellement exprimés ?
- Quelle est l'influence des facteurs "génotype" et "régime" sur l'expression des gènes ?

L'analyse de variance (ANOVA) permet d'apprécier l'effet d'une ou plusieurs variables qualitatives (les facteurs) sur une variable quantitative (la variable réponse, ici le niveau d'expression des gènes). Pour l'analyse de nos données, un modèle d'ANOVA à trois facteurs (génotype, régime, gène) permet de mettre en évidence des effets d'interaction d'ordre 3 très significatifs à l'aide du test de Fisher. Cela signifie qu'il existe des gènes régulés simultanément par le régime et le génotype, les effets du régime et du génotype étant non additifs. Le modèle d'ANOVA considéré s'écrit :

$$y_{ijkl} = g_i + r_j + G_k + gr_{ij} + gG_{ik} + rG_{jk} + grG_{ijk} + e_{ijkl} \tag{1.2}$$

où y_{ijkl} représente le logarithme du niveau d'expression du gène k ($k = 1, ..., 120$), pour le régime j ($j = 1, ..., 5$) et le génotype i ($i = 1, 2$), mesuré chez la souris l ($l = 1, ..., 4$) ;
g_i représente l'effet du génotype i ;
r_j représente l'effet du régime j ;
G_k représente l'effet du gène k ;
gr_{ij} représente l'effet de l'interaction du génotype i et du régime k ;
gG_{ik} représente l'effet de l'interaction du génotype i et du gène k ;
rG_{jk} représente l'effet de l'interaction du regime j et du gène k et
grG_{ijk} représente l'interaction d'ordre 3 combinant le génotype i, le régime j et le gène k.
On suppose que les résidus e_{ijkl} du modèle sont indépendants et identiquement distribués suivant une loi normale de moyenne nulle et de variance σ^2.

Les souris étant issues d'une lignée consanguine, elles ont été considérées dans un premier temps comme des répétitions indépendantes et identiquement distribuées. Cependant, à l'aide d'un modèle linéaire mixte, chaque souris peut être considérée comme un tirage aléatoire dans une population plus large de souris. Le modèle linéaire mixte mis en oeuvre est identique au modèle précédent en y ajoutant un effet aléatoire lié à l'individu (la souris) :

$$y_{ijkl} = g_i + r_j + G_k + gr_{ij} + gG_{ik} + rG_{jk} + grG_{ijk} + souris_l + e_{ijkl} \tag{1.3}$$

où $souris_l$ représente l'effet aléatoire de la souris l tel que $souris_l \sim \mathcal{N}(0, \sigma_s^2)$ iid et indépendants des e_{ijkl} ; les autres effets ayant la même définition que précédemment.

Dans ce cas, les estimations des composantes de la variance obtenues ici pour la variance "souris" est de 0.001 et pour la variance résiduelle de 0.007, la variance des observations est alors égale à 0.008. La variabilité indivuelle est très faible (12.5% de la variance totale). La variance des observations est dans ce cas identique à celle obtenue à l'aide d'une analyse de variance (modèle à effets fixes) puisque nous sommes dans le cadre d'un plan parfaitement équilibré et que la méthode d'estimation pour la modèle à effets mixtes est la méthode du maximum de vraisemblance restreinte (REML).

L'application du modèle linéaire mixte est beaucoup plus appropriée dans le cas où les variabilités dues à la technique, à la diversité génétique, aux gènes de la biopuce, ont un intérêt et que nous souhaitons les quantifier. Dans ce cas, nous considèrerons ces facteurs comme des effets aléatoires dans le modèle. C'est le cas dans l'étude transcriptomique décrite dans Bonnet *et al.* (2004), dans laquelle le logarithme du signal est modélisé en fonction des facteurs : membrane, truie, aiguille (ou bloc), jour d'hybridation, et des covariables tels que le logarithme de l'intensité du bruit de fond et de l'hybridation en sonde vecteur. Après une étape de choix de modèle pour conserver les effets significatifs, le modèle linéaire mixte a permis d'appréhender et de quantifier la part de variabilité due aux différentes sources de variation. La part de variabilité due à la diversité génétique représente 8%, celle due à la technique 4% et celle aux gènes 75%.

Toute inférence basée sur ce modèle sera valide pour tout animal, toute membrane, car l'échantillonnage des animaux, des membranes, ... de cette étude, dans une population plus large d'animaux, membranes, ... est pris en compte. Considérer les membranes (par exemple) comme effets fixes dans ce modèle aurait entrainé des conclusions valides uniquement sur les membranes de l'expérience. De plus, une structure de covariance non diagonale est prise en compte par ce modèle puisque deux signaux d'une même membrane seront corrélés, la corrélation étant égale à $\sigma^2_{membrane}/\sigma^2_{totale}$.

Chapitre 2

Estimation - Prédiction

L'analyse d'un modèle linéaire mixte consiste en deux problèmes d'estimation complémentaires : (1) l'estimation des effets fixes et aléatoires et (2) l'estimation des composantes de la variance et covariance du modèle. On parlera d'estimateurs pour les effets fixes et de prédicteurs pour les effets aléatoires.

2.1 Estimation des effets fixes

Nous supposerons dans un premier temps que la structure et les composantes de variance-covariance du modèle sont connus (ce qui est rarement le cas). D'autre part, nous avons vu dans la définition de Rao et Kleffé (approche structurale, partie 1.2) qu'un modèle linéaire mixte peut se mettre sous la forme d'un modèle linéaire : $y = X\beta+\varepsilon$ avec $\varepsilon \sim N(0, \Sigma)$ dans lequel la variable aléatoire ε est décomposée comme une combinaison linéaire de variables aléatoires structurales u_k, pour $k = 0, 1, ...K$, non observables : $\varepsilon = \sum_{k=0}^{K} Z_k u_k = Zu$.

La matrice de variance-covariance Σ étant supposée connue, les seuls paramètres inconnus sont le vecteur β et nécessitent d'être estimés à partir des données observées (y et X). Deux grandes méthodes sont généralement utilisées :

- la méthodes des moindres carrés pondérés ou généralisés, qui ne suppose connues que l'espérance et la variance de y ;
- la méthode du maximum de vraisemblance, qui suppose en plus que les résidus sont gaussiens.

Dans la méthode des moindres carrés généralisés, on cherche à minimiser la somme des carrés des écarts entre les valeurs observées et les valeurs ajustées (somme des carrés des résidus) selon une certaine métrique, soit :

$$\hat{\boldsymbol{\beta}} = Argmin(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}), \tag{2.1}$$

La solution est

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} \tag{2.2}$$

Dans l'approche de la méthode du maximum de vraisemblance, la densité de $\boldsymbol{y}$ s'écrit :

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right) \tag{2.3}$$

Le logarithme de la vraisemblance, multiplié par (-2), s'écrit $L(\boldsymbol{\beta}; \boldsymbol{y}) = -2.\ln f(\boldsymbol{y})$ et est égal à $(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$ à une constance près, pour $\boldsymbol{\Sigma}$ connu. Maximiser la vraisemblance

équivaut donc à faire des moindres carrés généralisés.

Remarque : quand la matrice de variance covariance se met sous la forme $\mathbf{\Sigma} = \sigma^2 \boldsymbol{I}$ avec σ^2 connu (les résidus sont supposés indépendants et identiques distribués), tout revient aux moindres carrés ordinaires (on se retrouve dans le cadre d'un modèle linéaire gaussien à effets fixes).

De manière générale, on ne connait ni la structure ni les composantes de la variance du modèle. Dans la suite de ce document, on cherchera à la fois à prédire les effets aléatoires et à estimer les effets fixes du modèle, en supposant dans un premier temps que les matrices de variance-covariance sont connues et dans un second temps on cherchera à estimer les composantes de la variance du modèle.

2.2 Prédiction des effets aléatoires

Le modèle considéré dans cette partie est le suivant

$$y = X\beta + Zu + e \tag{2.4}$$

avec y le vecteur $n \times 1$ des performances, β le vecteur $p \times 1$ des effets fixes, X la matrice d'incidence $n \times p$ associée aux effets fixes, u le vecteur $q \times 1$ des effets aléatoires : $u \sim N(0, G)$, (cas particulier usuel : $G = \sigma_u^2 I_q$) Z la matrice d'incidence $n \times q$ associée à u, e le vecteur $n \times 1$ des résidus : $e \sim N(0, R)$ (cas particulier usuel : $R = \sigma_e^2 I_n$).

On souhaite prédire les effets aléatoires u de ce modèle. On supposera pour l'instant que les composantes de la variance sont connues. C'est à dire qu'on connait les matrices G et R. L'espérance et la variance des observations sont :

$$\begin{aligned} E(y) &= X\beta \\ Var(y) &= \Sigma = ZVar(u)Z' + Var(e) = ZGZ' + R. \end{aligned}$$

Cas particulier où $G = \sigma_u^2 I_q$ et $R = \sigma_e^2 I_n$ alors $Var(y) = \sigma_u^2 ZZ' + \sigma_e^2 I_n$.

Faire une prédiction, c'est substituer à une variable aléatoire w (non observable) une variable aléatoire $\hat{w}$ fonction d'une variable aléatoire y observable : $\hat{w} = f(y)$, et telle que la distribution de $\hat{w}$ soit aussi proche que possible de celle de w selon un critère donné (distance, erreur quadratique moyenne, ...).

2.2.1 Le meilleur prédicteur (Best Predictor)

On parlera de meilleur prédicteur au sens de l'erreur quadratique moyenne $E(\tilde{u} - u)^2$.

On peut montrer que, dans le modèle linéaire mixte, le meilleur prédicteur de u est son espérance conditionnelle sachant les données (Cochran, 1951).

$$\tilde{u} = E(u|y). \tag{2.5}$$

Ce résultat est valable quelle que soit la densité conjointe $f(u, y)$ de u et de y.

L'espérance conditionnelle $\tilde{u} = E(u|y)$ minimise l'erreur quadratique moyenne

$$E(\tilde{u} - u)'A'(\tilde{u} - u)$$

au sens de la métrique A, pour toute matrice A symétrique définie positive.

Propriété 1 : Le meilleur prédicteur est sans biais :

$$E_y(\tilde{u}) = E(u).$$

Propriété 2 : La variance des erreurs de prédiction est égale à la valeur moyenne sur y de la variance conditionnelle de u :

$$Var(\tilde{u} - u) = E_y(Var(u|y))$$

On peut aussi montrer que

$$Cov(\tilde{u}, u') = Var(\tilde{u})$$
$$Cov(\tilde{u}, y') = Cov(u, y').$$

Lorsque la densité conjointe de u et de y est connue et gaussienne

$$\begin{pmatrix} u \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_u \\ \mu_y \end{pmatrix}, \begin{pmatrix} G & C \\ C' & V \end{pmatrix}\right),$$

le meilleur prédicteur vaut

$$\tilde{u} = \mu_u + CV^{-1}(y - \mu_y).$$

Il faut donc connaître les valeurs de μ_u, μ_y, C et V pour obtenir le meilleur prédicteur, ce qui n'est pas le cas en général.

2.2.2 Meilleur prédicteur linéaire (Best Linear Predictor)

On suppose ici que la distribution conjointe de y et u est inconnue, et que seuls les 2 premiers moments de la loi conjointe sont connus (on conait donc l'espérance et la variance). Une des possibilités est de se restreindre à une classe particulière de prédicteurs, en l'occurrence aux prédicteurs linéaires de la forme

$$\tilde{u} = a_0 + a'(y - E(y))$$

avec $E(y) = \mu_y$.

La minimisation de l'erreur quadratique $E(\tilde{u}-u)^2$ par rapport aux coefficients a_0 et $a = \{a_k\}$ est obtenue lorsque les dérivées partielles sont nulles et conduit à

$$\tilde{u} = \mu_u + Cov(u, y)Var(y)^{-1}(y - \mu_y). \tag{2.6}$$

On remarque que ce prédicteur a la même forme que le meilleur prédicteur résultant du cas gaussien.Il possède les mêmes propriétés (estimateur sans biais).

2.2.3 Meilleur prédicteur linéaire sans biais : le BLUP (Best Linear Unbiased Predictor)

Ce prédicteur a été proposé pour répondre à la levée de l'hypothèse de moments connus de premier ordre de la distribution de (u, y). La distribution conjointe et les espérances sont inconnues, seules les variances et les covariances sont supposées connues. Dans ce cas, le vecteur des effets fixes du modèle doit être estimé en même temps que le vecteur des effets aléatoires.

On suppose que le vecteur des observations y est généré par le modèle suivant : $y = X\beta + Zu + e$ où $\mu_y = X\beta$, $Var(y) = V = ZGZ' + R$, $u \sim N(0, G)$, $e \sim N(0, R)$ avec u et e indépendants.

On exprime la variable aléatoire à prédire sous la forme $w = L'\beta + u$.

Le prédicteur $\hat{w}$ de w est recherché
a) parmi les prédicteurs linéaires $a + By$, où ni a ni B ne dépendent de β
b) sans biais au sens $E(\hat{w}) = E(w)$
c) d'erreur quadratique moyenne $E(\hat{w} - w)'A(\hat{w} - w)$ minimum, pour A symétrique définie positive.

L'objectif est de minimiser l'erreur quadratique moyenne sous la contrainte de non biais. Après une écriture de la fonction à minimiser et le calcul des dérivées partielles, on obtient

$$BLUP(w) = \hat{w} = L'\hat{\beta} + C'V^{-1}(y - X\hat{\beta})$$

où $\hat{\beta}$ est l'estimateur des moindres carrés généralisés de β : $[X'V^{-1}X]\hat{\beta} = X'V^{-1}y$ et $CV^{-1}(y - X\hat{\beta}) = \hat{u}$ s'obtient à partir du meilleur prédicteur linéaire $\hat{u}$ de u.

Les anglosaxons écrivent $BLUP(w) = BLUE(L'\beta) + BLUP(u)$, où BLUE signifie Best Linear Unbiased Estimator.

Les équations du modèle mixte d'Henderson (MME = mixed model equations). Henderson a montré que l'estimateur (BLUE) de β et le prédicteur (BLUP) de u s'obtiennent à partir des équations suivantes :

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix} \tag{2.7}$$

Quelques propriétés

Dans ce cadre, les effets fixes et aléatoires sont estimés/prédits en même temps ; les matrices $\mathbf{G^{-1}}$ et $\mathbf{R^{-1}}$ ont en général une struture simple et la taille du système des éuqtions du modèle mixte à résoudre est égale à la dimension du vecteur β plus la dimension du vecteur $\mathbf{u}$ et non la dimension du vecteur des observations $\mathbf{y}$.

$$Var(L'\hat{\beta}) = L'(XV^{-1}X)^{-1}L \tag{2.8}$$
$$Var(\hat{u}) = C'PC \tag{2.9}$$

où $P = V^{-1}(I - Q)$ et $Q = X(X'V^{-1}X)^1X'V^{-1}$ est le projecteur orthogonal de y sur le sous espace vectoriel engendré par les colonnes de X, selon la métrique V^{-1}.

$$Cov(\hat{u}, u') = C'PC \tag{2.10}$$
$$Var(\hat{u}) = Cov(\hat{u}, u') \tag{2.11}$$
$$Var(\hat{u} - u) = G - C'PC \tag{2.12}$$
$$Cov(\hat{\beta}, \hat{u}') = 0 \tag{2.13}$$

A partir des équations du modèle mixte, si on note C^{-1} l'inverse de la matrice des coefficients (attention, ne pas confondre avec la matrice C de covariance entre u et y) :

$$C^{-1} = \begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} C^{\beta\beta} & C^{\beta u} \\ C^{u\beta} & C^{uu} \end{pmatrix} \tag{2.14}$$

On peut montrer que

$$Var(k'\hat{\beta}) = k'C^{\beta\beta}k \tag{2.15}$$
$$Cov(k'\hat{\beta}, \hat{u}) = 0 \tag{2.16}$$
$$Cov(k'\hat{\beta}, (\hat{u} - u)') = k'C^{\beta u} \tag{2.17}$$
$$Var(\hat{u} - u) = C^{uu} \tag{2.18}$$
$$Var(\hat{u}) = G - C^{uu} \tag{2.19}$$

2.3 Estimation des composantes de la variance

Une approche pour estimer les composantes de la variance est d'utiliser la méthode du maximum de vraisemblance. Sous l'hypothèse de normalité de la distribution des observations, la fonction de vraisemblance s'écrit :

$$f(y|\beta,\sigma) = (2n)^{-n/2}|V(\sigma)|^{-1/2} exp\left(-\frac{1}{2}(y-X\beta)'V^{-1}(\sigma)(y-X\beta)\right)$$

où σ représente le vecteur des composantes de la variance du modèle.

Cette fonction est strictement positive, on peut donc considérer son logarithme

$$lnf(y|\beta,\sigma) = L(\beta,\sigma;y) = -\frac{n}{2}ln(2\pi) - \frac{1}{2}ln|V| - \frac{1}{2}(y-X\beta)'V^{-1}(\sigma)(y-X\beta)$$

Le logarithme étant une fonction croissante, rechercher le maximum de la vraisemblance équivaut à rechercher le maximum de L ou le minimum de $-2L$. Les estimées du maximum de vraisemblance de β et σ sont des éléments de l'espace des paramètres, qui maximisant la fonction L.Toutefois, l'estimation des paramètres à l'aide du maximum de vraisemblance ne tient pas compte de la perte de degrés de liberté liée à l'estimation des effets fixes β. Les composantes de la variance estimées sont de ce fait biaisées.

Un cas simple pour étudier cela est l'estimation de la moyenne et de la variance à partir d'un échantillon de n observations $y_i \sim N(\mu,\sigma^2)$ $\quad iid \quad$ $i = 1,...n$ par maximum de vraisemblance. Comme

$$-2L = n(ln2\pi + ln\sigma^2) + \sum_{i=1}^{n}(y_i-\mu)^2/\sigma^2,$$

on dérive $-2L$ par rapport à μ et σ^2, on annule les dérivées et on obtient $\hat{\mu} = \bar{y}$ et $\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2$.

Or si on calcule l'espérance de $\hat{\sigma}^2_{ML}$, on obtient $E(\hat{\sigma}^2_{ML}) = \sigma^2(1-\frac{1}{n})$, indiquant que l'estimateur du maximum de vraisemblance de σ^2 est biaisé. Le biais est alors égal à $-\frac{\sigma^2}{n}$. Remarquons tout de même que ce bias tend vers 0 quand n tend vers l'infini : l'estimateur du maximum de vraisemblance de σ^2 est asymptotiquement sans biais.

C'est la constation de ce biais qui est à l'origine du développement du concept de vraisemblance restreinte (ou résiduelle). L'idée est la suivante : l'estimation de μ interférant avec celle de σ^2, on va faire en sorte d'éliminer μ.

Patterson et Thompson (1971) proposent d'estimer les composantes de la variance indépendamment des effets fixes. Ils s'intéressent à la fonction de vraisemblance de contrastes d'erreur, appelée également vraisemblance restreinte ou résiduelle, c'est-à-dire la vraisemblance d'une fonction linéaire des données $(k'y)$ d'espérance nulle et indépendante des effets fixes.

La vraisemblance de ces contrastes (appelée aussi REML) s'écrit :

$$L(\sigma;k'y) = -\frac{1}{2}\left(ln|V| + ln|X^{*'}V^{-1}X^*| + (y-X^*\hat{\beta})'V^{-1}(y-X^*\hat{\beta})\right) + const.$$

où $\hat{\beta}$ est solution du système

$$X^{*'}V^{-1}X^*\hat{\beta} = X^{*'}V^{-1}y$$

et X^* est une sous matrice formée de $rang(X)$ colonnes de X linéairement indépendantes.

En pratique, (1) on se donne des valeurs initiales pour les composantes de la variance, (2) on calcule les estimations $\hat{\beta}$ et les prédictions $\hat{u}$ à l'aide des équations du modèle mixte, en fixant les composantes de la variance, (3) on calcule les estimations des composantes de la variance par la méthode du REML, en fixant $\hat{\beta}$ et $\hat{u}$. On itère les étapres (2) et (3) jusqu'à convergence.

Chapitre 3

Tests d'hypothèses - Choix de modèles

3.1 Tests sur effets fixes

Nous avons vu que le modèle linéaire mixte peut s'écrire sous la forme générale

$$y = X\beta + \epsilon$$

où β représente l'ensemble des effets fixes du modèle, et ϵ représente l'ensemble des effets aléatoires. On supposera par la suite que

$$\epsilon \sim N(0, V)$$

où V est la matrice de variance - covariance du modèle.

Dans ce cadre, l'estimation des moindres carrés (généralisés ou pondérés) de β s'obtient par résolution de l'équation

$$(X'V^{-1}X)\hat{\beta} = X'V^{-1}y.$$

Rappels de quelques propriétés

(i) L'estimateur des moindres carrés d'une fonction estimable des paramètres est sans biais, c'est à dire : $E(k'\hat{\beta}) = k'\beta$.

(ii) Sous l'hypothèse $V = \sigma^2 I$ (indépendance et variances homogènes), $k'\hat{\beta}$ est le meilleur estimateur linéaire sans biais (BLUE) de $k'\beta$ et $Var(k'\hat{\beta}) = \sigma^2 k'(X'X)^{-1}k$.

(iii) Sous l'hypothèse de normalité des résidus, la distribution de l'estimateur est aussi normale :

$$k'\hat{\beta} \sim N(k'\beta, \sigma^2 k'(X'X)^{-1}k).$$

Test

On souhaite tester l'hypothèse suivante :

$$H_0 : k'\beta = m$$

contre son alternative

$$H_1 : k'\beta \neq m$$

avec k matrice $p \times r$ dont les r colonnes sont linéairement indépendantes, et m vecteur $r \times 1$.

3.1.1 V est connue

(a) Sous l'hypothèse $V = \sigma^2 I$ connue et sous H_0, la statistique

$$(k'\hat{\beta} - m)'(Var(k'\hat{\beta}))^{-1}(k'\hat{\beta} - m)$$

suit une loi du khi-deux à rang(k) degrés de liberté, avec $Var(k'\hat{\beta}) = \sigma^2 k'(X'X)^{-1}k$.

(b) Sous l'hypothèse V quelconque mais connue et sous H_0, la statistique

$$(k'\hat{\beta} - m)'(Var(k'\hat{\beta}))^{-1}(k'\hat{\beta} - m)$$

suit une loi du khi-deux à rang(k) degrés de liberté, avec

$$Var(k'\hat{\beta}) = k'(X'V^{-1}X)^{-1}k.$$

3.1.2 V est connue à un coefficient près

On suppose $V = \sigma^2 V_0$ avec V_0 matrice quelconque mais connue, et σ^2 paramètre inconnu.
On a :

$$\begin{aligned} Var(k'\hat{\beta}) &= k'(X'V^{-1}X)^{-1}k \\ &= \sigma^2 k'(X'V_0^{-1}X)^{-1}k \\ W &= (k'\hat{\beta} - m)'(Var(k'\hat{\beta}))^{-1}(k'\hat{\beta} - m) \\ &= \frac{1}{\sigma^2}(k'\hat{\beta} - m)'(X'V_0^{-1}X)^{-1}(k'\hat{\beta} - m) \\ &= \frac{1}{\sigma^2}Q \end{aligned}$$

Sous H_0, la statistique $Q = \sigma^2 W$ suit une loi $\sigma^2\chi^2_{rg(k)}$. Or, on ne connait pas σ^2.
On sait par contre que

$$\frac{(n - rg(X))\hat{\sigma}^2}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi^2_{n-rg(X)}$$

où n est le nombre total d'observations, et que Q et SSE sont indépendants. On peut donc former la statistique suivante (qui ne dépend pas de σ^2) :

$$F = \frac{[Q/\sigma^2]/rg(k)}{[SSE/\sigma^2]/[n - rg(X)]} = \frac{Q/rg(k)}{[SSE]/[n - rg(X)]} \sim F_{rg(k);n-rg(X)}$$

3.1.3 V est quelconque et inconnue

Quand la matrice V est inconnue, le moindre déséquilibre dans le plan d'échantillonnage est rédhibitoire pour l'obtention de tests exacts. On utilise alors des résultats asymptotiques. Les tests qui suivent ne sont alors pas des tests exacts.

Dans ce cadre, la matrice V est estimée par maximum de vraisemblance $\hat{V} = V(\hat{\gamma})$, où γ est le vecteur des composantes de la variance.
$\hat{\beta}$ est alors solution de l'équation

$$(X'\hat{V}^{-1}X)\hat{\beta} = X'\hat{V}^{-1}y.$$

Sous H_0, la statistique du **test de Wald**

$$W(\hat{\gamma}) = (k'\hat{\beta} - m)'(k'(X'\hat{V}^{-1}X)^{-1}k)^{-1}(k'\hat{\beta} - m)$$

tend vers une loi du khi-deux à rang(k) degrés de liberté, quand $n \to +\infty$.

Par analogie au cas précédent, on peut utiliser le **test de type F** : la statistique

$$F = \frac{W(\hat{\gamma})}{rg(k)}$$

suit asymptotiquement une loi $F_{rg(k),d}$ sous H_0, où d est le degré de liberté du dénominateur. Son calcul est délicat. Différentes méthodes ont été proposées dans la littérature pour le calcul de d (Satterthwaite, Kenward et Roger, etc).

Remarque : dans le cas du modèle linéaire à effets fixes classique, on n'estime que σ_e^2, donc $d = n - rg(X)$. Dans le cas du modèle linéaire mixte, on doit estimer σ_u^2 et σ_e^2 (au minimum).

D'autres test peuvent aussi être utilisés, comme le **test du rapport de vraisemblance**. Sous H_0, la statistique

$$\lambda = -2L_R + 2L_C$$

suit asymptotiquement une loi du khi-deux à r degrés de liberté, où L_R est le maximum du logarithme de la vraisemblance sous le modèle réduit (sous H_0), L_C est le maximum du logarithme de la vraisemblance sous le modèle complet (sous H_1), et r est la différence de paramètres à estimer sous H_0 et sous H_1.

Dans la cadre de ce test, les paramètres du modèle sont estimés par maximum de vraisemblance.

Enfin, il est important de souligner que les deux modèles contrastés vis-à-vis des effets fixes β doivent présenter la même structure de variance-covariance.

En résumé, pour comparer 2 modèles emboités avec le test du maximum de vraisemblance, nous comparons les log-vraisemblance des 2 modèles. Si la différence est grande, le fait de passer d'un modèle simple (modèle sous H_0) à un modèle plus complexe (modèle sous H_1) a apporté un écart de log-vraisemblance significatif. Donc le modèle sous H_1 est acceptable. Par contre si l'écrat est faible, cela veut dire que les 2 modèles sont voisins et par souci de parcimonie le modèle sous H_0 est conservé.

3.1.4 Test robuste

Cette approche est utile lorsque l'on ne connaît pas la structure de variance - covariance des observations.

La sélection d'un modèle global nécessite à la fois le choix des effets fixes (modèle de l'espérance E) et celui des effets aléatoires (modèle de variance - covariance V). En fait, ce choix est complexe puisque la comparaison de modèles sur l'espérance E depend de la structure de variance-covariance V et celle de modèles V depend de E. En pratique, on peut procéder de manière empirique : (i) on part d'une structure de variance - covariance V donnée, on choisit/teste un modèle sur l'espérance E, (ii) puis à modèle d'espérance E fixé, on compare différentes structures de variance-covariance V. On itére ensuite les étapes (i) et (ii).

Une alternative consiste à faire inférence sur les effets fixes par des estimateurs robustes (estimateur dit "sandwich" de Liang et Zeger) vis à vis de la structure V. L'idée de base est d'utiliser les propriétés de l'estimateur des moindres carrés pondérés

$$(X'WX)\hat{\beta} = X'Wy$$

sans avoir fait l'hypothèse que la matrice de variance - covariance V est bien spécifiée (c'est à dire que $W = V^{-1}$).

En effet, sous l'hypothèse

$$y \sim N(\mu, V)$$

avec $\mu = X\beta$, on a

$$E(\hat{\beta}) = \beta$$

$$Var(\hat{\beta}) = (X'WX)^{-1}X'Var(y)X(X'WX)^{-1}.$$

Dans le cas de données répétées sur la même unité expérimentale i,

$$Var(\hat{\beta}) = \sum_i (X_i'W_iX_i)^{-1}X_i'Var(y_i)X_i(X_i'W_iX_i)^{-1}$$

où $y_i \sim N(X_i\beta, V_i)$ et $Cov(y_i, y_{i'}) = 0$ pour $i \neq i'$.

Liang et Zeger proposent de remplacer $V_i = Var(y_i)$ par un estimateur convergent

$$\hat{V}_i = (y_i - \mu_i(\hat{\beta}))(y_i - \mu_i(\hat{\beta}))'.$$

Ces auteurs montrent que l'estimateur de $Var(\hat{\beta})$ ainsi obtenu est lui-même convergent pourvu que l'espérance $\mu_i = X_i\beta$ soit bien spécifiée.

Divers choix sont alors possibles pour la matrice W_i, le choix le plus simple étant l'identité, qui correspond aux moindres carrés simples. Mais on peut aussi choisir d'autres types de matrices pour W_i.

Les estimateurs robustes peuvent être obtenus à l'aide du logiciel SAS, procédure PROC MIXED et en utilisant l'option *empirical*.

3.2 Choix de modèles

Lorsque l'on souhaite comparer des modèles qui ne possèdent pas les mêmes effets fixes, ni la même structure de variance covariance ; c'est dire lorsque les modèles ne sont pas emboités, il est nécessaire de disposer d'autres stratégies pour guider le choix du meilleur modèle parmi tous les modèles candidats.

Par définition, on dit qu'un modèle est emboité dans un autre modèle plus général quand il est un cas particulier du modèle général.

3.2.1 Modèles linéaires classiques (modèles à effets fixes)

1. Quand on veut choisir entre 2 modèles concurrents comportant le même nombre de paramètres, on peut utiliser le coefficient de détermination R^2, qui mesure la qualité de l'ajustement :

$$R^2 = \frac{Var(\hat{y})}{Var(y)} = 1 - \frac{Var(\hat{e})}{Var(y)}$$

Ce coefficient est compris entre 0 et 1. Le meilleur modèle sera celui ayant le R^2 le plus élevé.

2. Quand le nombre de paramètres est différent d'un modèle à l'autre, on peut utiliser le coefficient de détermination modifié par une fonction de pénalité dependant du nombre de paramètres du modèle :

$$R^2_{ajust} = \frac{(n-1)R^2 - p}{n-p-1} = 1 - (1-R^2)\frac{n-1}{n-p-1}$$

où n est le nombre d'observation et p le nombre de paramètres. Le meilleur modèle sera celui ayant le R^2_{ajust} le plus élevé.

3. Le coefficient C_p de Mallows est aussi populaire :

$$C_p = \frac{SSE}{\hat{\sigma}^2} - n + 2p$$

où SSE est la somme des carrés résiduels du modèle, et $\hat{\sigma}^2$ la variance résiduelle estimée. Le meilleur modèle sera celui ayant le C_p le plus faible.

3.2.2 Modèles linéaires mixtes

Les critères présentés sont basés sur une estimation de la distance entre chacun des modèles candidats et le vrai modèle (inconnu), puis sur le choix de celui qui minimise cette distance. Une mesure de ressemblance entre deux modèles est donnée par l'information de Kullback-Leibler faisant intervenir des logarithmes des densités.

Le critère d'Akaike se calcule ainsi :

$$AIC = -2\sum_{i=1}^{n} \log f_{\hat{\theta}}(y_i) + 2k$$

où k est le nombre de paramètres du modèle dans le cas général.

Ce critère permet de comparer des modèles non nécessairement emboités. La seule condition est que les paramètres de ces modèles doivent être estimés par maximum de vraisemblance. Il s'agit d'un critère de choix, non d'un test statistique.

Le premier terme du critère peut être interprété comme une mesure de l'inadéquation du modèle aux données. Le second terme représente une fonction de pénalité mesurant la complexité du modèle à travers son nombre de paramètres. Le meilleur modèle est celui qui a le plus faible AIC.

La procédure MIXED de SAS donne

$$AIC_{SAS} = L - k$$

où L est le log de la vraisemblance et k est le nombre de composantes de la variance. Le meilleur modèle est celui qui a le plus grand AIC_{SAS}.

Le critère de Schwarz

Ce critère est basé sur un raisonnement bayésien.

$$BIC = -2\sum_{i=1}^{n} \log f_{\hat{\theta}}(y_i) + k \log n.$$

La proc MIXED donne

$$BIC = L - k \log(n^*)/2$$

où L est le logarithme de la vraisemblance, k est le nombre de composantes de la variance, et $n^* = n$ (resp. $n - p$) si la méthode d'estimation des composantes de la variance est la maximum de vraisemblance ML (resp. maximum de vraisemblance restreinte REML).

Conclusion

Ces critères sont en général assez simples à mettre en oeuvre. Ils doivent être utilisés à partir des estimations par la méthode ML ou REML.

Dans le cas REML, les estimations étant faites sur des données transformées de manière à annuler la partie fixe du modèle, les comparaisons ne peuvent porter que sur des modèles dont la partie fixe est identique.

Il faut également corriger le nombre d'observations n utilisé par les critères en remplaçant n par $n^* = n - p$, où p est le nombre de paramètres pour la partie fixe du modèle.

3.3 Tests sur composantes de la variance

En règle générale, lorsque les modèles sont emboités, le test le plus couramment utilisé est le test du rapport de vraisemblance (ML ou REML selon le cas).

Quand on veut tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$, on forme la statistique suivante :

$$\lambda_n = -2L_R + 2L_C$$

où L_R est la log vraisemblance sous le modèle réduit (sous H_0) et L_C est la log vraisemblance sous le modèle complet (sous H_1).

Sous certaines conditions de régularité, la statistique λ_n suit sous H_0 une loi du khi-deux à $r = ddl_C - ddl_R$ degrés de liberté (soit la différence entre les nombres de paramètres sous les 2 modèles).

La distribution asymptotique de λ_n repose sur la normalité asymptotique des estimateurs du maximum de vraisemblance. Si θ_0 est situé sur le bord de l'espace des paramètres, alors la loi asymptotique de $\hat{\theta}$ n'est généralement plus une gaussienne, et donc la loi de λ_n n'est plus une khi-deux.

Lorsque l'on teste $H_0 : \sigma^2 = 0$ contre $H_1 : \sigma^2 > 0$, la valeur 0 est à la frontière de l'espace des paramètres, puisqu'une variance σ^2 est définie sur $\mathbb{R}_+$.

Dans le cas d'un modèle linéaire mixte à un facteur à effets aléatoires autre que la résiduelle, si on veut tester la nullité de la variance de ce facteur (supprimer un effet aléatoire d'un modèle revient à tester si la variance de cet effet est nulle ou pas), la statistique λ_n du rapport de vraisemblance suit une loi qui est un mélange de deux khi-deux : $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$, où χ^2_0 est en fait la masse de Dirac en 0.

En conséquence, le test "naïf" (consistant à prendre une χ^2_1) est trop conservateur et le seuil s du test correct au niveau α correspond à $Pr[\chi^2_1 \geq s] = 2\alpha$ puisque sous H_0, la décision de rejet est prise lorsque la statistique est positive (c'est à dire une fois sur deux) et que celle-ci dépasse le seuil s.

Concrètement, la procédure correcte revient à effectuer un test unilatéral au lieu d'un test bilatéral.

Plus généralement, lorsque le modèle comporte q facteurs à effets aléatoires autres que la résiduelle, et que l'on souhaite tester la nullité d'une des q variances, la distribution sous l'hypothèse nulle de la statistique du rapport de vraisemblance suit un mélange de lois : $\frac{1}{2}\chi^2_{q-1} + \frac{1}{2}\chi^2_q$.

3.4 Discussion

Dans la pratique, quand on cherche un bon modèle pour ajuster les données, il faut trouver à la fois un bon modèle pour l'espérance ($E(y) = X\beta$) et un bon modèle sur la variance ($Var(y) = V = ZGZ' + R$).

On peut itérer la procédure suivante :

1) on se fixe la structure de variance, et on choisit un bon modèle pour l'espérance (tests des effets fixes ou critères de choix de modèles)

2) on se fixe la structure d'espérance, et on choisit un bon modèle pour la variance (tests des effets aléatoires ou critères de choix de modèles)

Les tests d'hypothèses présentés dans ce chapitre n'étant pas des tests exacts, les critères de choix de modèles étant des pis-aller, on peut aussi se donner l'heuristique suivante consistant à tester la significativité des facteurs dans un modèle à effets fixes. Le modèle étant choisi, certains facteurs sont alors considérés comme des effets aléatoires, selon la question posée et le mode d'échantillonnage des données.

Chapitre 4

Modèle linéaire mixte et approche bayésienne

4.1 Introduction à la statistique bayésienne

La suite s'inspire largement du très bon livre de Christian Robert (1992) "'L'analyse statistique bayésienne, Economica, Paris"'.

La démarche statistique est une démarche d'inversion : elle remonte des effets (les observations) aux causes (les paramètres du mécanisme générateur). Le modèle probabiliste s'intéresse à $y|\theta$, où y est l'observation et θ est le paramètre. Le modèle statistique est basé sur la vraisemblance $\ell(\theta|y)$, en écrivant $\ell(\theta|y) = f(y|\theta)$.

Le théorème de Bayes permet bien cette inversion. Soient A et E deux événements, on a

$$P(A|E) = \frac{P(E|A).P(A)}{P(E|A).P(A) + P(E|A^c).P(A^c)}. \tag{4.1}$$

Si $P(B) = P(A)$, alors

$$\frac{P(A|E)}{P(B|E)} = \frac{P(E|A)}{P(E|B)}. \tag{4.2}$$

Ce théorème a été démontré par Thomas Bayes (1702-1761) et indépendamment et de manière plus complète par Pierre Simon Laplace (1749-1827).

Cause et effets sont ici sur un même pied d'égalité, les deux pouvant être probabilisés.

La version continue du théorème de Bayes s'écrit :

$$f(\theta|y) = \frac{f(y|\theta).f(\theta)}{\int f(y|\theta).f(\theta)d\theta}. \tag{4.3}$$

Le dénominateur étant une constante en θ, il est souvent éliminé de l'écriture ci-dessus, pour n'en retenir que l'essentiel :

$$f(\theta|y) \propto f(y|\theta).f(\theta). \tag{4.4}$$

L'inférence sur θ combine donc l'information apportée par les observations y par le biais de la vraisemblance, et l'information a priori sur θ. $f(\theta|y)$ est appelée loi a posteriori de θ.

Le principe de vraisemblance s'énonce ainsi :

Toute l'information sur θ tirée de y est contenue dans la vraisemblance $\ell(\theta|y)$. De plus, si y_1 et y_2 sont tels qu'il existe une constante c telle que, quelque soit θ, $\ell(\theta|y_1) = c.\ell(\theta|y_2)$, ils

apportent la même information sur θ et doivent conduire à la même inférence. La statistique bayésienne respecte le principe de vraisemblance.

Résumer l'information a priori disponible sur le paramètre θ requiert aussi la quantification de l'incertitude sur cette information. C'est un point délicat sur lequel statisticiens "classiques" et bayésiens s'affrontent régulièrement.

Il est par ailleurs possible d'envisager des lois impropres pour l'a priori. Une mesure σ-finie est choisie telle que $\int f(\theta)d\theta = +\infty$. Par exemple, pour le modèle $y = \mu + e$, on peut écrire que $f(\mu) \in [-\infty, +\infty]$ uniformément.

Enfin, l'approche bayésienne permet d'accumuler les connaissances de manière logique :

Information a priori $f(\theta)$ sur les paramètres — Données y (expérience)

Théorème de Bayes

Information a posteriori $f(\theta|y)$ — Nouvelles données z

Théorème de Bayes

A posteriori $f(\theta|y, z)$

L'inférence finale sur θ est la même que celle fondée sur la réunion des deux jeux de données y et z.

4.2 Traitement bayésien du modèle linéaire

Un exemple simple. Considérons tout d'abord le modèle très simple suivant :

$$y_i|\mu, \sigma_e^2 \sim \mathcal{N}(\mu, \sigma_e^2) \qquad i = 1, ...n \tag{4.5}$$
$$\mu|\mu_0, \sigma_0^2 \sim \mathcal{N}(\mu_0, \sigma_0^2) \tag{4.6}$$

Les paramètres $\sigma_e^2, \mu_0, \sigma_0^2$ sont supposés connus.

On retrouve ici l'approche hiérarchique mentionnée dans les chapitres précédents de ce cours.

La loi a posteriori du paramètre μ est elle aussi gaussienne :

$$\mu|y \sim \mathcal{N}\left(\frac{n\bar{y}/\sigma_e^2 + \mu_0/\sigma_0^2}{n/\sigma_e^2 + 1/\sigma_0^2}, \frac{1}{n/\sigma_e^2 + 1/\sigma_0^2}\right). \tag{4.7}$$

On remarque que l'espérance a posteriori de μ est une somme pondérée de la moyenne a priori μ_0 et de la moyenne empirique $\bar{y}$. Le poids accordé aux données est d'autant plus grand que le nombre d'observations n est grand, et/ou que l'incertitude (σ_0^2) a priori sur μ est grande.

A la limite, quand l'information a priori est négligeable devant l'information apportée par les données, la distribution a posteriori est proche d'une $\mathcal{N}(\bar{y}, \sigma_e^2/n)$.

L'estimateur bayésien de μ est l'espérance a posteriori $E(\mu|y) = \frac{n\bar{y}/\sigma_e^2 + \mu_0/\sigma_0^2}{n/\sigma_e^2 + 1/\sigma_0^2}$. Il tend vers $\bar{y}$ quand n tend vers l'infini.

On aurait aussi pu envisager un a priori uniforme sur μ :

$$\mu|a, b \sim \mathcal{U}_{[a,b]}. \tag{4.8}$$

Si on fait tendre a vers $+\infty$ et b vers $-\infty$, on obtient un a priori "vague" et non-informatif : $f(\mu) \propto const.$ La distribution a posteriori de μ est alors proportionnelle à la vraisemblance : $f(\mu|y) \propto f(y|\mu)$. Dans ce cas, l'estimateur bayésien de μ vaut $E(\mu|y) = \bar{y}$ et la variance a posteriori vaut σ_e^2/n.

Bien que les philosophies bayésienne et fréquentiste (classique) soient radicalement différentes, il faut bien noter que leurs estimations ont ici les mêmes valeurs numériques.

Le modèle linéaire mixte général. Dans le modèle linéaire $y = X\beta + Zu + e$, que l'on peut encore écrire sous la forme distributionnelle suivante :

$$y|\beta, u, R \sim \mathcal{N}(X\beta + Zu, R) \tag{4.9}$$

les paramètres de position β et u sont traités sur un même plan. Si l'on n'a aucune connaissance a priori sur β, on choisit un a priori vague

$$f(\beta) \propto const. \tag{4.10}$$

Supposons que l'on ait une information a priori sur u que l'on puisse mettre sous la forme

$$u|G \sim \mathcal{N}(0, G). \tag{4.11}$$

Les hyperparamètres R et G sont supposés connus.

L'inférence sur $\theta = (\beta, u)$ est basée sur la distribution a posteriori

$$f(\beta, u|y, R, G) \propto f(y|\beta, u, R).f(\beta).f(u|G) \tag{4.12}$$

$$\propto f(y|\beta, u, R).f(u|G). \tag{4.13}$$

On peut montrer que l'estimateur bayésien de θ est la solution $\hat{\theta}$ des équations du modèle mixte, comme dans le cas classique. De plus,

$$\theta|y, G, R \sim \mathcal{N}(\hat{\theta}, \hat{C}) \tag{4.14}$$

avec $\hat{C}$ l'inverse de la matrice des coefficients.
En particulier, $\beta|y, G, R \sim \mathcal{N}(\hat{\beta}, \hat{C}^{\beta,\beta})$ et $u|y, G, R \sim \mathcal{N}(\hat{u}, \hat{C}^{u,u})$.

On notera le parallèle avec l'approche classique : $\hat{\beta} \sim \mathcal{N}(\beta, \hat{C}^{\beta,\beta})$ et $Var(\hat{u} - u) = \hat{C}^{u,u}$.

La distribution a posteriori de θ est gaussienne, donc le mode est égal à l'espérance. D'autre part, comme $f(y, u|\beta)$ (fonction maximisée par Henderson) est proportionnel à $f(\beta, u|y)$, il est logique que l'estimateur bayésien de θ (l'espérance a posteriori) soit aussi solution des équations du modèle mixte d'Henderson.

4.3 Estimation des composantes de la variance

Notons γ le vecteur des composantes de la variance : $R = R(\gamma)$ et $G = G(\gamma)$.

Quand on maximise la densité conjointe a posteriori $f(\theta, \gamma|y)$ de tous les paramètres (paramètres de position θ et paramètres de dispersion γ), les estimateurs sont ceux du maximum de vraisemblance. Notamment, avec $R = \sigma_e^2 I$, l'estimateur de σ_e^2, qui est l'estimateur du maximum de vraisemblance, est biaisé.

Si on s'intéresse à l'estimation des composantes de la variance en premier lieu, les paramètres de position θ sont inutiles en soit, on les appelle paramètres de nuisance. Pour les éliminer, le statisticien bayésien va les intégrer, et calculer la densité marginale des paramètres d'intérêt : $f(\gamma|y) = \int f(\theta, \gamma|y) d\theta$.

La maximisation de cette densité marginale conduit à des estimateurs de type REML (cf. chapitre 2).

Chapitre 5

Quelques Références

Bonnet A., Benne F., Dantec C., Gobert N., Frappart P.O., San Cristobal M., Hatey F., Tosser-Klopp G. (2044) Identification of genes and gene networks involved in pig ovarian follicular development, by using c-dna microarrays. In III International Workshop on the Development and Function of Reproductive organs, 2004.

Cochran, W. G. (1951), Testing a Linear Relation Among Variances, Biometrics, 7, 17-32.

Liang K.Y., Zeger S.L. (1986), Longitudinal data analysis using generalized linear models,Biometrika, 73, 13-22.

Lindley D.V., Smith A.F.M. (1972), Bayes Estimates for the Linear Model, Journal of the Royal Statistical Society B, 34, 1-41.

Miller J.J. (1977) Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. Ann. of Statistics, 5 :746-762.

Patterson H.D., Thompson R. (1971), Recovery of inter-block information when block sizes are unequal, Biometrika, 58, 545-554.

Rao C.R., Kleffe J. (1988), Estimation of variance components and applications, North Holland series in statistics and probability, Elsevier, Amsterdam.

Robert C. (1992) L'Analyse statistique bay¶esienne. Paris : Economica.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992) Variance Components, Wiley, New York.

Verbeke G., Molenberghs G. (1997) Linear Mixed Models in Practice : A SAS-Oriented Approach. Lecture Notes in Statistics 126. New York : Springer-Verlag.

Verbeke G., Molenberghs G. (2000) Linear mixed models for longitudinal data. Springer, New York.