MATH 60604
Modélisation statistique
§ 5b - Exemple de données longitudinales

Léo Belzile

HEC Montréal
Département de sciences de la décision

Choix de la structure de covariance: autres possibilités

- Avec des données longitudinales, il arrive parfois que la variance des observations varie en fonction du temps de mesure.
- Plusieurs des structures de covariance possibles dans la procédure `mixed` possèdent aussi une version hétérogène, c'est-à-dire une version où les variances peuvent être distinctes pour les différents temps de mesure.
- Par exemple, la structure $\mathsf{AR}(1)$ que nous venons de voir possèdent une version hétérogène, appelée $\mathsf{ARH}(1)$, dont la structure de covariance est

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 & \sigma_1\sigma_5\rho^4 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 & \sigma_2\sigma_5\rho^3 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho & \sigma_3\sigma_5\rho^2 \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 & \sigma_4\sigma_5\rho \\ \sigma_5\sigma_1\rho^4 & \sigma_5\sigma_2\rho^3 & \sigma_5\sigma_3\rho^2 & \sigma_5\sigma_2\rho & \sigma_5^2 \end{pmatrix}.$$

- Mais au lieu de supposer une variance commune σ^2 à tous les temps de mesure, on suppose plutôt que la variance au temps j est σ_j^2.

Syntaxe pour l'ajustement du modèle ARH(1)

Code SAS pour ajuster le modèle ARH(1)

```
proc mixed data=vengeance method=reml;
class id tcat;
model vengeance = sexe age vc wom t / solution;
repeated tcat / subject=id type=arh(1) r=1 rcorr=1;
run;
```

Matrice de corrélation et de covariance pour le sujet 1

Matrice R estimée pour id 1

Ligne	Col1	Col2	Col3	Col4	Col5
1	0.2937	0.1756	0.09260	0.05184	0.02271
2	0.1756	0.3937	0.2077	0.1162	0.05093
3	0.09260	0.2077	0.4109	0.2300	0.1008
4	0.05184	0.1162	0.2300	0.4829	0.2116
5	0.02271	0.05093	0.1008	0.2116	0.3477

Matrice de corrélation R estimée pour id 1

Ligne	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.5163	0.2666	0.1376	0.07107
2	0.5163	1.0000	0.5163	0.2666	0.1376
3	0.2666	0.5163	1.0000	0.5163	0.2666
4	0.1376	0.2666	0.5163	1.0000	0.5163
5	0.07107	0.1376	0.2666	0.5163	1.0000

- La matrice de covariance montre bien que la variance des observations est différente pour chaque temps de mesure
- la matrice de corrélation montre, comme pour la structure $\mathrm{AR}(1)$, que la corrélation entre 2 observation décroit avec le temps.

Estimés des paramètres de covariance du modèle ARH(1)

Valeur estimée du paramètre de covariance		
Param. de cov.	**Sujet**	**Estimation**
Var(1)	id	0.2937
Var(2)	id	0.3937
Var(3)	id	0.4109
Var(4)	id	0.4829
Var(5)	id	0.3477
ARH(1)	id	0.5163

- Ce modèle comporte six paramètres pour la structure de covariance.
- Dans le tableau, on voit les estimations des variances pour les cinq unités de temps.
- L'estimation du paramètre ρ est $\widehat{\rho} = 0{,}516$, très semblable à celui du modèle précédent (0,492).

Test du rapport de vraisemblance pour le modèle ARH(1)

Matrice R estimée pour id 1					
Ligne	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	0.3133	0.2165	0.07287	0.05930	0.03405
2	0.2165	0.4233	0.2258	0.1355	0.07025
3	0.07287	0.2258	0.4204	0.2371	0.1689
4	0.05930	0.1355	0.2371	0.4508	0.1444
5	0.03405	0.07025	0.1689	0.1444	0.3179

Matrice de corrélation R estimée pour id 1					
Ligne	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	1.0000	0.5946	0.2008	0.1578	0.1079
2	0.5946	1.0000	0.5354	0.3103	0.1915
3	0.2008	0.5354	1.0000	0.5446	0.4621
4	0.1578	0.3103	0.5446	1.0000	0.3815
5	0.1079	0.1915	0.4621	0.3815	1.0000

- La sortie SAS inclut les résultats du test de rapport de vraisemblance comparant le modèle ARH(1) (modèle complet) au modèle homoscédastique sans corrélation (modèle réduit).
- Les hypothèses de ce test sont $\mathscr{H}_0 : \rho = 0, \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_5^2$ et $\mathscr{H}_1 : \rho \neq 0$ ou au moins une des variances différente.
- La valeur-*p* est négligeable et on rejette $\mathscr{H}_0$; on conclut en faveur de l'utilisation de la structure de covariance ARH(1).

Choix de la structure de covariance: autre possibilité

- Une autre possibilité serait de ne spécifier aucune structure pour la covariance et estimer tous les paramètres du modèle

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54}^2 & \sigma_5^2 \end{pmatrix}$$

- Cette structure peut parfois être utile pour explorer la structure de covariance sans imposer un modèle rigide dès le départ. Mais son nombre de paramètres, $n_i(n_i - 1)/2$, restreint son utilisation aux cas où le nombre maximum d'observations par groupe est petit et le nombre de groupes m est grand.
- Dans notre exemple, on obtient 15 paramètres contrairement à deux pour les structures d'équicorrélation et AR(1), et à six pour la structure ARH(1).

Ajustement du modèle de covariance non structuré

Code SAS pour ajuster un modèle non structuré

```
proc mixed data=vengeance method=reml;
class id tcat;
model vengeance = sexe age vc wom t / solution;
repeated tcat / subject=id type=un r=1 rcorr=1;
run;
```

Matrice de corrélation et de covariance pour le sujet 1

Matrice R estimée pour id 1					
Ligne	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	0.3133	0.2165	0.07287	0.05930	0.03405
2	0.2165	0.4233	0.2258	0.1355	0.07025
3	0.07287	0.2258	0.4204	0.2371	0.1689
4	0.05930	0.1355	0.2371	0.4508	0.1444
5	0.03405	0.07025	0.1689	0.1444	0.3179

Matrice de corrélation R estimée pour id 1					
Ligne	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	1.0000	0.5946	0.2008	0.1578	0.1079
2	0.5946	1.0000	0.5354	0.3103	0.1915
3	0.2008	0.5354	1.0000	0.5446	0.4621
4	0.1578	0.3103	0.5446	1.0000	0.3815
5	0.1079	0.1915	0.4621	0.3815	1.0000

- On voit que les variances sont différentes pour chaque temps et il n'y a pas de structure spéciales pour la corrélation.
  1. Les variances semblent être à peu près les mêmes pour les mesures.
  2. La corrélation entre deux observations semble bel et bien décroître au fur et à mesure que le temps entre deux mesures augmente.
- Ceci suggère que la structure $\mathrm{AR}(1)$ est préférable à la structure d'équicorrélation.

Paramètres de covariance du modèle non structuré

Valeur estimée du paramètre de covariance		
Param. de cov.	**Sujet**	**Estimation**
UN(1,1)	id	0.3133
UN(2,1)	id	0.2165
UN(2,2)	id	0.4233
UN(3,1)	id	0.07287
UN(3,2)	id	0.2258
UN(3,3)	id	0.4204
UN(4,1)	id	0.05930
UN(4,2)	id	0.1355
UN(4,3)	id	0.2371
UN(4,4)	id	0.4508
UN(5,1)	id	0.03405
UN(5,2)	id	0.07025
UN(5,3)	id	0.1689
UN(5,4)	id	0.1444
UN(5,5)	id	0.3179

Critères d'information et test du rapport de vraisemblance

Tests d'ajustement	
-2 log-vraisemblance restreinte	659.3
AIC (préférer les petites valeurs)	689.3
AICC (préférer les petites valeurs)	690.6
BIC (préférer les petites valeurs)	725.0

Test du rapport de vraisemblance du modèle nul		
DDL	khi-2	Pr > khi-2
14	117.34	<.0001

- Comme d'ordinaire, la sortie inclut les critères AIC et BIC, que l'on pourra utiliser pour comparer les modèles de covariance non emboîtés.
- Le test du rapport de vraisemblance teste l'hypothèse nulle que tous les paramètres de variance (diagonale) sont égaux et que tous les termes hors diagonale sont nuls. Cette hypothèse est rejetée.