MATH 60604
Modélisation statistique
§ 5e - Modèle autorégressif d'ordre 1

HEC Montréal
Département de sciences de la décision

Choix de la structure de covariance

- Nous avons vu que modéliser la corrélation entre les mesures répétées d'une même personne en supposant l'équicorrélation semblait raisonnable.
- Cette structure est en effet plausible et, de plus, le paramètre qui mesurait cette covariance (ou corrélation) était significativement différent de zéro.
- Mais comment savoir si cette structure est adéquate? Une autre pourrait être préférable.
- Il existe plusieurs autres modèles de covariance et SAS permet d'en spécifier un très grand nombre. Nous en présenterons quelques-uns ici; une liste exhaustive se trouve dans le module d'aide de SAS.

Choix de la structure de covariance

- Dans beaucoup de cas, la structure de covariance est considérée comme un paramètre de nuisance. Plus précisément, souvent l'intérêt premier de l'étude concerne les effets des variables explicatives, $\boldsymbol{\beta}$.
- Dans ce cas, la structure de covariance ne nous intéresse pas en tant que tel mais nous savons qu'il est important de la modéliser adéquatement afin que l'inférence sur $\boldsymbol{\beta}$ soit valide.
- Dans ce cas, il est possible de baser le choix de la structure de covariance sur les critères d'information si les modèles ne sont pas emboîtés, pour autant qu'ils aient les mêmes variables explicatives (si ajusté avec la méthode REML).

Structure auto-régressive

- La structure équicorrélation utilisée plus tôt suppose que la corrélation entre deux observations est toujours la même.
- Lorsque nous avons à faire à des mesures répétées prises à différents moments dans le temps, comme ici, il est possible que la corrélation entre deux observations dépende du temps écoulé entre les mesures.
- C'est-à-dire, on pourrait croire que plus les observations sont rapprochées dans le temps, plus elles sont corrélées. La structure AR(1) (autorégressive d'ordre 1) est une structure simple qui permet de saisir en partie cet aspect.
- Tout comme la structure équicorrélation, la structure AR(1) comporte deux paramètres: un paramètre de corrélation ρ et un paramètre de variance σ^2.

Structure autorégressive

- Pour l'individu i avec cinq réponses, la matrice de corrélation est

$$\mathbf{R}_i = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}.$$

- La matrice de covariance est

$$\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{R}_i.$$

Corrélation du modèle autorégressif d'ordre un

- La corrélation (conditionnelle) entre deux observations espacées par une seule période de temps (deux semaines ici) est $\rho \in (-1, 1)$.
- La corrélation entre deux observations espacées par deux périodes de temps (quatre semaines) est ρ^2, et ainsi de suite.
- Lorsque $0 < \rho < 1$, la suite ρ, ρ^2, ρ^3, ρ^4, ..., est décroissante. Par conséquent, la corrélation entre deux observations décroît exponentiellement en fonction de la différence de temps entre les deux mesures.

Code SAS pour ajuster un modèle AR(1)

Code SAS pour ajuster un modèle AR(1)

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

Matrice de corrélation et de covariance pour le sujet 1 du modèle AR(1)

Matrice R estimée pour id 1					
Ligne	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	0.3770	0.1855	0.09128	0.04492	0.02210
2	0.1855	0.3770	0.1855	0.09128	0.04492
3	0.09128	0.1855	0.3770	0.1855	0.09128
4	0.04492	0.09128	0.1855	0.3770	0.1855
5	0.02210	0.04492	0.09128	0.1855	0.3770

Matrice de corrélation R estimée pour id 1					
Ligne	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
1	1.0000	0.4921	0.2421	0.1192	0.05863
2	0.4921	1.0000	0.4921	0.2421	0.1192
3	0.2421	0.4921	1.0000	0.4921	0.2421
4	0.1192	0.2421	0.4921	1.0000	0.4921
5	0.05863	0.1192	0.2421	0.4921	1.0000

- On voit que la corrélation entre deux observations diminue avec le temps écoulé entre elles.
- C'est bien ce que nous voulions faire en choisissant la structure de covariance AR(1).

Paramètres de la structure de covariance/corrélation du modèle AR(1)

Valeur estimée du paramètre de covariance		
Param. de cov.	**Sujet**	**Estimation**
AR(1)	id	0.4921
Residual		0.3770

- On voit que l'estimation du paramètre ρ est $\widehat{\rho} = 0{,}492$.
- On peut vérifier dans la matrice de corrélation du sujet 1 que la corrélation entre t_1 et t_2 est de $0{,}492$, que la corrélation entre t_1 et t_3 est de $0{,}492^2 = 0{,}24$, etc.
- Remarque: dans le modèle d'équicorrélation, la corrélation entre deux observations quelconques d'une même personne (peu importe le temps écoulé entre les mesures) était estimée à $0{,}356$.

Estimés des paramètres de la moyenne

Solution pour effets fixes

Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > |t|
Intercept	-0.1686	0.3201	75	-0.53	0.6000
sexe	0.1562	0.09791	75	1.60	0.1149
age	0.04562	0.006540	75	6.98	<.0001
vc	0.5209	0.02831	75	18.40	<.0001
wom	0.4002	0.03590	75	11.15	<.0001
t	-0.5686	0.02335	319	-24.35	<.0001

- Les estimés des paramètres β sont très similaires à ceux du modèle d'équicorrélation, mais pas identiques.
- L'effet de toutes les variables explicatives est significatif, hormis celui du `sexe`. Les conclusions sont les mêmes que précédemment.