

Régularisation dans les Modèles Linéaires Généralisés Mixtes avec effet aléatoire autorégressif

Jocelyn CHAUVET
Catherine TROTTIER, Xavier BRY

JdS 2017, Avignon
29 mai 2017



- 1 Introduction
- 2 Régularisation Ridge dans les GL2M avec effet aléatoire autorégressif d'ordre 1
- 3 Régularisation basée sur la construction de composantes supervisées
- 4 Conclusion & ouverture

L2M pour données répétées :

$$y = X\beta + U_1\xi^1 + \varepsilon$$

- ▶ $y = (\mathbf{y}_{11}, \dots, \mathbf{y}_{1R}, \mathbf{y}_{21}, \dots, \mathbf{y}_{2R}, \dots, \mathbf{y}_{N1}, \dots, \mathbf{y}_{NR})^T$
- ▶ β vecteur des effets fixes et X leur matrice de design
- ▶ $\xi^1 = (\xi_1^1, \xi_2^1, \dots, \xi_N^1)^T$ vecteur de l'effet aléatoire "individu" et $U_1 = I_N \otimes \mathbf{1}_R$ matrice de design correspondante
- ▶ $\varepsilon \sim \mathcal{N}(0, \sigma_0^2 \text{Id}_{NR}) \quad \perp \quad \xi^1 \sim \mathcal{N}(0, \sigma_1^2 \text{Id}_N)$

L2M pour données répétées : $y = X\beta + U_1\xi^1 + \varepsilon$

- ▶ $y = (\mathbf{y}_{11}, \dots, \mathbf{y}_{1R}, \mathbf{y}_{21}, \dots, \mathbf{y}_{2R}, \dots, \mathbf{y}_{N1}, \dots, \mathbf{y}_{NR})^T$
- ▶ β vecteur des effets fixes et X leur matrice de design
- ▶ $\xi^1 = (\xi_1^1, \xi_2^1, \dots, \xi_N^1)^T$ vecteur de l'effet aléatoire "individu" et $U_1 = I_N \otimes \mathbf{1}_R$ matrice de design correspondante
- ▶ $\varepsilon \sim \mathcal{N}(0, \sigma_0^2 \text{Id}_{NR}) \perp \xi^1 \sim \mathcal{N}(0, \sigma_1^2 \text{Id}_N)$

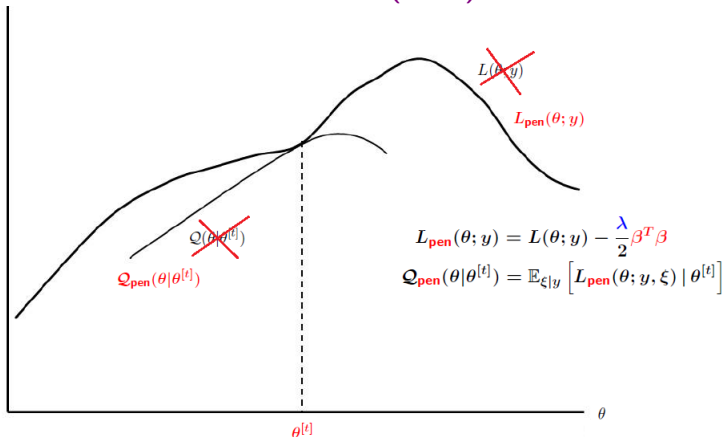
Estimation classique par algorithme EM, $\theta = (\beta, \sigma_0^2, \sigma_1^2)$

$$\mathbf{E} : Q(\theta | \theta^{[t]}) := \mathbb{E}_{\xi|y} [L(\theta; y, \xi) | \theta^{[t]}]$$

$$\mathbf{M} : \theta^{[t+1]} \leftarrow \arg \max_{\theta} Q(\theta | \theta^{[t]})$$

Algorithme EM avec pénalité RIDGE

Eliot et al (2011)



Estimation par algorithme EM pénalisé

- ▶ Comment calibrer le paramètre de régularisation λ ?
 - ↪ **Validation Croisée Généralisée (GCV)** à chaque étape de l'EM

Estimation par algorithme EM pénalisé

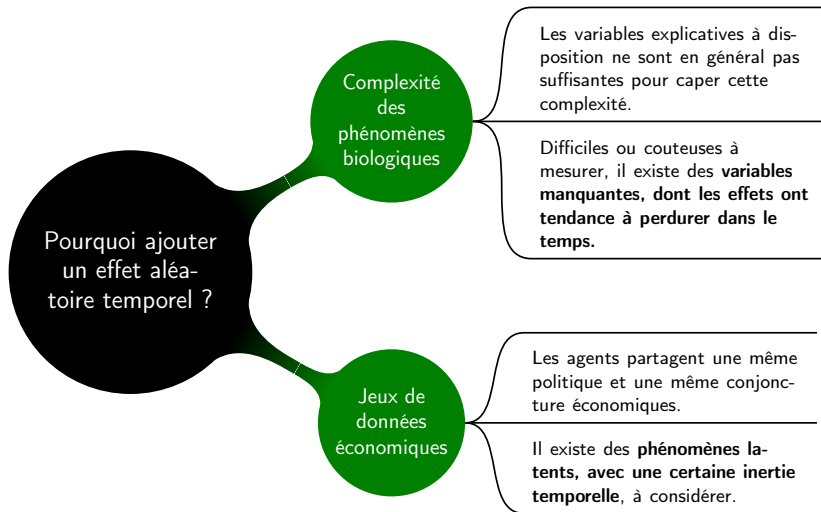
- ▶ Comment calibrer le paramètre de régularisation λ ?
 - ↳ **Validation Croisée Généralisée (GCV)** à chaque étape de l'EM

$$\text{GCV} : \lambda^{[t]} \leftarrow \arg \min_{\lambda} \left\{ \text{GCV}(\lambda) = \frac{n^{-1} \left\| y - H_{\lambda}^{[t]} y \right\|^2}{\left[1 - n^{-1} \text{tr} \left(H_{\lambda}^{[t]} \right) \right]^2} \right\}$$

$$\text{E} : Q_{\text{pen}} \left(\theta \mid \theta^{[t]} \right) := \mathbb{E}_{\xi \mid y} \left[L(\theta; y, \xi) - \frac{\lambda^{[t]}}{2} \beta^T \beta \mid \theta^{[t]} \right]$$

$$\text{M} : \theta^{[t+1]} \leftarrow \arg \max_{\theta} Q_{\text{pen}} \left(\theta \mid \theta^{[t]} \right)$$

- 2 Régularisation Ridge dans les GL2M avec effet aléatoire autorégressif d'ordre 1
 - Ajout d'un effet aléatoire temporel AR(1)
 - Extension aux GL2M
 - Tests sur données simulées



L2M pour données longitudinales :

$$y = X\beta + U_1\xi^1 + U_2\xi^2 + \varepsilon$$

$$y = (y_{11}, y_{12}, \dots, y_{1R}, \\ y_{21}, y_{22}, \dots, y_{2R}, \dots, \\ y_{N1}, y_{N2}, \dots, y_{NR})^T$$

- ▶ β vecteur des effets fixes
- ▶ $\xi^1 = (\xi_1^1, \xi_2^1, \dots, \xi_N^1)^T$ vecteur de l'effet aléatoire "individu", $U_1 = I_N \otimes \mathbf{1}_R$ matrice de design correspondante
- ▶ $\xi^2 = (\xi_1^2, \xi_2^2, \dots, \xi_R^2)^T$ vecteur de l'effet aléatoire "temporel", $U_2 = \mathbf{1}_N \otimes I_R$ matrice de design correspondante

Modélisation des effets aléatoires

- ▶ $\varepsilon \sim \mathcal{N}(0, \sigma_0^2 \text{Id}_{NR}) \perp \xi^1 \sim \mathcal{N}(0, \sigma_1^2 \text{Id}_N)$
- ▶ $\xi^2 \sim \mathcal{N}(0, \Sigma)$, avec $\Sigma_{ij} = \sigma_2^2 \frac{\rho^{|i-j|}}{1 - \rho^2}$, $\xi^2 \perp \xi^1$ et $\xi^2 \perp \varepsilon$

Modélisation des effets aléatoires

- ▶ $\varepsilon \sim \mathcal{N}(0, \sigma_0^2 \text{Id}_{NR}) \perp \xi^1 \sim \mathcal{N}(0, \sigma_1^2 \text{Id}_N)$
- ▶ $\xi^2 \sim \mathcal{N}(0, \Sigma)$, avec $\Sigma_{ij} = \sigma_2^2 \frac{\rho^{|i-j|}}{1-\rho^2}$, $\xi^2 \perp \xi^1$ et $\xi^2 \perp \varepsilon$

EM Généralisé, avec

$\theta = (\beta, \sigma_0^2, \sigma_1^2, \sigma_2^2, \rho)$ et $\xi = (\xi^1, \xi^2)$

GCV : $\lambda^{[t]} \leftarrow \arg \min_{\lambda} \text{GCV}(\lambda)$

$$E : \mathcal{Q}_{\text{pen}}(\theta | \theta^{[t]}) := \mathbb{E}_{\xi|y} \left[L(\theta; y, \xi) - \frac{\lambda^{[t]}}{2} \beta^T \beta | \theta^{[t]} \right]$$

$$M : \theta^{[t+1]} \text{ tel que : } \mathcal{Q}_{\text{pen}}(\theta^{[t+1]} | \theta^{[t]}) \geq \mathcal{Q}_{\text{pen}}(\theta^{[t]} | \theta^{[t]})$$

- ② Régularisation Ridge dans les GL2M avec effet aléatoire autorégressif d'ordre 1
 - Ajout d'un effet aléatoire temporel AR(1)
 - **Extension aux GL2M**
 - Tests sur données simulées

Extension aux GL2M

$$Y_i | \xi \stackrel{\text{iid}}{\sim} F \text{ de la famille exponentielle}$$
$$g(\underbrace{\mathbb{E}(Y | \xi)}_{\mu}) = \eta = X\beta + U_1\xi^1 + U_2\xi^2$$

Méthodes d'estimation dans les GL2M

- ▶ Approximations d'intégrales : Laplace, Gauss-Hermite
- ▶ Méthodes de Monte Carlo : MCMC, MCEM, MCML
- ▶ **Méthodes de linéarisation** : Schall (1991), PQL (Breslow et al. (1993))

La méthode

Étape de LINÉARISATION

- ▶ Linéarisation à l'ordre 1 de y_i au voisinage de μ_i :

$$y_i \simeq z_i = g(\mu_i) + (y_i - \mu_i)g'(\mu_i)$$
$$z_i = \eta_i + e_i$$

- ▶ Définition du modèle linéarisé :

$$\mathcal{M} : z = X\beta + U_1\xi^1 + U_2\xi^2 + e, \quad \text{avec } \mathbb{V}(e) = \Gamma$$

Étape d'ESTIMATION

Algorithme EM pénalisé sur le modèle \mathcal{M}

Itération générique

Linéarisation

$$\mathcal{M} : z^{[t]} = X\beta + U_1\xi^1 + U_2\xi^2 + e, \quad \text{avec } \mathbb{V}(e) = \Gamma^{[t]}$$

Estimation

$$\text{GCV} : \lambda^{[t]} \leftarrow \arg \min_{\lambda} \text{GCV}(\lambda)$$

$$\text{E} : \mathcal{Q}_{\text{pen}}(\theta | \theta^{[t]}) := \mathbb{E}_{\xi|z} \left[L(\theta; z^{[t]}, \xi) - \frac{\lambda^{[t]}}{2} \beta^T \beta | \theta^{[t]} \right]$$

$$\text{M} : \theta^{[t+1]} \text{ tel que : } \mathcal{Q}_{\text{pen}}(\theta^{[t+1]} | \theta^{[t]}) \geq \mathcal{Q}_{\text{pen}}(\theta^{[t]} | \theta^{[t]})$$

Mise à jour

$$\text{Définir } \xi^{[t+1]}, z^{[t+1]}, \Gamma^{[t+1]}$$

- ② Régularisation Ridge dans les GL2M avec effet aléatoire autorégressif d'ordre 1
 - Ajout d'un effet aléatoire temporel AR(1)
 - Extension aux GL2M
 - Tests sur données simulées

Plan de simulation

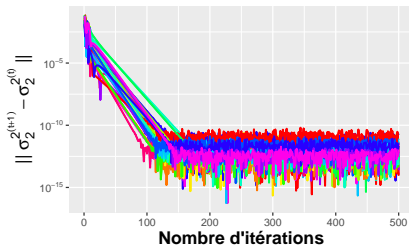
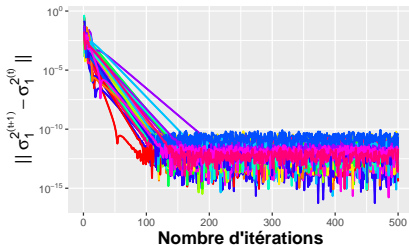
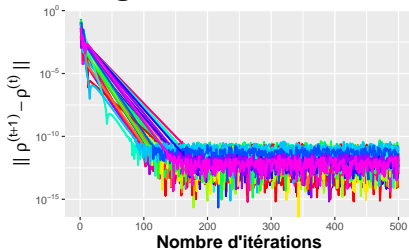
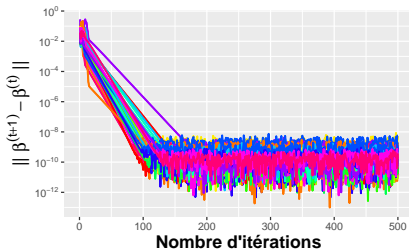
Données Poissonniennes

- ▶ $y \sim \mathcal{P}(\exp(X\beta + U_1\xi^1 + U_2\xi^2))$
- ▶ $X = [\underbrace{x^1, x^2, x^3, x^4, x^5}_{\text{corr. 2 à 2} = 0.5} \mid \underbrace{x^6, x^7, x^8, x^9, x^{10}}_{\text{indépendantes}}]$

3 questions :

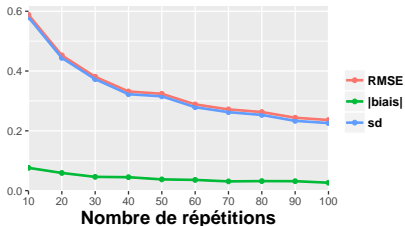
- ▶ La méthode est-elle convergente ?
- ▶ Les paramètres sont-ils bien estimés ?
- ▶ La méthode est-elle sensible à la valeur de ρ ?

La méthode est-elle convergente ?

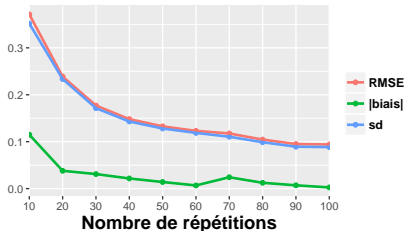


Les paramètres sont-ils bien estimés ?

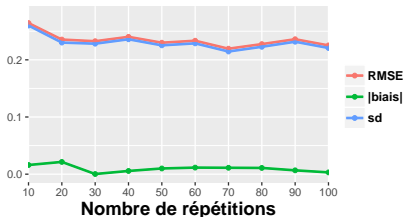
β



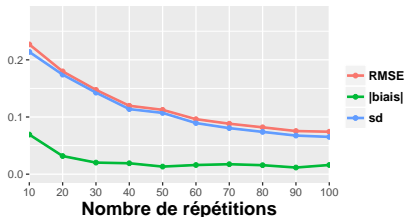
ρ



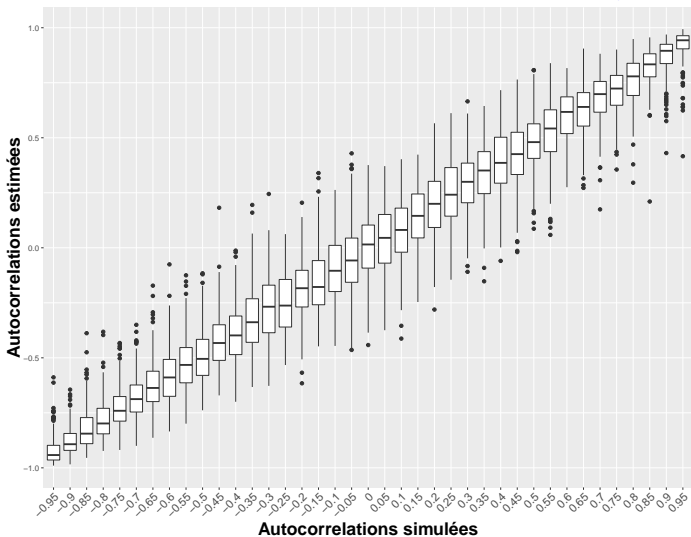
σ_1^2



σ_2^2



La méthode est-elle sensible à la valeur de ρ ?



- 1 Introduction
- 2 Régularisation Ridge dans les GL2M avec effet aléatoire autorégressif d'ordre 1
- 3 Régularisation basée sur la construction de composantes supervisées**
- 4 Conclusion & ouverture

- 3 Régularisation basée sur la construction de composantes supervisées
 - Modèle et méthode d'estimation
 - Résultats numériques

Modèle à composantes

$Y_i | \xi \stackrel{\text{iid}}{\sim} F$ de la famille exponentielle

$$g(\mathbb{E}(Y | \xi)) = \eta = \cancel{\mathbf{X}\boldsymbol{\beta} + U_1\xi^1 + U_2\xi^2}$$

$$\eta = \begin{cases} (\mathbf{X}\mathbf{u})\gamma + U_1\xi^1 + U_2\xi^2 & \text{pour } \mathbf{1} \text{ composante} \\ \sum_{k=1}^{K^*} (\mathbf{X}\mathbf{u}_k)\gamma_k + U_1\xi^1 + U_2\xi^2 & \text{pour } K^* \text{ composantes} \end{cases}$$

Log-Vraisemblance régularisée

On note $\theta = (u, \gamma, \sigma_1^2, \sigma_2^2, \rho)$ et $s \in [0, 1]$

$$L_{\text{reg}}(\theta; y) = (1 - s) L(\theta; y) + s \phi(u)$$

Log-Vraisemblance régularisée

On note $\theta = (u, \gamma, \sigma_1^2, \sigma_2^2, \rho)$ et $s \in [0, 1]$

$$L_{\text{reg}}(\theta; \mathbf{y}) = (1 - s) L(\theta; \mathbf{y}) + s \phi(u)$$

● **Log-Vraisemblance** : mesure (entre autre) la probabilité que les observations \mathbf{y} aient été générées à partir de la composante $\mathbf{f} = \mathbf{X}u$

Log-Vraisemblance régularisée

On note $\theta = (u, \gamma, \sigma_1^2, \sigma_2^2, \rho)$ et $s \in [0, 1]$

$$L_{\text{reg}}(\theta; y) = (1 - s) L(\theta; y) + s \phi(u)$$

● **Log-Vraisemblance** : mesure (entre autre) la probabilité que les observations y aient été générées à partir de la composante $f = Xu$

● **Critère de pertinence structurelle** : mesure la proximité de la composante aux structures fortes de X

À propos de $\phi(u)$...

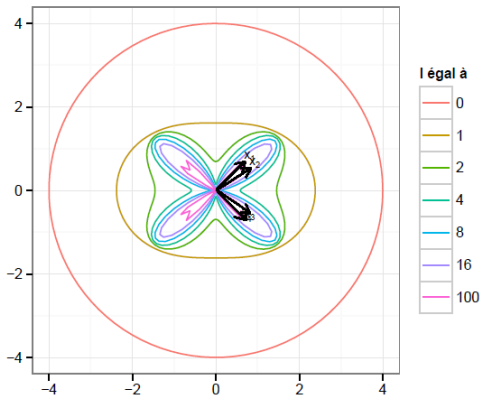
- ▶ Expression générale :

$$\phi(u) = \left(\sum_{j=1}^J [u' N_j u]^l \right)^{\frac{1}{l}}$$

- ▶ Cas particulier d'intérêt :

$$\phi(u) = \left(\sum_{j=1}^p [\text{cor}^2(Xu, x^j)]^l \right)^{\frac{1}{l}}$$

Courbes d'iso-valeur de ϕ



Itération générique pour le modèle à 1 composante

Linéarisation

$$\mathcal{M} : z^{[t]} = (\mathbf{X}\mathbf{u})\gamma + U_1\xi^1 + U_2\xi^2 + e, \quad \text{avec } \mathbb{V}(e) = \Gamma^{[t]}$$

Estimation

$$\mathbf{E} : \mathcal{Q}_{\text{reg}}(\theta, \theta^{[t]}) := \mathbb{E}_{\xi|z} \left[(\mathbf{1} - s)L(\theta; z^{[t]}, \xi) + s\phi(\mathbf{u}) \mid \theta^{[t]} \right]$$

$$\mathbf{M} : \blacktriangleright \text{À } \gamma^{[t]}, \sigma_1^{2[t]}, \sigma_2^{2[t]}, \rho^{[t]} \text{ fixés,}$$

$$u^{[t+1]} \longleftarrow \arg \max_{u: \|u\|=1} \mathcal{Q}_{\text{reg}}(\theta \mid \theta^{[t]})$$

\blacktriangleright Une fois $u^{[t+1]}$ construite,

$$\left(\gamma^{[t+1]}, \sigma_1^{2[t+1]}, \sigma_2^{2[t+1]}, \rho^{[t+1]} \right) \longleftarrow \arg \max_{\gamma, \sigma_1^2, \sigma_2^2, \rho} \mathcal{Q}_{\text{reg}}(\theta \mid \theta^{[t]})$$

Mise à jour : Définir $\xi^{[t+1]}, z^{[t+1]}, \Gamma^{[t+1]}$

- 3 Régularisation basée sur la construction de composantes supervisées
 - Modèle et méthode d'estimation
 - Résultats numériques

Plan de simulation

▶ $y \sim \mathcal{P}(\exp(X\beta + U_1\xi^1 + U_2\xi^2))$

▶ $X =$

$$\left[\underbrace{x^1 \dots x^{10}}_{\substack{\text{Gros faisceau} \\ \hookrightarrow \text{bruit}}} \quad \underbrace{x^{11} \dots x^{14}}_{\substack{\text{Var. indép.} \\ \hookrightarrow \text{bruit}}} \quad \underbrace{x^{15} \dots x^{19}}_{\substack{\text{Petit faisceau} \\ \hookrightarrow \text{explique } y}} \quad \underbrace{x^{20}}_{\substack{\text{Var. isolée} \\ \hookrightarrow \text{explique } y}} \right]$$

Paramètres à calibrer pour la méthode à composantes

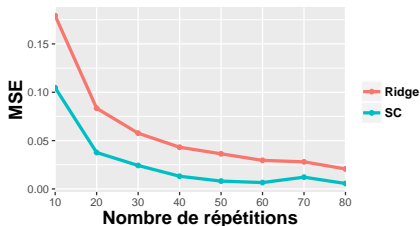
- ▶ le paramètre de compromis s
- ▶ le paramètre de localité des faisceaux visés l
- ▶ le nombre de composantes K^*

Validation croisée... résumé en vidéo !

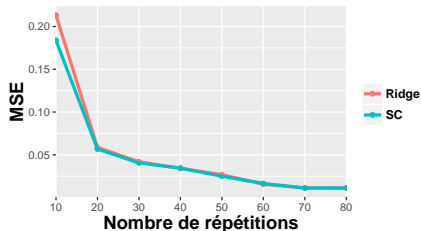
(Chargement...)

Comparaison des résultats

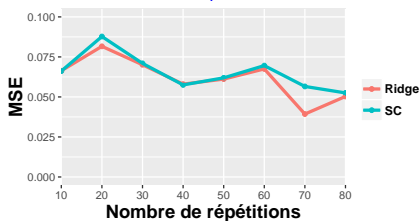
β



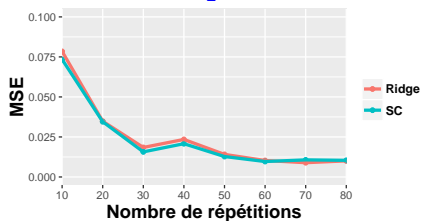
ρ



σ_1^2



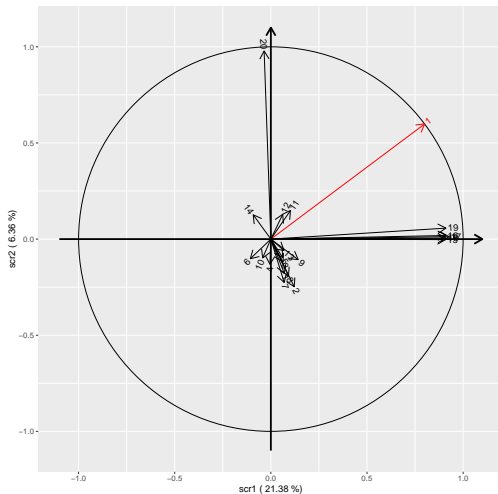
σ_2^2



Aide à l'interprétation : cercles de corrélation

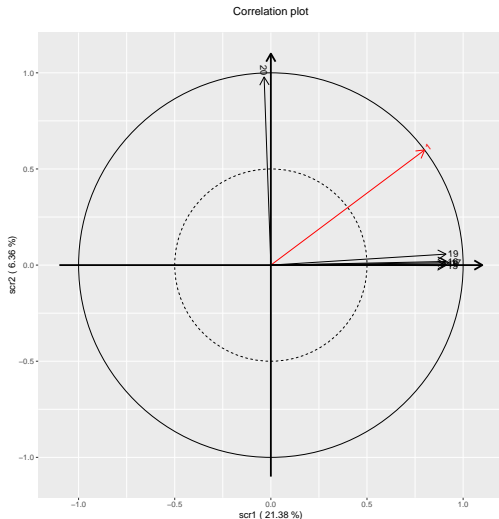
PLAN (1,2)

Correlation plot



Aide à l'interprétation : cercles de corrélation

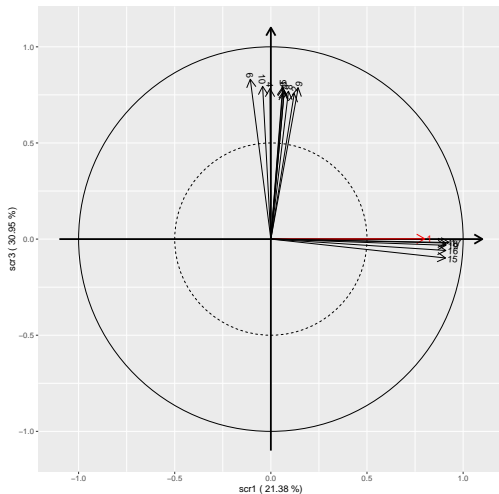
PLAN (1,2)... en évitant les variables mal représentées



Aide à l'interprétation : cercles de corrélation

PLAN (1,3)

Correlation plot



- 1 Introduction
- 2 Régularisation Ridge dans les GL2M avec effet aléatoire autorégressif d'ordre 1
- 3 Régularisation basée sur la construction de composantes supervisées
- 4 Conclusion & ouverture

