

Modèles à Variables Latentes application en neuroimagerie

Brice Ozenne

Hospices Civils de Lyon - 14 décembre 2016



UNIVERSITY OF
COPENHAGEN



Remarque linguistique

Pour me simplifier l'exposition, je vais conserver les acronymes anglais:

ex: LVM modèles à variables latentes

PET tomographie par émission de positons

Contexte clinique: l'étude de la dépression

La dépression saisonnière (hivernale):

- troubles dépressifs en hiver seulement
- lié au manque de soleil
- impact notablement les pays nordiques
(prévalence ~ 5 – 10%, Dam et al. 1998)



→ motivation pour mieux comprendre les mécanismes de la dépression

Comment étudier la dépression ?

- mesurer les symptômes
(tests psychologiques)
 - reconnaissance d'émotions
 - biais de mémorisation



- étudier les causes
(imagerie médicale)

• étude de l'anatomie et de la physiologie du cerveau
• étude de l'imagerie médicale (IRM, SPECT, PET, etc.)
• étude de l'activité cérébrale (EEG, fMRI, etc.)

Comment étudier la dépression ?

- mesurer les symptômes
(tests psychologiques)
 - reconnaissance d'émotions
 - biais de mémorisation

- étudier les causes
(imagerie médicale)
 - sérotonine:
"régulation des émotions"

smil **ben**
drab

serotonine
régulation des émotions

Comment étudier la dépression ?

- mesurer les symptômes
(tests psychologiques)
 - reconnaissance d'émotions
 - biais de mémorisation

- étudier les causes
(imagerie médicale)

- sérotonine:
"régulation des émotions"

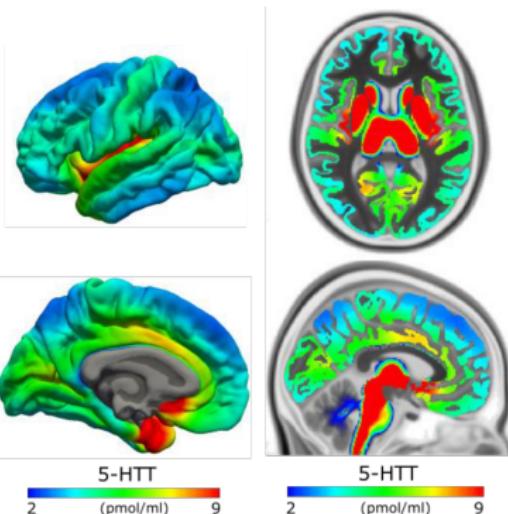
- signal BOLD¹:
"activité cérébrale"

meurtre
jambe
sourire

¹dépendant du niveau d'oxygène sanguin

Comment étudier la dépression ?

- mesurer les symptômes
(tests psychologiques)
 - reconnaissance d'émotions
 - biais de mémorisation



- étudier les causes
(imagerie médicale)
 - sérotonine:
"régulation des émotions"
 - signal BOLD¹:
"activité cérébrale"

Beliveau et al. 2016

¹dépendant du niveau d'oxygène sanguin

Comment étudier la dépression ?

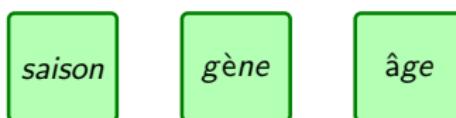
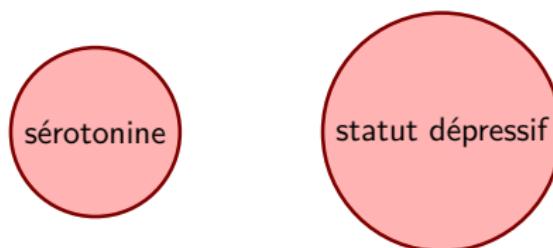
- mesurer les symptômes
(tests psychologiques)
 - reconnaissance d'émotions
 - biais de mémorisation

- étudier les causes
(imagerie médicale)
 - sérotonine:
"régulation des émotions"

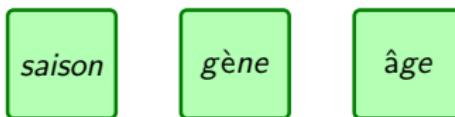
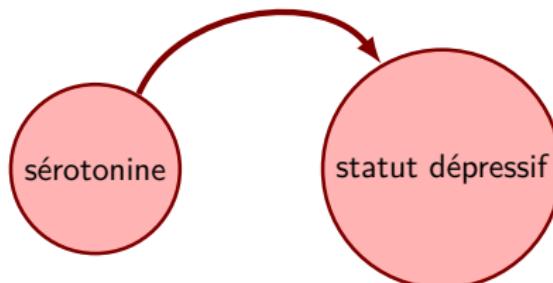
 - signal BOLD¹:
"activité cérébrale"

¹dépendant du niveau d'oxygène sanguin

Exemples de questions cliniques



Exemples de questions cliniques

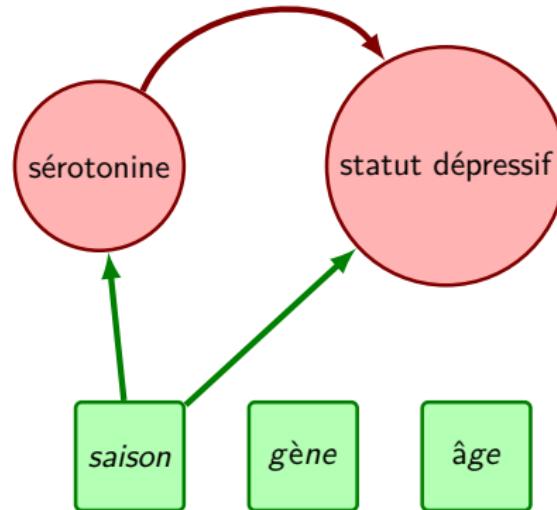


Une défaillance du système de régulation de la sérotonine peut-elle expliquer la dépression ?

→ traitement SSRI²

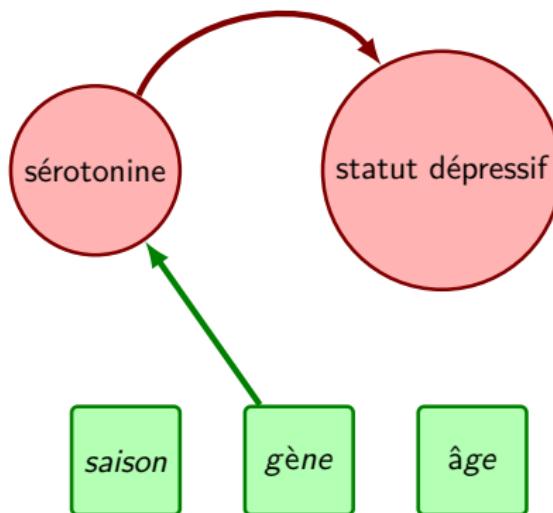
²inhibiteurs sélectifs de la recapture de la sérotonine

Exemples de questions cliniques



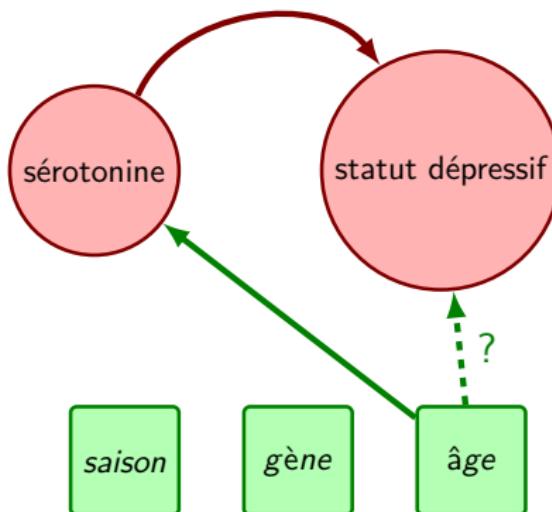
Quelle influence de la saison sur la dépression ?
Effet direct ? Indirect impactant le système sérotoninergique.

Exemples de questions cliniques



Peut-on identifier des patients particulièrement à risque ?

Exemples de questions cliniques



Quelles variables confondantes ?

Que mesure-t-on réellement ?

Tests psychologiques

- reconnaissance d'émotions: vitesse de réaction
- biais de mémorisation: taux d'erreur

→ mesures très **indirectes** du statut dépressif du patient

Imagerie médicale:

- on ne mesure pas la concentration en sérotonine !

on mesure la sérotonine dans le liquide céphalorachidien

on mesure la sérotonine dans les urines (émission nocturne)

on mesure la sérotonine dans les tissus (cerveau, cervelet, etc.)

→ mesure **indirecte** de la concentration en sérotonine

Que mesure-t-on réellement ?

Tests psychologiques

- reconnaissance d'émotions: vitesse de réaction
 - biais de mémorisation: taux d'erreur
- mesures très **indirectes** du statut dépressif du patient

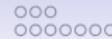
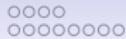
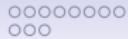
Imagerie médicale:

- on ne mesure pas la concentration en sérotonine !
- on injecte un traceur

qui va se fixer sur les sites de sérotonine dans le cerveau

on peut alors voir où il y a moins de sérotonine

- mesure **indirecte** de la concentration en sérotonine



Que mesure-t-on réellement ?

Tests psychologiques

- reconnaissance d'émotions: vitesse de réaction
 - biais de mémorisation: taux d'erreur
- mesures très **indirectes** du statut dépressif du patient

Imagerie médicale:

- on ne mesure pas la concentration en sérotonine !
 - on injecte un traceur
 - qui va prendre la place la sérotonine sur les récepteurs
 - on mesurer via imagerie PET la concentration en traceur
- mesure **indirecte** de la concentration en sérotonine

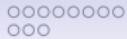
Que mesure-t-on réellement ?

Tests psychologiques

- reconnaissance d'émotions: vitesse de réaction
 - biais de mémorisation: taux d'erreur
- mesures très **indirectes** du statut dépressif du patient

Imagerie médicale:

- on ne mesure pas la concentration en sérotonine !
 - on injecte un traceur
- qui va prendre la place la sérotonine sur les récepteurs
- on mesurer via imagerie PET la concentration en traceur
- mesure **indirecte** de la concentration en sérotonine



Que mesure-t-on réellement ?

Tests psychologiques

- reconnaissance d'émotions: vitesse de réaction
 - biais de mémorisation: taux d'erreur
- mesures très **indirectes** du statut dépressif du patient

Imagerie médicale:

- on ne mesure pas la concentration en sérotonine !
 - on injecte un traceur
 - qui va prendre la place la sérotonine sur les récepteurs
 - on mesurer via imagerie PET la concentration en traceur
- mesure **indirecte** de la concentration en sérotonine

Que mesure-t-on réellement ?

Tests psychologiques

- reconnaissance d'émotions: vitesse de réaction
 - biais de mémorisation: taux d'erreur
- mesures très **indirectes** du statut dépressif du patient

Imagerie médicale:

- on ne mesure pas la concentration en sérotonine !
 - on injecte un traceur
 - qui va prendre la place la sérotonine sur les récepteurs
 - on mesurer via imagerie PET la concentration en traceur
- mesure **indirecte** de la concentration en sérotonine

Que mesure-t-on réellement ?

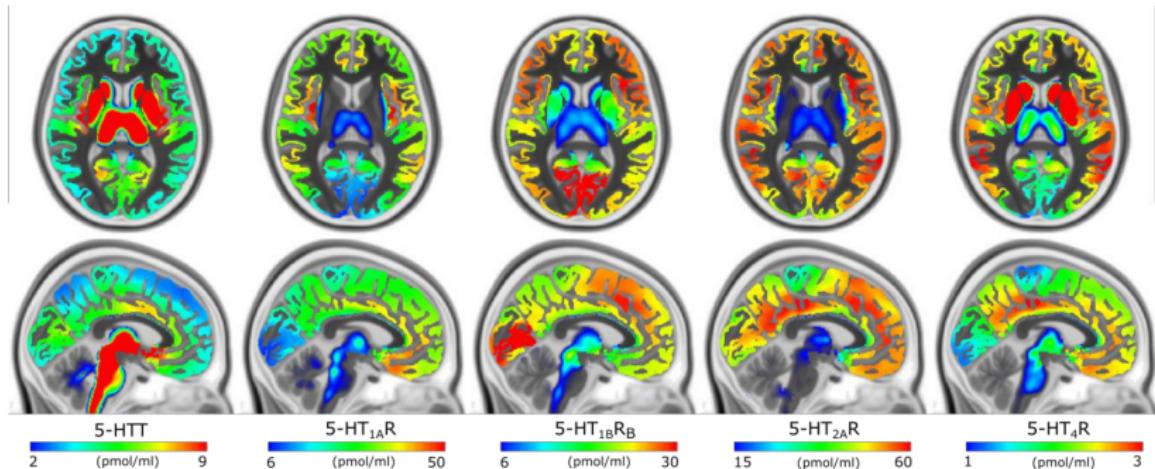
Tests psychologiques

- reconnaissance d'émotions: vitesse de réaction
 - biais de mémorisation: taux d'erreur
- mesures très **indirectes** du statut dépressif du patient

Imagerie médicale:

- on ne mesure pas la concentration en sérotonine !
 - on injecte un traceur
 - qui va prendre la place la sérotonine sur les récepteurs
 - on mesurer via imagerie PET la concentration en traceur
- mesure **indirecte** de la concentration en sérotonine

Des mesures de la sérotonine



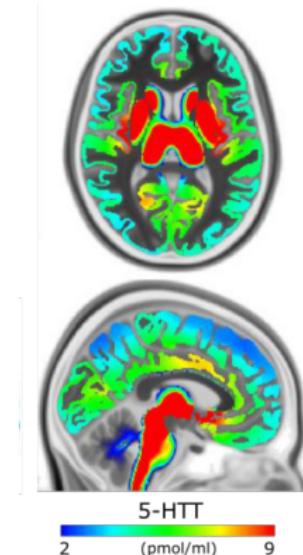
Beliveau et al. 2016

Ici 5 récepteurs différents

PET - En pratique

Le signal est approximativement constant dans chaque région:

- pre-processing (e.g. normalisation)
- définition des régions
- moyenne du signal PET par région



Beliveau et al.
2016

Étude d'un seul récepteur à la fois:

- contraintes pratiques:

acquisition multi-récepteurs complexe et coûteuse

PET - En pratique

Le signal est approximativement constant dans chaque région:

- pre-processing (e.g. normalisation)
- définition des régions
- moyenne du signal PET par région

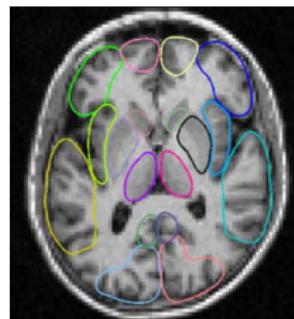
Étude d'un seul récepteur à la fois:

- contraintes pratiques:
acquisition multi-récepteurs complexe et coûteuse
- contraintes liées à l'analyse:
modèle multi-récepteurs ?

PET - En pratique

Le signal est approximativement constant dans chaque région:

- pre-processing (e.g. normalisation)
- définition des régions
- moyenne du signal PET par région



Étude d'un seul récepteur à la fois:

- contraintes pratiques:
acquisition multi-récepteurs complexe et coûteuse
- contraintes liées à l'analyse:
modèle multi-récepteurs ?

PET - En pratique

Le signal est approximativement constant dans chaque région:

- pre-processing (e.g. normalisation)
- définition des régions
- moyenne du signal PET par région

Étude d'un seul récepteur à la fois:

- contraintes pratiques:
acquisition multi-récepteurs complexe et coûteuse
- contraintes liées à l'analyse:
modèle multi-récepteurs ?

PET - En pratique

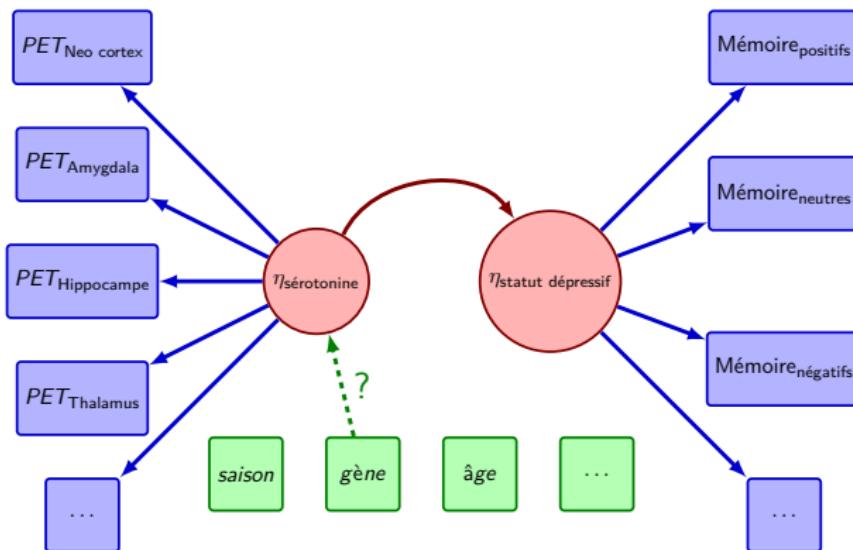
Le signal est approximativement constant dans chaque région:

- pre-processing (e.g. normalisation)
- définition des régions
- moyenne du signal PET par région

Étude d'un seul récepteur à la fois:

- contraintes pratiques:
acquisition multi-récepteurs complexe et coûteuse
- contraintes liées à l'analyse:
modèle multi-récepteurs ?

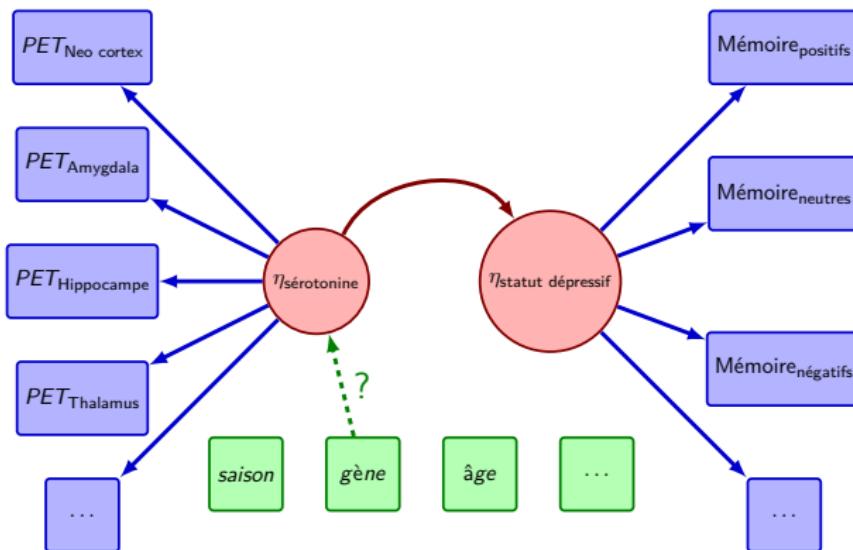
Une représentation plus proche de la réalité



η indique une variable non observée i.e. latente

→ système complexe

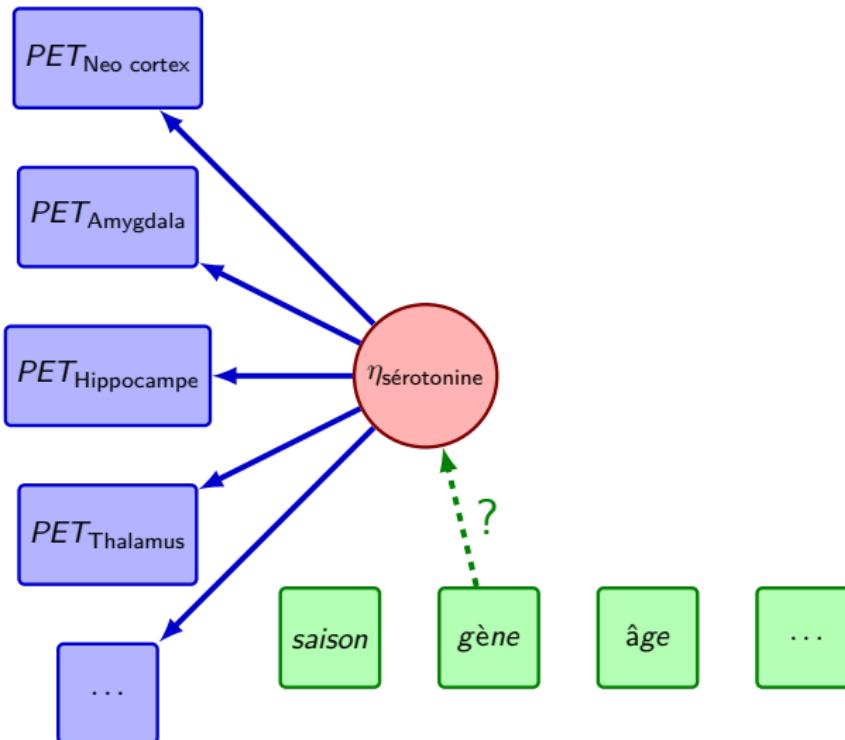
Une représentation plus proche de la réalité



η indique une variable non observée i.e. latente

→ système complexe

Pour aujourd'hui



Notations

Pour chaque individu $i \in [1, n]$, nous disposons de:

- R outcomes: $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iR})$
ex: mesures PET dans chaque région pour chaque individu
 \mathbf{Y} : matrice (n, R)
 $Y = \text{vec}(\mathbf{Y})$: vecteur $n \times R$
- p variables explicatives: $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$
ex: age, sexe, type de scanner, ...

Pourquoi pas un modèle linéaire ?

Un modèle par région ?

- ✓ simple
- ✓ peu d'hypothèses
- ✗ difficile à interpréter:
→ effet génétique: une estimation par région (peu intéressant)

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ \dots \end{bmatrix} = \begin{bmatrix} X_1 \beta_1 \\ X_2 \beta_2 \\ X_3 \beta_3 \\ X_4 \beta_4 \\ \dots \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \dots \end{bmatrix}$$

Pourquoi pas un modèle linéaire ?

Un modèle par région ?

- ✓ simple
- ✓ peu d'hypothèses
- ✗ difficile à interpréter:
→ effet génétique: une estimation par région (peu intéressant)

Un modèle sur l'ensemble des régions

- ✓ test d'un effet global possible
- ✗ modèle complexe
- ✗ hypothèse: effet des variables explicatives identiques pour chaque région

modèle linéaire global

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

Σ matrice (n, n): covariance entre chaque paire d'observations

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} & \dots \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} & \dots \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} & \sigma_{3,5} & \dots \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 & \sigma_{4,5} & \dots \\ \sigma_{1,5} & \sigma_{2,5} & \sigma_{3,5} & \sigma_{4,5} & \sigma_5^2 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

- modèle non identifiable
- il faut faire des hypothèses supplémentaires

modèle linéaire global

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

Σ matrice (n, n) : covariance entre chaque paire d'observations

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} & \dots \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} & \dots \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} & \sigma_{3,5} & \dots \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 & \sigma_{4,5} & \dots \\ \sigma_{1,5} & \sigma_{2,5} & \sigma_{3,5} & \sigma_{4,5} & \sigma_5^2 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

- modèle non identifiable
- il faut faire des hypothèses supplémentaires

modèle linéaire global

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

Σ matrice (n, n): covariance entre chaque paire d'observations

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} & \dots \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} & \dots \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} & \sigma_{3,5} & \dots \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 & \sigma_{4,5} & \dots \\ \sigma_{1,5} & \sigma_{2,5} & \sigma_{3,5} & \sigma_{4,5} & \sigma_5^2 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

- modèle non identifiable
- il faut faire des hypothèses supplémentaires

modèle linéaire global - résidus iid

L'hypothèse la plus courante:

(\mathcal{H}_{iid}) les résidus sont indépendants et identiquement distribués

$$\Rightarrow \Sigma = \sigma^2 Id = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & \dots \\ 0 & \sigma^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma^2 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma^2 & 0 & \dots \\ 0 & 0 & 0 & 0 & \sigma^2 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Alors $\hat{\beta} = \hat{\beta}^{OLS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$

Théorème de Gauss Markov

$\hat{\beta}^{OLS}$ est Best Linear Unbiased Predictor

modèle linéaire global - résidus iid

L'hypothèse la plus courante:

(\mathcal{H}_{iid}) les résidus sont indépendants et identiquement distribués

$$\Rightarrow \Sigma = \sigma^2 Id = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & \dots \\ 0 & \sigma^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma^2 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma^2 & 0 & \dots \\ 0 & 0 & 0 & 0 & \sigma^2 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Alors $\hat{\beta} = \hat{\beta}^{OLS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$

Théorème de Gauss Markov

$\hat{\beta}^{OLS}$ est Best Linear Unbiased Predictor

modèle linéaire global - résidus iid

L'hypothèse la plus courante:

(\mathcal{H}_{iid}) les résidus sont indépendants et identiquement distribués

$$\Rightarrow \Sigma = \sigma^2 Id = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & \dots \\ 0 & \sigma^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma^2 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma^2 & 0 & \dots \\ 0 & 0 & 0 & 0 & \sigma^2 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Alors $\hat{\beta} = \hat{\beta}^{OLS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$

Théorème de Gauss Markov

$\hat{\beta}^{OLS}$ est Best Linear Unbiased Predictor

résidus iid - est-ce bien raisonnable ?

Cohorte: 15 individus contrôles et 17 individus ayant des troubles dépressifs en hiver (Larsen et al. n.d., non publié).

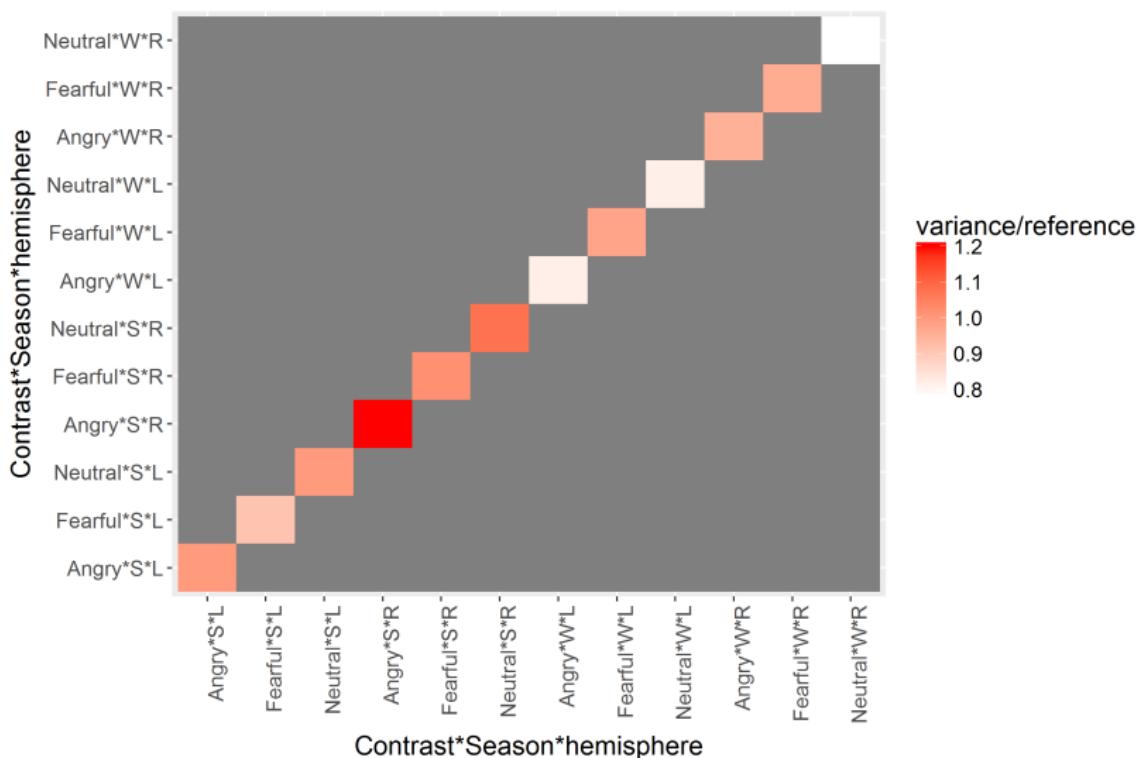
Mesure: signal fMRI de l'amygdale suite à un stimulus visuel (image).

Design: 2 hémisphères cérébraux (droit, gauche)
3 types de stimuli (positif, négatif, neutre)
2 temps de mesures (été, hiver)
⇒ 12 mesures par patient

Modèle: réponse ~ groupe × saison + hémisphère

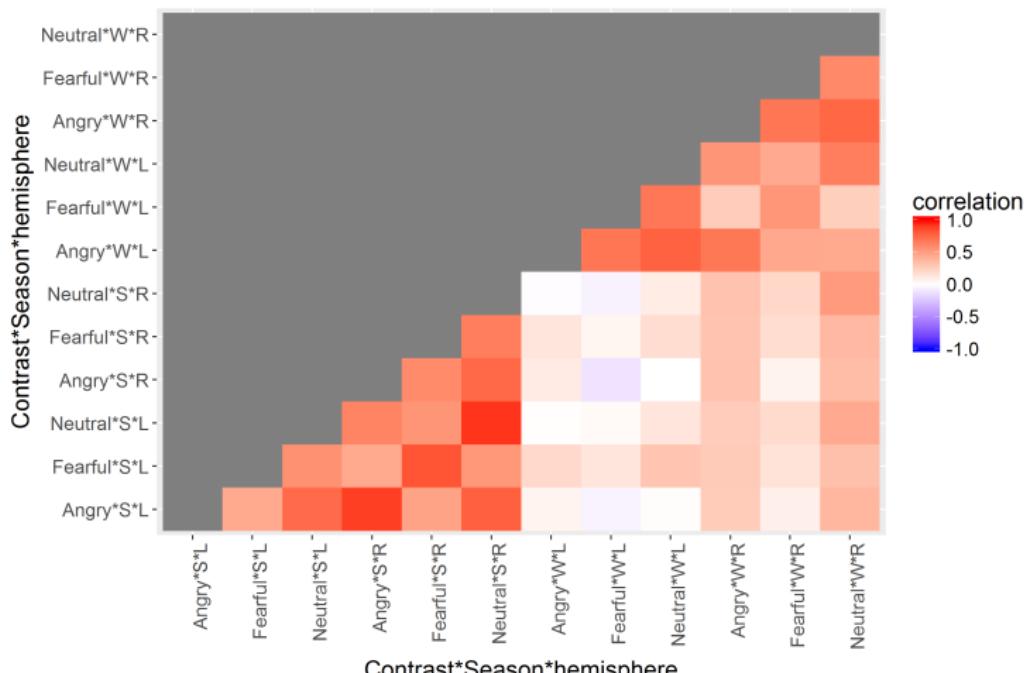
résidus iid - est-ce bien raisonnable ?

Variance des résidus d'un même individu (après modélisation) :



résidus iid - est-ce bien raisonnable ?

Corrélation entre les résidus d'un même individu
(après modélisation) :



modèle linéaire global - résidus non iid

Que se passe-t-il si (\mathcal{H}_{iid}) est fausse ? On suppose toujours :

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\text{i.e. } \mathbb{E}[Y] = \mathbf{X}\beta$$

$$\begin{aligned} \mathbb{E}[\hat{\beta}^{OLS}] &= \mathbb{E}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Y] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbb{E}[Y] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\beta \\ &= \beta \end{aligned}$$

→ β est non biaisé
 (en l'absence de données manquantes)

modèle linéaire global - résidus non iid

Que se passe-t-il si (\mathcal{H}_{iid}) est fausse ? On suppose toujours :

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\text{i.e. } \mathbb{E}[Y] = \mathbf{X}\beta$$

$$\begin{aligned} \mathbb{E}[\hat{\beta}^{OLS}] &= \mathbb{E}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y] \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[Y] \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\beta \\ &= \beta \end{aligned}$$

→ β est non biaisé
(en l'absence de données manquantes)

modèle linéaire global - résidus non iid

Que se passe-t-il si (\mathcal{H}_{iid}) est fausse ? On suppose toujours :

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\text{i.e. } \mathbb{E}[Y] = \mathbf{X}\beta$$

$$\begin{aligned} \mathbb{E}[\hat{\beta}^{OLS}] &= \mathbb{E}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y] \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[Y] \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\beta \\ &= \beta \end{aligned}$$

→ β est non biaisé
 (en l'absence de données manquantes)

modèle linaire global - résidus non iid

Pourquoi ne pas utiliser β_{OLS} ?

- ✗ pas de formule pour σ_β dans le cas non-iid
 - ✓ bootstrap
- ✗ β_{OLS} n'est pas un estimateur efficace
 - perte de puissance

Alternative: spécifier Σ avec un nombre raisonnable de paramètres

- moindres carrés généralisés
- modèles mixtes
- modèles à variables latentes

modèle linaire global - résidus non iid

Pourquoi ne pas utiliser β_{OLS} ?

- ✗ pas de formule pour σ_β dans le cas non-iid
- ✓ bootstrap
- ✗ β_{OLS} n'est pas un estimateur efficace
→ perte de puissance

Alternative: spécifier Σ avec un nombre raisonnable de paramètres

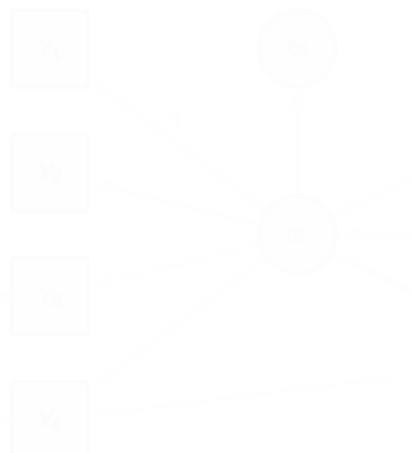
- moindres carrés généralisés
- modèles mixtes
- modèles à variables latentes

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

$$\text{modèle de mesure: } Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \quad \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

$$\text{modèle structurel: } \eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \quad \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$$

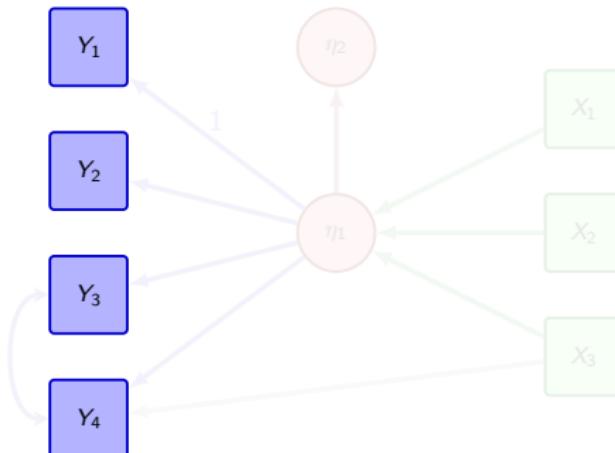


Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



Y : outcomes

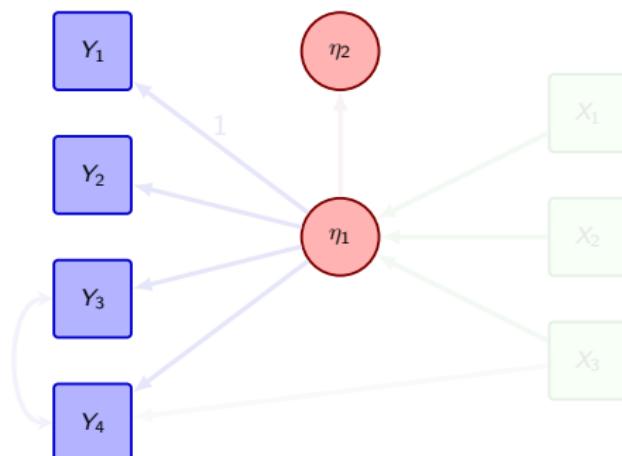
ex: mesures PET

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



η : variables latentes

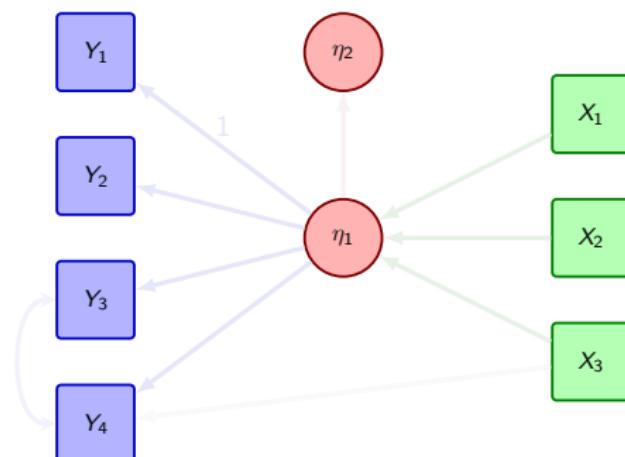
ex: sérotonine

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



X : variables explicatives

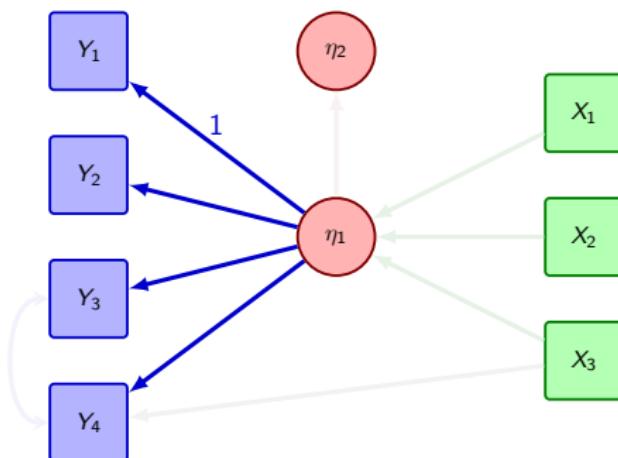
ex: age, sexe, gène

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



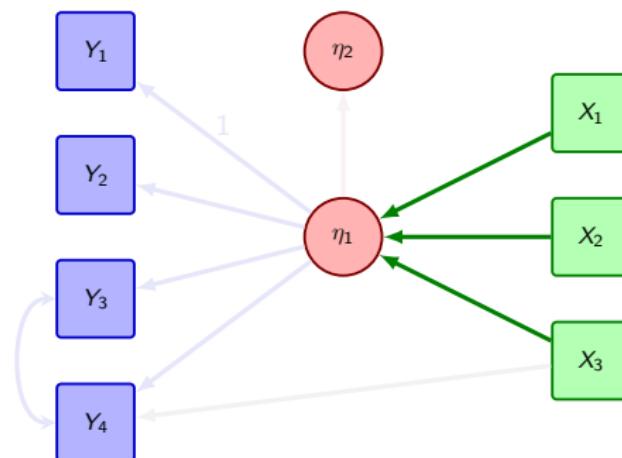
Λ : corrélation entre les outcomes

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



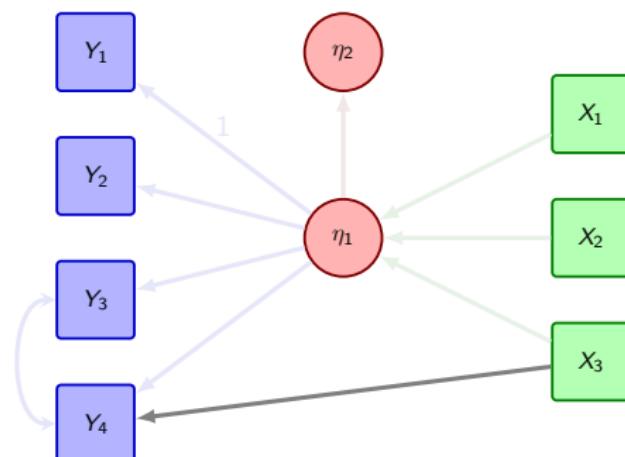
Γ : effets des variables explicatives communs entre les outcomes

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



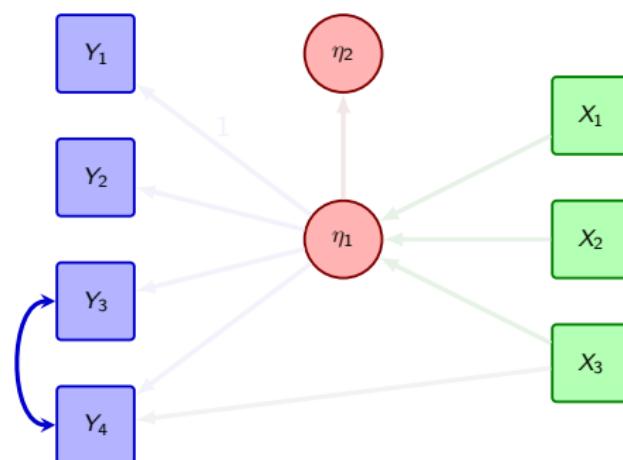
K : effets des variables explicatives outcome-spécifique

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



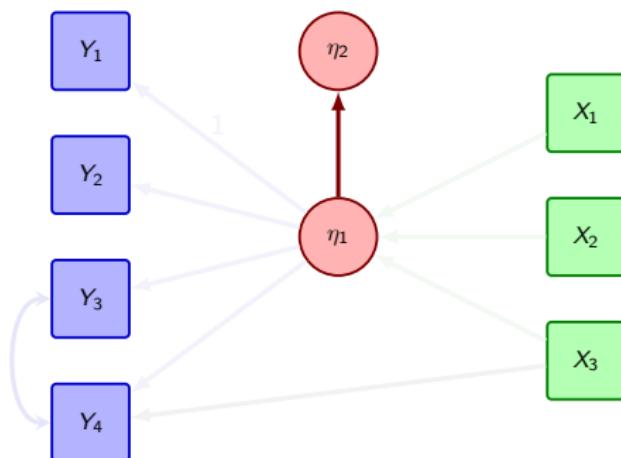
$\Sigma_\varepsilon - diag(\Sigma_\varepsilon)$: corrélation spécifique entre 2 outcomes

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

modèle de mesure: $Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$

modèle structurel: $\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$



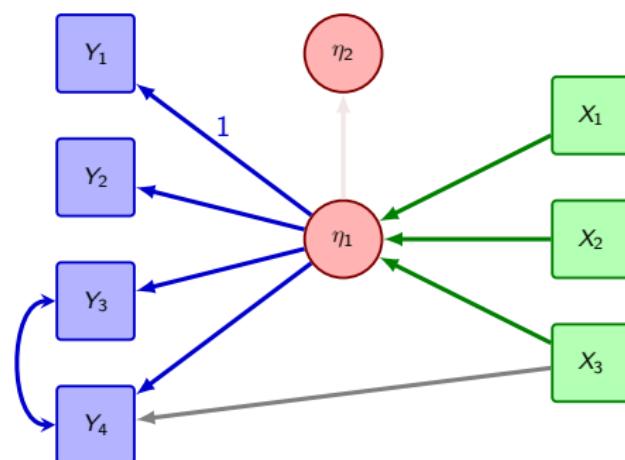
B : relation entre les variables latentes

Extensions du modèle linéaire - LVM

Modèles à variables latentes (LVM):

$$\text{modèle de mesure: } Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \quad \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

$$\text{modèle structurel: } \eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \quad \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$$



Un modèle possible (\mathcal{M}_{vrai})

LVM - Propriétés

$$\begin{aligned}Y_{id} &= \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \quad \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon) \\ \eta_{id} &= \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \quad \zeta \sim \mathcal{N}(0, \Sigma_\zeta)\end{aligned}$$

Donc en notant $\theta = (\nu, \Lambda, K, \Sigma_\varepsilon, \alpha, B, \Gamma, \Sigma_\zeta)$

$$Y_{id}|X_{id} \sim \mathcal{N}(\mu_{id}(\theta, X_{id}), \Sigma(\theta))$$

→ estimation par maximum de vraisemblance
ex: algorithme Newton Raphson

LVM - Propriétés

$$\begin{aligned}Y_{id} &= \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \quad \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon) \\ \eta_{id} &= \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \quad \zeta \sim \mathcal{N}(0, \Sigma_\zeta)\end{aligned}$$

Donc en notant $\theta = (\nu, \Lambda, K, \Sigma_\varepsilon, \alpha, B, \Gamma, \Sigma_\zeta)$

$$Y_{id}|X_{id} = \mathcal{N}(\mu_{id}(\theta, X_{id}), \Sigma(\theta))$$

→ estimation par maximum de vraisemblance
ex: algorithme Newton Raphson

LVM - Propriétés

$$Y_{id} = \nu + \Lambda \eta_{id} + K X_{id} + \varepsilon_{id}, \quad \varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

$$\eta_{id} = \alpha + B \eta_{id} + \Gamma X_{id} + \zeta_{id}, \quad \zeta \sim \mathcal{N}(0, \Sigma_\zeta)$$

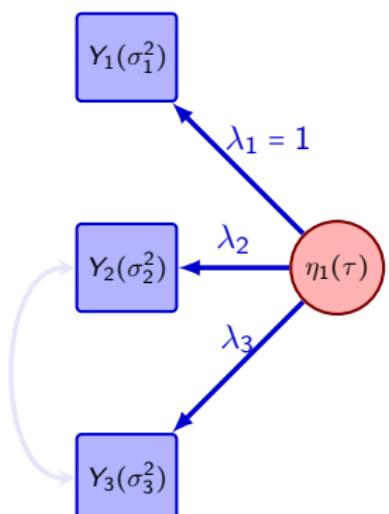
Donc en notant $\theta = (\nu, \Lambda, K, \Sigma_\varepsilon, \alpha, B, \Gamma, \Sigma_\zeta)$

$$Y_{id}|X_{id} = \mathcal{N}(\mu_{id}(\theta, X_{id}), \Sigma(\theta))$$

- estimation par maximum de vraisemblance
ex: algorithme Newton Raphson

LVM - Exemple de matrice de variance covariance

$$Y_{i,id} = \Lambda \eta_{id} + \varepsilon_{i,id}$$



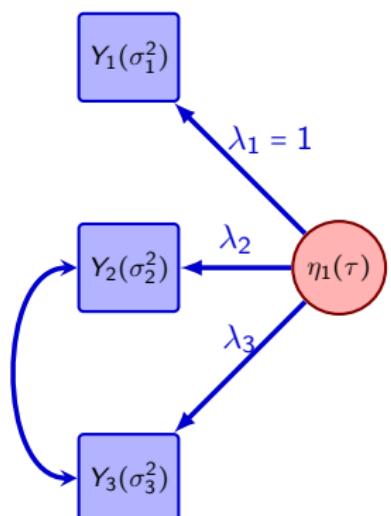
$$\Sigma = \text{diag}(\Sigma_{id})$$

$$= \begin{bmatrix} \Sigma_{id} & 0 & 0 & \dots \\ 0 & \Sigma_{id} & 0 & \dots \\ 0 & 0 & \Sigma_{id} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

$$\Sigma_{id} = \begin{bmatrix} \sigma_1^2 + \tau & \lambda_2\tau & \lambda_3\tau \\ \lambda_2\tau & \sigma_2^2 + \lambda_2^2\tau & \lambda_2\lambda_3\tau \\ \lambda_3\tau & \lambda_2\lambda_3\tau & \sigma_3^2 + \lambda_3^2\tau \end{bmatrix}$$

LVM - Exemple de matrice de variance covariance

$$Y_{i,id} = \Lambda \eta_{id} + \varepsilon_{i,id}$$



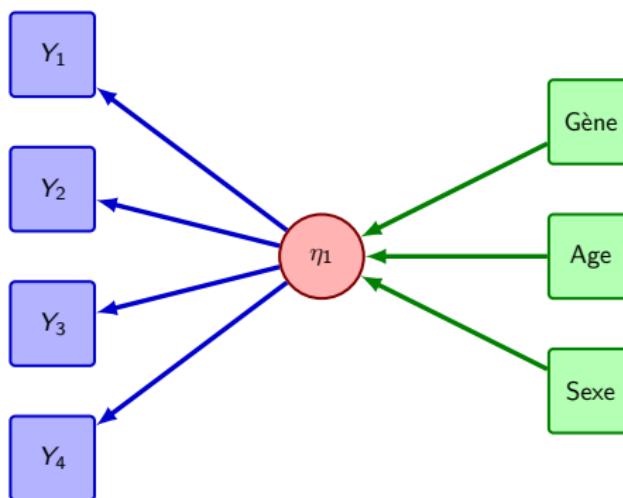
$$\Sigma = \text{diag}(\Sigma_{id})$$

$$= \begin{bmatrix} \Sigma_{id} & 0 & 0 & \dots \\ 0 & \Sigma_{id} & 0 & \dots \\ 0 & 0 & \Sigma_{id} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

$$\Sigma_{id} = \begin{bmatrix} σ₁² + τ & λ₂τ & λ₃τ \\ λ₂τ & σ₂² + λ₂²τ & σ_{2,3} \\ λ₃τ & σ_{2,3} & σ₃² + λ₃²τ \end{bmatrix}$$

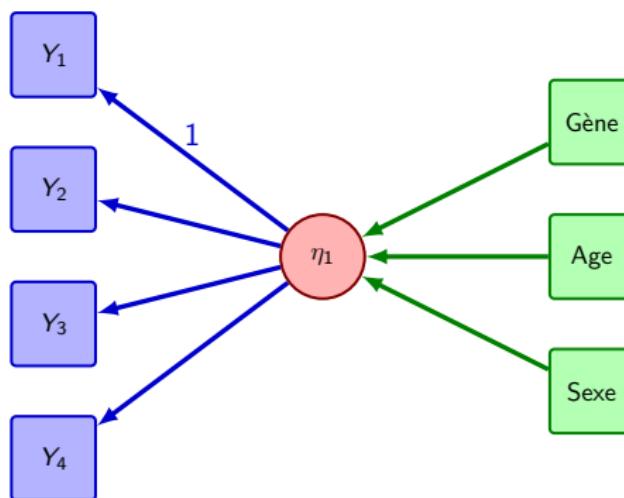
LVM - construction du modèle

Modèle a priori:



LVM - construction du modèle

Contrainte d'identifiabilité:

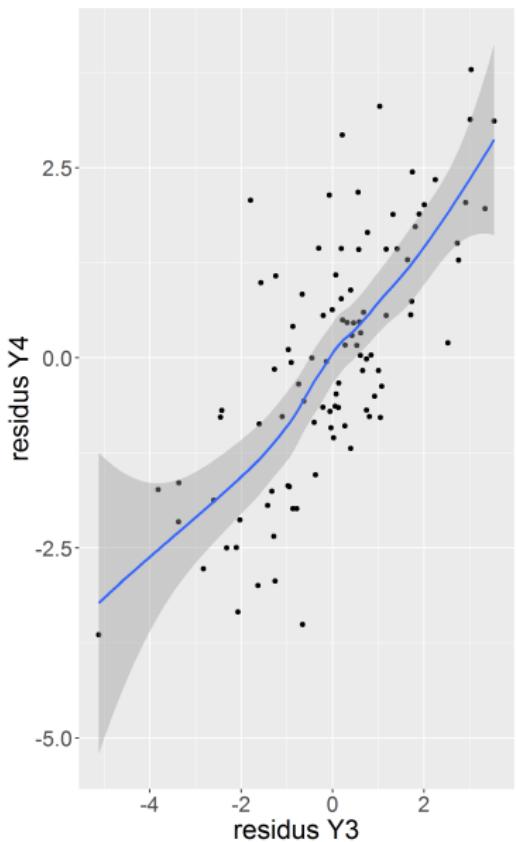


LVM - estimation du modèle

Package lava dans  (Holst and Budtz-Jørgensen 2013):

```
> library(lava)
>
> m.lvm <- lvm(c(Y1,Y2,Y3,Y4)~eta_1,
> eta_1 ~ Gene + Age + Sexe)
> latent(m.lvm) <- ~eta_1
>
> # estimation
> e.lvm <- estimate(m.lvm, data = d)
>
> # diagnostics
> res.lvm <- as.data.frame(residuals(e.lvm))
> gg.res <- ggplot(res.lvm, aes(x = Y3, y = Y4)) + geom_point()
> gg.res <- gg.res + geom_smooth()
```

LVM - diagnostiques



```
> modelsearch(e.lvm)
```

Score:	S	P(S>s)	Index	holm
13.57	0.0002296	Y4~~Gene	0.01102	
13.57	0.0002296	Y4~Gene	0.01102	
13.57	0.0002296	Gene~Y4	0.01102	
13.81	0.0002021	Y3~~Sexe	0.01031	
13.81	0.0002021	Y3~Sexe	0.01031	
13.81	0.0002021	Sexe~Y3	0.01031	
34.63	3.997e-09	Y4~~Sexe	2.158e-07	
34.63	3.997e-09	Y4~Sexe	2.158e-07	
34.63	3.997e-09	Sexe~Y4	2.158e-07	

LVM - diagnostiques

```
> covariance(m.lvm) <- Y3~Y4  
> regression(m.lvm) <- Y4~Sexe  
> e.lvm <- estimate(m.lvm, data = d)  
> gof(e.lvm)
```

```
Number of observations = 100  
BIC = 1359.753  
AIC = 1291.898  
log-Likelihood of model = -628.9489
```

```
log-Likelihood of saturated model = -624.0851  
Chi-squared statistic: q = 9.727685 , df = 9  
P(Q>q) = 0.3729772
```

```
RMSEA (90% CI): 0.0284 (0;0.1183)  
P(RMSEA<0.05)=0.5609202
```

```
rank(Information) = 17 (p=17)  
condition(Information) = 0.0113861  
mean(score^2) = 8.90522e-10
```

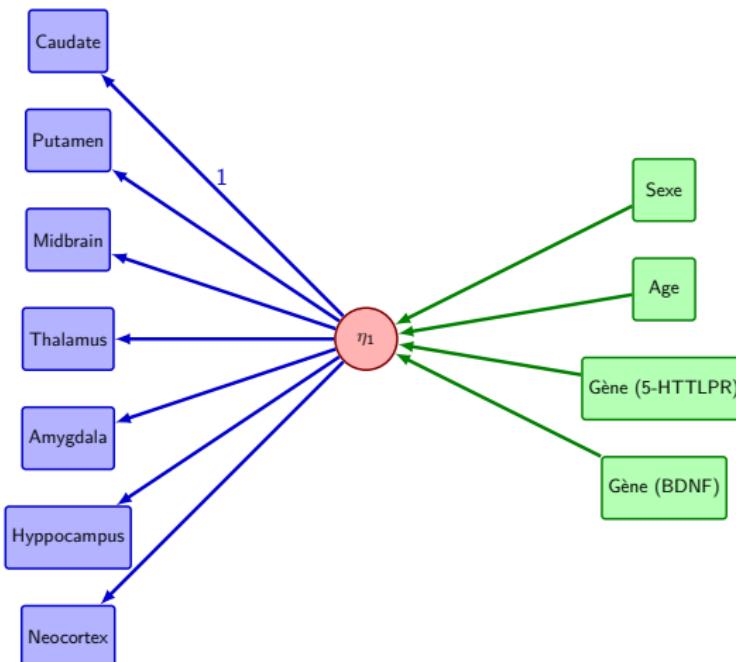
Exemple d'application

Cohorte: 144 individus sans symptômes dépressifs (Fisher et al. n.d., no publié).

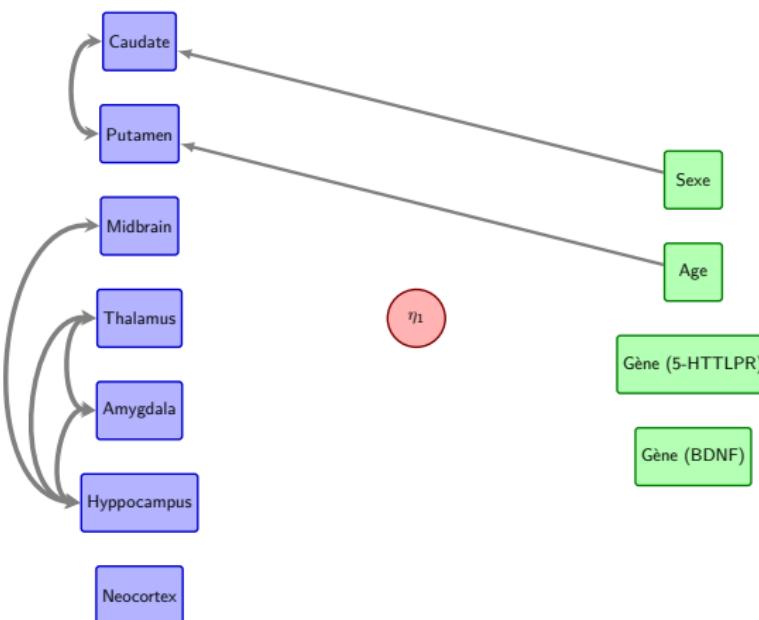
Mesure: signal PET dans différentes régions du cerveau.

Objectif: déterminer s'il existe une association entre le génotype (BDNF, htllpr2) et le niveau de sérotonine.

Application - modèle a priori

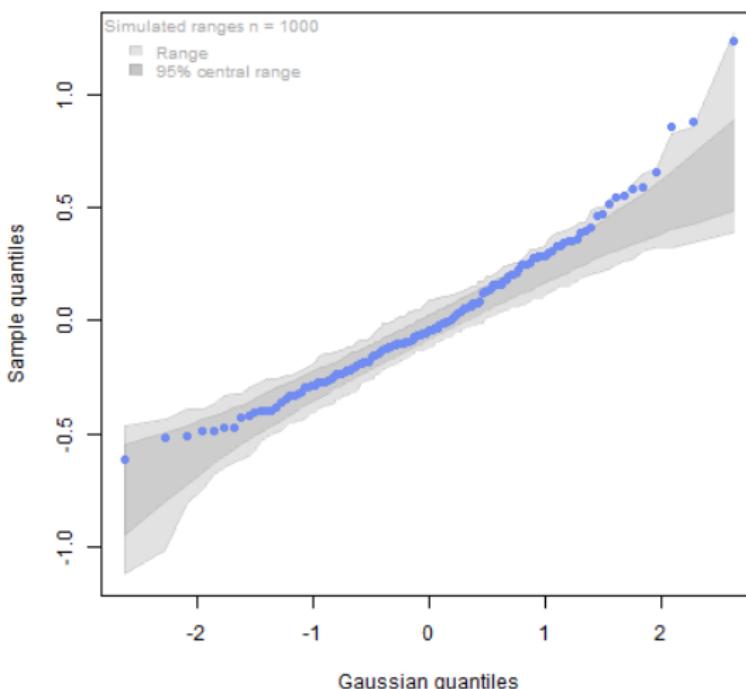


Application - dépendance locale



LVM - hypothèse de normalité

Amygdala



LVM - hypothèse de normalité

Léger écart à la normalité:

- transformer la variable de réponse
- utiliser une estimation "robuste" de la matrice de variance-covariance

```
> eRobust <- estimate(m, data = dd, id = dd[,"cimbi.id"],  
control=list(constrain=T,start=coef(e)))  
> summary(eRobust)
```

	Estimate	Robust SE	Naive SE	SE
u~age0	-0.01892	0.00471	0.00537	
u~sexMale	-0.25708	0.08293	0.07752	
u~httlpr2sx	-0.01448	0.07848	0.08049	
u~bdnf2mx	0.17021	0.08115	0.07919	
...				

Rappel sur la méthode de maximum de vraisemblance

Notons $\mathcal{L}(\theta, Z)$ la vraisemblance du modèle en θ avec les données Z

$\mathcal{S}(\theta, Z)$ la dérivée première de la log-vraisemblance

$\mathcal{H}(\theta, Z)$ la dérivée seconde de la log-vraisemblance

Afin d'estimer θ , nous cherchons à maximiser $\mathcal{L}(\theta, Z)$

i.e., à résoudre $\mathcal{S}(\theta, Z) = 0$

Sous certaines conditions de régularité,
nous pouvons effectuer un développement de Taylor autour de θ :

$$\begin{aligned}\mathcal{S}(\hat{\theta}^{MLE}, Z) &= \sum_{id=1}^n \mathcal{S}_{id}(\hat{\theta}^{MLE}, Z_{id}) = 0 \\ &= \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) + \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} (\hat{\theta}^{MLE} - \theta)\end{aligned}$$

avec $\theta^* \in [\hat{\theta}^{MLE}; \theta]$

Rappel sur la méthode de maximum de vraisemblance

Notons $\mathcal{L}(\theta, Z)$ la vraisemblance du modèle en θ avec les données Z

$\mathcal{S}(\theta, Z)$ la dérivée première de la log-vraisemblance

$\mathcal{H}(\theta, Z)$ la dérivée seconde de la log-vraisemblance

Afin d'estimer θ , nous cherchons à maximiser $\mathcal{L}(\theta, Z)$
i.e., à résoudre $\mathcal{S}(\theta, Z) = 0$

Sous certaines conditions de régularité,
nous pouvons effectuer un développement de Taylor autour de θ :

$$\begin{aligned}\mathcal{S}(\hat{\theta}^{MLE}, Z) &= \sum_{id=1}^n \mathcal{S}_{id}(\hat{\theta}^{MLE}, Z_{id}) = 0 \\ &= \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) + \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} (\hat{\theta}^{MLE} - \theta)\end{aligned}$$

avec $\theta^* \in [\hat{\theta}^{MLE}; \theta]$

Rappel sur la méthode de maximum de vraisemblance

Notons $\mathcal{L}(\theta, Z)$ la vraisemblance du modèle en θ avec les données Z

$\mathcal{S}(\theta, Z)$ la dérivée première de la log-vraisemblance

$\mathcal{H}(\theta, Z)$ la dérivée seconde de la log-vraisemblance

Afin d'estimer θ , nous cherchons à maximiser $\mathcal{L}(\theta, Z)$
i.e., à résoudre $\mathcal{S}(\theta, Z) = 0$

Sous certaines conditions de régularité,
nous pouvons effectuer un développement de Taylor autour de θ :

$$\begin{aligned}\mathcal{S}(\hat{\theta}^{MLE}, Z) &= \sum_{id=1}^n \mathcal{S}_{id}(\hat{\theta}^{MLE}, Z_{id}) = 0 \\ &= \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) + \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} (\hat{\theta}^{MLE} - \theta)\end{aligned}$$

avec $\theta^* \in [\hat{\theta}^{MLE}; \theta]$

Estimateur sandwich²

$$\begin{aligned}
 0 &= \sum_{i=1}^n \mathcal{S}_{id}(\theta, Z_{id}) + \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} (\hat{\theta}^{MLE} - \theta) \\
 \sqrt{n}(\hat{\theta}^{MLE} - \theta) &= - \left(\frac{1}{n} \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right) \\
 &\xrightarrow[n \rightarrow \infty]{D} -\mathcal{H}[\theta, Z]^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right)
 \end{aligned}$$

Si l'on dispose d'un échantillon $\mathbf{Z} = (Z_1, \dots, Z_n)$ d'observations iid alors d'après le théorème limite central:

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta) \sim \mathcal{N}\left(0, \mathcal{H}[\theta, Z]^{-1} \frac{1}{n} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id})^2 \mathcal{H}[\theta, Z]^{-1}\right)$$

²les conditions de regularité sur la vraisemblance sont ici omises

Estimateur sandwich²

$$\begin{aligned}
 0 &= \sum_{i=1}^n \mathcal{S}_{id}(\theta, Z_{id}) + \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} (\hat{\theta}^{MLE} - \theta) \\
 \sqrt{n}(\hat{\theta}^{MLE} - \theta) &= - \left(\frac{1}{n} \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right) \\
 &\xrightarrow[n \rightarrow \infty]{D} -\mathcal{H}[\theta, Z]^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right)
 \end{aligned}$$

Si l'on dispose d'un échantillon $\mathbf{Z} = (Z_1, \dots, Z_n)$ d'observations iid alors d'après le théorème limite central:

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta) \sim \mathcal{N}\left(0, \mathcal{H}[\theta, Z]^{-1} \frac{1}{n} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id})^2 \mathcal{H}[\theta, Z]^{-1}\right)$$

²les conditions de regularité sur la vraisemblance sont ici omises

Estimateur sandwich²

$$\begin{aligned}
 0 &= \sum_{i=1}^n \mathcal{S}_{id}(\theta, Z_{id}) + \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} (\hat{\theta}^{MLE} - \theta) \\
 \sqrt{n}(\hat{\theta}^{MLE} - \theta) &= - \left(\frac{1}{n} \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right) \\
 &\xrightarrow[n \rightarrow \infty]{D} -\mathcal{H}[\theta, Z]^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right)
 \end{aligned}$$

Si l'on dispose d'un échantillon $\mathbf{Z} = (Z_1, \dots, Z_n)$ d'observations iid alors d'après le théorème limite central:

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta) \sim \mathcal{N}\left(0, \mathcal{H}[\theta, Z]^{-1} \frac{1}{n} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id})^2 \mathcal{H}[\theta, Z]^{-1}\right)$$

²les conditions de regularité sur la vraisemblance sont ici omises

Estimateur sandwich²

$$\begin{aligned}
 0 &= \sum_{i=1}^n \mathcal{S}_{id}(\theta, Z_{id}) + \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} (\hat{\theta}^{MLE} - \theta) \\
 \sqrt{n}(\hat{\theta}^{MLE} - \theta) &= - \left(\frac{1}{n} \sum_{id=1}^n \frac{\partial \mathcal{S}_{id}(\theta^*, Z_{id})}{\partial \theta} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right) \\
 &\xrightarrow[n \rightarrow \infty]{D} -\mathcal{H}[\theta, Z]^{-1} \left(\frac{1}{\sqrt{n}} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id}) \right)
 \end{aligned}$$

Si l'on dispose d'un échantillon $\mathbf{Z} = (Z_1, \dots, Z_n)$ d'observations iid alors d'après le théorème limite central:

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta) \sim \mathcal{N}\left(0, \mathcal{H}[\theta, Z]^{-1} \frac{1}{n} \sum_{id=1}^n \mathcal{S}_{id}(\theta, Z_{id})^2 \mathcal{H}[\theta, Z]^{-1}\right)$$

²les conditions de régularité sur la vraisemblance sont ici omises

Estimateur sandwich

Estimation de l'incertitude autour de θ :

- ✓ asymptotiquement valide sans hypothèse de normalité
- ✓ asymptotiquement valide même si Σ est mal spécifié
(les observations au niveau cluster doivent être indépendantes)
- ✓ coïncide avec la matrice d'information si le modèle est valide
- ✗ Pas nécessairement efficace

Estimateur sandwich

Estimation de l'incertitude autour de θ :

- ✓ asymptotiquement valide sans hypothèse de normalité
- ✓ asymptotiquement valide même si Σ est mal spécifié
(les observations au niveau cluster doivent être indépendantes)
- ✓ coïncide avec la matrice d'information si le modèle est valide
- ✗ Pas nécessairement efficace

LVM - Extensions

Déjà implémentées par Klaus et Esben:

- données manquantes (Missing At Random)

$$\begin{aligned}\mathcal{L}(\theta, y^{obs}|x) &= \int \mathbb{P}_\theta [y^{obs}, y^{manquant}|x] dy^{manquant} \\ &= \mathbb{P}_\theta [y^{obs}|x] \int \mathbb{P}_\theta [y^{manquant}|y^{obs}, x] dy^{manquant}\end{aligned}$$

→ Full Information Maximum Likelihood

- données binaires ou censurées (`lava.tobit`)

En cours:

- modèles non-linéaires (`lava.nlin`)
- temps jusqu'à événement
- intégration d'un terme de régularisation

LVM - Extensions

Déjà implémentées par Klaus et Esben:

- données manquantes (Missing At Random)

$$\begin{aligned}\mathcal{L}(\theta, y^{obs}|x) &= \int \mathbb{P}_\theta [y^{obs}, y^{manquant}|x] dy^{manquant} \\ &= \mathbb{P}_\theta [y^{obs}|x] \int \mathbb{P}_\theta [y^{manquant}|y^{obs}, x] dy^{manquant}\end{aligned}$$

→ Full Information Maximum Likelihood

- données binaires ou censurées (`lava.tobit`)

En cours:

- modèles non-linéaires (`lava.nlin`)
- temps jusqu'à événement
- intégration d'un terme de régularisation

LVM - Pourquoi régulariser

- pour pouvoir estimer le modèle si $p > n$
e.g. analyse au niveau voxel des images PET
- pour sélectionner les variables de façon pertinente
- pour étudier la sensibilité du processus de sélection sur l'estimation de l'effet

Simulation (risque α)

Design:

- X variable d'intérêt
- C_1 variable confondante connue
- C_2 to C_9 variables confondantes potentielles
→ ici aucune n'a d'effet sur Y

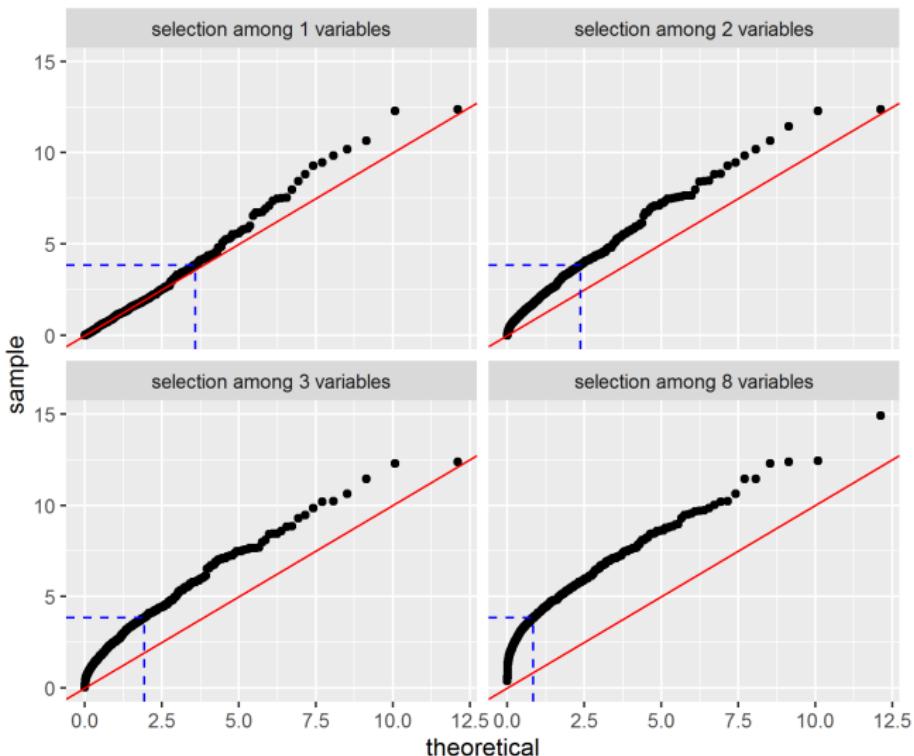
Modèle

- $\mathcal{M}_0 : Y \sim X + C_1$
- $\mathcal{M}_i : Y \sim X + C_1 + C_i$ pour $i \in \{2, \dots, 9\}$
- $\mathcal{M}_j : Y \sim X + C_1 + C_j$ où j est la variable confondante la plus significative

Test: test du ratio de vraisemblance (LRT)

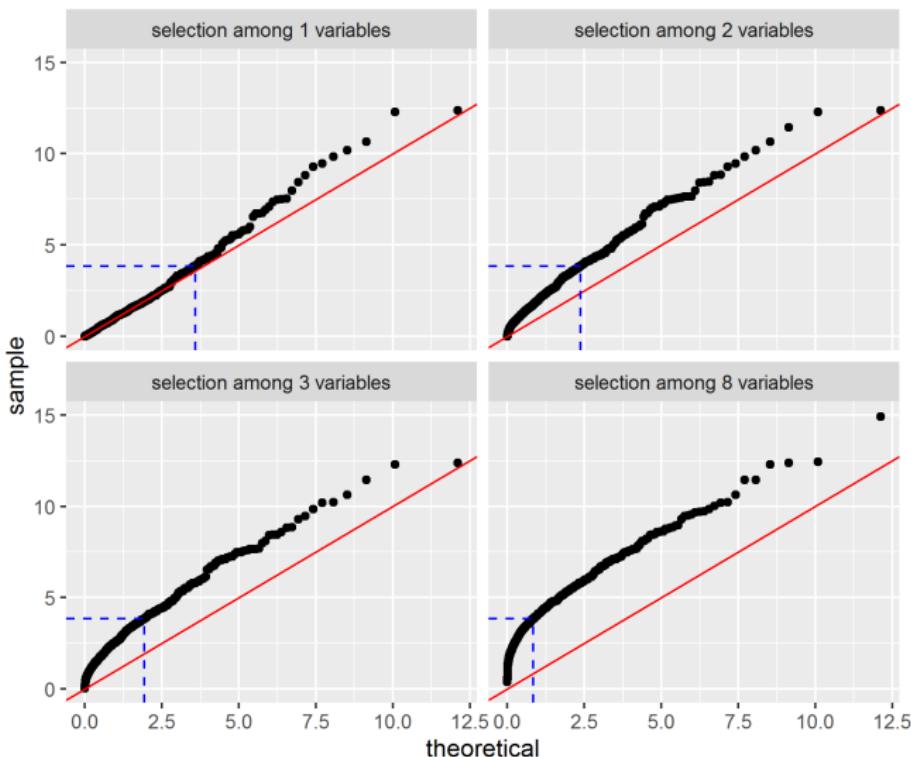
- $LRT(\mathcal{M}_0, \mathcal{M}_i) \sim \chi^2(1)$ mais $LRT(\mathcal{M}_0, \mathcal{M}_j)$?

Simulation (risque α)



avec 8 variables $\alpha = 0.3595$

Simulation (risque α)



avec 8 variables $\alpha = 0.3595$

Pénalisation lasso - point de vue bayésien

$$\underbrace{\mathbb{P}[\beta, \sigma^2 | Y, X]}_{posterior} \propto \underbrace{\mathbb{P}[Y, X | \beta, \sigma^2]}_{vraisemblance} \underbrace{\mathbb{P}[\beta, \sigma^2]}_{prior}$$

$$\propto \mathbb{P}[Y, X | \beta, \sigma^2] \mathbb{P}[\beta | \sigma^2] \mathbb{P}[\sigma^2]$$

On utilise une distribution de Laplace comme prior pour les paramètres:

$$\mathbb{P}(\beta | \sigma^2) = \prod_{j=1}^p \frac{1}{2\lambda} \exp^{-\frac{|\beta_j|}{\lambda}}$$

→ poids important en 0 pour l'effet des variables régularisées

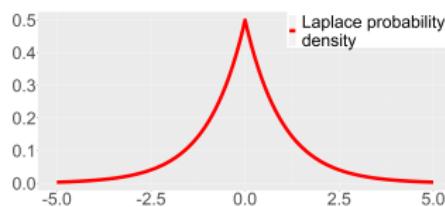
Pénalisation lasso - point de vue bayésien

$$\underbrace{\mathbb{P}[\beta, \sigma^2 | Y, X]}_{posterior} \propto \underbrace{\mathbb{P}[Y, X | \beta, \sigma^2]}_{vraisemblance} \underbrace{\mathbb{P}[\beta, \sigma^2]}_{prior}$$

$$\propto \mathbb{P}[Y, X | \beta, \sigma^2] \mathbb{P}[\beta | \sigma^2] \mathbb{P}[\sigma^2]$$

On utilise une distribution de Laplace comme prior pour les paramètres:

$$\mathbb{P}(\beta | \sigma^2) = \prod_{j=1}^p \frac{1}{2\lambda} \exp^{-\frac{|\beta_j|}{\lambda}}$$



→ poids important en 0 pour l'effet des variables régularisées

Pénalisation lasso - point de vue bayésien

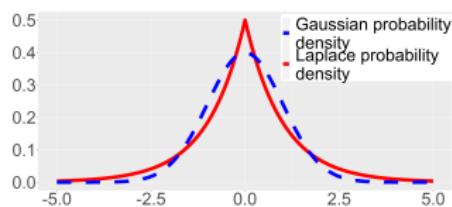
$$\underbrace{\mathbb{P}[\beta, \sigma^2 | Y, X]}_{posterior} \propto \underbrace{\mathbb{P}[Y, X | \beta, \sigma^2]}_{vraisemblance} \underbrace{\mathbb{P}[\beta, \sigma^2]}_{prior}$$

$$\propto \mathbb{P}[Y, X | \beta, \sigma^2] \mathbb{P}[\beta | \sigma^2] \mathbb{P}[\sigma^2]$$

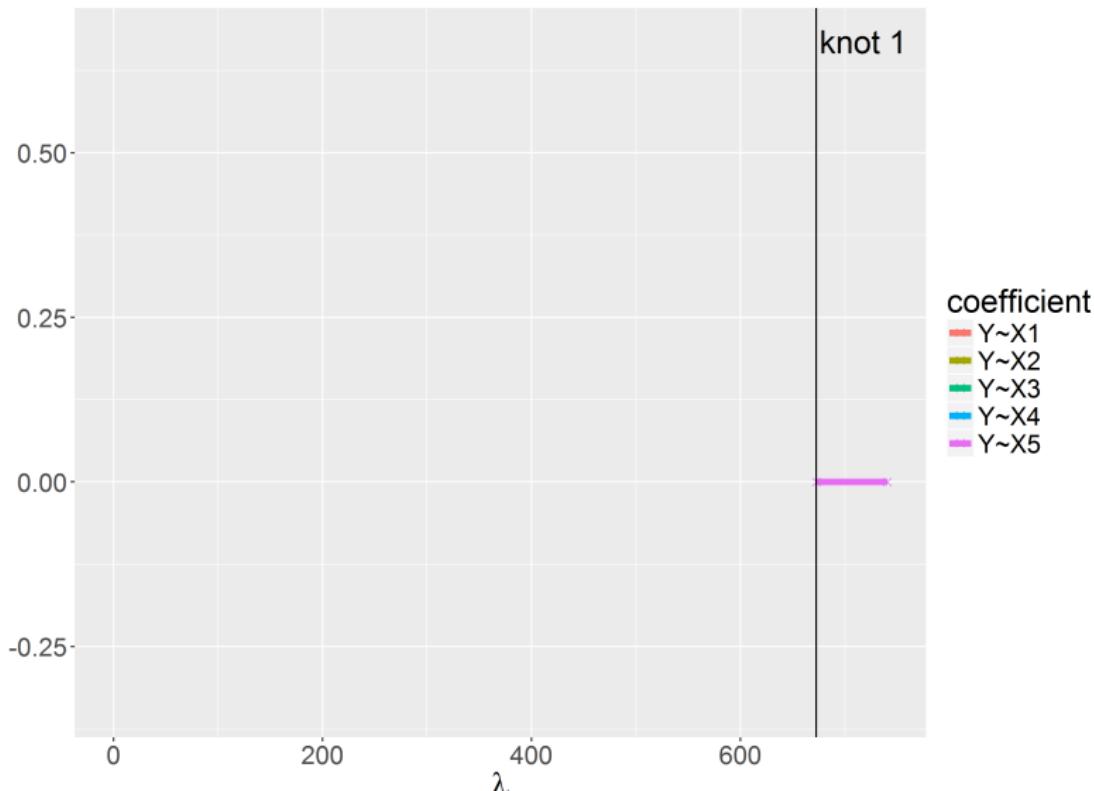
On utilise une distribution de Laplace comme prior pour les paramètres:

$$\mathbb{P}(\beta | \sigma^2) = \prod_{j=1}^p \frac{1}{2\lambda} \exp^{-\frac{|\beta_j|}{\lambda}}$$

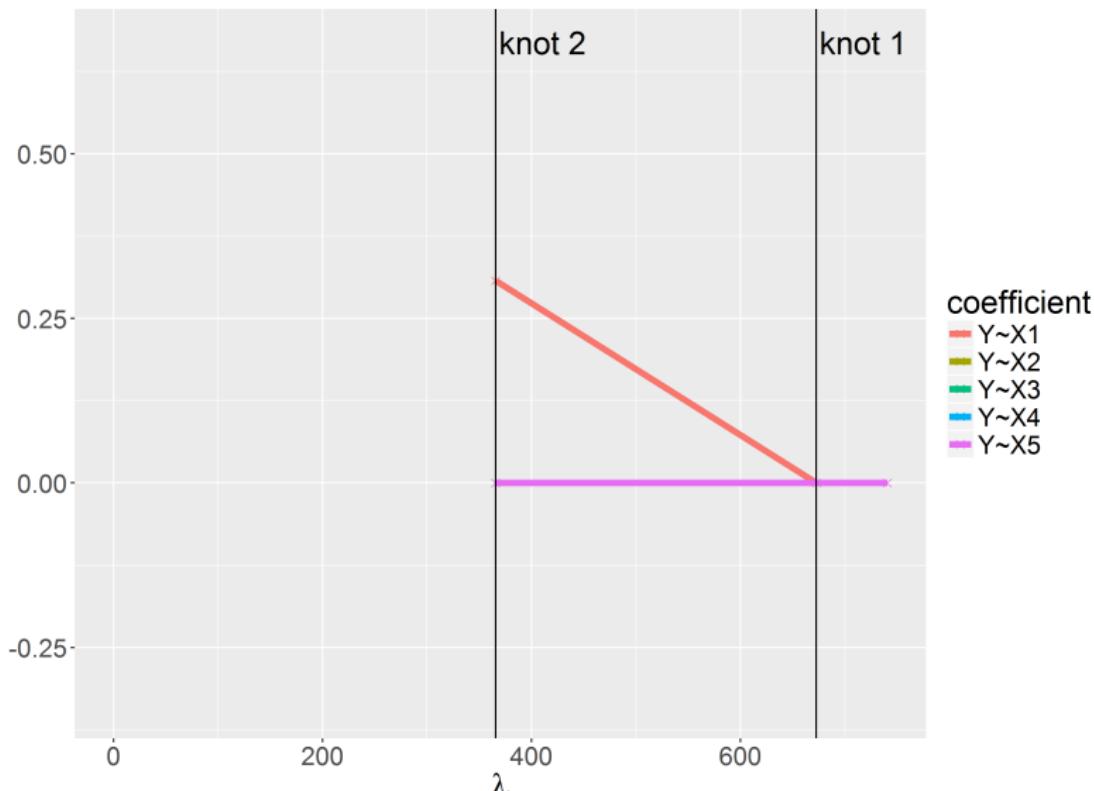
→ poids important en 0 pour l'effet des variables régularisées



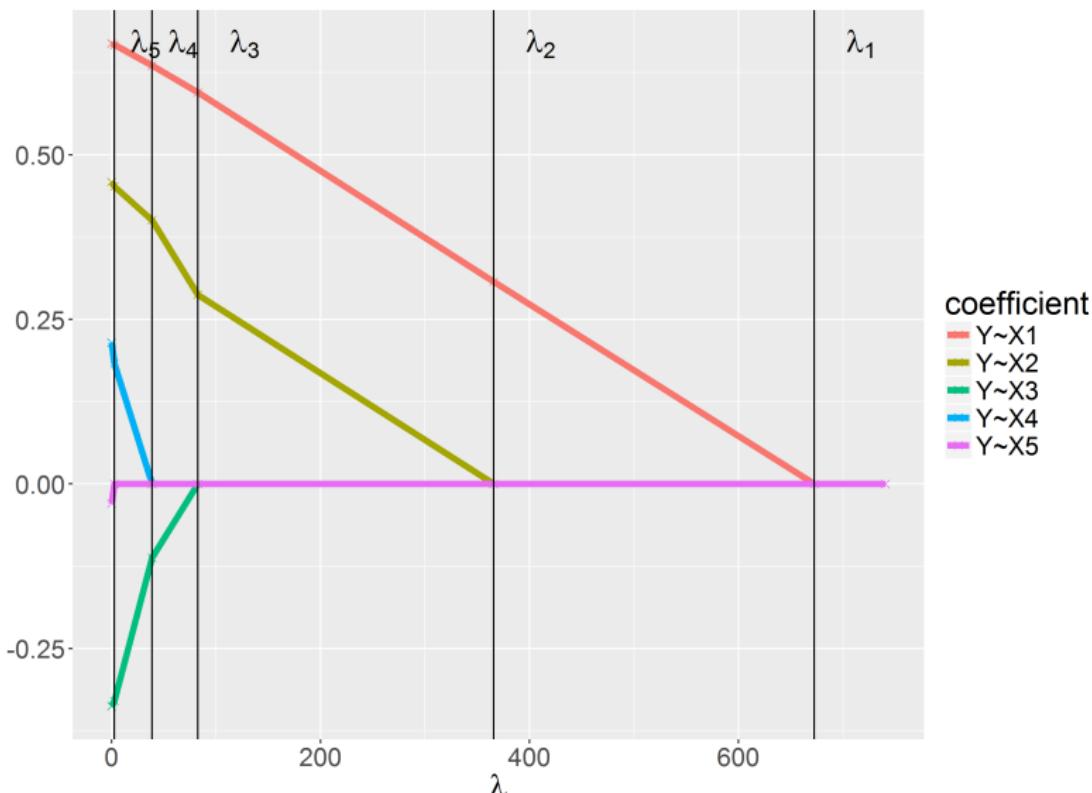
Cas linéaire - observations iid



Cas linéaire - observations iid



Cas linéaire - observations iid



Cas linéaire - covariance test (Lockhart et al. 2014)

$\mathcal{M}(\lambda)$ une régression lasso avec p variables pénalisées:

→ p nœuds où l'on va inclure une nouvelle variable dans le modèle

Au nœud k ,

la k -ième variable confondante est inclue dans le modèle

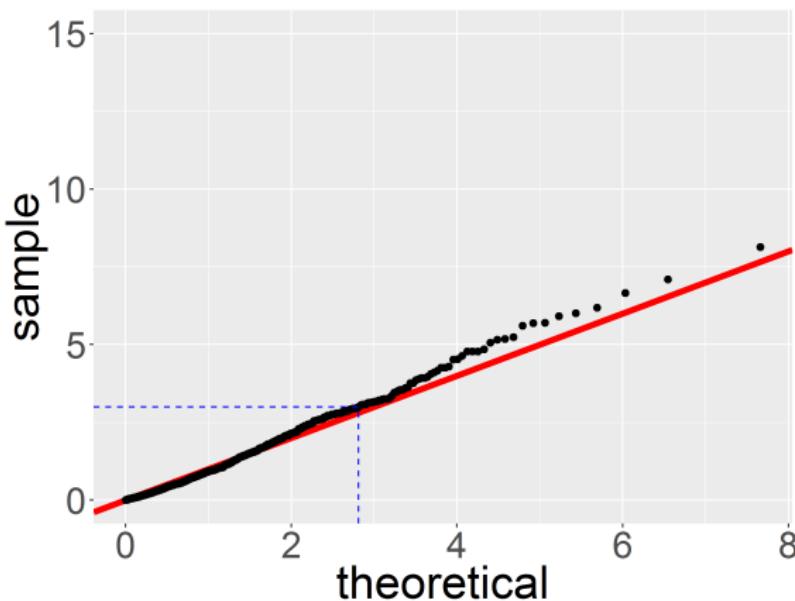
Test: devrait-on continuer à inclure d'autres variables ?

\mathcal{H}_0 : non toutes les variables confondantes sont déjà dans $\mathcal{M}(\lambda_{k-1})$

\mathcal{H}_1 : oui la nouvelle variable à inclure à λ_k est pertinente

Retour à la simulation

Lasso sélectionnant la première variable confondante parmi 8



Généralisation aux LVM

LVM:

- ✓ Estimation d'un LVM avec regularization lasso/ridge
(lava.penalty)

Estimation de la solution pour l'ensemble des λ

- entre deux nœuds, β évolue non-linéairement
- équation différentielle à résoudre
- travail en cours ...

Conclusion

Les modèles à variables latentes sont un outils de modélisation flexible

- ✓ vue d'ensemble des relations entre les variables
- ✓ gère plusieurs variables de réponses
- ✓ repose sur un maximum de vraisemblance
- ✗ interactions possibles mais non naturelles

lava est une librairie fantastique

- facile d'utilisation
- intègre écart-types robustes
- gestion des données manquante (Y)
- progrès possibles pour la sélection de variables

Tak for jeres opmærksomhed
merci pour votre attention

et merci à mes collègues:

- Esben Budtz-Jørgensen et Klaus Kähler Holst (LVM)
- Thomas Alexander Gerdts (Analyse de survie)
- Gitte Moos Knudsen et Patrick Fisher (NRU)

Questions ?

Bibliography I



Beliveau, V. et al. (2016). 'A high-resolution in vivo atlas of the human brain's serotonin system'. In: *Journal of Neuroscience* In press.



Dam, H. et al. (1998). 'Prevalence of winter depression in Denmark.' In: *Acta psychiatica Scandinavica* 97.1, pp. 1–4. ISSN: 0001-690X. URL: <http://www.ncbi.nlm.nih.gov/pubmed/9504695>.



Fisher, P. M. et al. 'BDNF val66met association with serotonin transporter binding in healthy humans'. In: *Under review in Translational Psychiatry*.



Holst, K. K. and E. Budtz-Jørgensen (2013). *Linear latent variable models: The lava-package*. Vol. 28. 4, pp. 1385–1452. ISBN: 0018001203.



Larsen, C. B. et al. 'Amygdala response to emotional faces in seasonal affective disorder'. In: *To be submitted to Journal of Affective Disorders*.



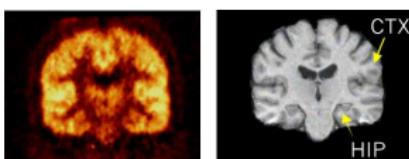
Lockhart, R. et al. (2014). *A significance test for the lasso*.

Pipeline from PET scan to BP

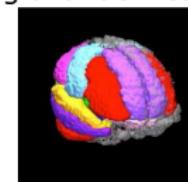
PET scan



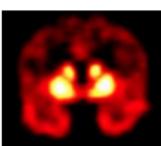
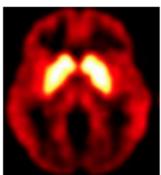
Co-register PET with MRI



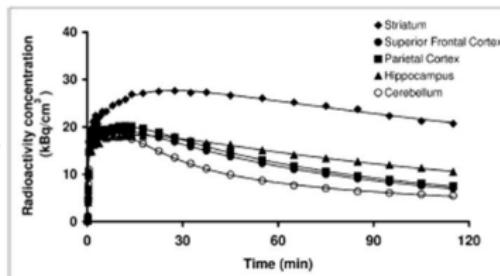
Regional delineation



Binding potential



Time-activity curve



Extensions du modèle linéaire - GLS

Moindres carrés généralisés (GLS):

- Spécification directe de Σ (par bloc)

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = diag(\Sigma_{id}) = \begin{bmatrix} \Sigma_{id} & 0 & 0 & \dots \\ 0 & \Sigma_{id} & 0 & \dots \\ 0 & 0 & \Sigma_{id} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

$$\Sigma_{id} = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho & \dots \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \dots \\ \sigma^2\rho & \sigma^2\rho & \sigma^2 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

Compound symmetry : trop simpliste

Extensions du modèle linéaire - GLS

Moindres carrés généralisés (GLS):

- Spécification directe de Σ (par bloc)

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = diag(\Sigma_{id}) = \begin{bmatrix} \Sigma_{id} & 0 & 0 & \dots \\ 0 & \Sigma_{id} & 0 & \dots \\ 0 & 0 & \Sigma_{id} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

$$\Sigma_{id} = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho & \dots \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho^2 & \dots \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

Auto-regressive(1) : pas d'ordre naturel entre les régions

Extensions du modèle linéaire - GLS

Moindres carrés généralisés (GLS):

- Spécification directe de Σ (par bloc)

$$Y = \mathbf{X}\beta + \varepsilon \text{ où } \varepsilon \sim \mathcal{N}(0, \Sigma)$$

$$\Sigma = diag(\Sigma_{id}) = \begin{bmatrix} \Sigma_{id} & 0 & 0 & \dots \\ 0 & \Sigma_{id} & 0 & \dots \\ 0 & 0 & \Sigma_{id} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

$$\Sigma_{id} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \dots \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \dots \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

Unstructured : trop complexe, peu interprétable

Extensions du modèle linéaire - LMM

Modèle linéaire à effet mixte (LMM):

- effet aléatoire: modélisation de la corrélation iter-observations

$$Y_{id} = \mathbf{X}_{id}\beta + Z_{id}u_{id} + \varepsilon_{id}$$

avec $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ et $u_{id} \sim \mathcal{N}(0, \tau)$

$$\mathbb{V}ar [Y_{id}] = \begin{bmatrix} \sigma^2 + \tau & \tau & \tau & \dots \\ \tau & \sigma^2 + \tau & \tau & \dots \\ \tau & \tau & \sigma^2 + \tau & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

intercepte aléatoire : trop simpliste

Extensions du modèle linéaire - LMM

Modèle linéaire à effet mixte (LMM):

- effet aléatoire: modélisation de la corrélation iter-observations

$$Y_{id} = \mathbf{X}_{id}\beta + Z_{id}u_{id} + \varepsilon_{id}$$

avec $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ et $u_{id} \sim \mathcal{N}\left(0, \begin{bmatrix} \tau_{1,1} & \tau_{1,2} \\ \tau_{1,2} & \tau_{2,2} \end{bmatrix}\right)$

$$\text{Cov}[Y_{it_1}, Y_{it_2}] = \tau_1 + \tau_{1,2}(t_1 + t_2) + \tau_{2,2}t_1t_2 + \mathbb{1}_{t_1=t_2}\sigma^2$$

intercepte et pente aléatoire : pas d'ordre naturel entre les régions