Bayesian nonparametric methods for clustering

Part 2: Outcome-guided analysis

Anaïs Rouanet - Postdoctoral fellow

INSERM U1219, Bordeaux Population Health research Center, Bordeaux, France

CNC 2021 - April 2021, 7th

Summary of Part 1:

Nonparametric clustering

- Heterogeneous population with unknown clusters
- Cluster number estimated from the data
- Unconstrained number of clusters

Characterized by

- One or several variables (Gaussian, skew t-distributions...)

Bayesian framework

- Dirichlet process prior
- MCMC estimation

Introduction of Part 2:

Nonparametric clustering

- Heterogeneous population with unknown clusters
- Cluster number estimated from the data
- Unconstrained number of clusters

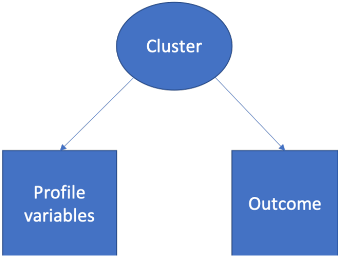
Characterized by

- One or several variables (Gaussian, skew t-distributions...)
- An outcome used as noisy surrogate for clusters (Bair, 2013)

Bayesian framework

- Dirichlet process prior
- MCMC estimation

Outcome-guided clustering



→ Subgroups with specific outcome distributions and variable profiles

→ Independence between profile variables and outcome conditionally on the clusters

Motivational example - Dementia study

Dementia

- Syndrome (chronic or progressive nature)
- Deterioration in cognitive functions
- Impairment of the social and occupational functioning.

Precision medicine

- Improve disease prevention and treatment
- Refine patient characterisation
- Predict disease risk and treatment effect at the individual level

Profile regression: methodology

Notations:

Y_i vector of individual outcome measurements
$W_i = (w_{i1}, ..., w_{iP})^\top$ individual vector of profile variables
c_i allocation variable

Likelihood:

$$L(\theta|Y,W) = \prod_{i=1}^{N} \sum_{g}^{\infty} P(c_i = g|\theta^c)\ f(W_i|c_i = g; \theta_g^W)\ f(Y_i|c_i = g; \theta_g^Y)$$

Independence assumption between W and Y given the cluster allocations.

Profile regression: membership probability

Notations:

Y_i vector of individual outcome measurements
$W_i = (w_{i1}, ..., w_{iP})^\top$ individual vector of profile variables
c_i allocation variable

Likelihood:

$$L(\theta|Y,W) = \prod_{i=1}^{N} \sum_{g}^{\infty} \underbrace{P(c_i = g|\theta^c)}_{\text{Membership prob.}} \; f(W_i|c_i = g; \theta_g^W) \; f(Y_i|c_i = g; \theta_g^Y)$$

Independence assumption between W and Y given the cluster allocations.

Sub-model: Dirichlet Process mixture model

$\mathcal{A} = \sum_{g=1}^{G} \pi_g \; \delta_{\theta_g}$

α : Concentration parameter
$\mathcal{A}_0$: Base distribution

$\boldsymbol{\pi}|\alpha \sim Dirichlet(\alpha/G, ..., \alpha/G)$
$c_i|\boldsymbol{\pi} \sim Multinom(\boldsymbol{\pi})$
$\theta_{c_i}|\mathcal{A}_0 \sim \mathcal{A}_0$
$Y_i, W_i|c_i \sim f(\cdot; \theta_{c_i})$

Profile regression: membership probability

Notations:

Y_i vector of individual outcome measurements
$W_i = (w_{i1}, ..., w_{iP})^\top$ individual vector of profile variables
c_i allocation variable

Likelihood:

$$L(\theta|Y, W) = \prod_{i=1}^{N} \sum_{g}^{\infty} f(W_i|c_i = g; \theta_g^W) \; f(Y_i|c_i = g; \theta_g^Y)$$

Independence assumption between W and Y given the cluster allocations.

Sub-model: Dirichlet Process mixture model

$\mathcal{A} = \sum_{g=1}^{+\infty} \pi_g \; \delta_{\theta_g}$

α : Concentration parameter
$\mathcal{A}_0$: Base distribution

$\boldsymbol{\pi}|\alpha \sim GEM(\alpha)$
$c_i|\boldsymbol{\pi} \sim Multinom(\boldsymbol{\pi})$
$\theta_{c_i}|\mathcal{A}_0 \sim \mathcal{A}_0$
$Y_i, W_i|c_i \sim f(\cdot; \theta_{c_i})$

Profile regression: profile variables

Notations:

Y_i vector of individual outcome measurements
$W_i = (w_{i1}, ..., w_{iP})^\top$ individual vector of profile variables
c_i allocation variable

Likelihood:

$$L(\theta|Y, W) = \prod_{i=1}^{N} \sum_{g}^{\infty} P(c_i = g|\theta^c) \underbrace{f(W_i|c_i = g; \theta_g^W)}_{\text{Profile variables}} f(Y_i|c_i = g; \theta_g^Y)$$

Independence assumption between W and Y given the cluster allocations.

Sub-model for profile variables: given $c_i = g$,

Discrete case: $W_i^{(1)} \sim Multinom(\phi_{\mathbf{g}})$
Continuous case: $W_i^{(2)} \sim \mathcal{N}_P(\mu_{\mathbf{g}}; V_g)$

Profile regression: outcome

Notations:

Y_i vector of individual outcome measurements
$W_i = (w_{i1}, ..., w_{iP})^\top$ individual vector of profile variables
c_i allocation variable

Likelihood:

$$L(\theta|Y, W) = \prod_{i=1}^{N} \sum_{g}^{\infty} P(c_i = g|\theta^c)\ f(W_i|c_i = g; \theta_g^W)\ \underbrace{f(Y_i|c_i = g; \theta_g^Y)}_{\text{Outcome}}$$

Sub-model for outcome: given cluster g,

$$Y_i \sim \mathcal{N}(\theta_g + \beta^\top X_i, \sigma_g^2)$$

Other types of outcomes:
Bernoulli, Poisson, Binomial, Categorical, Quantile, Survival

Estimated cluster-specific parameters

- MCMC Gibbs sampler: $\boldsymbol{\Theta} = (\alpha, [\phi]_g, [\mu]_g, [V]_g, c, \beta)$

- Posterior similarity matrix S from pairwise co-clustering probability

- Selection of the representative clustering P^*

  Partitioning around medoids on dissimilarity matrix $1 - S$
  Average silhouette width

Estimated cluster-specific parameters

- MCMC Gibbs sampler: $\mathbf{\Theta} = (\alpha, [\phi]_g, [\mu]_g, [V]_g, c, \beta)$

$$\alpha \sim f(\alpha \mid [\phi]_g, [\mu]_g, [V]_g, c, \beta)$$

- Posterior similarity matrix S from pairwise co-clustering probability

- Selection of the representative clustering P^*

  Partitioning around medoids on dissimilarity matrix $1 - S$
  Average silhouette width

Estimated cluster-specific parameters

- MCMC Gibbs sampler: $\boldsymbol{\Theta} = (\alpha, [\phi]_g, [\mu]_g, [V]_g, c, \beta)$

$$\alpha \sim f(\alpha \mid [\phi]_g, [\mu]_g, [V]_g, c, \beta)$$

$$[\phi]_g \sim f([\phi]_g \mid \alpha, [\mu]_g, [V]_g, c, \beta) \text{ etc...}$$

- Posterior similarity matrix S from pairwise co-clustering probability

- Selection of the representative clustering P^*

  Partitioning around medoids on dissimilarity matrix $1 - S$
  Average silhouette width

Estimation

Cluster-specific parameters as the average over the MCMC iterations:

$$\bar{\Theta}_g^{(k)} = \frac{1}{N_g} \sum_{i | c_i^* = g} \Theta_{c_i^{(k)}}$$

$\Theta_{c_i^{(k)}}$ parameter for the cluster subject i is allocated to, at iteration k
N_g the number of subjects in cluster g of P^*.

Estimation

Cluster-specific parameters as the average over the MCMC iterations:

$$\hat{\Theta}_g = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N_g} \sum_{i|c_i^*=g} \Theta_{c_i^{(k)}}$$

$\Theta_{c_i^{(k)}}$ parameter for the cluster subject i is allocated to, at iteration k
N_g the number of subjects in cluster g of P^*.

Estimation

Cluster-specific parameters as the average over the MCMC iterations:

$$\hat{\Theta}_g = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N_g} \sum_{i|c_i^*=g} \Theta_{c_i^{(k)}}$$

$\Theta_{c_i^{(k)}}$ parameter for the cluster subject i is allocated to, at iteration k
N_g the number of subjects in cluster g of P^*.

$\rightarrow$ Integration of the clustering uncertainty in the parameters estimation.
$\rightarrow$ Empirical credible intervals computed from MC sample of $\bar{\Theta}_g^{(k)}$

Convergence

Difficult to assess convergence based on parameters traces:

- Fixed effects converge quickly
- Cluster-specific parameters not traceable
- If stuck in local mode, α not good indicator (Hastie et al., 2014)

$\rightarrow$ Marginal partition posterior $p(C|Y, W$

- identify runs that were significantly different from others, perhaps due to convergence issues
- which run explored the higher posterior probability regions

Application to ADNI data - (Mueller et al., 2005)

Objective : Identify subgroups of the population associated with specific cognitive evolution patterns and brain imaging profiles.

- North American Alzheimer's Disease Neuroimaging Initiative
- Sample of 199 subjects, 55 years old and over, 8-year follow-up
- Normalised Mini-Mental State Examination (Philipps et al., 2014)

Application to ADNI data - (Mueller et al., 2005)

Objective : Identify subgroups of the population associated with specific cognitive evolution patterns and brain imaging profiles.

- North American Alzheimer's Disease Neuroimaging Initiative
- Sample of 199 subjects, 55 years old and over, 8-year follow-up
- Normalised Mini-Mental State Examination (Philipps et al., 2014)
- Profile variables:
  - → Standardised MRI brain volumes: whole brain, ventricles, hippocampus, entorhinal cortex, fusiform gyrus and middle temporal gyrus, standardised by the intracerebral volume
  - → Gender (1 for women, 0 for men) and Education (1 if greater than or equal to 16 years of education, 0 otherwise), APOE4 status (1 if 1 or 2 APOE4 alleles, 0 otherwise)

Model specification

Profile variables, given cluster g:

Gender, Sex, Education:
$W_i^{(1)} \sim \mathit{Bernouilli}(\phi_g)$ with $\phi_g \sim \mathit{Dirichlet}(a_j)$

Standardized MRI brain volumes:
$W_i^{(2)} \sim \mathcal{N}_P(\mu_g; V_g)$ with $\mu_g \sim \mathcal{N}_P(\mu_0; V_0)$, $V_g \sim \mathit{InvWishart}(R_0, K_0)$

Model specification

Profile variables, given cluster g:

Gender, Sex, Education:

$W_i^{(1)} \sim Bernouilli(\phi_g)$ with $\phi_g \sim Dirichlet(a_j)$

Standardized MRI brain volumes:

$W_i^{(2)} \sim \mathcal{N}_P(\mu_g; V_g)$ with $\mu_g \sim \mathcal{N}_P(\mu_0; V_0)$, $V_g \sim InvWishart(R_0, K_0)$

Longitudinal normalized MMSE, given cluster g:

$Y(t_i) = f_g(t_i) + \epsilon_g(t_i)$ with $\epsilon_g \sim \mathcal{N}(0, L_{1,g})$

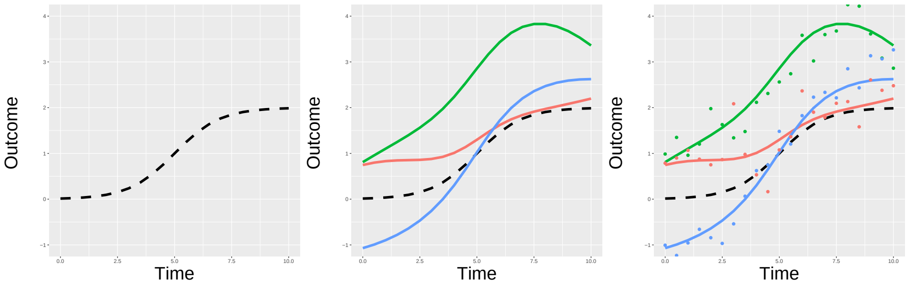
$f_g(\cdot) \sim GP(\mathbf{0}, \mathcal{K}_g(\cdot, \cdot))$ with $\mathcal{K}_g(s,t) = L_{g,2} \exp\left(-\frac{(s-t)^2}{L_{g,3}}\right)$

$$\exp(L_{g,l}) \sim \mathcal{N}(0, \sigma_l^2)$$

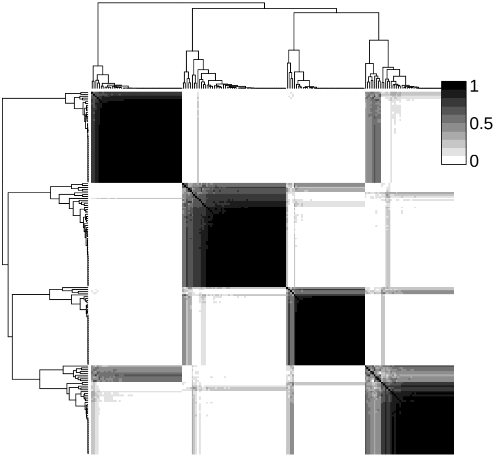
Gaussian Process prior

Given the cluster g :

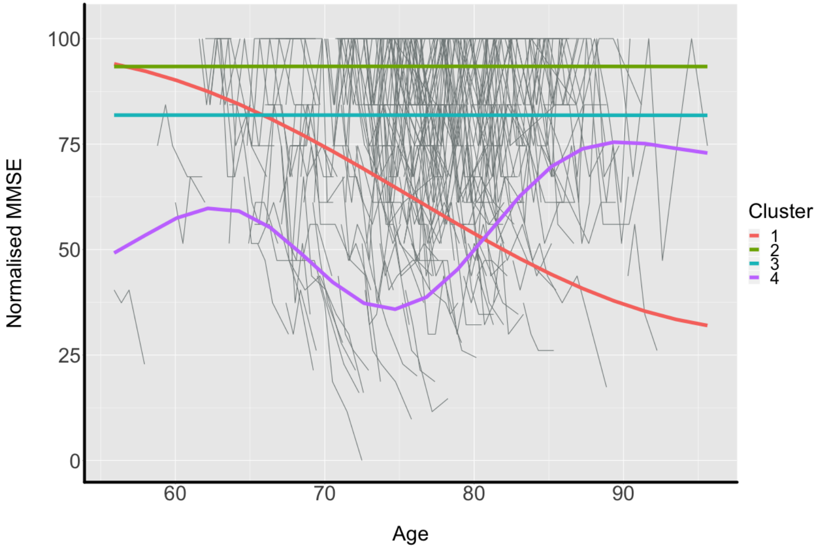
$$Y_{ij} = f_g(t_{ij}) + \epsilon_{ijg} \text{ with } \epsilon_{ijg} \sim \mathcal{N}(0, L_{g,1})$$
$$f_g(\cdot) \sim GP(m_g(\cdot),\ \mathcal{K}_g(\cdot,\cdot))$$



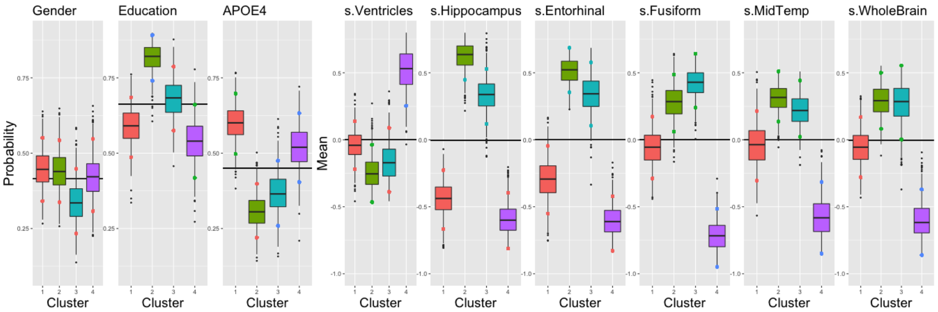
Posterior similarity matrix



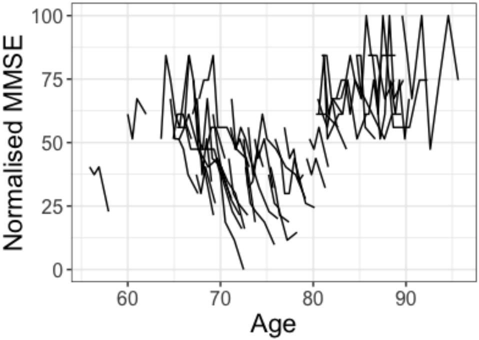
Cluster-specific cognitive patterns

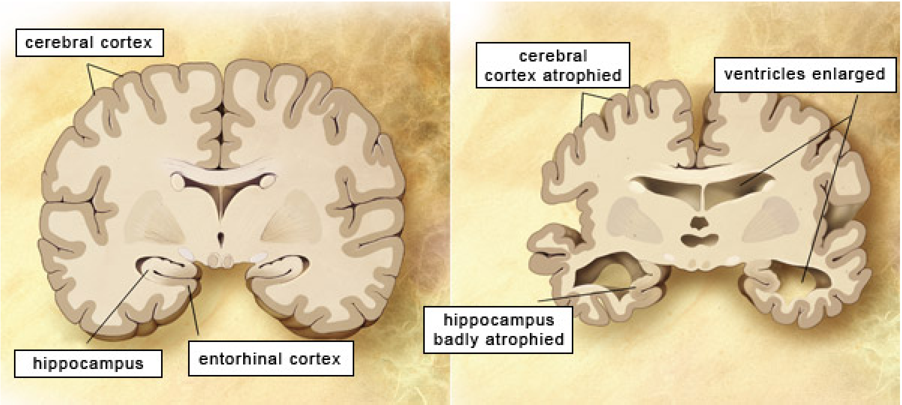


Cluster profiles



Cognitive pattern in cluster 4




cerebral cortex
hippocampus
entorhinal cortex
cerebral cortex atrophied
ventricles enlarged
hippocampus badly atrophied

Discussion

Conclusion

- Mixture model with unconstrained number of clusters
- Outcome-guided clustering analysis
- Integration of different types of profile variables
- Extension of the package **PReMiuM**

Practical

- Hands-on exercises on the PReMiuM package (profRegr function)
- Data on cognitive decline from lcmm package
- Details on DP parameters

References

Rouanet A, et al. (2020) Benefit of Bayesian clustering of longitudinal data: study of cognitive decline for precision medicine. In Bayesian Methods in Pharmaceutical Research (eds: Lesaffre E, Baio G, and Boulanger B), Chapman & Hall/CRC Biostatistics Series, ISBN: 978-1-138-74848-4 (pp. 223–242).

Rouanet A, et al. Bayesian profile regression with a longitudinal response.

Liverani S, et al. (2015). PReMiuM: An R Package for Profile Regression Mixture Models Using Dirichlet Processes. Journal of Statistical Software, 64(7),1-30.

Mueller SG, et al. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers and Dementia, 1(1):55-66.

Philipps, V, et al. (2014). Normalized mini-mental state examination for assessing cognitive change in population-based brain aging studies. Neuroepidemiology, 43(1): 15–25.

Bair, E. (2013). Semi-supervised clustering methods. Wiley Interdisciplinary Reviews: Computational Statistics, 5(5): 349–361.