

Multiple imputation for clustering on incomplete data

V. Audigier, N. Niang

CNAM, CEDRIC-MSDMA, Paris

29èmes Rencontres de la SFC, September 12th, 2024

Clustering

Data $\mathbf{X}_{n \times p} = (x_{ij})$ a matrix of numeric values

Each individual i belongs to a unique cluster $\mathcal{C}_i \in \{1, \dots, K\}$.

Aim identify \mathcal{C}_i for each i based on individual profiles $(\mathbf{x}_i)_{1 \leq i \leq n}$

Methods

Distance-based

- k-means
- fuzzy C-means
- hierarchical clustering
- partitioning around medoids

Model-based

- gaussian mixture models
- mixture of multivariate t -distributions
- mixture of skew normal distributions

Clustering with missing values

However, \mathbf{X} is frequently **incomplete**... $\mathbf{x}_i = (\mathbf{x}_i^{obs}, \mathbf{x}_i^{miss})$

Ad-hoc methods

- removing incomplete observations
- removing incomplete variables
- single imputation

Direct methods

- k-means (Chi et al., 2016; Honda et al., 2011; Wagstaff, 2004)
- fuzzy C-means (Zhang et al., 2016; Hathaway and Bezdek, 2001)
- gaussian mixture (Miao et al., 2016; Marbac et al., 2019; McCaw et al., 2022)

k-POD (Chi et al., 2016)

Chi et al. (2016) proposed a direct method for k-means clustering

k-means

$$\arg \min_{\mathbf{A} \in \mathcal{H}, \mathbf{B}} \|\mathbf{X} - \mathbf{A}\mathbf{B}\|_F^2$$

- \mathcal{H} set of membership matrices ($n \times K$), $\mathbf{B}_{K \times p}$ matrix of centers coordinates
- $\|\cdot\|_F$ Frobenius norm
- $\Omega \subset \{1, \dots, n\} \times \{1, \dots, p\} \rightarrow$ subset of the indices for observed entries
- P_Ω projection operator so that $[P_\Omega(\mathbf{X})]_{ij} = \begin{cases} x_{ij} & \text{if } (i, j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$

k-POD

$$\arg \min_{\mathbf{A} \in \mathcal{H}, \mathbf{B}} \|P_\Omega(\mathbf{X}) - P_\Omega(\mathbf{A}\mathbf{B})\|_F^2$$

The criterion is optimised by alternating imputation by b_k and kmeans clustering

Available in the [kpodclustr](#) R package

FCM by optimal completion strategy (Hathaway and Bezdek, 2001)

fuzzy c-means

$$\underset{\Gamma, \mathbf{B}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{k=1}^K \gamma_{ki}^{\alpha} \| \mathbf{x}_i - \mathbf{b}_k \|_2^2$$

fuzzy c-means OCS

$$\underset{\Gamma, \mathbf{B}, \mathbf{X}^{\text{miss}}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{k=1}^K \gamma_{ki}^{\alpha} \| (\mathbf{x}_i^{\text{obs}}, \mathbf{x}_i^{\text{miss}}) - \mathbf{b}_k \|_2^2$$

with $\Gamma_{K \times n} = (\gamma_{ki})$ degrees of membership; α fuzzification parameter

The criterion is optimised by alternating FCM and imputation by weighted centers

$$\gamma_{ki}^{\alpha} \leftarrow 1 / \sum_{\ell=1}^K \left(\frac{\| \mathbf{x}_i - \mathbf{b}_k \|_2^2}{\| \mathbf{x}_i - \mathbf{b}_{\ell} \|_2^2} \right)^{\frac{1}{\alpha-1}}$$

$$\mathbf{b}_{kj} \leftarrow \left(\sum_{i=1}^n \gamma_{ki}^{\alpha} \mathbf{x}_{ij} \right) / \left(\sum_{i=1}^n \gamma_{ki}^{\alpha} \right)$$

$$\mathbf{x}_{ij}^{\text{miss}} \leftarrow \left(\sum_{k=1}^K \gamma_{ki}^{\alpha} \mathbf{b}_{kj} \right) / \left(\sum_{k=1}^K \gamma_{ki}^{\alpha} \right)$$

Ignorable-GMM (Marbac et al., 2019)

Gaussian mixture models (GMM)

$$f(x; \theta) = \sum_{k=1}^K \pi_k f_k(x; \theta_k) \quad \theta = (\theta_k)_{1 \leq k \leq K}, \theta_k = (\mu_k, \Sigma_k)$$

Log-likelihood GMM

$$\sum_{i=1}^n \log \sum_{k=1}^K \pi_k \prod_{j=1}^p f_{kj}(x_{ij}; \theta_{kj})$$

Log-likelihood ignorable-GMM

$$\sum_{i=1}^n \log \sum_{k=1}^K \pi_k \prod_{j \in O_i} f_{kj}(x_{ij}; \theta_{kj})$$

$O_i \subseteq 1, \dots, p$ the subset of variables indices that are observed for individual i

The criterion is optimised by using an EM algorithm

Available in the [VarSelLCM](#) R package

Direct methods: pros and cons

Direct methods provide an elegant way to address missing values

However, the approach

- is not versatile
- has not been developed for all clustering methods

Multiple Imputation (MI)

- a popular method to address missing values
- could be potentially used for any clustering method

Direct methods: pros and cons

Direct methods provide an elegant way to address missing values

However, the approach

- is not versatile
- has not been developed for all clustering methods

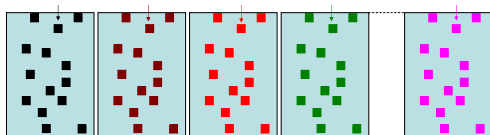
Multiple Imputation (MI)

- a popular method to address missing values
- could be potentially used for any clustering method

Multiple imputation (Rubin, 1987)

- 1 Generate a set of M parameters $(\zeta_m)_{1 \leq m \leq M}$ of an **imputation model** to generate M plausible imputed data sets

$$f(X^{miss} | X^{obs}, \zeta_1) \quad \dots \quad \dots \quad \dots \quad f(X^{miss} | X^{obs}, \zeta_M)$$



- 2 Fit the **analysis model** on each imputed data set: $\hat{\psi}_m, \widehat{\text{Var}}(\hat{\psi}_m)$
- 3 Combine the results using Rubin's rules

- 1 $\hat{\psi} = \frac{1}{M} \sum_{m=1}^M \hat{\psi}_m$

- 2 $T = \frac{1}{M} \sum_{m=1}^M \widehat{\text{Var}}(\hat{\psi}_m) + \frac{1}{M-1} \sum_{m=1}^M (\hat{\psi}_m - \hat{\psi})^2$

Challenges in clustering

MI is not tailored for cluster analysis

- How to impute the incomplete data set?
- How to assess a “variability” accounting for missing values?
- How to “average” partitions?

Some works on the “averaging” step

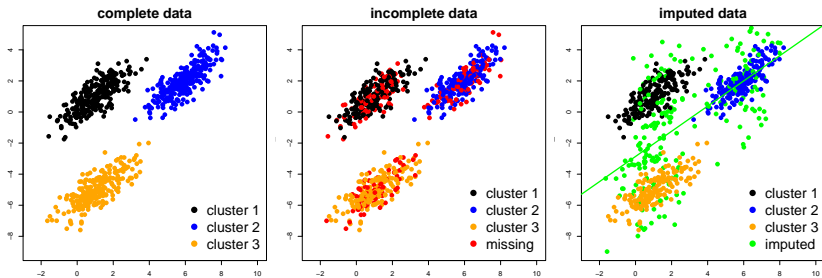
- by stacking (Plaehn, 2019)
- by using consensus clustering methods (Faucheux et al., 2020; Bruckers et al., 2017; Basagana et al., 2013; Aschenbruck et al., 2022)

Aim: highlighting how imputation, analysis and pooling steps should be carried out

Outline

- ① Introduction
- ② MI for clustering
 - Imputation
 - Analysis
 - Pooling
- ③ Simulation study
 - Simulated data
 - Real data
- ④ Conclusion

Imputation model for clustering: the issue



JM-DP (Kim et al., 2014)

Joint Modeling based on Dirichlet Process mixture of products of multivariate normal distributions

Based on a Bayesian formulation

$$\mu_k | \Sigma_k \sim \mathcal{N}(\mu_0, h^{-1} \Sigma_k) \quad \Sigma_k \sim \mathcal{W}^{-1}(df, G)$$

$$\text{with } \text{diag}(G) = (g_1, \dots, g_p) \quad g_j \sim \mathcal{G}(a_0, b_0)$$

$$\pi_k = v_k \prod_{\ell < k} (1 - v_\ell)$$

$$\text{with } \begin{cases} v_k \sim \text{Beta}(1, \alpha) \text{ and } \alpha \sim \mathcal{G}(a_\alpha, b_\alpha) \text{ for } k < K \\ v_K = 1 \end{cases}$$

- parameters: $\zeta = (\pi, \mu, \Sigma)$
- hyperparameters: $h, \mu_0, df, a_0, b_0, a_\alpha, b_\alpha$

$(\zeta_m)_{1 \leq m \leq M}$ is generated using a Data-Augmentation algorithm

Properties

JM-DP

- accounts for the heterogeneity
- accounts for the heteroscedasticity
- the number of clusters is only bounded

Modeling assumptions

- based on the normality assumption

Available in the **DPImputeCont** (Kim, 2020) and **clusterMI** (Audigier, 2024) R packages

Fully conditional specification

Instead of specifying one joint distribution $f(X; \zeta)$, a conditional distribution is specified for each (incomplete) variable $f(X_j | X_{-j}; \zeta_j)$

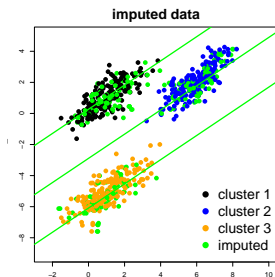
$$\text{Ex : } f(X_j | X_{-j}; \zeta_j) = \mathcal{N}(X_{-j}\beta, \sigma^2) \quad \zeta_j = (\beta, \sigma)$$

To impute the m th data set

- initialize $\mathbf{x}_i^{\text{miss}}$ for all i
- for j in $1 \dots p$
 - a generate ζ_j based on observed individuals on X_j
 - b impute X_j^{miss} according to $f(X_j | X_{-j}; \zeta_j)$
- repeat until convergence

FCS-homo (Audigier et al., 2021)

Addressing the issue by using regression models including the class variable W as explanatory variable



FCS-homo

- generating X_j^{miss} given W is performed using regression models including a intercept specific to each cluster

$$f(X_j|X_{-j}, W; \zeta_j) = \mathcal{N}(X_{-j}\beta + \mu_w, \sigma^2) \quad \zeta_j = (\beta, \sigma, \mu_w)$$

- generating W given X by linear discriminant analysis

$$\mathbb{P}(W = w|X; \zeta_w) \propto \exp(\delta_{w,x}) \quad \zeta_w = (\pi, \mu, \Sigma)$$

Properties

FCS-homo

- addresses the cluster structure
- assumes homoscedastic regression models
- required a pre-defined number of clusters

Can be easily modified

- to account for heteroscedasticity (van Buuren, 2011)
- to improve sparsity (Zahid and Heumann, 2019)
- to address outliers (Templ et al., 2011)
- to use semi-parametric models (Morris et al., 2014)
- ...

Available in the R package [clusterMI](#) (Audigier, 2024)

Analysis

Aim: From each imputed data set \mathbf{X}_m , compute a partition Ψ_m as well as an associated instability measure V_m

- Ψ_m can be obtained by any clustering algorithm
- V_m assesses how the Voronoi tessellation varies with the train sample for any clustering algorithm

Following Fang and Wang (2012)

- generate C bootstrap pairs $(\mathbf{X}_c, \tilde{\mathbf{X}}_c)_{1 \leq c \leq C}$ from \mathbf{X}
- perform cluster analysis from $(\mathbf{X}_c, \tilde{\mathbf{X}}_c)_{1 \leq c \leq C}$ to obtain $(\Psi_c, \tilde{\Psi}_c)$
- classify individuals of \mathbf{X} from Ψ_c and $\tilde{\Psi}_c$ to obtain $(\Psi'_c, \tilde{\Psi}'_c)$
- assess the instability $V = \frac{1}{C} \sum_{c=1}^C \frac{\delta(\Psi'_c, \tilde{\Psi}'_c)}{n^2}$

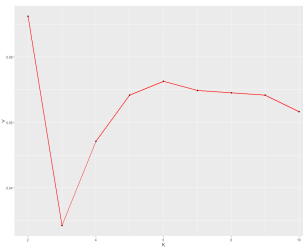


Figure: Instability (V) according to the number of clusters (K)

Partitions pooling

Ψ_m the partition from (X^{obs}, X_m^{miss}) , which average $\hat{\Psi}$ for $(\Psi_m)_{1 \leq m \leq M}$?

With **complete data** Jain (2017) extended to partitions

- the expected mean

$$\arg \min_{\Psi \in \mathcal{P}_{n,K}} \int_{\mathcal{P}_{n,K}} \delta^\alpha(\Psi^*, \Psi) d\pi(\Psi^*)$$

- the mean estimate

$$\arg \min_{\Psi \in \mathcal{P}_{n,K}} \sum_{j=1}^J \delta^\alpha(\Psi, \Psi_j) \quad (1)$$

with δ a dissimilarity, $\alpha \in \mathbb{N}$, $\mathcal{P}_{n,K}$ the set of partitions of n observations in K clusters

with $(\Psi_j)_{1 \leq j \leq J}$ a set of observed partitions

After MI

$$\hat{\Psi} = \arg \min_{\Psi \in \mathcal{P}_{n,K}} \sum_{m=1}^M \delta^\alpha(\Psi, \Psi_m) \quad (\text{median partition problem})$$

Properties

- Theoretically appealing, but solving (1) is highly challenging
- Iterative algorithms are required (Vega-Pons and Ruiz-Shulcloper, 2011)

Partitions pooling

Ψ_m the partition from (X^{obs}, X_m^{miss}) , which average $\widehat{\Psi}$ for $(\Psi_m)_{1 \leq m \leq M}$?

With **complete data** Jain (2017) extended to partitions

- the expected mean

$$\arg \min_{\Psi \in \mathcal{P}_{n,K}} \int_{\mathcal{P}_{n,K}} \delta^\alpha(\Psi^*, \Psi) d\pi(\Psi^*)$$

- the mean estimate

$$\arg \min_{\Psi \in \mathcal{P}_{n,K}} \sum_{j=1} \delta^\alpha(\Psi, \Psi_j) \quad (1)$$

with δ a dissimilarity, $\alpha \in \mathbb{N}$, $\mathcal{P}_{n,K}$ the set of partitions of n observations in K clusters

with $(\Psi_j)_{1 \leq j \leq J}$ a set of observed partitions

After MI

$$\widehat{\Psi} = \arg \min_{\Psi \in \mathcal{P}_{n,K}} \sum_{m=1}^M \delta^\alpha(\Psi, \Psi_m) \quad (\text{median partition problem})$$

Properties

- Theoretically appealing, but solving (1) is highly challenging
- Iterative algorithms are required (Vega-Pons and Ruiz-Shulcloper, 2011)

Mirkin-based methods

δ chosen as the number of disagreements between partitions

$$\delta(\Psi, \Psi') = \sum_{(i,i')} \delta_{ii'} \quad \delta_{ii'} = \begin{cases} 1 & \text{if } i \text{ and } i' \text{ belong to the same cluster} \\ & \text{in one partition and not in the other} \\ 0 & \text{otherwise} \end{cases}$$

Two methods can be exhibited

- ① BOK: the space of solutions is constrained to $(\Psi_m)_{1 \leq m \leq M}$ instead of $\mathcal{P}_{n,K}$
- ② SAOM: the BOK solution is improved by using stochastic relabeling of individuals

Properties

- The error for the BOK solution does not exceed two times the error of the optimal partition (Filkov and Skiena, 2004)

$$\sum_{m=1}^M \delta(\Psi_{BOK}, \Psi_m) \leq 2 \sum_{m=1}^M \delta(\Psi_{opt}, \Psi_m)$$

- SAOM provides a better solution, but computationally intensive

NMF-based methods

Non negative matrix factorization is powerful method widely used for solving many optimization problems

Principle

- consider the Mirkin distance for δ
- rewrite the optimization problem in terms of connectivity matrices $(\mathbf{U}_m)_{1 \leq m \leq M}$ instead of partitions $(\Psi_m)_{1 \leq m \leq M}$

$$\operatorname{argmin}_{\Psi \in \mathcal{P}_n^K} \sum_{m=1}^M \delta(\Psi, \Psi_m) \iff \operatorname{argmin}_{\mathbf{U} \in \mathcal{U}} \|\bar{\mathbf{U}} - \mathbf{U}\|_F^2$$

$$\text{with } \bar{\mathbf{U}} = \frac{1}{M} \sum_{m=1}^M \mathbf{U}_m$$

Properties

- can be solved using various algorithms (Lee and Seung, 2001; Li et al., 2007)
- monotone convergence
- no label switching problem, various choices for K_m are available

Instability after MI (Audigier and Niang, 2022)

Following Fang and Wang (2012), the **within instability** can be assessed by

$$\frac{1}{M} \sum_{m=1}^M V_m$$

while the **between instability** can be computed by averaging the proportions of disagreements

$$\frac{1}{M^2} \sum_{m=1}^M \sum_{m'=1}^M \delta(\Psi_m, \Psi_{m'}) / n^2$$

the **total instability** T is

$$T = \frac{1}{M} \sum_{m=1}^M V_m + \frac{1}{M^2} \sum_{m=1}^M \sum_{m'=1}^M \delta(\Psi_m, \Psi_{m'}) / n^2$$

Properties

Based on a simulation study

- pooling partitions using NMF-based methods is less time consuming and more accurate than Mirkin-based methods
- a larger value for M improves the partition accuracy ($M \approx 20$)
- T provides an accurate estimate for the number of clusters with missing values

Outline

- ① Introduction
- ② MI for clustering
 - Imputation
 - Analysis
 - Pooling
- ③ Simulation study
 - Simulated data
 - Real data
- ④ Conclusion

Simulation design: data generation

Complete data generation

- A base-case configuration: GMM with $K = 3$ components

$$\begin{aligned} \mu_1 &= (0, 0, 0, 0, \Delta, \Delta, 0, \Delta^2) \\ \mu_2 &= (0, 0, 0, 0, -\Delta, -\Delta, -\Delta, 0) \\ \mu_3 &= (0, 0, 0, 0, -\Delta, \Delta, \Delta, -\Delta^2) \end{aligned} \quad \Sigma_k = \begin{pmatrix} I_4 & & & 0 \\ & 1 & \rho & \rho & \rho \\ & 0 & \rho & 1 & \rho & \rho \\ & & \rho & \rho & 1 & \rho \\ & & & \rho & \rho & \rho & 1 \end{pmatrix}$$

$$n_k = 250 \text{ (for all } k \text{ in } \{1, 2, 3\}), \Delta = 2 \rho = 0.3$$

- 10 other configurations varying: the separability between clusters, the number of clusters, the cluster size, the balance between clusters sizes and the heteroscedasticity.

Missing data generation

- MCAR: $\mathbb{P}(r_{ij} = 0) = \tau \quad \forall i, j$
- MAR: $\mathbb{P}(r_{ij} = 0) = \Phi(a_\tau + x_{i1}) \quad \forall j \neq 1$
- $\tau \in \{10\%, 25\%, 40\%\}$

Evaluation

For each incomplete data set (200 per configuration), 3 **clustering methods** are applied: k-means, fuzzy c-means, clustering by GMM

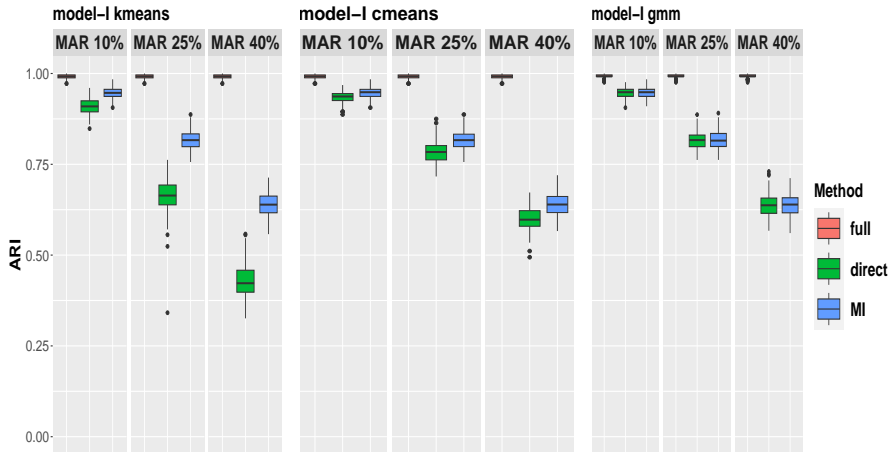
MI¹ is compared to **competitive direct methods**

- **k-POD** (Chi et al., 2016, *The American Statistician*) implemented in the *kpodclustr* R package
- **FCM by optimal completion strategy** (Hathaway and Bezdek, 2001, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*)
- **Ignorable-GMM** (Marbac et al., 2019, *Journal of Classification*) implemented in the *VarSelLCM* R package

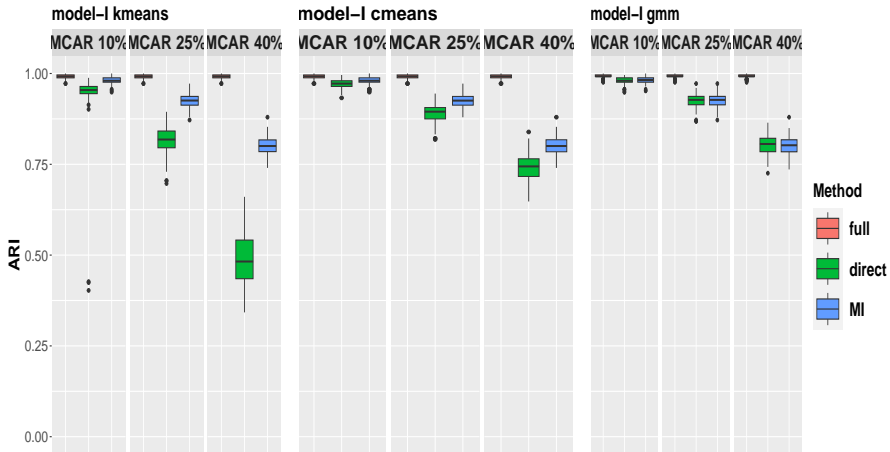
Criteria: Adjusted Rand Index

¹(JM-DP, $M = 20$, pooling by NMF)

Results: base-case, MAR



Results: base-case, MCAR



Summary

Based on **simulated data** under mixture model distribution

- MI outperforms direct methods for kmeans and fuzzy c means
- MI and direct methods provide similar ARI for GMM
- Differences between MI and direct method highlighted for kmeans with more separated clusters
- Similar results by modifying the number of clusters, the cluster size, the balance between clusters sizes and the heteroscedasticity

Real data sets

Data

	n	p	Type	Variables		K	Silhouette Index	Size of cluster	
				Number for which Shapiro rejects normality				(min)	(max)
wine	178	13	Real	7		3	0.57	48	71
ovarian	216	100	Real	64		2	0.50	95	121
iris	150	4	Real	1		3	0.52	50	50
glass	214	9	Real	9		2	0.56	51	163
breast cancer	699	9	Discrete	9		2	0.59	241	458

- Gaussian assumption seems not observed
- p large
- n_k small compared to p
- partitions not obvious

Simulation design

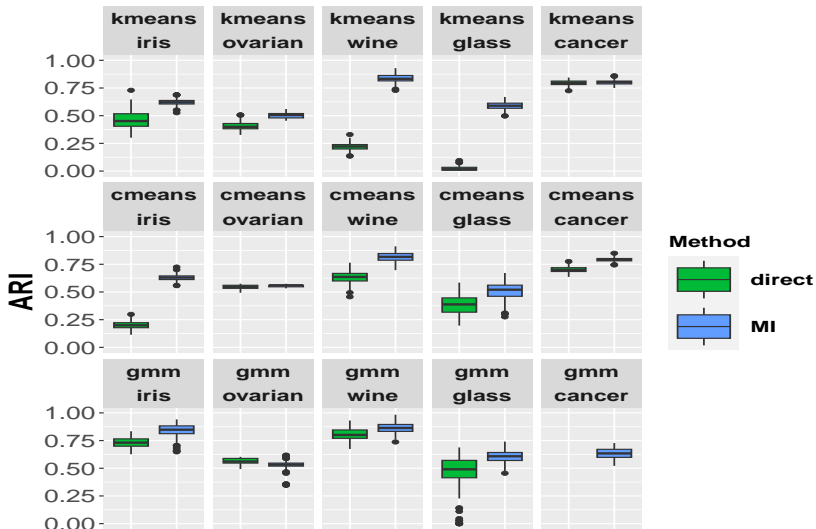
Data sets generation

- 25% missing values (MCAR or a MAR mechanism)
- 200 missing data patterns per mechanism

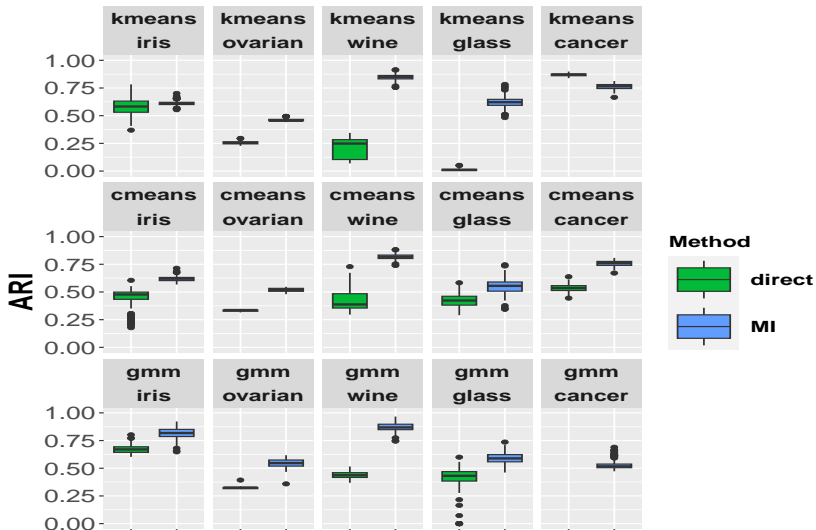
Data sets analysis

- missing values are addressed by MI (FCS-homo, $M = 20$)
- cluster analysis by k-means, fuzzy c-means or GMM
- pooling using NMF

Real data, MCAR



Real data, MAR



Take-home message

MI is a **competitive method** for addressing missing values in clustering

- for model-based or distance-based methods
- good performances on real data

In practice

- A **suitable imputation model** is required
- A large value for M is recommended
- The number of clusters can be easily estimated
- Mixed data can be handled
- Available in the [clusterMI](#) R package

Some perspectives

- Investigating other imputation models dedicated to large data sets
- Developing indices for clustering with missing values

References I

- Aschenbruck, R., Szepannek, G., and Wilhelm, A. F. X. (2022). Imputation strategies for clustering mixed-type data with missing values. *Journal of Classification*, 40(1):2–24.
- Audigier, V. (2024). *clusterMI*. R package version 1.2.1.
- Audigier, V. and Niang, N. (2022). Clustering with missing data: which equivalent for Rubin's rules? *Advances in Data Analysis and Classification*.
- Audigier, V., Niang, N., and Resche-Rigon, M. (2021). Clustering with missing data: which imputation model for which cluster analysis method?
- Basagana, X., Barrera-Gomez, J., Benet, M., Anto, J. M., and Garcia-Aymerich, J. (2013). A Framework for Multiple Imputation in Cluster Analysis. *American Journal of Epidemiology*, 177(7):718–725.
- Bruckers, L., Molenberghs, G., and Dendale, P. (2017). Clustering multiply imputed multivariate high-dimensional longitudinal profiles. *Biometrical Journal*, 59(5):998–1015.
- Chi, J. T., Chi, E. C., and Baraniuk, R. G. (2016). k-pod: A method for k-means clustering of missing data. *The American Statistician*, 70(1):91–99.
- Fang, Y. and Wang, J. (2012). Selection of the number of clusters via the bootstrap method. *Comput. Stat. Data Anal.*, 56(3):468–477.

References II

- Faucheux, L., Resche-Rigon, M., Curis, E., Soumelis, V., and Chevret, S. (2020). Clustering with missing and left-censored data: A simulation study comparing multiple-imputation-based procedures. *Biometrical Journal*, n/a(n/a).
- Filkov, V. and Skiena, S. (2004). Integrating microarray data by consensus clustering. *International Journal on Artificial Intelligence Tools*, 13(04):863–880.
- Hathaway, R. J. and Bezdek, J. C. (2001). Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(5):735–744.
- Honda, K., Nonoguchi, R., Notsu, A., and Ichihashi, H. (2011). Pca-guided k-means clustering with incomplete data. In *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, pages 1710–1714.
- Jain, B. J. (2017). Consistency of mean partitions in consensus clustering. *Pattern Recognition*, 71:26 – 35.
- Kim, H. J. (2020). *DPIImputeCont*. R package version 1.2.2.
- Kim, H. J., Reiter, J. P., Wang, Q., Cox, L. H., and Karr, A. (2014). Multiple imputation of missing or faulty values under linear constraints. *Journal of Business & Economic Statistics*, 32(3):375–386.

References III

- Lee, D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In Leen, T., Dietterich, T., and Tresp, V., editors, *Advances in Neural Information Processing Systems*, volume 13. MIT Press.
- Li, T., Ding, C., and Jordan, M. I. (2007). Solving consensus and semi-supervised clustering problems using nonnegative matrix factorization. In *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, ICDM '07*, page 577–582, USA. IEEE Computer Society.
- Marbac, M., Sedki, M., and Patin, T. (2019). Variable selection for mixed data clustering: application in human population genomics. *Journal of Classification*, pages 1–19.
- McCaw, Z. R., Aschard, H., and Julienne, H. (2022). Fitting gaussian mixture models on incomplete data. *BMC bioinformatics*, 23(1):1–20.
- Miao, W., Ding, P., and Geng, Z. (2016). Identifiability of normal and normal mixture models with nonignorable missing data. *Journal of the American Statistical Association*, 111(516):1673–1683.
- Morris, T. P., White, I. R., and Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC medical research methodology*, 14(1):75.

References IV

- Plaehn, D. (2019). Revisiting french tomato data: Cluster analysis with incomplete data. *Food Quality and Preference*, 76:146 – 159.
- Rubin, D. (1987). *Multiple Imputation for Non-Response in Survey*. Wiley, New-York.
- Templ, M., Kowarik, A., and Filzmoser, P. (2011). Iterative stepwise regression imputation using standard and robust methods. *Computational Statistics & Data Analysis*, 55(10):2793 – 2806.
- van Buuren, S. (2011). Multiple imputation of multilevel data. In *The Handbook of Advanced Multilevel Analysis*, chapter forthcoming. Routledge, Milton Park, UK.
- Vega-Pons, S. and Ruiz-Shulcloper, J. (2011). A survey of clustering ensemble algorithms. *IJPRAI*, 25(3):337–372.
- Wagstaff, K. (2004). Clustering with missing values: No imputation required. In Banks, D., McMorris, F. R., Arabie, P., and Gaul, W., editors, *Classification, Clustering, and Data Mining Applications*, pages 649–658, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Zahid, F. M. and Heumann, C. (2019). Multiple imputation with sequential penalized regression. *Statistical Methods in Medical Research*, 28(5):1311–1327. PMID: 29451087.
- Zhang, L., Lu, W., Liu, X., Pedrycz, W., and Zhong, C. (2016). Fuzzy c-means clustering of incomplete data based on probabilistic information granules of missing values. *Knowledge-Based Systems*, 99:51–70.