Multiple imputation with principal component methods

Vincent Audigier

Inserm ECSTRA team, Saint-Louis Hospital, Paris

IRMA, March 29, 2016

Missing values

NA					*NA*	*NA*	
			NA				
	NA						*NA*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		NA	*NA*	*NA*			

- Aim: inference on a quantity θ from incomplete data
  $\rightarrow$ point estimate $\hat{\theta}$ and associated variability T

Missing values

NA					*NA*	*NA*	
			NA				
	NA						*NA*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		NA	*NA*	*NA*			

- Aim: inference on a quantity θ from incomplete data
  $\rightarrow$ point estimate $\hat{\theta}$ and associated variability T
- R: response indicator $\rightarrow f(R;\psi)$
  $X = (X^{obs}, X^{miss})$: data $\rightarrow f(X;\gamma)$

Missing values

NA					NA	NA	
			NA				
	NA						NA
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		NA	NA	NA			

- Aim: inference on a quantity θ from incomplete data
  $\rightarrow$ point estimate $\hat{\theta}$ and associated variability T
- R: response indicator $\rightarrow f(R;\psi)$
  $X = (X^{obs}, X^{miss})$: data $\rightarrow f(X;\gamma)$
- Likelihood approaches: $\theta = \gamma$
  - Frequentist $\rightarrow L(\gamma, \psi | X^{obs}, R)$
  - Bayesian $\rightarrow p(\gamma, \psi | X^{obs}, R)$

Missing values

NA					*NA*	*NA*	
			NA				
	NA						*NA*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		NA	*NA*	*NA*			

- Aim: inference on a quantity θ from incomplete data
  $\rightarrow$ point estimate $\hat{\theta}$ and associated variability T
- R: response indicator $\rightarrow f(R;\psi)$
  $X = (X^{obs}, X^{miss})$: data $\rightarrow f(X;\gamma)$
- Likelihood approaches: $\theta = \gamma$
  - Frequentist $\rightarrow L(\gamma, \psi | X^{obs}, R)$
  - Bayesian $\rightarrow p(\gamma, \psi | X^{obs}, R)$

Require a modelisation effort

Ignorability

The missing data mechanism is ignorable if:

1. MAR assumption

$$
\begin{aligned}
f\left(X^{obs}, R; \gamma, \psi\right) &= \int f\left(X^{obs}, X^{miss}, R; \gamma, \psi\right) dX^{miss} \\
&= \int f\left(R|X^{obs}, X^{miss}; \psi\right) f\left(X^{obs}, X^{miss}; \gamma\right) dX^{miss} \\
&\overset{MAR}{=} \int f\left(R|X^{obs}; \psi\right) f\left(X^{obs}, X^{miss}; \gamma\right) dX^{miss} \\
&= f\left(R|X^{obs}; \psi\right) f\left(X^{obs}; \gamma\right)
\end{aligned}
$$

2. Distinctness between γ and ψ

Ignorability

The missing data mechanism is ignorable if:

1. MAR assumption

$$
\begin{aligned}
f\left(X^{obs}, R; \gamma, \psi\right) &= \int f\left(X^{obs}, X^{miss}, R; \gamma, \psi\right) dX^{miss} \\
&= \int f\left(R|X^{obs}, X^{miss}; \psi\right) f\left(X^{obs}, X^{miss}; \gamma\right) dX^{miss} \\
&\overset{MAR}{=} \int f\left(R|X^{obs}; \psi\right) f\left(X^{obs}, X^{miss}; \gamma\right) dX^{miss} \\
&= f\left(R|X^{obs}; \psi\right) f\left(X^{obs}; \gamma\right)
\end{aligned}
$$

2. Distinctness between γ and ψ

$\Rightarrow$ Inference on γ can be done without knowing ψ

How to make inference under ignorability?

Aim: inference on a quantity θ from incomplete data under ignorability $\left(f\left(X^{obs}, R; \gamma, \psi\right) \propto f\left(X^{obs}; \gamma\right)\right)$

- Frequentist $\rightarrow$ EM
- Bayesian $\rightarrow$ Data Augmentation

How to make inference under ignorability?

Aim: inference on a quantity θ from incomplete data under ignorability $\left(f\left(X^{obs}, R; \gamma, \psi\right) \propto f\left(X^{obs}; \gamma\right)\right)$

- Frequentist $\rightarrow$ EM
- Bayesian $\rightarrow$ Data Augmentation
- Constraint: $\theta = \gamma$

  $\rightarrow$ can be laborious for several quantities of interest
  $\rightarrow$ difficulties to include auxiliary variables
  (MAR: $f\left(R|X^{obs}, X^{miss}; \psi\right) = f\left(R|X^{obs}; \psi\right)$)

How to make inference under ignorability?

Aim: inference on a quantity θ from incomplete data under ignorability $\left(f\left(X^{obs}, R; \gamma, \psi\right) \propto f\left(X^{obs}; \gamma\right)\right)$

- Frequentist $\rightarrow$ EM
- Bayesian $\rightarrow$ Data Augmentation
- Constraint: $\theta = \gamma$

  $\rightarrow$ can be laborious for several quantities of interest
  $\rightarrow$ difficulties to include auxiliary variables
  (MAR: $f\left(R|X^{obs}, X^{miss}; \psi\right) = f\left(R|X^{obs}; \psi\right)$)

$\Rightarrow$ Multiple imputation can be used when $\theta = \gamma$ or $\theta \neq \gamma$

Multiple imputation (Rubin, 1987)

$$
\begin{aligned}
p\left(\theta|X^{obs},R\right) &= \int p\left(\theta|X^{obs},X^{miss}\right) p\left(X^{miss}|X^{obs},R\right) dX^{miss} \\
&= \int p\left(\theta|X^{obs},X^{miss}\right) p\left(X^{miss}|X^{obs}\right) dX^{miss} \\
&\approx \frac{1}{M}\sum_{m=1}^{M} p\left(\theta|X_m^{miss},X^{obs}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\theta|X^{obs}\right] = \mathbb{E}\left[\mathbb{E}\left[\theta|X^{obs},X^{miss}\right]|X^{obs}\right] &\approx \hat{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m \\
\mathbb{V}\left[\theta|X^{obs}\right] = \mathbb{E}\left[\mathbb{V}\left[\theta|X^{obs},X^{miss}\right]|X^{obs}\right] &\approx T = \frac{1}{M}\sum_{m=1}^{M}\widehat{Var}\left(\hat{\theta}_m\right) \\
+\mathbb{V}\left[\mathbb{E}\left[\theta|X^{obs},X^{miss}\right]|X^{obs}\right] &\quad +\frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{\theta}_m-\mathbb{E}\left[\theta|X^{obs}\right]\right)^2
\end{aligned}
$$

Multiple imputation (Rubin, 1987)

$$
\begin{aligned}
p\left(\theta|X^{obs},R\right) &= \int p\left(\theta|X^{obs},X^{miss}\right)p\left(X^{miss}|X^{obs},R\right)dX^{miss} \\
&= \int p\left(\theta|X^{obs},X^{miss}\right)p\left(X^{miss}|X^{obs}\right)dX^{miss} \\
&\approx \frac{1}{M}\sum_{m=1}^{M} p\left(\theta|X_m^{miss},X^{obs}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\theta|X^{obs}\right] = \mathbb{E}\left[\mathbb{E}\left[\theta|X^{obs},X^{miss}\right]|X^{obs}\right] &\approx \hat{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m \\
\mathbb{V}\left[\theta|X^{obs}\right] = \mathbb{E}\left[\mathbb{V}\left[\theta|X^{obs},X^{miss}\right]|X^{obs}\right] &\approx T = \frac{1}{M}\sum_{m=1}^{M}\widehat{Var}\left(\hat{\theta}_m\right) \\
+\mathbb{V}\left[\mathbb{E}\left[\theta|X^{obs},X^{miss}\right]|X^{obs}\right] &\quad +\frac{\left(1+\frac{1}{M}\right)}{M-1}\sum_{m=1}^{M}\left(\hat{\theta}_m-\hat{\theta}\right)^2
\end{aligned}
$$

Generating imputed data sets

$$p\left(X^{miss}|X^{obs}\right) = \int p\left(X^{miss}|X^{obs},\gamma\right) p\left(\gamma|X^{obs}\right) d\gamma$$

To simulate $p\left(X^{miss}|X^{obs},\gamma\right)$: Joint modelling or Fully conditional specification:

- JM: define $p(X,\gamma)$, draw from $p\left(X^{miss}|X^{obs},\hat{\gamma}_1\right)$, $p\left(X^{miss}|X^{obs},\hat{\gamma}_2\right)$, $\ldots$, $p\left(X^{miss}|X^{obs},\hat{\gamma}_M\right)$
- FCS: define $p\left(X_k|X_{-k},\gamma_{-k}\right)$, draw from $p\left(X_k^{miss}|X_{-k}^{obs},\hat{\gamma}_{-k}\right)$ for all k. Repeat with $\left(\hat{\gamma}_{-k}^2\right)_{1\leq k\leq K}$, $\ldots$, $\left(\hat{\gamma}_{-k}^M\right)_{1\leq k\leq K}$.

However... $I < K$? high dependence? high dimensionality?

Generating imputed data sets

$$p\left(X^{miss}|X^{obs}\right) = \int p\left(X^{miss}|X^{obs},\gamma\right) p\left(\gamma|X^{obs}\right) d\gamma$$

To simulate $p\left(X^{miss}|X^{obs},\gamma\right)$: Joint modelling or Fully conditional specification:

- JM: define $p(X,\gamma)$, draw from $p\left(X^{miss}|X^{obs},\hat{\gamma}_1\right)$, $p\left(X^{miss}|X^{obs},\hat{\gamma}_2\right)$, ..., $p\left(X^{miss}|X^{obs},\hat{\gamma}_M\right)$
- FCS: define $p\left(X_k|X_{-k},\gamma_{-k}\right)$, draw from $p\left(X_k^{miss}|X_{-k}^{obs},\hat{\gamma}_{-k}\right)$ for all k. Repeat with $\left(\hat{\gamma}_{-k}^2\right)_{1\leq k\leq K}$, ..., $\left(\hat{\gamma}_{-k}^M\right)_{1\leq k\leq K}$.

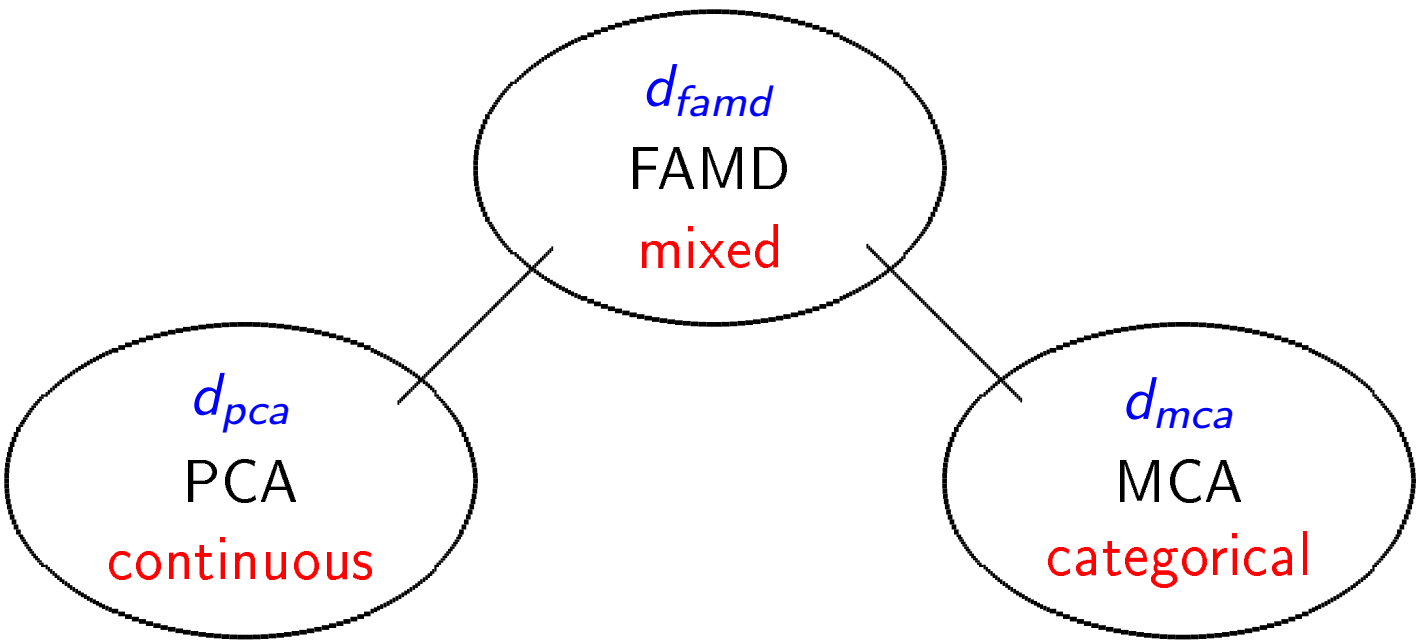
However... $I < K$? high dependence? high dimensionality?

Could principal component methods provide another way to deal with missing values?

Principal component methods

Dimensionality reduction:

- individuals are seen as elements of $\mathbb{R}^K$
- a distance d on $\mathbb{R}^K$
- $Vect(v_1, ..., v_S)$ maximising the projected inertia



$$d^2_{famd} = d^2_{pca} + d^2_{mca}$$

How to perform FAMD?

FAMD can be seen as the SVD of $\mathbf{X}$ with weights for

- the continuous variables and categories: $(\mathbf{D}_\Sigma)^{-1}$
- the individuals: $\frac{1}{I}\mathbb{1}_I$

$$\longrightarrow SVD\left(\mathbf{X}, (\mathbf{D}_\Sigma)^{-1}, \frac{1}{I}\mathbb{1}_I\right)$$

$$X = \begin{array}{|ccc|c|ccc|ccc|}
\hline
11.04 & \dots & 2.07 & 1 & 0 & \dots & 1 & 0 & 0 \\
10.76 & \dots & 1.86 & 1 & 0 & \dots & 1 & 0 & 0 \\
11.02 & \dots & 2.04 & 1 & 0 & \dots & 1 & 0 & 0 \\
11.02 & \dots & 1.92 & 0 & 1 & \dots & 0 & 1 & 0 \\
 & & & & & & & & \\
 & & & & & & & & \\
 & & & & & & & & \\
11.06 & & 2.01 & 0 & 1 & & 0 & 0 & 1 \\
10.95 & & 1.67 & 0 & 1 & & 0 & 1 & 0 \\
\hline
\end{array}$$

$$D_\Sigma = \begin{array}{|cc|}
\hline
\sigma_{x_1} & \\
 \ddots & 0 \\
 & \sigma_{x_k} \\
\hline
 & I_{k+1} \\
0 & \ddots \\
 & I_K \\
\hline
\end{array}$$

How to perform FAMD?

$$SVD\left(\mathbf{X}, (\mathbf{D}_\Sigma)^{-1}, \frac{1}{I}\mathbb{1}_I\right) \longrightarrow \begin{array}{ll} \mathbf{X}_{I\times K} & = \mathbf{U}_{I\times K}\Lambda^{1/2}_{K\times K}\mathbf{V}^\top_{K\times K} \\ \text{with} & \mathbf{U}^\top\left(\frac{1}{I}\mathbb{1}_I\right)\mathbf{U} = \mathbb{1}_K \\ & \mathbf{V}^\top\mathbf{D}_\Sigma^{-1}\mathbf{V} = \mathbb{1}_K \end{array}$$

- principal components: $\hat{\mathbf{F}}_{I\times S} = \hat{\mathbf{U}}_{I\times S}\hat{\Lambda}^{1/2}_{S\times S}$
- loadings: $\hat{\mathbf{V}}^\top_{K\times S}$
- fitted matrix: $\hat{\mathbf{X}}_{I\times K} = \hat{\mathbf{U}}_{I\times S}\hat{\Lambda}^{1/2}_{S\times S}\hat{\mathbf{V}}^\top_{K\times S}$

$$\| \hat{\mathbf{X}} - \mathbf{X} \|^2_{\mathbf{D}_\Sigma^{-1}\otimes\frac{1}{I}\mathbb{1}} = tr\left(\left(\hat{\mathbf{X}} - \mathbf{X}\right)\mathbf{D}_\Sigma^{-1}\left(\hat{\mathbf{X}} - \mathbf{X}\right)^\top\frac{1}{I}\mathbb{1}_I\right)$$

minimized under the constraint of rank S

Properties of the method

- The distance between individuals is:

$$d^2(i,i') = \sum_{j=1}^{k} \frac{(x_{ij} - x_{i'j})^2}{\sigma^2_{x_j}} + \sum_{j=k+1}^{K} \frac{1}{l_j}(x_{ij} - x_{i'j})^2$$

- The principal component $\mathbf{F}_s$ maximises:

$$\sum_{var \in continuous} r^2(\mathbf{F}_s, var) + \sum_{var \in categorical} \eta^2(\mathbf{F}_s, var)$$

FAMD with missing values

$\Rightarrow$ FAMD: least squares

$$\|\mathbf{X}_{I\times K} - \mathbf{U}_{I\times S}\mathbf{\Lambda}^{\frac{1}{2}}_{S\times S}\mathbf{V}^{\top}_{K\times S}\|^2$$

$\Rightarrow$ FAMD with missing values: weighted least squares

$$\|\mathbf{W}_{I\times K} * (\mathbf{X}_{I\times K} - \mathbf{U}_{I\times S}\mathbf{\Lambda}^{\frac{1}{2}}_{S\times S}\mathbf{V}^{\top}_{K\times S})\|^2$$

with $w_{ij} = 0$ if x_{ij} is missing, $w_{ij} = 1$ otherwise

Many algorithms developed for PCA such as NIPALS (Christoffersson, 1970) or iterative PCA (Kiers, 1997)

FAMD with missing values

Iterative FAMD algorithm:

1. initialization: imputation by mean/proportion
2. iterate until convergence
   (a) estimation of the parameters of FAMD
   $\rightarrow$ SVD of $\left(\mathbf{X}, (\mathbf{D}_\Sigma)^{-1}, \frac{1}{I}\mathbb{1}_I\right)$
   (b) imputation of the missing values with
   $\hat{\mathbf{X}}_{I\times K} = \hat{\mathbf{U}}_{I\times S}\hat{\Lambda}^{1/2}_{S\times S}\hat{\mathbf{V}}^{\top}_{K\times S}$
   (c) $\mathbf{D}_\Sigma$ is updated

NA	...	2.07	A	...	A
10.76	...	1.86	A	...	A
11.02	...	NA	A	...	NA
11.02	...	1.92	B	...	B
11.06		2.01	NA	...	C
NA		1.67	B	...	B

$\rightarrow$

NA	...	2.07	1	0	...	1	0	0
10.76	...	1.86	1	0	...	1	0	0
11.02	...	NA	1	0	...	NA	NA	NA
11.02	...	1.92	0	1	...	0	1	0
11.06		2.01	NA	NA		0	0	1
NA		1.67	0	1		0	1	0

FAMD with missing values

Iterative FAMD algorithm:

1. initialization: imputation by mean/proportion
2. iterate until convergence
   (a) estimation of the parameters of FAMD
   $\rightarrow$ SVD of $\left(\mathbf{X}, (\mathbf{D}_\Sigma)^{-1}, \frac{1}{I}\mathbb{1}_I\right)$
   (b) imputation of the missing values with
   $\hat{\mathbf{X}}_{I\times K} = \hat{\mathbf{U}}_{I\times S}\hat{\Lambda}^{1/2}_{S\times S}\hat{\mathbf{V}}^{\top}_{K\times S}$
   (c) $\mathbf{D}_\Sigma$ is updated

NA	...	2.07	A	...	A
10.76	...	1.86	A	...	A
11.02	...	NA	A	...	NA
11.02	...	1.92	B	...	B
11.06		2.01	NA	...	C
NA		1.67	B	...	B

$\rightarrow$

11.01	...	2.07	1	0	...	1	0	0
10.76	...	1.86	1	0	...	1	0	0
11.02	...	1.89	1	0	...	0.61	0.19	0.20
11.02	...	1.92	0	1	...	0	1	0
11.06		2.01	0.32	0.68		0	0	1
11.01		1.67	0	1		0	1	0

FAMD with missing values

Iterative FAMD algorithm:

1. initialization: imputation by mean/proportion
2. iterate until convergence
   (a) estimation of the parameters of FAMD
   $\rightarrow$ SVD of $\left(\mathbf{X}, (\mathbf{D}_\Sigma)^{-1}, \frac{1}{I}\mathbb{1}_I\right)$
   (b) imputation of the missing values with
   $\hat{\mathbf{X}}_{I\times K} = \hat{\mathbf{U}}_{I\times S}\hat{\Lambda}^{1/2}_{S\times S}\hat{\mathbf{V}}^{\top}_{K\times S}$
   (c) $\mathbf{D}_\Sigma$ is updated

NA	...	2.07	A	...	A
10.76	...	1.86	A	...	A
11.02	...	NA	A	...	NA
11.02	...	1.92	B	...	B
11.06		2.01	NA	...	C
NA		1.67	B	...	B

$\rightarrow$

11.04	...	2.07	1	0	...	1	0	0
10.76	...	1.86	1	0	...	1	0	0
11.02	...	2.04	1	0	...	0.81	0.05	0.14
11.02	...	1.92	0	1	...	0	1	0
11.06		2.01	0.25	0.75		0	0	1
10.95		1.67	0	1		0	1	0

Single imputation with FAMD (Audigier et al., 2014)

Iterative FAMD algorithm:

1. initialization: imputation by mean/proportion
2. iterate until convergence
   (a) estimation of the parameters of FAMD
   $\rightarrow$ SVD of $\left(\mathbf{X}, (\mathbf{D}_\Sigma)^{-1}, \frac{1}{I}\mathbb{1}_I\right)$
   (b) imputation of the missing values with
   $\hat{\mathbf{X}}_{I\times K} = \hat{\mathbf{U}}_{I\times S}\hat{\Lambda}^{1/2}_{S\times S}\hat{\mathbf{V}}^{\top}_{K\times S}$
   (c) $\mathbf{D}_\Sigma$ is updated

11.04	...	2.07	A	...	A
10.76	...	1.86	A	...	A
11.02	...	2.04	A	...	A
11.02	...	1.92	B	...	B
11.06		2.01	B	...	C
10.95		1.67	B	...	B

$\leftarrow$

11.04	...	2.07	1	0	...	1	0	0
10.76	...	1.86	1	0	...	1	0	0
11.02	...	2.04	1	0	...	0.81	0.05	0.14
11.02	...	1.92	0	1	...	0	1	0
11.06		2.01	0.25	0.75		0	0	1
10.95		1.67	0	1		0	1	0

$\Rightarrow$ the imputed values can be seen as degree of membership

Single imputation with FAMD (Audigier et al., 2014)

Iterative FAMD algorithm:

1. initialization: imputation by mean/proportion
2. iterate until convergence
   (a) estimation of the parameters of FAMD
   $\rightarrow$ SVD of $\left(\mathbf{X}, (\mathbf{D}_\Sigma)^{-1}, \frac{1}{I}\mathbb{1}_I\right)$
   (b) imputation of the missing values with
   $\hat{\mathbf{X}}_{I\times K} = \hat{\mathbf{U}}_{I\times S} f\left(\hat{\Lambda}^{1/2}_{S\times S}\right)\hat{\mathbf{V}}^{\top}_{K\times S}$ $\qquad f\left(\lambda_s^{1/2}\right) = \lambda_s^{1/2} - \frac{\hat{\sigma}^2}{\lambda_s^{1/2}}$
   (c) $\mathbf{D}_\Sigma$ is updated

11.04	...	2.07	A	...	A
10.76	...	1.86	A	...	A
11.02	...	2.04	A	...	A
11.02	...	1.92	B	...	B
11.06		2.01	B	...	C
10.95		1.67	B	...	B

$\leftarrow$

11.04	...	2.07	1	0	...	1	0	0
10.76	...	1.86	1	0	...	1	0	0
11.02	...	2.04	1	0	...	0.81	0.05	0.14
11.02	...	1.92	0	1	...	0	1	0
11.06		2.01	0.25	0.75		0	0	1
10.95		1.67	0	1		0	1	0

$\Rightarrow$ the imputed values can be seen as degree of membership

Simulation results

Single imputation with FAMD shows a high quality of prediction compared to random forests (Stekhoven and Bühlmann, 2012)

- on real data
- when the relationships between continuous variables are linear
- for rare categories
- with MAR/MCAR mechanism

Can impute mixed, continuous or categorical data

Simulation results

Single imputation with FAMD shows a high quality of prediction compared to random forests (Stekhoven and Bühlmann, 2012)

- on real data
- when the relationships between continuous variables are linear
- for rare categories
- with MAR/MCAR mechanism

Can impute mixed, continuous or categorical data

But a single imputation method only

From single imputation to multiple imputation

$$p\left(X^{miss}|X^{obs}\right) = \int p\left(X^{miss}|X^{obs},\gamma\right) p\left(\gamma|X^{obs}\right) d\gamma$$

1. Reflect the variability on the parameters of the imputation model
   $\rightarrow \left(\left(\hat{\mathbf{U}}_{I\times S}, \hat{\Lambda}^{1/2}_{S\times S}, \hat{\mathbf{V}}^{\top}_{K\times S}\right)_1, \ldots, \left(\hat{\mathbf{U}}_{I\times S}, \hat{\Lambda}^{1/2}_{S\times S}, \hat{\mathbf{V}}^{\top}_{K\times S}\right)_M\right)$
   Bayesian or Bootstrap
2. Add a disturbance on the prediction by $\hat{X}_m = \hat{\mathbf{U}}_m \hat{\Lambda}^{1/2}_m \hat{\mathbf{V}}^{\top}_m$
   $\rightarrow$ need to distinguish continuous and categorical data

PCA model (Caussinus, 1986)

Model

$$\begin{aligned}\mathbf{X}_{I\times K} &= \tilde{\mathbf{X}}_{I\times K} + \varepsilon_{I\times K} \\ &= \mathbf{U}_{I\times S}\mathbf{\Lambda}^{\frac{1}{2}}_{S\times S}\mathbf{V}^{\top}_{K\times S} + \varepsilon_{I\times K} \text{ with } \varepsilon \sim \mathcal{N}\left(0, \sigma^2 \mathbb{1}_K\right)\end{aligned}$$

Maximum Likelihood:

$$\hat{\mathbf{X}}^S = \widehat{\mathbf{U}}_{I\times S}\widehat{\mathbf{\Lambda}}^{\frac{1}{2}}_{S\times S}\widehat{\mathbf{V}}^{\top}_{K\times S} \quad \rightarrow \quad \widehat{\sigma^2} = \| \mathbf{X} - \hat{\mathbf{X}}^S \|^2 / \text{degrees of f.}$$

Bayesian formulation:

- Hoff (2007): Uniform prior for $\mathbf{U}$ and $\mathbf{V}$, Gaussian on $(\lambda_s)_{s=1\ldots S}$
- Verbanck et al. (2013): Prior on $\tilde{\mathbf{X}}$

Bayesian PCA (Verbanck et al., 2013)

$$\begin{aligned} \text{Model: } \mathbf{X}_{I\times K} &= \tilde{\mathbf{X}}_{I\times K} + \varepsilon_{I\times K} \\ x_{ik} &= \tilde{x}_{ik} + \varepsilon_{ik} \ , \ \varepsilon_{ik} \sim \mathcal{N}(0, \sigma^2) \\ &= \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js} + \varepsilon_{ik} \\ &= \sum_{s=1}^{S} \tilde{x}_{ik}^{(s)} + \varepsilon_{ik} \end{aligned}$$

Prior: $\tilde{x}_{ik}^{(s)} \sim \mathcal{N}(0, \tau_s^2)$

Posterior: $\left(\tilde{x}_{ik}^{(s)} | x_{ik}^{(s)}\right) \sim \mathcal{N}(\Phi_s x_{ik}^{(s)}, \Phi_s \sigma^2)$ with $\Phi_s = \frac{\tau_s^2}{\tau_s^2 + \sigma^2}$

Empirical Bayes for τ_s^2: $\hat{\tau}_s^2 = \left(\hat{\lambda}_s - \hat{\sigma}^2\right)$

$$\hat{\Phi}_s = \frac{\hat{\lambda}_s - \hat{\sigma}^2}{\hat{\lambda}_s} = \frac{\text{signal variance}}{\text{total variance}} \quad \text{(Efron and Morris, 1972)}$$

Multiple imputation with Bayesian PCA (Audigier et al., 2015)

1. Variability of the parameters, M plausible $(\tilde{x}_{ij})^1, \ldots, (\tilde{x}_{ij})^M$
   - Posterior distribution: Bayesian PCA

$$\left(\tilde{x}_{ij}^{(s)} | x_{ij}^{(s)}\right) = \mathcal{N}(\Phi_s x_{ij}^{(s)}, \Phi_s \sigma^2)$$

2. Imputation according to the PCA model using the set of M parameters $x_{ij}^{miss} \leftarrow \mathcal{N}(\hat{x}_{ij}, \hat{\sigma}^2)$

Multiple imputation with Bayesian PCA (Audigier et al., 2015)

1. Variability of the parameters, M plausible $(\tilde{x}_{ij})^1, \ldots, (\tilde{x}_{ij})^M$
   - Posterior distribution: Bayesian PCA

$$\left(\tilde{x}_{ij}^{(s)} | x_{ij}^{(s)}\right) = \mathcal{N}(\Phi_s x_{ij}^{(s)}, \Phi_s \sigma^2)$$

   - Data Augmentation (Tanner and Wong, 1987)

2. Imputation according to the PCA model using the set of M parameters $x_{ij}^{miss} \leftarrow \mathcal{N}(\hat{x}_{ij}, \hat{\sigma}^2)$

Multiple imputation with Bayesian PCA (Audigier et al., 2015)

Data augmentation

- a Gibbs sampler
- simulate $\left(\psi, X^{miss} | X^{obs}\right)$ from
  - (I) $\left(X^{miss} | X^{obs}, \psi\right)$: imputation
  - (P) $\left(\psi | X^{obs}, X^{miss}\right)$: draw from the posterior
- convergence checked by graphical investigations

For Bayesian PCA:

- initialisation: ML estimate for $\tilde{\mathbf{X}}$
- for ℓ in $1...L$
  - (I) Given $\tilde{\mathbf{X}}$, $x_{ij}^{miss} \leftarrow \mathcal{N}(\tilde{x}_{ij}, \hat{\sigma}^2)$
  - (P) $\tilde{x}_{ij} \leftarrow \mathcal{N}\left(\sum_s \hat{\Phi}_s x_{ij}^{(s)}, \frac{\hat{\sigma}^2 \sum_s \hat{\Phi}_s)}{I-1}\right)$

MI methods for continuous data

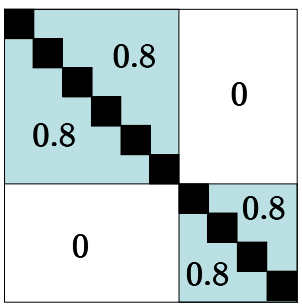
Generally based on normal distribution:

- JM: $X_{I\times K} : x_{i.} \sim \mathcal{N}(\mu, \Sigma)$ (Honaker et al., 2011)
  1. Bootstrap rows: $X^1, \ldots, X^M$
     EM algorithm: $(\mu^1, \Sigma^1), \ldots, (\mu^M, \Sigma^M)$
  2. Imputation: $x_{i.}^m$ drawn from $\mathcal{N}(\mu^m, \Sigma^m)$

- FCS: $\mathcal{N}\left(\mu_{X_k|X_{(-k)}}, \Sigma_{X_k|X_{(-k)}}\right)$ (Van Buuren, 2012)
  1. Bayesian approach: (β^m, σ^m)
  2. Imputation: stochastic regression x_{ij}^m drawn from $\mathcal{N}\left(X_{(-k)}\beta^m, \sigma^m\right)$

Simulations

- Quantities of interest: $\theta_1 = \mathbb{E}[Y], \theta_2 = \beta_1, \theta_3 = \rho$
- 1000 simulations
  - data set drawn from $\mathcal{N}_p(\mu, \Sigma)$ with a two-block structure, varying I (30 or 200), K (6 or 60) and ρ (0.3 or 0.9)



  - 10% or 30% of missing values using a MCAR mechanism
  - multiple imputation using $M = 20$ imputed arrays
- Criteria
  - bias
  - CI width, coverage

Results for the expectation

	parameters				confidence interval width			coverage		
	I	K	ρ	%	*JM*	*FCS*	*BayesMIPCA*	*JM*	*FCS*	*BayesMIPCA*
1	30	6	0.3	0.1	0.803	0.805	0.781	0.955	0.953	0.950
2	30	6	0.3	0.3		1.010	0.898		0.971	0.949
3	30	6	0.9	0.1	0.763	0.759	0.756	0.952	0.95	0.949
4	30	6	0.9	0.3		0.818	0.783		0.965	0.953
5	30	60	0.3	0.1			0.775			0.955
6	30	60	0.3	0.3			0.864			0.952
7	30	60	0.9	0.1			0.742			0.953
8	30	60	0.9	0.3			0.759			0.954
9	200	6	0.3	0.1	0.291	0.294	0.292	0.947	0.947	0.946
10	200	6	0.3	0.3	0.328	0.334	0.325	0.954	0.959	0.952
11	200	6	0.9	0.1	0.281	0.281	0.281	0.953	0.95	0.952
12	200	6	0.9	0.3	0.288	0.289	0.288	0.948	0.951	0.951
13	200	60	0.3	0.1		0.304	0.289		0.957	0.945
14	200	60	0.3	0.3		0.384	0.313		0.981	0.958
15	200	60	0.9	0.1		0.282	0.279		0.951	0.948
16	200	60	0.9	0.3		0.296	0.283		0.958	0.952

Properties for BayesMIPCA

A MI method based on a Bayesian treatment of the PCA model

advantages

- captures the structure of the data: good inferences for regression coefficient, correlation, mean
- a dimensionality reduction method: ($I < K$ or $I > K$, low or high percentage of missing values)
- no inversion issue: strong or weak relationships
- a regularization strategy improving stability

remains competitive if:

- the low rank assumption is not verified
- the Gaussian assumption is not true

Multiple imputation for categorical data using MCA

MI for categorical data is very challenging for a moderate number of variables

- estimation issues
- storage issues

Multiple imputation for categorical data using MCA

MI for categorical data is very challenging for a moderate number of variables

- estimation issues
- storage issues

MI with MCA

1. Variability on the parameters of the imputation model

$$\left(\left(\hat{\mathbf{U}}_{I\times S}, \hat{\Lambda}^{1/2}_{S\times S}, \hat{\mathbf{V}}^{\top}_{K\times S}\right)_1, \ldots, \left(\hat{\mathbf{U}}_{I\times S}, \hat{\Lambda}^{1/2}_{S\times S}, \hat{\mathbf{V}}^{\top}_{K\times S}\right)_M\right)$$

$\rightarrow$ A non-parametric bootstrap approach

2. Add a disturbance on the MCA prediction $\hat{X}_m = \hat{\mathbf{U}}_m \hat{\Lambda}^{1/2}_m \hat{\mathbf{V}}^{\top}_m$

Multiple imputation with MCA (Audigier et al., 2015)

1. Variability of the parameters of MCA $(\hat{\mathbf{U}}_{I\times S}, \hat{\Lambda}^{1/2}_{S\times S}, \hat{\mathbf{V}}^{\top}_{K\times S})$ using a non-parametric bootstrap:
   - define M weightings $(R_m)_{1\leq m\leq M}$ for the individuals
   - estimate MCA parameters using SVD of $\left(\mathbf{X}, \frac{1}{K}(\mathbf{D}_\Sigma)^{-1}, R_m\right)$
2. Imputation:

$\hat{\mathbf{X}}_1$

1	0	...	1	0
1	0	...	1	0
1	0	...	0.81	0.19
0.25	0.75		0	1
0	1		0	1

$\hat{\mathbf{X}}_2$

1	0	...	1	0
1	0	...	1	0
1	0	...	0.60	0.40
0.26	0.74		0	1
0	1		0	1

...

$\hat{\mathbf{X}}_M$

1	0	...	1	0
1	0	...	1	0
1	0	...	0.74	0.16
0.20	0.80		0	1
0	1		0	1

Draw categories from the values of $\left(\hat{\mathbf{X}}_m\right)_{1\leq m\leq M}$

A	...	A
A	...	A
A	...	B
B	...	C
B	...	B

A	...	A
A	...	A
A	...	A
B	...	C
B	...	B

...

A	...	A
A	...	A
A	...	B
B	...	C
B	...	B

Properties

MCA address the categorical data challenge by

- requiring a small number of parameters
- preserving the essential data structure
- using a regularisation strategy

MIMCA can be applied on various data sets

- small or large number of variables/categories
- small or large number of individuals

MI methods for categorical data

- Log-linear model (Schafer, 1997)
  - Hypothesis on $X = (x_{ijk})_{i,j,k}$: $X|\gamma \sim \mathcal{M}(n, \gamma)$

$$log(\gamma_{ijk}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

  1. Variability of the parameter γ: Bayesian formulation
  2. Imputation using the set of M parameters

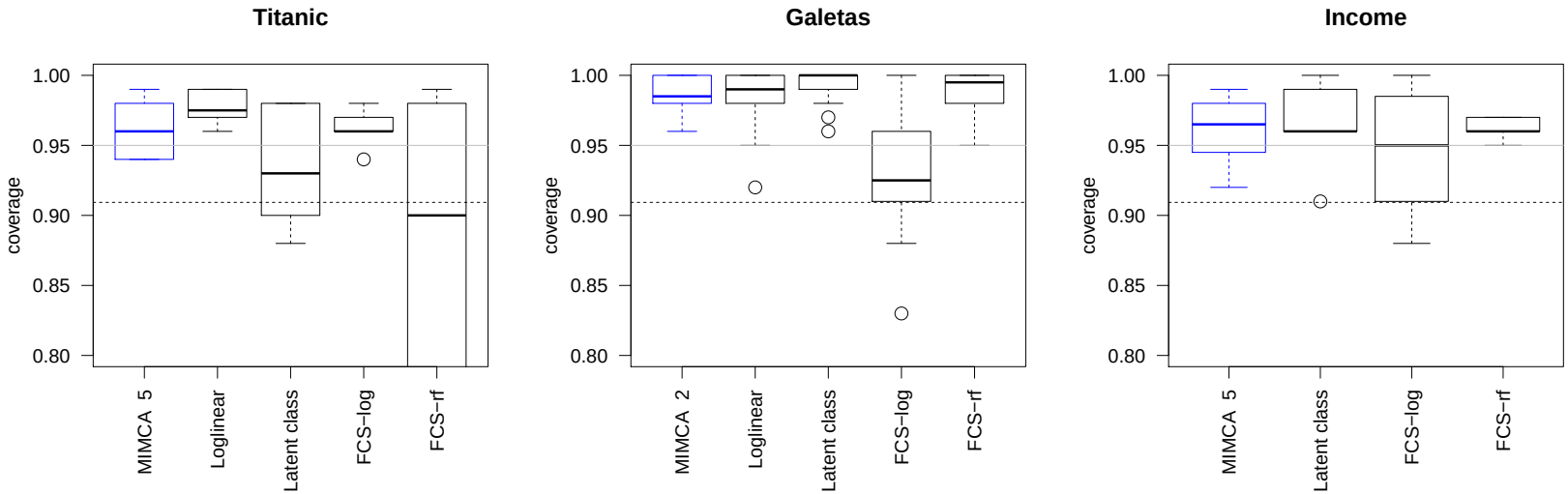
- Latent class model (Si and Reiter, 2013)
  - Hypothesis: $\mathbb{P}(X = (x_1, \ldots, x_K); \gamma) = \sum_{\ell=1}^{L} \left(\gamma_\ell \prod_{k=1}^{K} \gamma_{x_k}^{(\ell)} \right)$
  1. Variability of the parameters γ_L and γ_X : Bayesian formulation
  2. Imputation using the set of M parameters

- FCS: GLM (Van Buuren, 2012) or Random Forests (Doove et al., 2014; Shah et al., 2014)

Simulations from real data sets

- Quantities of interest: θ = parameters of a logistic model
- Simulation design (repeated 200 times)
  - the real data set is considered as a population
  - drawn one sample from the data set
  - generate 20% of missing values
  - multiple imputation using $M = 5$ imputed arrays
- Criteria
  - bias
  - CI width, coverage
- Comparison with :
  - JM: log-linear model, latent class model
  - FCS: logistic regression, random forests

Results - Inference



	Titanic	Galetas	Income
Number of variables	4	4	14
Number of categories	≤ 4	≤ 11	≤ 9

Results - Time

	Titanic	Galetas	Income
MIMCA	2.750	8.972	**58.729**
Loglinear	0.740	4.597	NA
Latent class model	10.854	17.414	143.652
FCS logistic	4.781	38.016	881.188
FCS forests	265.771	112.987	6329.514

Table: Time consumed in second

	Titanic	Galetas	Income
Number of individuals	2201	1192	6876
Number of variables	4	4	14

Conclusion

MI methods using dimensionality reduction method

- captures the relationships between variables
- captures the similarities between individuals
- requires a small number of parameters

Address some imputation issues:

- can be applied on various data sets
- provide correct inferences for analysis model based on relationships between pairs of variables

Available in the R package missMDA

Perspectives

To go further:

- require a modelisation effort when categorical variables occur
  - for a deeper understanding of the methods
  - for an extension of the current methods
  - for a MI method based on FAMD

  $\rightarrow$ some lines of research:

  - link between CA and log-linear model
  - link between log-linear model and general locator model

- uncertainty on the number of dimensions S

References I

V. Audigier, F. Husson, and J. Josse. MIMCA: Multiple imputation for categorical variables with multiple correspondence analysis. *Statistics and Computing*, 2016.

V. Audigier, F. Husson, and J. Josse. Multiple imputation for continuous variables using a bayesian principal component analysis. *Journal of Statistical Computation and Simulation*, 2015.

V. Audigier, F. Husson, and J. Josse. A principal component method to impute missing values for mixed data. *Advances in Data Analysis and Classification*, pages 1–22, 2014.

D. B. Rubin. *Multiple Imputation for Non-Response in Survey*. Wiley, New-York, 1987.

J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC, London, 1997.