Multiple Imputation with Bayesian PCA

Vincent Audigier, Julie Josse, François Husson

Applied mathematics department, Agrocampus Ouest, Rennes, France

Bordeaux, 6 March 2014

Outline

Missing values

			X				Y
NA					*NA*	*NA*	
			NA				
	NA						*NA*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		NA	*NA*	*NA*			

Missing values

			X				Y
NA					*NA*	*NA*	
			NA				
	NA						*NA*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		NA	*NA*	*NA*			

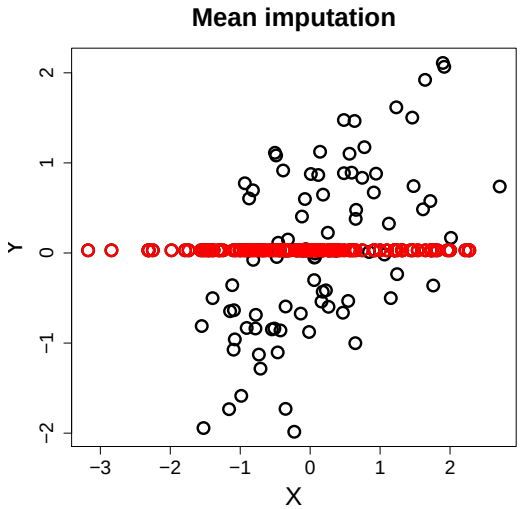
- Deletion of individuals: complete case

Missing values

			X				Y
NA					*NA*	*NA*	
			NA				
	NA						*NA*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		NA	*NA*	*NA*			

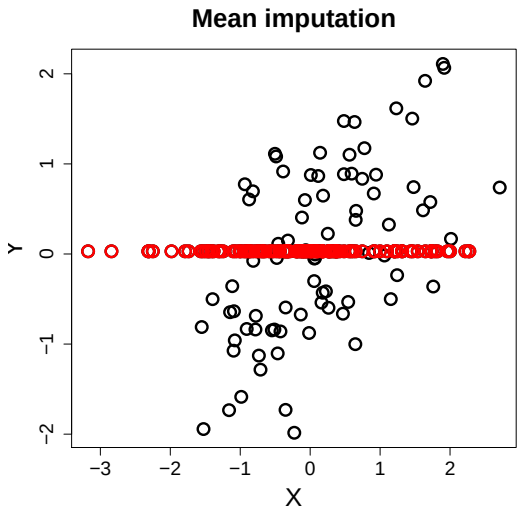
- Deletion of individuals: complete case
- Imputation

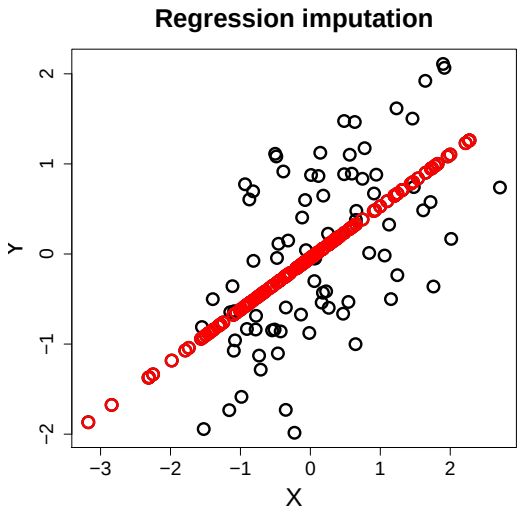
Single imputation methods



$\mu_y = 0$	0.01
$\sigma_y = 1$	0.5
$\rho = 0.6$	0.30
$CI\,\mu_y 95\%$	39.4

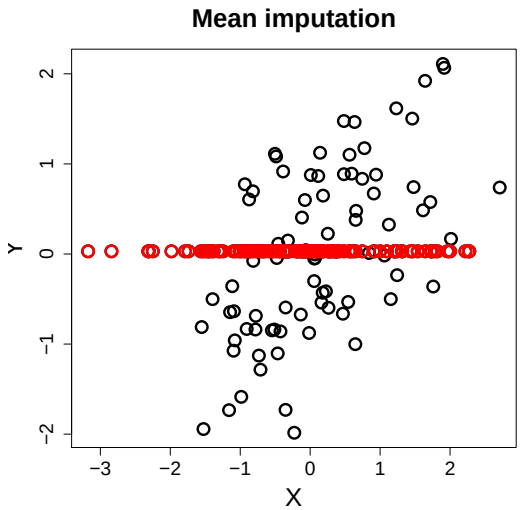
Single imputation methods

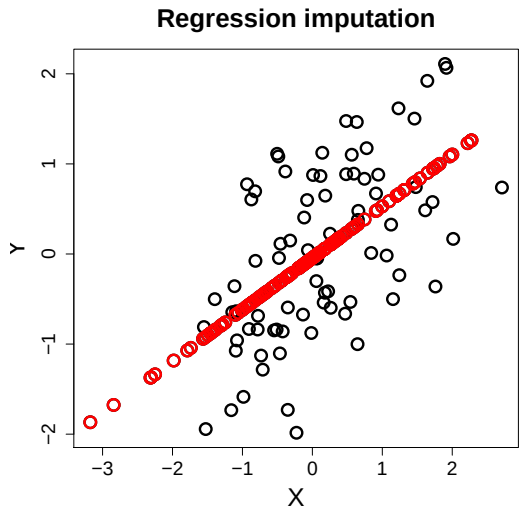


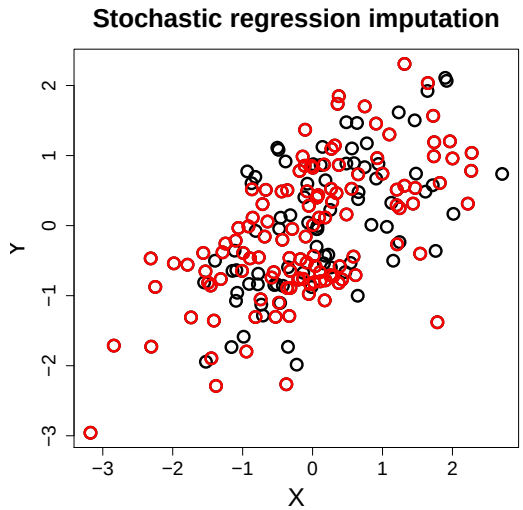


$\mu_y = 0$	0.01	0.01
$\sigma_y = 1$	0.5	0.72
$\rho = 0.6$	0.30	0.78
$CI\,\mu_y 95\%$	39.4	61.6

Single imputation methods





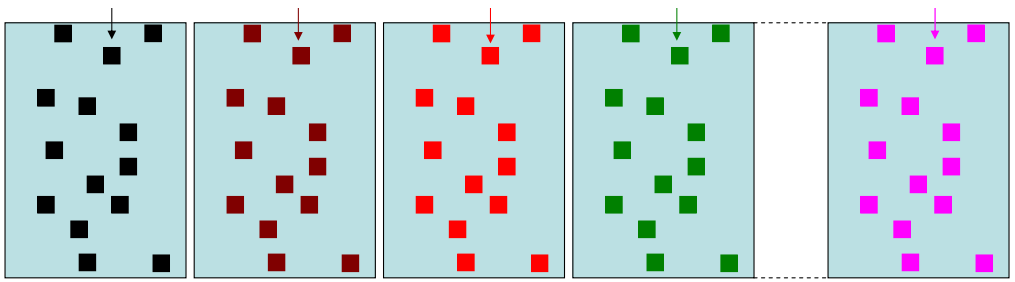


$\mu_y = 0$	0.01	0.01	0.01
$\sigma_y = 1$	0.5	0.72	0.99
$\rho = 0.6$	0.30	0.78	0.59
$CI\,\mu_y 95\%$	39.4	61.6	70.8

$\Rightarrow$ Standard errors of the parameters ($\hat{\sigma}_{\hat{\mu}_y}$) calculated from the imputed data set are underestimated

Multiple imputation (Rubin, 1987)

- Generate M plausible values for each missing value



- Perform the analysis on each imputed data set: $\hat{\theta}_m, \widehat{Var}\left(\hat{\theta}_m\right)$
- Combine the results: $\hat{\theta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_m$

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

$\Rightarrow$ Aim: provide estimation of the parameters and of their variability (taken into account the variability due to missing values)

A multiple imputation procedure requires a single imputation method

1. Introduction
2. Single imputation with PCA
3. Multiple imputation with Bayesian PCA

Outline

PCA

$\Rightarrow$ Geometrical point of view

$$\|\mathbf{X}_{n\times p} - \hat{\mathbf{X}}^S_{n\times p}\|^2 \qquad \hat{\mathbf{X}}^S = \mathbf{U}_{n\times S}\mathbf{\Lambda}^{\frac{1}{2}}_{S\times S}\mathbf{V}'_{p\times S}$$

- $\mathbf{F} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$ principal components - scores
- $\mathbf{V}$ principal axes - loadings

$\Rightarrow$ Model point of view: $\mathbf{X}_{n\times p} = \tilde{\mathbf{X}}_{n\times p} + \varepsilon_{n\times p}$

$$\begin{aligned} x_{ij} &= \tilde{x}_{ij} + \varepsilon_{ij} \ , \ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ &= \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js} + \varepsilon_{ij} \end{aligned}$$

Max Likelihood = Min least squares

Imputation with PCA

$\Rightarrow$ PCA: least squares

$$\|\mathbf{X}_{n\times p} - \mathbf{U}_{n\times S}\mathbf{\Lambda}^{\frac{1}{2}}_{S\times S}\mathbf{V}'_{p\times S}\|^2$$

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

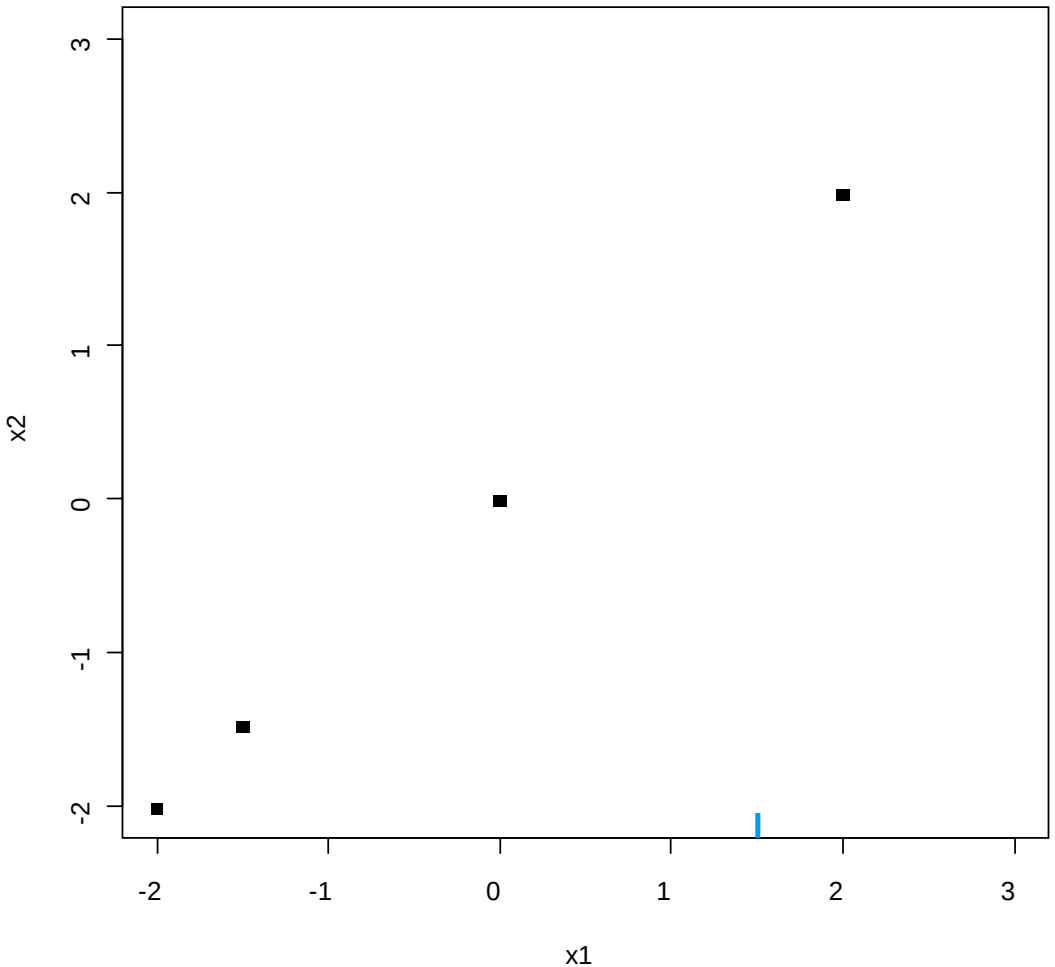
$$\|\mathbf{W}_{n\times p} * (\mathbf{X}_{n\times p} - \mathbf{U}_{n\times S}\mathbf{\Lambda}^{\frac{1}{2}}_{S\times S}\mathbf{V}'_{p\times S})\|^2$$

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

Many algorithms: weighted alternating least squares (Gabriel & Zamir, 1979); iterative PCA (Kiers, 1997)

Iterative PCA

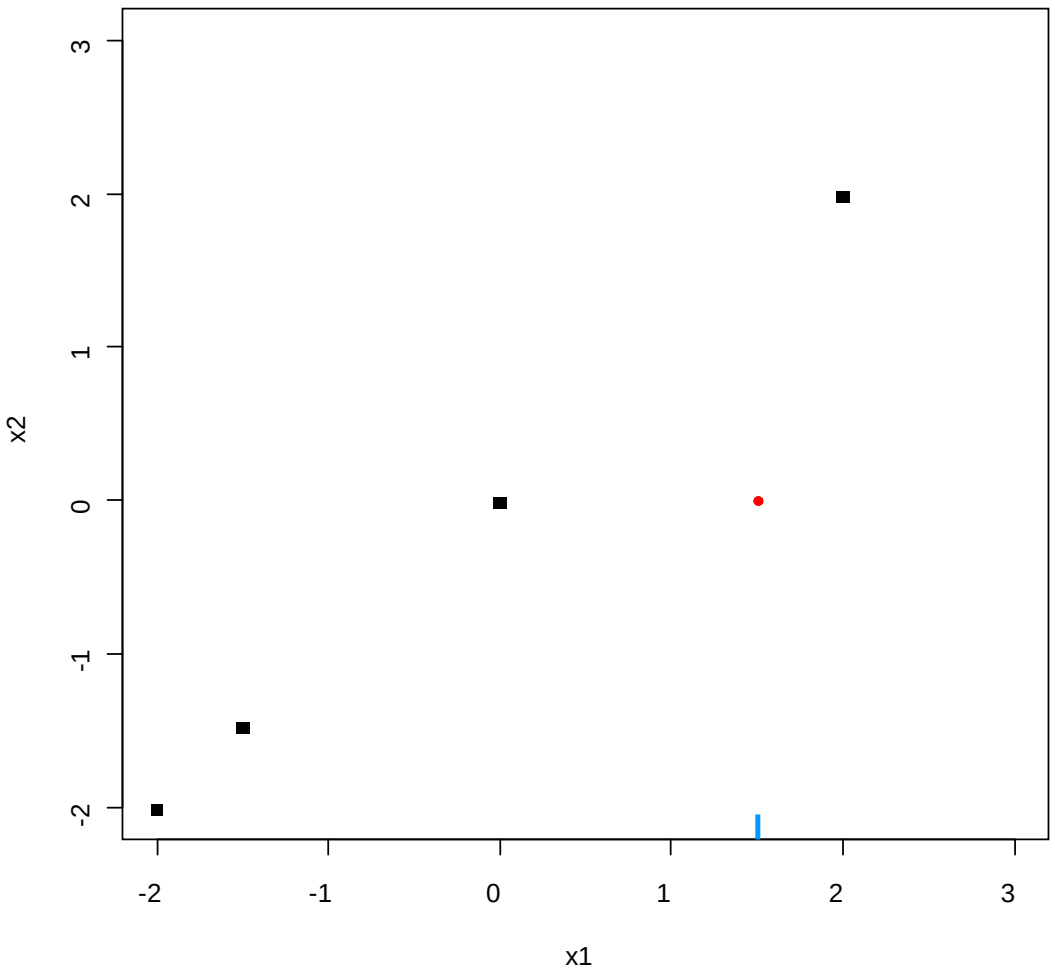
x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	NA
2.0	1.98



Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	NA
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	0.00
2.0	1.98



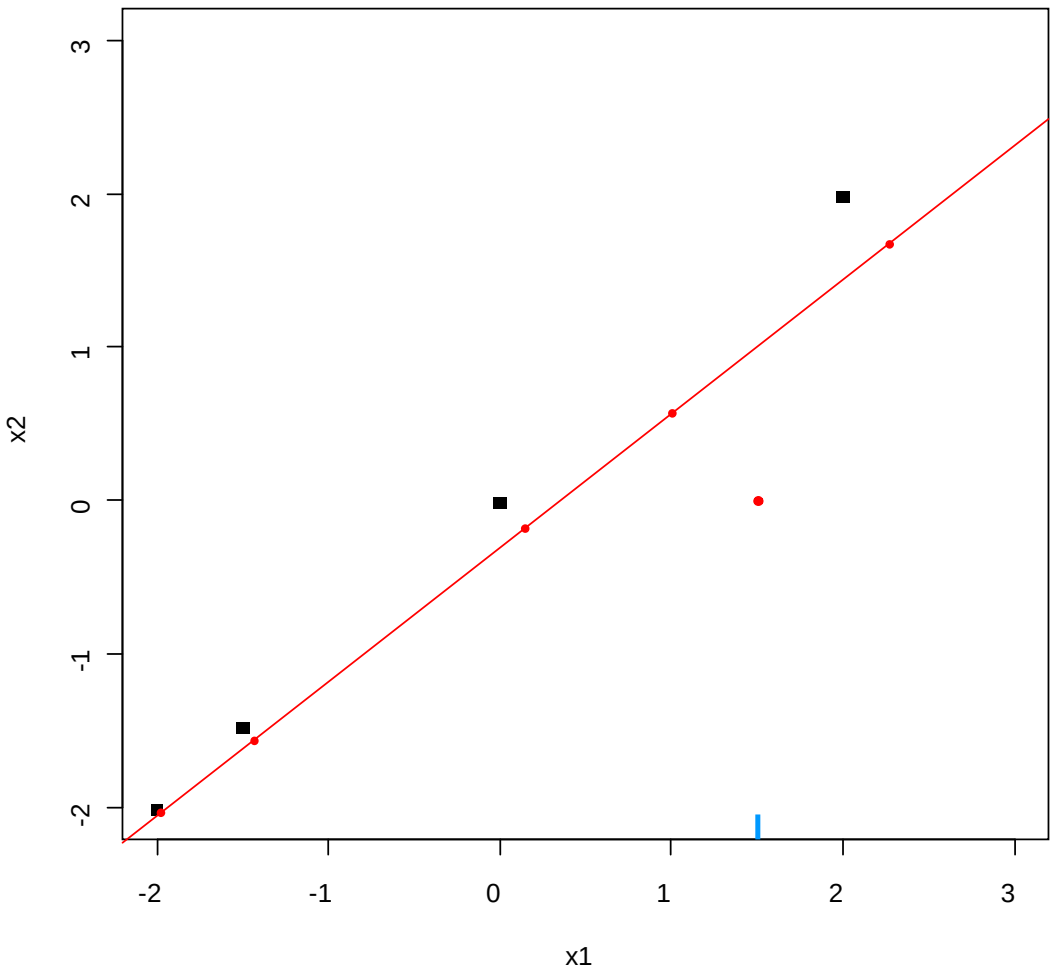
Initialization $\ell = 0$: $\mathbf{X}^0$ (mean imputation)

Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	NA
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.00**
2.0	1.98

$\widehat{x1}$	$\widehat{x2}$
-1.98	-2.04
-1.44	-1.56
0.15	-0.18
1.00	0.57
2.27	1.67



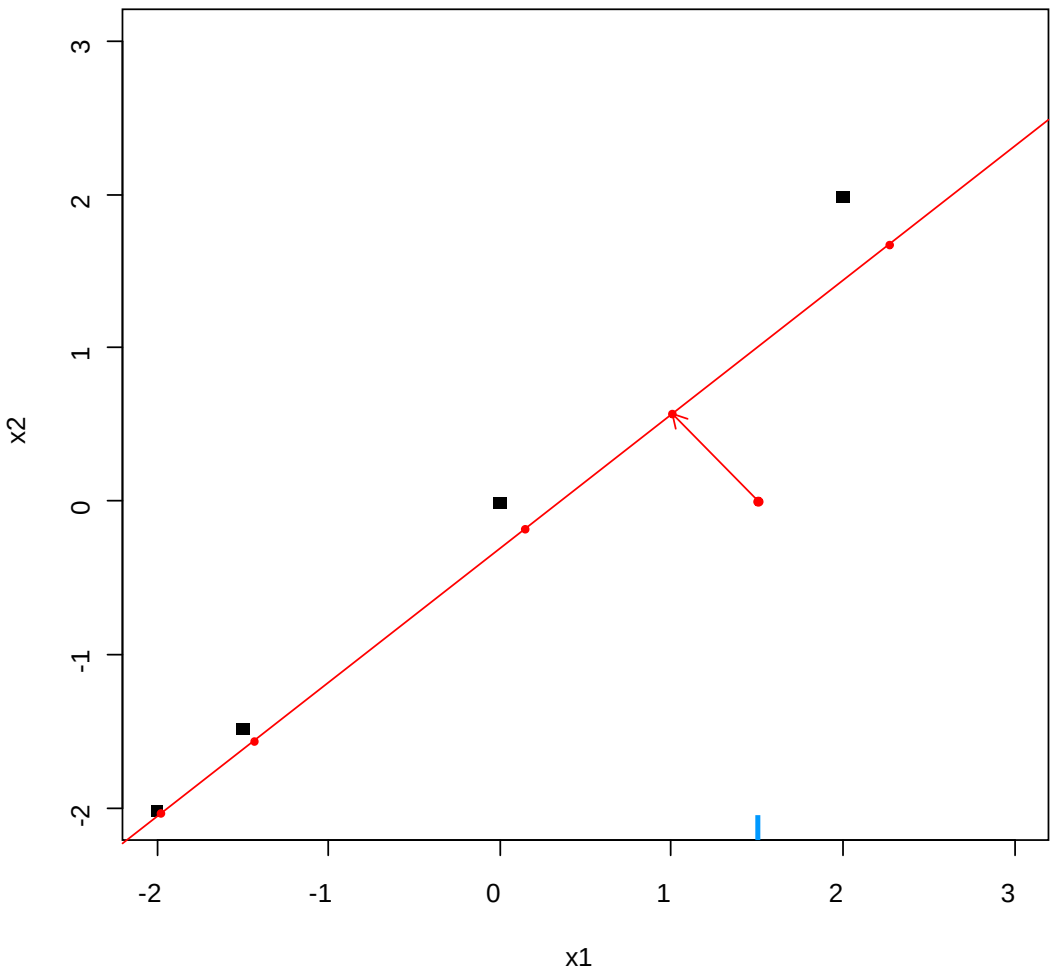
PCA on the completed data set $\rightarrow (\mathbf{U}^\ell, \mathbf{\Lambda}^\ell, \mathbf{V}^\ell)$;

Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	NA
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.00**
2.0	1.98

$\hat{x}1$	$\hat{x}2$
-1.98	-2.04
-1.44	-1.56
0.15	-0.18
1.00	**0.57**
2.27	1.67



Missing values are imputed with the model matrix $\hat{\mathbf{X}}^{\ell} = \mathbf{U}^{\ell}\mathbf{\Lambda}^{\mathbf{1/2}^{\ell}}\mathbf{V}^{\ell\prime}$

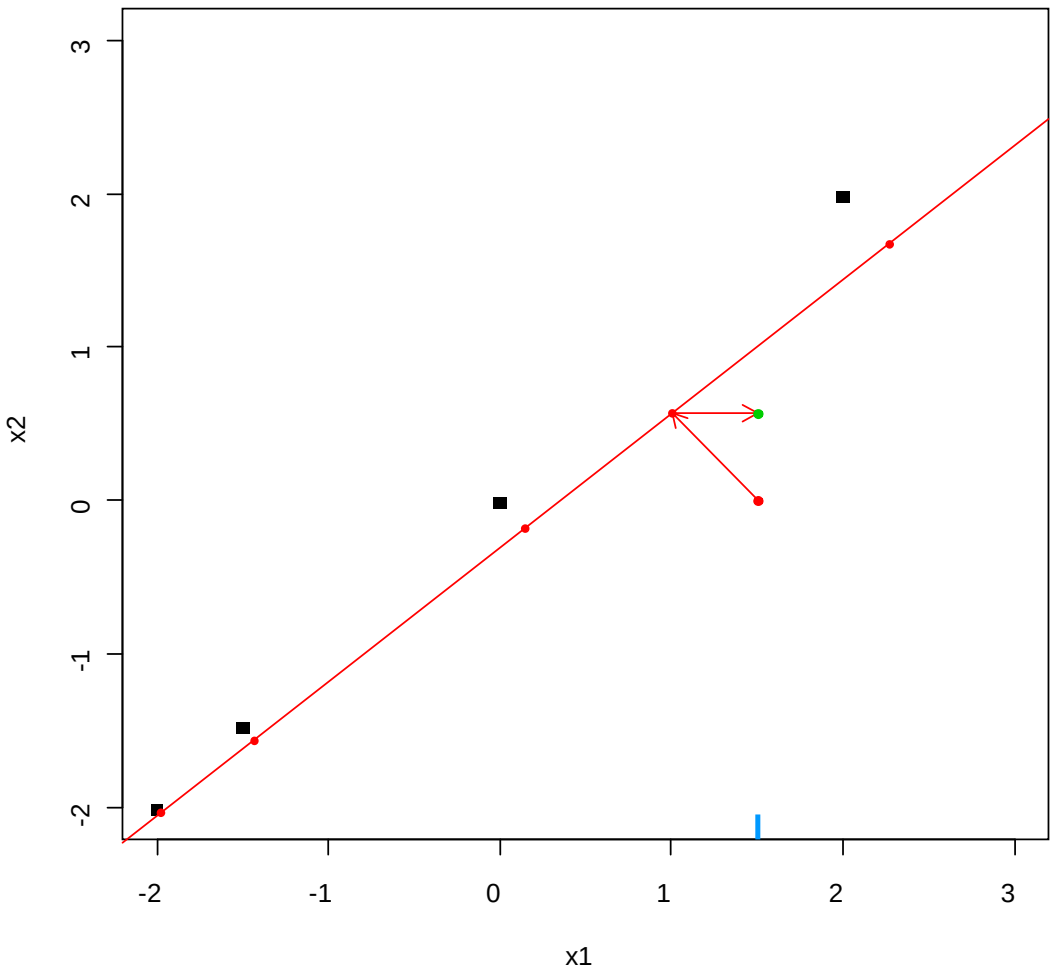
Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**NA**
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.00**
2.0	1.98

$\widehat{x1}$	$\widehat{x2}$
-1.98	-2.04
-1.44	-1.56
0.15	-0.18
1.00	**0.57**
2.27	1.67

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.57**
2.0	1.98



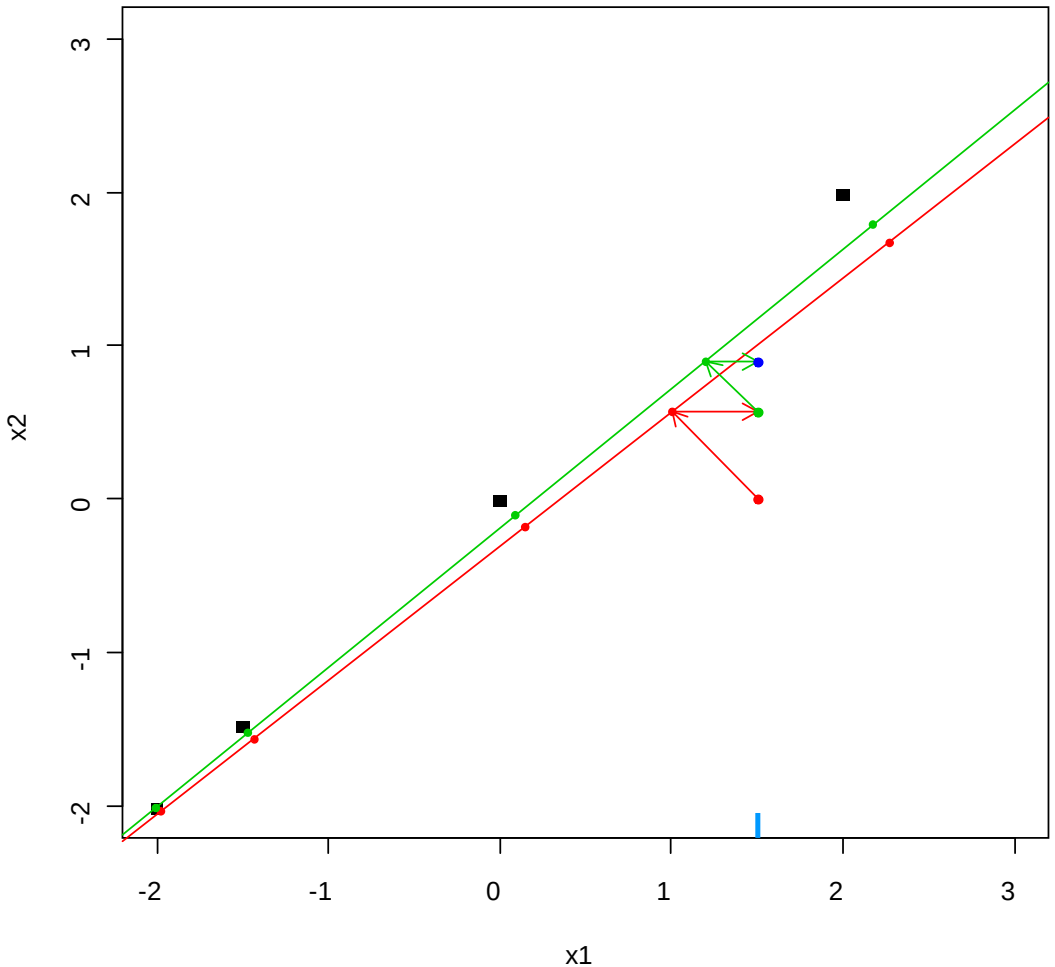
The new imputed dataset is $\mathbf{X}^{\ell} = \mathbf{W} * \mathbf{X} + (1 - \mathbf{W}) * \hat{\mathbf{X}}^{\ell}$

Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**NA**
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.57**
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.57**
2.0	1.98



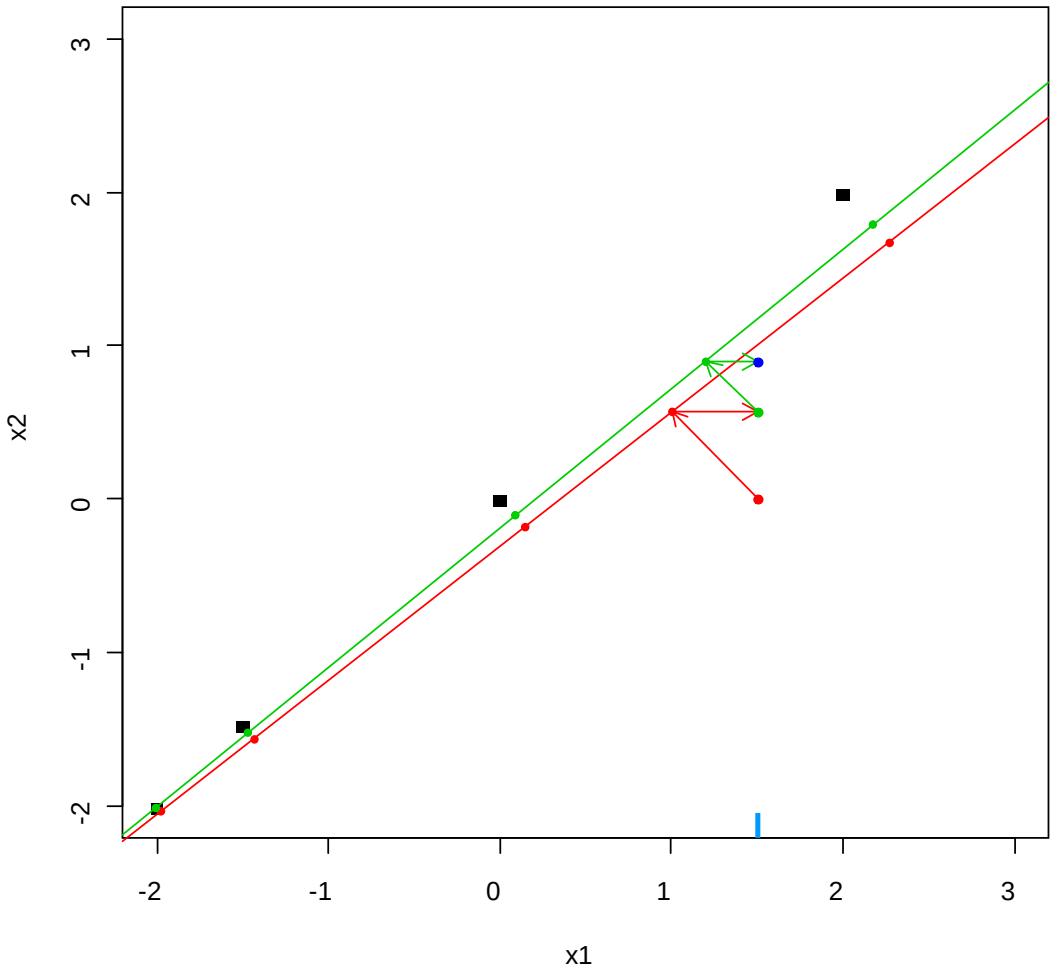
Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**NA**
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.57**
2.0	1.98

$\hat{x}1$	$\hat{x}2$
-2.00	-2.01
-1.47	-1.52
0.09	-0.11
1.20	**0.90**
2.18	1.78

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.90**
2.0	1.98



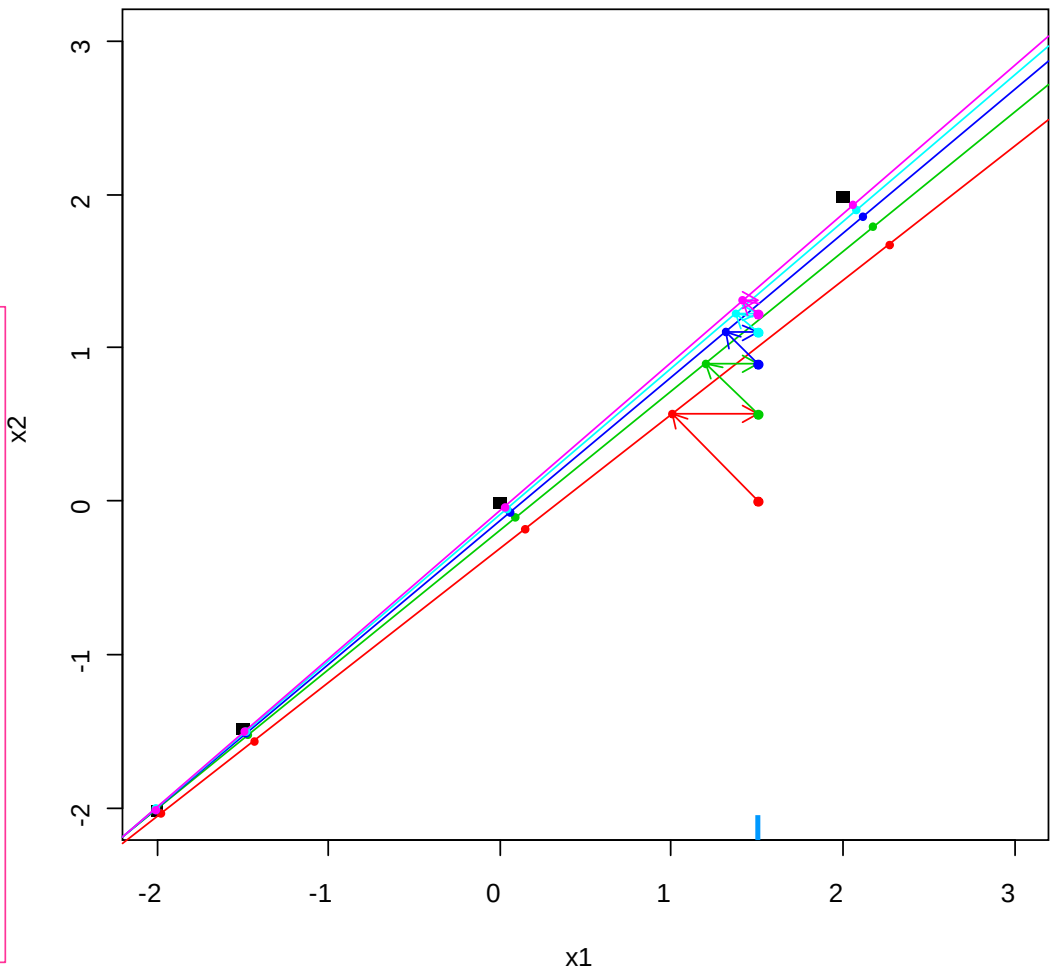
Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	NA
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.00**
2.0	1.98

$\widehat{x1}$	$\widehat{x2}$
-1.98	-2.04
-1.44	-1.56
0.15	-0.18
1.00	0.57
2.27	1.67

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**0.57**
2.0	1.98

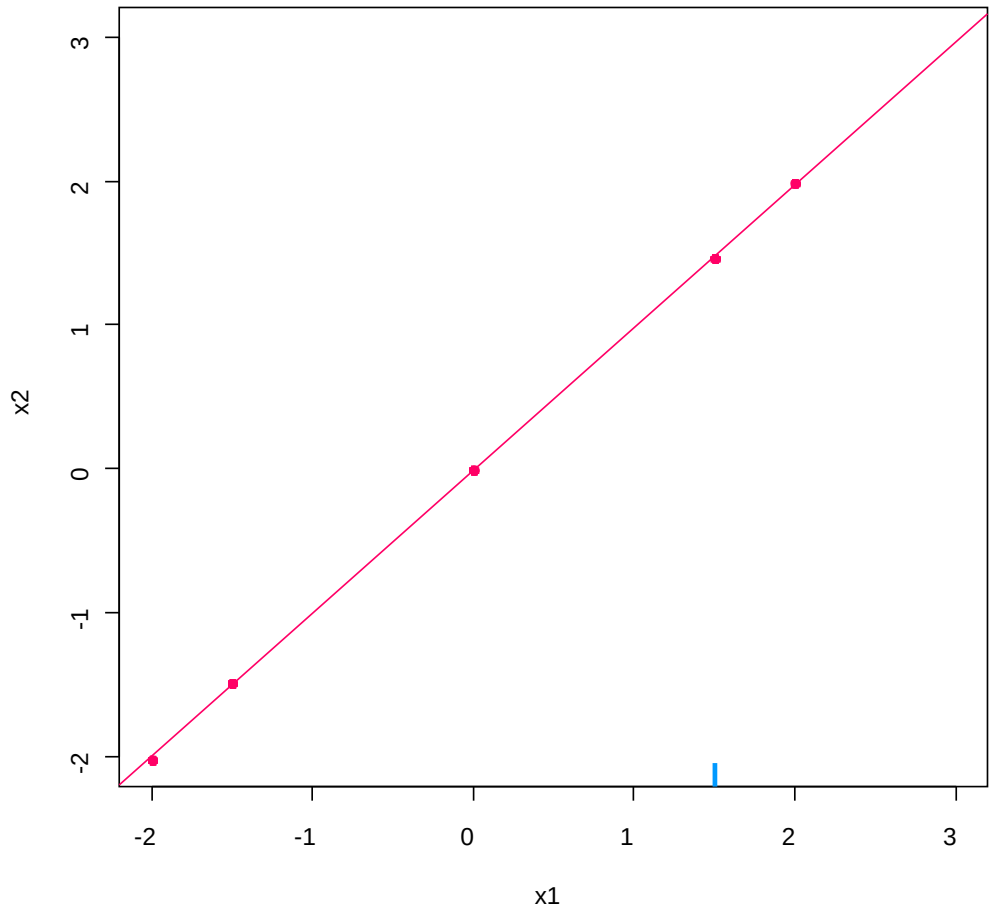


Steps are repeated until convergence

Iterative PCA

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**NA**
2.0	1.98

x1	x2
-2.0	-2.01
-1.5	-1.48
0.0	-0.01
1.5	**1.46**
2.0	1.98



PCA on the completed data set $\rightarrow (\mathbf{U}^\ell, \mathbf{\Lambda}^\ell, \mathbf{V}^\ell)$

Missing values are imputed with the model matrix $\hat{\mathbf{X}}^\ell = \mathbf{U}^\ell \mathbf{\Lambda}^{\mathbf{1/2}^\ell} \mathbf{V}^{\ell\prime}$

Iterative PCA

1. initialization $\ell = 0$: $\mathbf{X}^0$ (mean imputation)
2. step ℓ:
   (a) PCA on the completed data $\rightarrow (\mathbf{U}^\ell, \mathbf{\Lambda}^\ell, \mathbf{V}^\ell)$; S dim. kept
   (b) $\mathbf{X}^\ell = \mathbf{W} * \mathbf{X} + (1 - \mathbf{W}) * (\hat{\mathbf{X}}^\ell = \mathbf{U}^\ell \mathbf{\Lambda}^{1/2^\ell} \mathbf{V}^{\ell\prime})$
3. steps of estimation and imputation are repeated

Iterative PCA

1. initialization $\ell = 0$: $\mathbf{X}^0$ (mean imputation)
2. step ℓ:
   (a) PCA on the completed data $\rightarrow (\mathbf{U}^\ell, \mathbf{\Lambda}^\ell, \mathbf{V}^\ell)$; S dim. kept
   (b) $\mathbf{X}^\ell = \mathbf{W} * \mathbf{X} + (1 - \mathbf{W}) * (\hat{\mathbf{X}}^\ell = \mathbf{U}^\ell \mathbf{\Lambda}^{1/2^\ell} \mathbf{V}^{\ell\prime})$
3. steps of estimation and imputation are repeated

Iterative PCA

1. initialization $\ell = 0$: $\mathbf{X}^0$ (mean imputation)
2. step ℓ:
   (a) PCA on the completed data $\rightarrow (\mathbf{U}^\ell, \mathbf{\Lambda}^\ell, \mathbf{V}^\ell)$; S dim. kept
   (b) $\mathbf{X}^\ell = \mathbf{W} * \mathbf{X} + (1 - \mathbf{W}) * (\hat{\mathbf{X}}^\ell = \mathbf{U}^\ell \mathbf{\Lambda}^{1/2^\ell} \mathbf{V}^{\ell\prime})$
3. steps of estimation and imputation are repeated

$\Rightarrow$ Imputation (matrix completion framework, Netflix)

$\Rightarrow$ Reduce the variability

$\Rightarrow$ EM algorithm

Iterative PCA

1. initialization $\ell = 0$: $\mathbf{X}^0$ (mean imputation)
2. step ℓ:
   (a) PCA on the completed data $\rightarrow (\mathbf{U}^\ell, \mathbf{\Lambda}^\ell, \mathbf{V}^\ell)$; S dim. kept
   (b) $\mathbf{X}^\ell = \mathbf{W} * \mathbf{X} + (1 - \mathbf{W}) * (\hat{\mathbf{X}}^\ell = \mathbf{U}^\ell \mathbf{\Lambda}^{1/2^\ell} \mathbf{V}^{\ell\prime})$
3. steps of estimation and imputation are repeated

$\Rightarrow$ Imputation (matrix completion framework, Netflix)

$\Rightarrow$ Reduce the variability

$\Rightarrow$ EM algorithm

$\Rightarrow$ Overfitting problems

- many parameters / observed values (S large - many NA)
- data are very noisy

Trust too much the relationships between variables

Regularized iterative PCA (Josse *et al.*, 2009)

$\Rightarrow$ Initialization - estimation step - imputation step

The imputation step:

$$\hat{x}_{ij}^{\text{PCA}} = \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js}$$

is replaced by a "shrunk" imputation step:

$$\hat{x}_{ij}^{\text{rPCA}} = \sum_{s=1}^{S} \left(\frac{\lambda_s - \hat{\sigma}^2}{\lambda_s} \right) \sqrt{\lambda_s} u_{is} v_{js} = \sum_{s=1}^{S} \left(\sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}} \right) u_{is} v_{js}$$

Compromise hard/ soft thresholding (Mazumder, Hastie & Tibshirani, 2010)
σ^2 small $\rightarrow$ regularized PCA $\approx$ PCA
σ^2 large $\rightarrow$ mean imputation

Properties of the imputation

- Good imputation quality when the structure is strong (imputation using similarities between individuals and relationship between variables)
- Regularization improves cases with noise and NA
- Close to the mean imputation in extreme cases
- Regularization softens the impact of a wrong choice for S
- Better imputation than random forests (Stekhoven & Bühlmann, 2011) / soft thresholding
- Possibility to impute categorical data (MCA) and mixed data (FAMD)

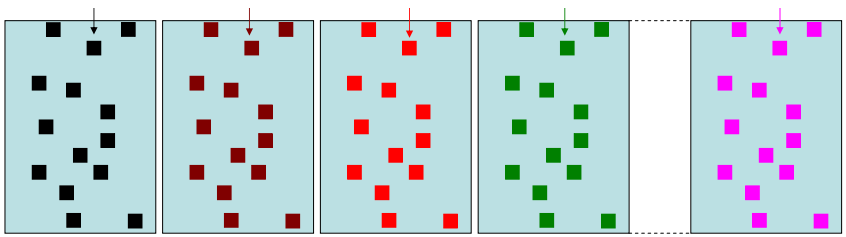
Outline

Multiple imputation

Single imputation: using usual methods on the completed data leads to underestimate the standard errors

$\Rightarrow$ a single value can't reflect the uncertainty of prediction

1. Generating B imputed data sets



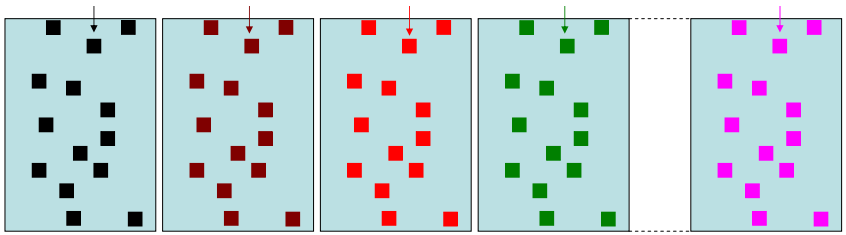
2. Performing the analysis on each imputed data set
3. Combining: variance = within + between imputation variance

Multiple imputation

Single imputation: using usual methods on the completed data leads to underestimate the standard errors

$\Rightarrow$ a single value can't reflect the uncertainty of prediction

1. Generating B imputed data sets



for $b = 1, ..., B$, missing values x_{ij}^b are imputing by drawing from the predictive distribution $\mathcal{N}(\hat{x}_{ij}, \hat{\sigma}^2)$

$\Rightarrow$ "improper" imputation

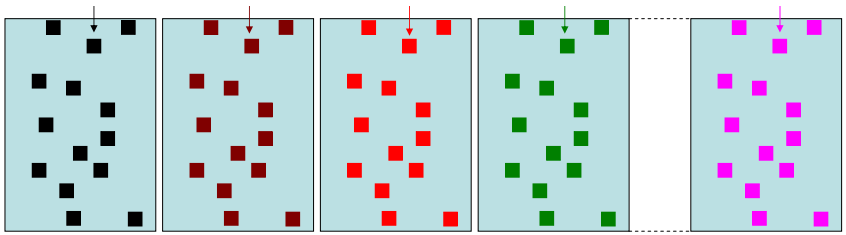
2. Performing the analysis on each imputed data set
3. Combining: variance = within + between imputation variance

Multiple imputation

Single imputation: using usual methods on the completed data leads to underestimate the standard errors

$\Rightarrow$ a single value can't reflect the uncertainty of prediction

1. Generating B imputed data sets



for $b = 1, ..., B$, missing values x_{ij}^b are imputing by drawing from the predictive distribution $\mathcal{N}(\hat{x}_{ij}, \hat{\sigma}^2)$

$\Rightarrow$ "improper" imputation

$\Rightarrow$ Prediction variance = estimation variance plus noise

2. Performing the analysis on each imputed data set
3. Combining: variance = within + between imputation variance

Proper multiple imputation

$$x_{ij} = \tilde{x}_{ij} + \varepsilon_{ij}$$

1. Variability of the parameters, B plausible: $(\hat{x}_{ij})^1, ..., (\hat{x}_{ij})^B$

   $\Rightarrow$ Bootstrap
   $\Rightarrow$ Posterior distribution: Bayesian PCA

2. Noise: for $b = 1, ..., B$, missing values x_{ij}^b are imputing by drawing from the predictive distribution $\mathcal{N}(\hat{x}_{ij}^b, \hat{\sigma}^2)$

Bayesian PCA complete case

$$\begin{aligned} x_{ij} &= \tilde{x}_{ij} + \varepsilon_{ij} \ , \ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ &= \sum_{s=1}^{S} \tilde{x}_{ij}^{(s)} + \varepsilon_{ij} = \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js} + \varepsilon_{ij} \end{aligned}$$

$\Rightarrow$ Priors on $\mathbf{U}, \mathbf{\Lambda}, \mathbf{V}$: overparametrization issue

• Hoff (2009): Uniform prior for $\mathbf{U}$ and $\mathbf{V}$ von Mises-Fisher distributions (Stiefel manifold); $(\lambda_s)_{s=1\ldots S} \sim \mathcal{N}\left(0, s_\lambda^2\right)$

• Josse & Denis (2013): disregarding the constraints; $u_{is} \sim \mathcal{N}(0,1)$, $v_{js} \sim \mathcal{N}(0,1)$, $(\lambda_s)_{s=1\ldots S} \sim \mathcal{N}\left(0, s_\lambda^2\right)$

$\Rightarrow$ Priors on $\mathbf{U}, \mathbf{\Lambda}, \mathbf{V}$: priors on $\tilde{\mathbf{X}}$

Bayesian PCA (Josse *et al.*, 2013) complete case

Model: $x_{ij} = \sum_{s=1}^{S} \tilde{x}_{ij}^{(s)} + \varepsilon_{ij}$, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$

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

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

Empirical Bayes: max the likelihood of $\left(x_{ij}^{(s)}\right) \sim \mathcal{N}(0, \tau_s^2 + \sigma^2)$ as a function of τ_s^2 to obtain: $\hat{\tau_s}^2 = \left(\lambda_s - \hat{\sigma}^2\right)$

$$\hat{\Phi}_s = \frac{\lambda_s - \hat{\sigma}^2}{\lambda_s} = \frac{\text{signal variance}}{\text{total variance}}$$

Bayesian PCA (Josse *et al.*, 2013) complete case

Point estimate: expectation of the posterior distribution

$$\hat{x}_{ij}^{\text{rPCA}} = \sum_{s=1}^{S} \left(\frac{\lambda_s - \hat{\sigma}^2}{\lambda_s} \right) \sqrt{\lambda_s} u_{is} v_{js} = \sum_{s=1}^{S} \left(\sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}} \right) u_{is} v_{js}$$

$$\hat{x}_{ij}^{\text{PCA}} = \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js}$$

$\Rightarrow$ Regularized version of the Max. Likelihood

Bayesian PCA (Josse *et al.*, 2013) complete case

Point estimate: expectation of the posterior distribution

$$\hat{x}_{ij}^{\text{rPCA}} = \sum_{s=1}^{S} \left(\frac{\lambda_s - \hat{\sigma}^2}{\lambda_s} \right) \sqrt{\lambda_s} u_{is} v_{js} = \sum_{s=1}^{S} \left(\sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}} \right) u_{is} v_{js}$$

$$\hat{x}_{ij}^{\text{PCA}} = \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js}$$

$\Rightarrow$ Regularized version of the Max. Likelihood

$$Y = \mathbf{X}\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2)$$
$$\beta \sim \mathcal{N}(0,\ \tau^2)$$
$$\hat{\beta}^{\text{MAP}} = (\mathbf{X}'\mathbf{X} + \kappa)^{-1}\mathbf{X}'Y, \text{ with } \kappa = \sigma^2/\tau^2$$

Bayesian PCA (Josse *et al.*, 2013) complete case

Point estimate: expectation of the posterior distribution

$$\hat{x}_{ij}^{\text{rPCA}} = \sum_{s=1}^{S} \left(\frac{\lambda_s - \hat{\sigma}^2}{\lambda_s} \right) \sqrt{\lambda_s} u_{is} v_{js} = \sum_{s=1}^{S} \left(\sqrt{\lambda_s} - \frac{\hat{\sigma}^2}{\sqrt{\lambda_s}} \right) u_{is} v_{js}$$

$$\hat{x}_{ij}^{\text{PCA}} = \sum_{s=1}^{S} \sqrt{\lambda_s} u_{is} v_{js}$$

$\Rightarrow$ Regularized version of the Max. Likelihood

$$Y = \mathbf{X}\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2)$$
$$\beta \sim \mathcal{N}(0,\ \tau^2)$$
$$\hat{\beta}^{\text{MAP}} = (\mathbf{X}'\mathbf{X} + \kappa)^{-1}\mathbf{X}'Y, \text{ with } \kappa = \sigma^2/\tau^2$$

$\Rightarrow$ Other priors: other regularizations (Todeschini *et al*, 2013)

Multiple imputation Bayes-PCA

1. Variability of the parameters, B plausible $(\hat{x}_{ij})^1, ..., (\hat{x}_{ij})^B$
   $\Rightarrow$ Posterior distribution: Bayesian PCA

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

   $\Rightarrow$ Data Augmentation (Tanner and Wong, 1987)

2. Noise: for $b = 1, ..., B$, missing values x_{ij}^b are imputing by drawing from the predictive distribution $\mathcal{N}(\hat{x}_{ij}, \hat{\sigma}^2)$

Data augmentation

$\Rightarrow$ Augmenting data by prediction on missing data

(I) given $\tilde{X}$ and σ^2, imputing the missing values x_{ij} by a draw from the predictive distribution $\mathcal{N}\left(\tilde{x}_{ij}, \sigma^2\right)$

(P) drawing $\tilde{x}_{ij}$ from $\mathcal{N}\left(\hat{x}_{ij}^{rPCA}, \frac{\hat{\sigma}^2 \sum_s \hat{\phi}_s}{\min n-1, p}\right)$

$\Rightarrow$ Observed posterior distribution

Outline

Joint modeling

- Hypothesis $X_{n\times p}$: $x_{i.} \sim \mathcal{N}(\mu, \Sigma)$
- Algorithm:
  1. Bootstrap rows: $X^1, \ldots, X^B$
     EM algorithm: $(\mu^1, \Sigma^1), \ldots, (\mu^B, \Sigma^B)$
  2. Imputation: x_{ij}^b draw from $\mathcal{N}(\mu^b, \Sigma^b)$
- Implemented: `R` package `Amelia` (J. Honaker, G. King, M. Blackwell)

Conditional modeling

- Hypothesis: one model/variable
  all variables continuous and a regression/variable: $\mathcal{N}(\mu, \Sigma)$

- Algorithm:

  One variable with missing values:

  1. Bootstrap - Bayesian approaches: (β^1, σ^1), ... , (β^B, σ^B)
  2. Imputation: stochastic regression x_{ij}^b draw from $\mathcal{N}\left(X_{-j}\beta^b, \sigma^b\right)$

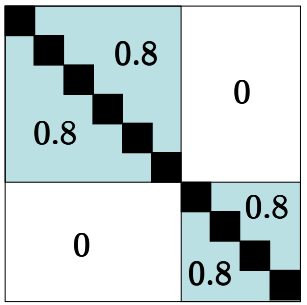
  Many variables with missing values: cycles through variables

- Implemented: `R` package `MICE` (Stef van Buuren)

Simulations

The simulated data $\mathcal{N}(\mu, \Sigma)$:

- $n = 30,200$ individuals
- $p = 6,60$ variables with a two-block structure
  $\Rightarrow$ 2 underlying dimensions

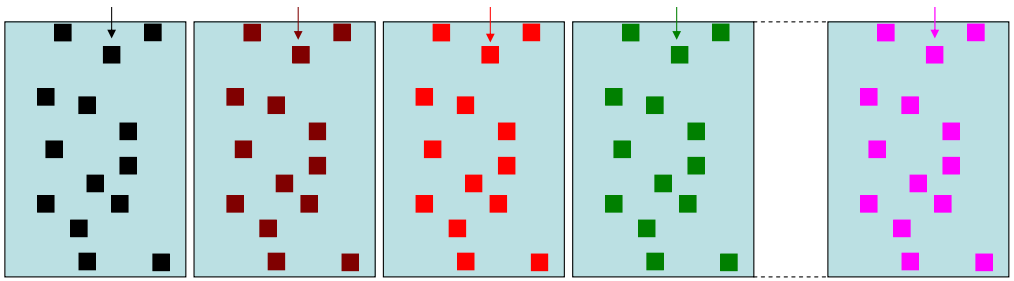


Many scenarios are considered:

- relationship between variables: low (0.3) or strong (0.8)
- missing values mechanism: MCAR (Rubin, 1976)
- percentage of missing values: 10%, 30%

$\Rightarrow$ Imputation with $M = 20$ Amelia, MICE, MIPCA, deletion
$\Rightarrow$ 1000 simulations

Simulations



$\Rightarrow$ Parameters:

$\theta_1 = \mu_{X_1}$

$\theta_2 = \beta_1$, regression: $X_1 \sim X_2 + ... + X_{10}$

$\theta_3 = \rho(X_{p-1}, X_p)$

$\Rightarrow$ For each method:

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

$T = \frac{1}{M} \sum_m \widehat{Var}\left(\hat{\theta}_m\right) + \left(1 + \frac{1}{M}\right) \frac{1}{M-1} \sum_m \left(\hat{\theta}_m - \hat{\theta}\right)^2$

$\Rightarrow$ Biais, RMSE θ_i - coverage and width CI over 1000 simulations

Results for the expectation

	parameters				confidence interval width				coverage			
	n	p	ρ	%	*LD*	*Amelia*	*BayesMI*	*BayesMIPCA*	*LD*	*Amelia*	*BayesMI*	*BayesMIPCA*
1	30	6	0.3	0.1	1.034	0.803	0.805	0.781	0.936	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	1.048	0.763	0.759	0.756	0.951	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	0.383	0.291	0.294	0.292	0.938	0.947	0.947	0.946
10	200	6	0.3	0.3	0.864	0.328	0.334	0.325	0.942	0.954	0.959	0.952
11	200	6	0.9	0.1	0.385	0.281	0.281	0.281	0.945	0.953	0.95	0.952
12	200	6	0.9	0.3	0.862	0.288	0.289	0.288	0.942	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

Results for the correlation coefficient

$\Rightarrow$ Fisher transformation
$\Rightarrow$ Compare the CI width to the one of the complete case

- BayesMI and Amelia similar
- MIPCA smaller width and better coverage
- strong relationships: MIPCA » BayesMI (width 3 times larger than complete case versus 1. 4 for MIPCA; coverage 68%)

Conclusion - Perspectives

$\Rightarrow$ A new multiple imputation method based on PCA
$\Rightarrow$ Bayesian PCA: Shrinkage point estimates - Posterior distribution

$\Rightarrow$ Good results and works when n<p
$\Rightarrow$ Requires to define the number of dimensions
$\Rightarrow$ Linear relationships (X_1^2)

$\Rightarrow$ Categorical data?
$\Rightarrow$ MI: good theory for regression parameters. Other parameters?
$\Rightarrow$ Alternative: modify the estimation procedure

Implementation

$\Rightarrow$ R package `missMDA`

$\Rightarrow$ Missing values in principal components methods (PCA, MCA, MIXPCA,multi-tables methods: MFA)

$\Rightarrow$ Single imputation for continuous - categorical - mixed data
$\Rightarrow$ Multiple imputation

$\Rightarrow$ Video on Youtube! FactoMineR package

Multiple imputation for PCA

⇒ Uncertainty around the position of the individuals and variables