Separable Covariance Models for Multiway Array Data

Peter Hoff

Statistics and Biostatistics
University of Washington

Outline

Examples of multiway data

Separable covariance arrays

Trade example

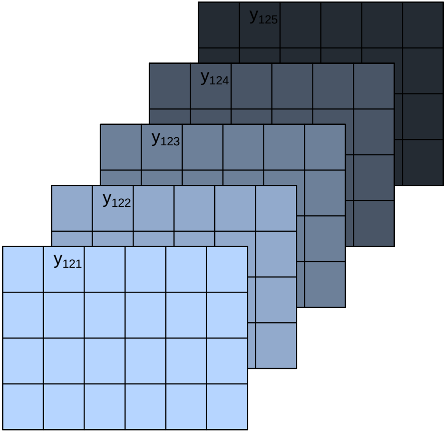
Factor analysis

Deep interactions

Array-valued data

$y_{i,j,k} =$

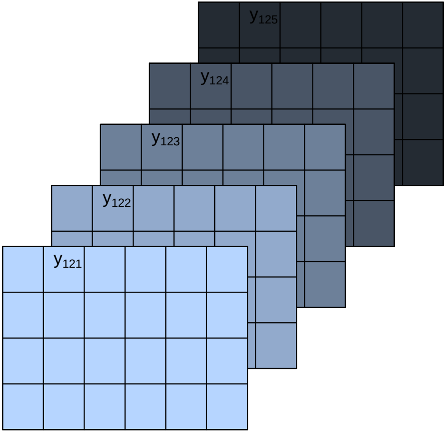
- jth measurement on ith subject under condition k (psychometrics)
- sample mean of variable i for group j in state k (cross-classified data)
- type-k relationship between i and j (multivariate relational data)
- time-k relationship between i and j (dynamic relational data)



Array-valued data

$y_{i,j,k} =$

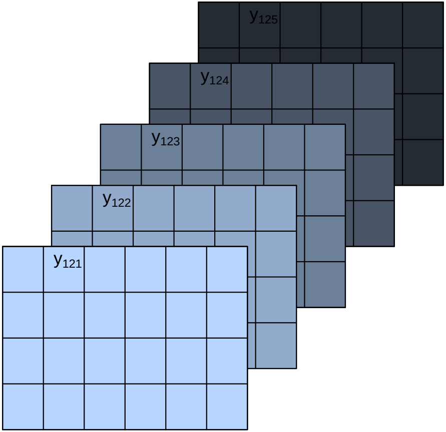
- jth measurement on ith subject under condition k (psychometrics)
- sample mean of variable i for group j in state k (cross-classified data)
- type-k relationship between i and j (multivariate relational data)
- time-k relationship between i and j (dynamic relational data)



Array-valued data

$y_{i,j,k} =$

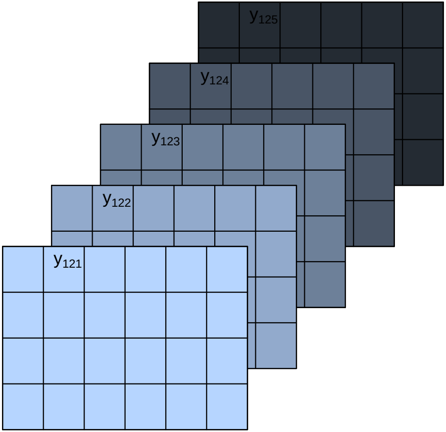
- jth measurement on ith subject under condition k (psychometrics)
- sample mean of variable i for group j in state k (cross-classified data)
- type-k relationship between i and j (multivariate relational data)
- time-k relationship between i and j (dynamic relational data)



Array-valued data

$y_{i,j,k} =$

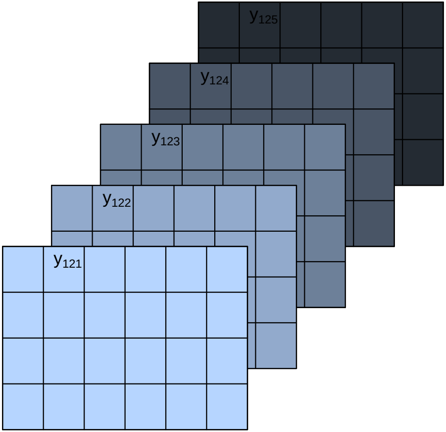
- jth measurement on ith subject under condition k (psychometrics)
- sample mean of variable i for group j in state k (cross-classified data)
- type-k relationship between i and j (multivariate relational data)
- time-k relationship between i and j (dynamic relational data)



Array-valued data

$y_{i,j,k} =$

- jth measurement on ith subject under condition k (psychometrics)
- sample mean of variable i for group j in state k (cross-classified data)
- type-k relationship between i and j (multivariate relational data)
- time-k relationship between i and j (dynamic relational data)



Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \boldsymbol{\Theta} + \mathbf{E}$$

$\boldsymbol{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\boldsymbol{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\boldsymbol{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the "main features" (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the "main features" of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Mean and variance structure

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{\Theta}$ describes the “main features” (the mean),

$\mathbf{E}$ describes deviations from main features (the residual).

Questions:

- How do we define and estimate the “main features” of an array?
- How can we summarize the residual variance?

$\mathbf{\Theta}$ can be defined and estimated using

- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from $\mathbf{\Theta}$?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \mathbf{\Theta} + \mathbf{E}_t$

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{\Theta} + \mathbf{E}_1 \\ &\vdots \\ \mathbf{Y}_{10} &= \mathbf{\Theta} + \mathbf{E}_{10} \end{aligned}$$

- $\mathbf{\Theta} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times30\times6\times10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \mathbf{\Theta} + \mathbf{E}_t$

$$
\begin{aligned}
\mathbf{Y}_1 &= \mathbf{\Theta} + \mathbf{E}_1 \\
&\vdots \\
\mathbf{Y}_{10} &= \mathbf{\Theta} + \mathbf{E}_{10}
\end{aligned}
$$

- $\mathbf{\Theta} \in \mathbb{R}^{30\times30\times6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times30\times6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \mathbf{\Theta} + \mathbf{E}_t$

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{\Theta} + \mathbf{E}_1 \\ &\vdots \\ \mathbf{Y}_{10} &= \mathbf{\Theta} + \mathbf{E}_{10} \end{aligned}$$

- $\mathbf{\Theta} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1, \ldots, \mathbf{E}_{10}\}$ be described?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \mathbf{\Theta} + \mathbf{E}_t$

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{\Theta} + \mathbf{E}_1 \\ &\vdots \\ \mathbf{Y}_{10} &= \mathbf{\Theta} + \mathbf{E}_{10} \end{aligned}$$

- $\mathbf{\Theta} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1, \ldots, \mathbf{E}_{10}\}$ be described?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \mathbf{\Theta} + \mathbf{E}_t$

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{\Theta} + \mathbf{E}_1 \\ &\vdots \\ \mathbf{Y}_{10} &= \mathbf{\Theta} + \mathbf{E}_{10} \end{aligned}$$

- $\mathbf{\Theta} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \mathbf{\Theta} + \mathbf{E}_t$

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{\Theta} + \mathbf{E}_1 \\ &\vdots \\ \mathbf{Y}_{10} &= \mathbf{\Theta} + \mathbf{E}_{10} \end{aligned}$$

- $\mathbf{\Theta} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \boldsymbol{\Theta} + \mathbf{E}_t$

$$\begin{aligned}\mathbf{Y}_1 &= \boldsymbol{\Theta} + \mathbf{E}_1 \\ &\vdots \\ \mathbf{Y}_{10} &= \boldsymbol{\Theta} + \mathbf{E}_{10}\end{aligned}$$

- $\boldsymbol{\Theta} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

Covariance structure of multivariate relational arrays

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

"Replications" over time: $\mathbf{Y} = \{\mathbf{Y}_1,\ldots,\mathbf{Y}_{10}\}$, $\mathbf{Y}_t = \mathbf{\Theta} + \mathbf{E}_t$

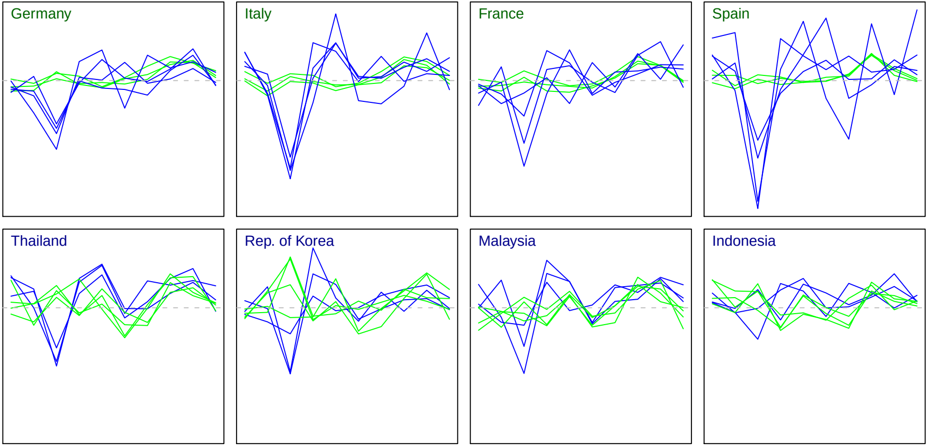
$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{\Theta} + \mathbf{E}_1 \\ \vdots & \qquad \vdots \\ \mathbf{Y}_{10} &= \mathbf{\Theta} + \mathbf{E}_{10} \end{aligned}$$

- $\mathbf{\Theta} \in \mathbb{R}^{30\times 30\times 6}$, constant over time;
- $\mathbf{E}_t \in \mathbb{R}^{30\times 30\times 6}$, changing over time.

How should the covariance among $\mathbf{E} = \{\mathbf{E}_1,\ldots,\mathbf{E}_{10}\}$ be described?

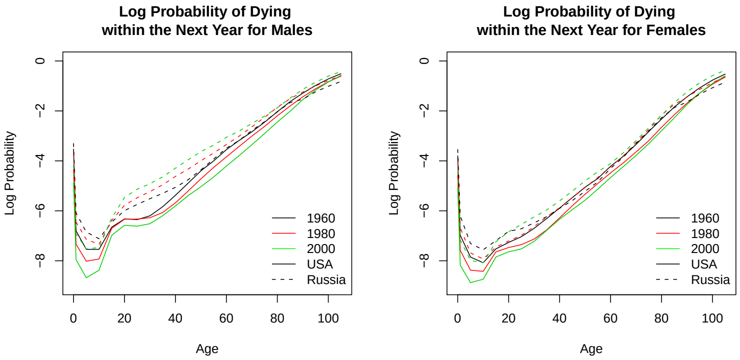
Longitudinal trade relations

Yearly change in log-trade averaged over commodity types



Mortality tables

(Joint work with Bailey Fosdick)



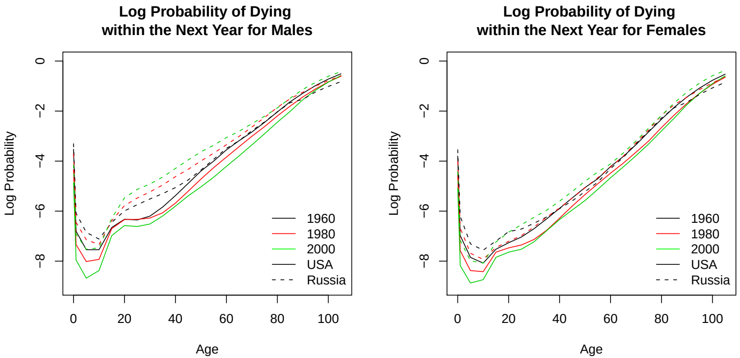
Human Mortality Database: (log) probability of dying in the next year

- 38 countries
- 23 age levels (0, 1 and then every 5 years)
- 9 times periods (1960 to 2000 every 5 years)
- 2 sexes

A $39 \times 23 \times 9 \times 2$-dimensional table.

Mortality tables

(Joint work with Bailey Fosdick)



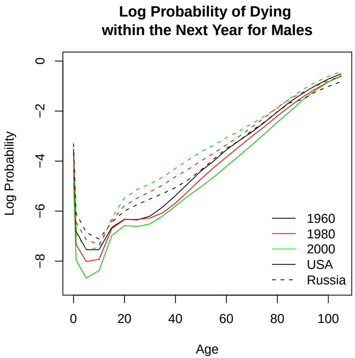
Human Mortality Database: (log) probability of dying in the next year

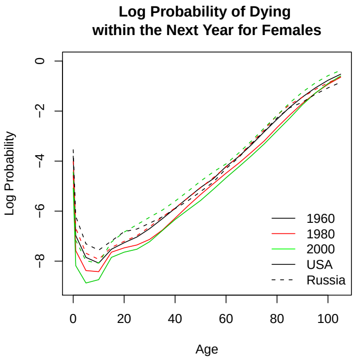
- 38 countries
- 23 age levels (0, 1 and then every 5 years)
- 9 times periods (1960 to 2000 every 5 years)
- 2 sexes

A $39 \times 23 \times 9 \times 2$-dimensional table.

Mortality tables

(Joint work with Bailey Fosdick)





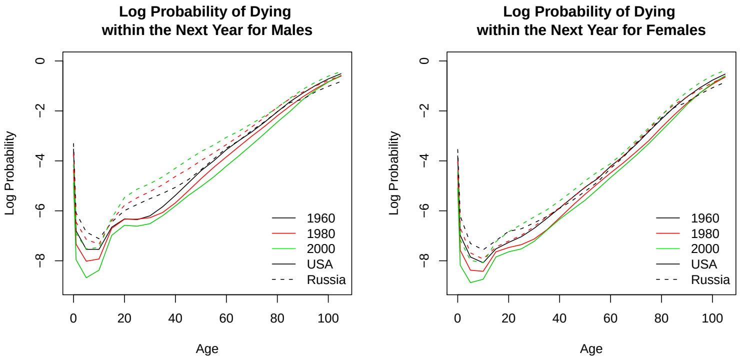
Human Mortality Database: (log) probability of dying in the next year

- 38 countries
- 23 age levels (0, 1 and then every 5 years)
- 9 times periods (1960 to 2000 every 5 years)
- 2 sexes

A $39 \times 23 \times 9 \times 2$-dimensional table.

Mortality tables

(Joint work with Bailey Fosdick)



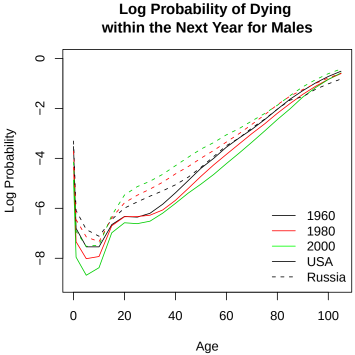
Human Mortality Database: (log) probability of dying in the next year

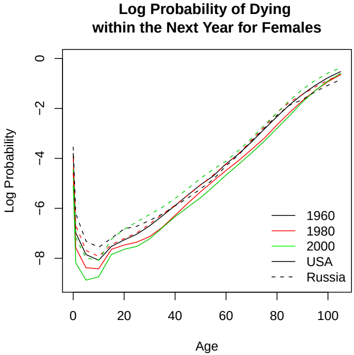
- 38 countries
- 23 age levels (0, 1 and then every 5 years)
- 9 times periods (1960 to 2000 every 5 years)
- 2 sexes

A $39 \times 23 \times 9 \times 2$-dimensional table.

Mortality tables

(Joint work with Bailey Fosdick)





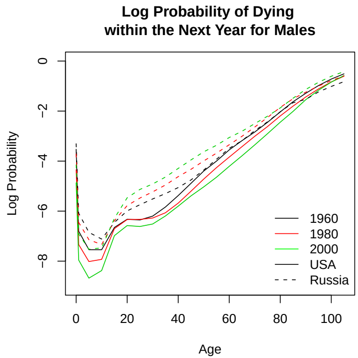
Human Mortality Database: (log) probability of dying in the next year

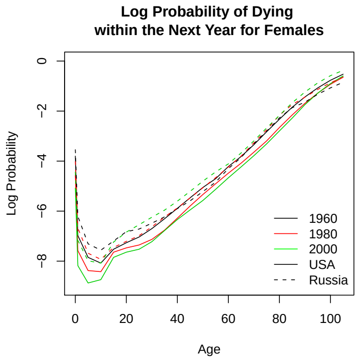
- 38 countries
- 23 age levels   (0, 1 and then every 5 years)
- 9 times periods (1960 to 2000 every 5 years)
- 2 sexes

A $39 \times 23 \times 9 \times 2$-dimensional table.

Mortality tables

(Joint work with Bailey Fosdick)





Human Mortality Database: (log) probability of dying in the next year

- 38 countries
- 23 age levels (0, 1 and then every 5 years)
- 9 times periods (1960 to 2000 every 5 years)
- 2 sexes

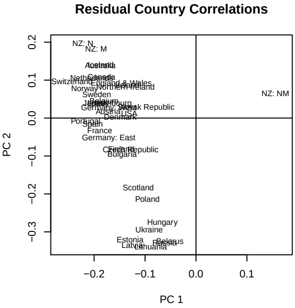
A $39 \times 23 \times 9 \times 2$-dimensional table.

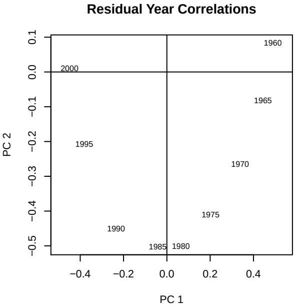
Mortality tables

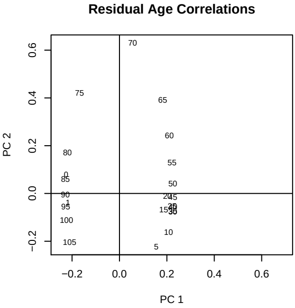
Preliminary model fitting:

$$y_{\text{age},i,j,k} = \sum_{r=0}^{4} (a_{i,r} + b_{j,r} + c_{k,r}) \times \text{age}^r + \epsilon_{\text{age},i,j,k}$$

Examine the residual array $\mathbf{E} \in \mathbb{R}^{38 \times 23 \times 9 \times 2}$ for dependence: $\mathbf{\Sigma}_k \approx \mathbf{E}_{(k)} \mathbf{E}_{(k)}^T$







Deep interaction priors

(Joint work with Alex Volfovsky)

Consider the usual three-factor "ANOVA decomposition" model:

$$\begin{aligned} y_{i,j,k,l} &= \mu_{j,k,l} + \epsilon_{i,j,k,l} \\ &= \mu + \left[a_j + b_k + c_l\right] + \left[(ab)_{j,k} + (ac)_{j,l} + (bc)_{k,l}\right] + \left[(abc)_{j,k,l}\right] + \epsilon_{i,j,k,l} \end{aligned}$$

Parameters are vectors, matrices and arrays based on three index sets.

Estimation methods:

- OLS estimation
- OLS with reduced model
- Bayes/penalized estimation

For the latter, how should priors on the parameters be specified?

Deep interaction priors

(Joint work with Alex Volfovsky)

Consider the usual three-factor "ANOVA decomposition" model:

$$\begin{aligned} y_{i,j,k,l} &= \mu_{j,k,l} + \epsilon_{i,j,k,l} \\ &= \mu + \left[a_j + b_k + c_l\right] + \left[(ab)_{j,k} + (ac)_{j,l} + (bc)_{k,l}\right] + \left[(abc)_{j,k,l}\right] + \epsilon_{i,j,k,l} \end{aligned}$$

Parameters are vectors, matrices and arrays based on three index sets.

Estimation methods:

- OLS estimation
- OLS with reduced model
- Bayes/penalized estimation

For the latter, how should priors on the parameters be specified?

Deep interaction priors

(Joint work with Alex Volfovsky)

Consider the usual three-factor "ANOVA decomposition" model:

$$
\begin{aligned}
y_{i,j,k,l} &= \mu_{j,k,l} + \epsilon_{i,j,k,l} \\
&= \mu + \Big[a_j + b_k + c_l\Big] + \Big[(ab)_{j,k} + (ac)_{j,l} + (bc)_{k,l}\Big] + \Big[(abc)_{j,k,l}\Big] + \epsilon_{i,j,k,l}
\end{aligned}
$$

Parameters are vectors, matrices and arrays based on three index sets.

Estimation methods:

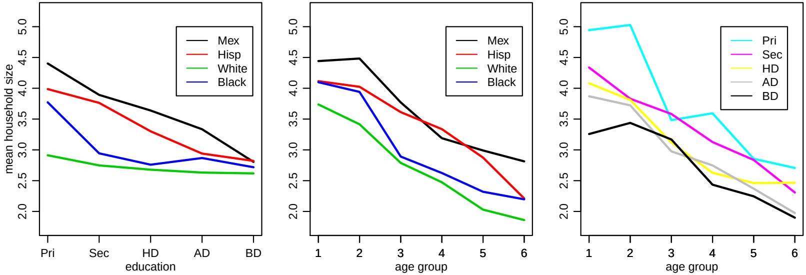
- OLS estimation
- OLS with reduced model
- Bayes/penalized estimation

For the latter, how should priors on the parameters be specified?

Deep interaction priors

NHANES 2007-08

- 4823 respondents
- asked about household size, education, ethnicity and age.
- sample size per edu×ethn×age category ranged between 1 and 214.



We see general similarities between certain levels of the factors.

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\mathrm{Cov}[\mathbf{E}] = \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2$$

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\mathrm{Cov}[\mathbf{E}] = \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2$$

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\mathrm{Cov}[\mathbf{E}] \quad = \quad \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2$$

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\begin{aligned} \mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \\ \mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \end{aligned}$$

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{E}\mathbf{E}^T] &= \mathbf{\Sigma}_1 \times \mathrm{tr}(\mathbf{\Sigma}_2)
\end{aligned}$$

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\begin{aligned} \mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \\ \mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\ \mathrm{E}[\mathbf{E}\mathbf{E}^T] &= \mathbf{\Sigma}_1 \times \mathrm{tr}(\mathbf{\Sigma}_2) \\ \mathrm{E}[\mathbf{E}^T\mathbf{E}] &= \mathbf{\Sigma}_2 \times \mathrm{tr}(\mathbf{\Sigma}_1) \end{aligned}$$

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{E}\mathbf{E}^T] &= \mathbf{\Sigma}_1 \times \mathrm{tr}(\mathbf{\Sigma}_2) \\
\mathrm{E}[\mathbf{E}^T\mathbf{E}] &= \mathbf{\Sigma}_2 \times \mathrm{tr}(\mathbf{\Sigma}_1) \\
\mathrm{Cov}[e_{i,j} e_{k,l}] &= \sigma_{1,i,k}\sigma_{2,j,l}
\end{aligned}$$

Separable covariance structure for matrices

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\}$$

Often the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance structure:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{E}\mathbf{E}^T] &= \mathbf{\Sigma}_1 \times \mathrm{tr}(\mathbf{\Sigma}_2) \\
\mathrm{E}[\mathbf{E}^T\mathbf{E}] &= \mathbf{\Sigma}_2 \times \mathrm{tr}(\mathbf{\Sigma}_1) \\
\mathrm{Cov}[e_{i,j} e_{k,l}] &= \sigma_{1,i,k}\sigma_{2,j,l}
\end{aligned}$$

This is the covariance structure of the "matrix normal" model (Dawid, 1981)

Generating the matrix normal class

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \quad \overset{\text{iid}}{\sim} \quad \text{normal}(0, 1)$$

Generating the matrix normal class

Multivariate normal model:

$$\begin{aligned} \mathbf{z} = \{z_j : j = 1, \ldots, m\} &\stackrel{\text{iid}}{\sim} \text{normal}(0,1) \\ \mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} &\sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T) \end{aligned}$$

Generating the matrix normal class

Multivariate normal model:

$$\begin{aligned} \mathbf{z} = \{z_j : j = 1, \ldots, m\} &\overset{\text{iid}}{\sim} \text{normal}(0,1) \\ \mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} &\sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T) \end{aligned}$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$

Generating the matrix normal class

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T)$$

Evaluating separability

Stein (2005) argues against separability for space-time data: smoothness at origin does not imply smoothness away from the origin.

Genton (2007) argues that separable approximations to nonseparable covariance can be useful for space-time data.

An evaluation of separability for the smooth space-time domain is possible and useful because separability can be tested and parsimonious alternatives exist:

- Separable : $\Sigma = \Sigma_2(\theta) \otimes \Sigma_1(\gamma)$
- Matérn class: Handcock and Stein (1993)
- Markov models

Evaluating separability

Stein (2005) argues against separability for space-time data: smoothness at origin does not imply smoothness away from the origin.

Genton (2007) argues that separable approximations to nonseparable covariance can be useful for space-time data.

An evaluation of separability for the smooth space-time domain is possible and useful because separability can be tested and parsimonious alternatives exist:

- Separable : $\Sigma = \Sigma_2(\theta) \otimes \Sigma_1(\gamma)$
- Matérn class: Handcock and Stein (1993)
- Markov models

Evaluating separability

Stein (2005) argues against separability for space-time data: smoothness at origin does not imply smoothness away from the origin.

Genton (2007) argues that separable approximations to nonseparable covariance can be useful for space-time data.

An evaluation of separability for the smooth space-time domain is possible and useful because separability can be tested and parsimonious alternatives exist:

- Separable : $\Sigma = \Sigma_2(\theta) \otimes \Sigma_1(\gamma)$
- Matérn class: Handcock and Stein (1993)
- Markov models

Non-smooth domains

Unordered index sets

- country
- ethnicity
- generic sets of variables

Limitations of separability: Separable = log additive

$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = \log \sigma_{1,i,j} + \log \sigma_{2,k,l}$$
$$= a_{i,j} + b_{k,l}$$

Alternatives to separability: Nonseparable = log additive + interactions?

$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k} + d_{i,j,l} + e_{i,k,l} + f_{j,k,l}$$
$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k,l}$$

Non-smooth domains

Unordered index sets

- country
- ethnicity
- generic sets of variables

Limitations of separability: Separable = log additive

$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = \log \sigma_{1,i,j} + \log \sigma_{2,k,l} = a_{i,j} + b_{k,l}$$

Alternatives to separability: Nonseparable = log additive + interactions?

$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k} + d_{i,j,l} + e_{i,k,l} + f_{j,k,l}$$
$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k,l}$$

Non-smooth domains

Unordered index sets

- country
- ethnicity
- generic sets of variables

Limitations of separability: Separable = log additive

$$\log \text{Cov}(y_{i,k}, y_{j,l}) = \log \sigma_{1,i,j} + \log \sigma_{2,k,l}$$
$$= a_{i,j} + b_{k,l}$$

Alternatives to separability: Nonseparable = log additive + interactions?

$$\log \text{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k} + d_{i,j,l} + e_{i,k,l} + f_{j,k,l}$$
$$\log \text{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k,l}$$

Non-smooth domains

Unordered index sets

- country
- ethnicity
- generic sets of variables

Limitations of separability: Separable = log additive

$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = \log \sigma_{1,i,j} + \log \sigma_{2,k,l} = a_{i,j} + b_{k,l}$$

Alternatives to separability: Nonseparable = log additive + interactions?

$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k} + d_{i,j,l} + e_{i,k,l} + f_{j,k,l}$$

$$\log \mathrm{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k,l}$$

Non-smooth domains

Unordered index sets

- country
- ethnicity
- generic sets of variables

Limitations of separability: Separable = log additive

$$\log \text{Cov}(y_{i,k}, y_{j,l}) = \log \sigma_{1,i,j} + \log \sigma_{2,k,l}$$
$$= a_{i,j} + b_{k,l}$$

Alternatives to separability: Nonseparable = log additive + interactions?

$$\log \text{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k} + d_{i,j,l} + e_{i,k,l} + f_{j,k,l}$$
$$\log \text{Cov}(y_{i,k}, y_{j,l}) = a_{i,j} + b_{k,l} + c_{i,j,k,l}$$

Non-smooth domains

Unordered index sets

- country
- ethnicity
- generic sets of variables

Limitations of separability: Separable = log additive

$$\begin{aligned}\log \mathrm{Cov}(y_{i,k}, y_{j,l}) &= \log \sigma_{1,i,j} + \log \sigma_{2,k,l} \\ &= a_{i,j} + b_{k,l}\end{aligned}$$

Alternatives to separability: Nonseparable = log additive + interactions?

$$\begin{aligned}\log \mathrm{Cov}(y_{i,k}, y_{j,l}) &= a_{i,j} + b_{k,l} + c_{i,j,k} + d_{i,j,l} + e_{i,k,l} + f_{j,k,l} \\ \log \mathrm{Cov}(y_{i,k}, y_{j,l}) &= a_{i,j} + b_{k,l} + c_{i,j,k,l}\end{aligned}$$

Data limitations

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \stackrel{\text{iid}}{\sim} \text{mnorm}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$,

$$-2 \log p(\mathbf{Y}_1, \ldots, \mathbf{Y}_n | \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) + c = \text{tr}(\sum_{i=1}^{n} \mathbf{\Sigma}_1{}^{-1} \mathbf{Y}_i \mathbf{\Sigma}_2{}^{-1} \mathbf{Y}_i^T) + nm_2 \ln |\mathbf{\Sigma}_1| + nm_1 \ln |\mathbf{\Sigma}_2|.$$

An MLE must satisfy

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{nm_2} \sum \mathbf{Y}_i \hat{\mathbf{\Sigma}}_2{}^{-1} \mathbf{Y}_i^T \text{ and } \hat{\mathbf{\Sigma}}_2 = \frac{1}{nm_1} \sum \mathbf{Y}_i^T \hat{\mathbf{\Sigma}}_1{}^{-1} \mathbf{Y}_i.$$

Consider the following block coordinate descent algorithm of Dutilleul (1999): Given $\hat{\mathbf{\Sigma}}_2{}^s$,

$$\begin{aligned} \hat{\mathbf{\Sigma}}_1{}^{s+1} &= \frac{1}{nm_2} \sum \mathbf{Y}_i (\hat{\mathbf{\Sigma}}_2{}^s)^{-1} \mathbf{Y}_i^T \\ \hat{\mathbf{\Sigma}}_2{}^{s+1} &= \frac{1}{nm_1} \sum \mathbf{Y}_i^T (\hat{\mathbf{\Sigma}}_1{}^{s+1})^{-1} \mathbf{Y}_i \end{aligned}$$

For what values of n will this converge to an MLE?

Data limitations

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \overset{\text{iid}}{\sim} \text{mnorm}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$,

$$-2\log p(\mathbf{Y}_1, \ldots, \mathbf{Y}_n | \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) + c = \text{tr}(\sum_{i=1}^{n} \mathbf{\Sigma}_1^{-1} \mathbf{Y}_i \mathbf{\Sigma}_2^{-1} \mathbf{Y}_i^T) + nm_2 \ln |\mathbf{\Sigma}_1| + nm_1 \ln |\mathbf{\Sigma}_2|.$$

An MLE must satisfy

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{nm_2} \sum \mathbf{Y}_i \hat{\mathbf{\Sigma}}_2^{-1} \mathbf{Y}_i^T \text{ and } \hat{\mathbf{\Sigma}}_2 = \frac{1}{nm_1} \sum \mathbf{Y}_i^T \hat{\mathbf{\Sigma}}_1^{-1} \mathbf{Y}_i.$$

Consider the following block coordinate descent algorithm of Dutilleul (1999): Given $\hat{\mathbf{\Sigma}}_2^s$,

$$\hat{\mathbf{\Sigma}}_1^{s+1} = \frac{1}{nm_2} \sum \mathbf{Y}_i (\hat{\mathbf{\Sigma}}_2^s)^{-1} \mathbf{Y}_i^T$$
$$\hat{\mathbf{\Sigma}}_2^{s+1} = \frac{1}{nm_1} \sum \mathbf{Y}_i^T (\hat{\mathbf{\Sigma}}_1^{s+1})^{-1} \mathbf{Y}_i$$

For what values of n will this converge to an MLE?

Data limitations

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \stackrel{\text{iid}}{\sim} \text{mnorm}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$,

$$-2 \log p(\mathbf{Y}_1, \ldots, \mathbf{Y}_n | \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) + c = \text{tr}(\sum_{i=1}^{n} \mathbf{\Sigma}_1^{-1} \mathbf{Y}_i \mathbf{\Sigma}_2^{-1} \mathbf{Y}_i^T) + nm_2 \ln |\mathbf{\Sigma}_1| + nm_1 \ln |\mathbf{\Sigma}_2|.$$

An MLE must satisfy

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{nm_2} \sum \mathbf{Y}_i \hat{\mathbf{\Sigma}}_2^{-1} \mathbf{Y}_i^T \quad \text{and} \quad \hat{\mathbf{\Sigma}}_2 = \frac{1}{nm_1} \sum \mathbf{Y}_i^T \hat{\mathbf{\Sigma}}_1^{-1} \mathbf{Y}_i.$$

Consider the following block coordinate descent algorithm of Dutilleul (1999): Given $\hat{\mathbf{\Sigma}}_2^{s}$,

$$\hat{\mathbf{\Sigma}}_1^{s+1} = \frac{1}{nm_2} \sum \mathbf{Y}_i (\hat{\mathbf{\Sigma}}_2^{s})^{-1} \mathbf{Y}_i^T$$

$$\hat{\mathbf{\Sigma}}_2^{s+1} = \frac{1}{nm_1} \sum \mathbf{Y}_i^T (\hat{\mathbf{\Sigma}}_1^{s+1})^{-1} \mathbf{Y}_i$$

For what values of n will this converge to an MLE?

Data limitations

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \overset{\text{iid}}{\sim} \text{mnorm}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$,

$$-2 \log p(\mathbf{Y}_1, \ldots, \mathbf{Y}_n | \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) + c = \text{tr}(\sum_{i=1}^{n} \mathbf{\Sigma}_1{}^{-1} \mathbf{Y}_i \mathbf{\Sigma}_2{}^{-1} \mathbf{Y}_i^T) + nm_2 \ln |\mathbf{\Sigma}_1| + nm_1 \ln |\mathbf{\Sigma}_2|.$$

An MLE must satisfy

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{nm_2} \sum \mathbf{Y}_i \hat{\mathbf{\Sigma}}_2{}^{-1} \mathbf{Y}_i^T \quad \text{and} \quad \hat{\mathbf{\Sigma}}_2 = \frac{1}{nm_1} \sum \mathbf{Y}_i^T \hat{\mathbf{\Sigma}}_1{}^{-1} \mathbf{Y}_i.$$

Consider the following block coordinate descent algorithm of Dutilleul (1999): Given $\hat{\mathbf{\Sigma}}_2{}^s$,

$$\begin{aligned} \hat{\mathbf{\Sigma}}_1{}^{s+1} &= \frac{1}{nm_2} \sum \mathbf{Y}_i (\hat{\mathbf{\Sigma}}_2{}^s)^{-1} \mathbf{Y}_i^T \\ \hat{\mathbf{\Sigma}}_2{}^{s+1} &= \frac{1}{nm_1} \sum \mathbf{Y}_i^T (\hat{\mathbf{\Sigma}}_1{}^{s+1})^{-1} \mathbf{Y}_i \end{aligned}$$

For what values of n will this converge to an MLE?

Data limitations

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \stackrel{\text{iid}}{\sim} \text{mnorm}(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$,

$$-2 \log p(\mathbf{Y}_1, \ldots, \mathbf{Y}_n | \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) + c = \text{tr}\left(\sum_{i=1}^{n} \boldsymbol{\Sigma}_1{}^{-1} \mathbf{Y}_i \boldsymbol{\Sigma}_2{}^{-1} \mathbf{Y}_i^T\right) + nm_2 \ln |\boldsymbol{\Sigma}_1| + nm_1 \ln |\boldsymbol{\Sigma}_2|.$$

An MLE must satisfy

$$\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{nm_2} \sum \mathbf{Y}_i \hat{\boldsymbol{\Sigma}}_2{}^{-1} \mathbf{Y}_i^T \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_2 = \frac{1}{nm_1} \sum \mathbf{Y}_i^T \hat{\boldsymbol{\Sigma}}_1{}^{-1} \mathbf{Y}_i.$$

Consider the following block coordinate descent algorithm of Dutilleul (1999):
Given $\hat{\boldsymbol{\Sigma}}_2{}^s$,

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_1{}^{s+1} &= \frac{1}{nm_2} \sum \mathbf{Y}_i (\hat{\boldsymbol{\Sigma}}_2{}^s)^{-1} \mathbf{Y}_i^T \\ \hat{\boldsymbol{\Sigma}}_2{}^{s+1} &= \frac{1}{nm_1} \sum \mathbf{Y}_i^T (\hat{\boldsymbol{\Sigma}}_1{}^{s+1})^{-1} \mathbf{Y}_i \end{aligned}$$

For what values of n will this converge to an MLE?

Data limitations

$\mathbf{n \geq m_1 m_2}$: An unconstrained MLE exists, so the separable likelihood is bounded and an MLE exists.

$\mathbf{n \geq \max\{m_1, m_2\}}$: Srivastiva et al. (2008) show the MLE exists and is essentially unique.

$\mathbf{n \geq \max\{m_1/m_2, m_2/m_1\}}$: Each step of the BCD yields a full-rank covariance matrix if $nm_2 \geq m_1$ and $nm_1 \geq m_2$.

- Dutilleul (1999) claims this is sufficient for the existence of an MLE.
- However, it is easy to show that the likelihood is generally unbounded.
- We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$.

Sadly, our sample size is generally $n = 1$. Estimation requires priors/penalties:

- Bayesian inference: Quintana and West (1988), Carvalho and West (2007)
- Penalized likelihood: Allen and Tibshirani (2011)

Data limitations

$\mathbf{n \geq m_1 m_2}$: An unconstrained MLE exists, so the separable likelihood is bounded and an MLE exists.

$\mathbf{n} \geq \max\{\mathbf{m_1}, \mathbf{m_2}\}$: Srivastiva et al. (2008) show the MLE exists and is essentially unique.

$\mathbf{n} \geq \max\{\mathbf{m_1}/\mathbf{m_2}, \mathbf{m_2}/\mathbf{m_1}\}$: Each step of the BCD yields a full-rank covariance matrix if $nm_2 \geq m_1$ and $nm_1 \geq m_2$.

- Dutilleul (1999) claims this is sufficient for the existence of an MLE.
- However, it is easy to show that the likelihood is generally unbounded.
- We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$.

Sadly, our sample size is generally $n = 1$. Estimation requires priors/penalties:

- Bayesian inference: Quintana and West (1988), Carvalho and West (2007)
- Penalized likelihood: Allen and Tibshirani (2011)

Data limitations

$\mathbf{n \geq m_1 m_2}$: An unconstrained MLE exists, so the separable likelihood is bounded and an MLE exists.

$\mathbf{n} \geq \max\{\mathbf{m_1}, \mathbf{m_2}\}$: Srivastiva et al. (2008) show the MLE exists and is essentially unique.

$\mathbf{n} \geq \max\{\mathbf{m_1/m_2}, \mathbf{m_2/m_1}\}$: Each step of the BCD yields a full-rank covariance matrix if $nm_2 \geq m_1$ and $nm_1 \geq m_2$.

- Dutilleul (1999) claims this is sufficient for the existence of an MLE.
- However, it is easy to show that the likelihood is generally unbounded.
- We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$.

Sadly, our sample size is generally $n = 1$. Estimation requires priors/penalties:

- Bayesian inference: Quintana and West (1988), Carvalho and West (2007)
- Penalized likelihood: Allen and Tibshirani (2011)

Data limitations

$\mathbf{n \geq m_1 m_2}$: An unconstrained MLE exists, so the separable likelihood is bounded and an MLE exists.

$\mathbf{n} \geq \max\{\mathbf{m_1}, \mathbf{m_2}\}$: Srivastiva et al. (2008) show the MLE exists and is essentially unique.

$\mathbf{n} \geq \max\{\mathbf{m_1}/\mathbf{m_2}, \mathbf{m_2}/\mathbf{m_1}\}$: Each step of the BCD yields a full-rank covariance matrix if $nm_2 \geq m_1$ and $nm_1 \geq m_2$.

- Dutilleul (1999) claims this is sufficient for the existence of an MLE.
- However, it is easy to show that the likelihood is generally unbounded.
- We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$.

Sadly, our sample size is generally $n = 1$. Estimation requires priors/penalties:

- Bayesian inference: Quintana and West (1988), Carvalho and West (2007)
- Penalized likelihood: Allen and Tibshirani (2011)

Data limitations

$\mathbf{n \geq m_1 m_2}$: An unconstrained MLE exists, so the separable likelihood is bounded and an MLE exists.

$\mathbf{n} \geq \max\{\mathbf{m_1, m_2}\}$: Srivastiva et al. (2008) show the MLE exists and is essentially unique.

$\mathbf{n} \geq \max\{\mathbf{m_1/m_2, m_2/m_1}\}$: Each step of the BCD yields a full-rank covariance matrix if $nm_2 \geq m_1$ and $nm_1 \geq m_2$.

- Dutilleul (1999) claims this is sufficient for the existence of an MLE.
- However, it is easy to show that the likelihood is generally unbounded.
- We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$.

Sadly, our sample size is generally $n = 1$. Estimation requires priors/penalties:

- Bayesian inference: Quintana and West (1988), Carvalho and West (2007)
- Penalized likelihood: Allen and Tibshirani (2011)

Data limitations

$\mathbf{n \geq m_1 m_2}$: An unconstrained MLE exists, so the separable likelihood is bounded and an MLE exists.

$\mathbf{n} \geq \max\{\mathbf{m_1}, \mathbf{m_2}\}$: Srivastiva et al. (2008) show the MLE exists and is essentially unique.

$\mathbf{n} \geq \max\{\mathbf{m_1/m_2}, \mathbf{m_2/m_1}\}$: Each step of the BCD yields a full-rank covariance matrix if $nm_2 \geq m_1$ and $nm_1 \geq m_2$.

- Dutilleul (1999) claims this is sufficient for the existence of an MLE.
- However, it is easy to show that the likelihood is generally unbounded.
- We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$.

Sadly, our sample size is generally $n = 1$. Estimation requires priors/penalties:

- Bayesian inference: Quintana and West (1988), Carvalho and West (2007)
- Penalized likelihood: Allen and Tibshirani (2011)

Data limitations

$\mathbf{n \geq m_1 m_2}$: An unconstrained MLE exists, so the separable likelihood is bounded and an MLE exists.

$\mathbf{n} \geq \max\{\mathbf{m_1}, \mathbf{m_2}\}$: Srivastiva et al. (2008) show the MLE exists and is essentially unique.

$\mathbf{n} \geq \max\{\mathbf{m_1}/\mathbf{m_2}, \mathbf{m_2}/\mathbf{m_1}\}$: Each step of the BCD yields a full-rank covariance matrix if $nm_2 \geq m_1$ and $nm_1 \geq m_2$.

- Dutilleul (1999) claims this is sufficient for the existence of an MLE.
- However, it is easy to show that the likelihood is generally unbounded.
- We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$.

Sadly, our sample size is generally $n = 1$. Estimation requires priors/penalties:

- Bayesian inference: Quintana and West (1988), Carvalho and West (2007)
- Penalized likelihood: Allen and Tibshirani (2011)

Separable covariance structure for arrays

$$\mathbf{Y} = \boldsymbol{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2) \times (m_3 \times m_3)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}]\}$$

A parsimonious alternative is an "array normal" model:

$$\mathrm{Cov}[\mathbf{E}] = \boldsymbol{\Sigma}_1 \circ \boldsymbol{\Sigma}_2 \circ \boldsymbol{\Sigma}_3$$

Separable covariance structure for arrays

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2) \times (m_3 \times m_3)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}]\}$$

A parsimonious alternative is an “array normal” model:

$$\mathrm{Cov}[\mathbf{E}] \quad = \quad \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3$$

Separable covariance structure for arrays

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2) \times (m_3 \times m_3)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}]\}$$

A parsimonious alternative is an "array normal" model:

$$\begin{aligned} \mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 \\ \mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \end{aligned}$$

Separable covariance structure for arrays

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2) \times (m_3 \times m_3)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}]\}$$

A parsimonious alternative is an "array normal" model:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{E}_{(k)}\mathbf{E}_{(k)}^T] &= \mathbf{\Sigma}_k \times \prod_{j \neq k} \mathrm{tr}(\mathbf{\Sigma}_j)
\end{aligned}$$

Separable covariance structure for arrays

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2) \times (m_3 \times m_3)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathsf{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}]\}$$

A parsimonious alternative is an "array normal" model:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{E}_{(k)}\mathbf{E}_{(k)}^T] &= \mathbf{\Sigma}_k \times \prod_{j \neq k} \mathrm{tr}(\mathbf{\Sigma}_j) \\
\log \mathsf{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}] &= \log \sigma_{1,j_1,j_2} + \log \sigma_{2,k_1,k_2} + \log \sigma_{3,l_1,l_2}
\end{aligned}$$

Separable covariance structure for arrays

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}$$

$\mathbf{E} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, so $\mathrm{Cov}[\mathbf{E}]$ is an $(m_1 \times m_1) \times (m_2 \times m_2) \times (m_3 \times m_3)$ array:

$$\mathrm{Cov}[\mathbf{E}] = \{\mathrm{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}]\}$$

A parsimonious alternative is an "array normal" model:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{E}] &= \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3 \\
\mathrm{Cov}[\mathrm{vec}(\mathbf{E})] &= \mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1 \\
\mathrm{E}[\mathbf{E}_{(k)}\mathbf{E}_{(k)}^T] &= \mathbf{\Sigma}_k \times \prod_{j \neq k} \mathrm{tr}(\mathbf{\Sigma}_j) \\
\log \mathrm{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}] &= \log \sigma_{1,j_1,j_2} + \log \sigma_{2,k_1,k_2} + \log \sigma_{3,l_1,l_2}
\end{aligned}$$

Generating separable covariance structure

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$

Generating separable covariance structure

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \stackrel{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Generating separable covariance structure

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \stackrel{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \stackrel{\text{iid}}{\sim} \text{normal}(0, 1)$$

Generating separable covariance structure

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T)$$

Generating separable covariance structure

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T)$$

Array normal model:

$$\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$

Generating separable covariance structure

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T)$$

Array normal model:

$$\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{\text{iid}}{\sim} \text{normal}(0,1)$$
$$\mathbf{Y} = \mathbf{M} + \left[\begin{array}{c} \mathbf{A}\ \mathbf{Z}\ \mathbf{B}^T \\ \mathbf{C} \end{array} \right] ?$$

Reduced-rank array decompositions

What are the main features of a data array **Y**?

Features are "main" if they lie in a low-rank subspace.

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E}\ , \quad \text{rank}(\mathbf{\Theta}) < \text{rank}(\mathbf{Y})$$

What does "rank" mean for tensors? How do you compute an SVD?

Reduced-rank array decompositions

What are the main features of a data array $\mathbf{Y}$?

Features are "main" if they lie in a low-rank subspace.

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E} \ , \ \ \text{rank}(\mathbf{\Theta}) < \text{rank}(\mathbf{Y})$$

What does "rank" mean for tensors? How do you compute an SVD?

Reduced-rank array decompositions

What are the main features of a data array $\mathbf{Y}$?

Features are “main” if they lie in a low-rank subspace.

$$\mathbf{Y} = \boldsymbol{\Theta} + \mathbf{E} \ , \quad \text{rank}(\boldsymbol{\Theta}) < \text{rank}(\mathbf{Y})$$

What does “rank” mean for tensors? How do you compute an SVD?

Reduced-rank array decompositions

CANDECOMP-PARAFAC (Carroll and Chang 1970, Harshman 1970):

$$\boldsymbol{\Theta} = \sum_{r=1}^{R} \lambda_r (\mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r) \qquad \theta_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}$$

HOSVD (Tucker 1964, De Lathauwer et al. 2000, Kolda 2006):

$$\boldsymbol{\Theta} = \mathbf{D} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} \quad \theta_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}$$

- **D** is the $R_1 \times R_2 \times R_3$ *core array*
- **U** , **V** , **W** are $m_1 \times R_1$, $m_2 \times R_2$, $m_3 \times R_3$ orthogonal matrices.
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

Reduced-rank array decompositions

CANDECOMP-PARAFAC (Carroll and Chang 1970, Harshman 1970):

$$\boldsymbol{\Theta} = \sum_{r=1}^{R} \lambda_r (\mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r) \qquad \theta_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}$$

HOSVD (Tucker 1964, De Lathauwer et al. 2000, Kolda 2006):

$$\boldsymbol{\Theta} = \mathbf{D} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} \quad \theta_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}$$

- **D** is the $R_1 \times R_2 \times R_3$ *core array*
- **U** , **V** , **W** are $m_1 \times R_1$, $m_2 \times R_2$, $m_3 \times R_3$ orthogonal matrices.
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

Reduced-rank array decompositions

CANDECOMP-PARAFAC (Carroll and Chang 1970, Harshman 1970):

$$\boldsymbol{\Theta} = \sum_{r=1}^{R} \lambda_r (\mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r) \quad \theta_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}$$

HOSVD (Tucker 1964, De Lathauwer et al. 2000, Kolda 2006):

$$\boldsymbol{\Theta} = \mathbf{D} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} \quad \theta_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}$$

- **D** is the $R_1 \times R_2 \times R_3$ *core array*
- **U** , **V** , **W** are $m_1 \times R_1$, $m_2 \times R_2$, $m_3 \times R_3$ orthogonal matrices.
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

Reduced-rank array decompositions

CANDECOMP-PARAFAC (Carroll and Chang 1970, Harshman 1970):

$$\boldsymbol{\Theta} = \sum_{r=1}^{R} \lambda_r (\mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r) \qquad \theta_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}$$

HOSVD (Tucker 1964, De Lathauwer et al. 2000, Kolda 2006):

$$\boldsymbol{\Theta} = \mathbf{D} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} \quad \theta_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}$$

- $\mathbf{D}$ is the $R_1 \times R_2 \times R_3$ *core array*
- $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ are $m_1 \times R_1$, $m_2 \times R_2$, $m_3 \times R_3$ orthogonal matrices.
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

Reduced-rank array decompositions

CANDECOMP-PARAFAC (Carroll and Chang 1970, Harshman 1970):

$$\boldsymbol{\Theta} = \sum_{r=1}^{R} \lambda_r (\mathbf{u}_r \circ \mathbf{v}_r \circ \mathbf{w}_r) \quad \theta_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}$$

HOSVD (Tucker 1964, De Lathauwer et al. 2000, Kolda 2006):

$$\boldsymbol{\Theta} = \mathbf{D} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} \quad \theta_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}$$

- **D** is the $R_1 \times R_2 \times R_3$ *core array*
- **U** , **V** , **W** are $m_1 \times R_1$, $m_2 \times R_2$, $m_3 \times R_3$ orthogonal matrices.
- "$\times$" is array-matrix multiplication (De Lathauwer et al., 2000)

The full rank multilinear Tucker product

$$y_{i,j,k} = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} z_{i',j',k'} a_{i',i} b_{j',j} c_{k',k}$$

$$\mathbf{Y} = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$

$$= \mathbf{Z} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$

Array-matrix multiplication: $\mathbf{Z} \times_1 \mathbf{A}$

1. Matricizize: $\mathbf{Z}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$
2. Multiply: $\mathbf{A}\mathbf{Z}_{(1)}$
3. Reform: $\mathbf{Z} \times_1 \mathbf{A} = \text{array}(\text{vec}(\mathbf{A}\mathbf{Z}_{(1)}), m_1, m_2, m_3)$

$$\mathbf{Z} \times_j (\mathbf{F} + \mathbf{G}) = \mathbf{Z} \times_j \mathbf{F} + \mathbf{Z} \times_j \mathbf{G}$$

$$(\mathbf{Z} \times_j \mathbf{F}) \times_k \mathbf{G} = (\mathbf{Z} \times_k \mathbf{G}) \times_j \mathbf{F} = \mathbf{Z} \times_j \mathbf{F} \times_k \mathbf{G}$$

$$(\mathbf{Z} \times_j \mathbf{F}) \times_j \mathbf{G} = \mathbf{Z} \times_j (\mathbf{G}\mathbf{F})$$

If $\mathbf{Y} = \mathbf{Z} \times \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$, then

$$\mathbf{Y}_{(k)} = \mathbf{A}_k \mathbf{Z}_{(k)} (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^T.$$

The full rank multilinear Tucker product

$$
\begin{aligned}
y_{i,j,k} &= \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} z_{i',j',k'} a_{i',i} b_{j',j} c_{k',k} \\
\mathbf{Y} &= \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \\
&= \mathbf{Z} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}
\end{aligned}
$$

Array-matrix multiplication: $\mathbf{Z} \times_1 \mathbf{A}$

1. Matricizize: $\mathbf{Z}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$
2. Multiply: $\mathbf{A}\mathbf{Z}_{(1)}$
3. Reform: $\mathbf{Z} \times_1 \mathbf{A} = \text{array}(\text{vec}(\mathbf{A}\mathbf{Z}_{(1)}), m_1, m_2, m_3)$

$$
\begin{aligned}
\mathbf{Z} \times_j (\mathbf{F} + \mathbf{G}) &= \mathbf{Z} \times_j \mathbf{F} + \mathbf{Z} \times_j \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_k \mathbf{G} &= (\mathbf{Z} \times_k \mathbf{G}) \times_j \mathbf{F} = \mathbf{Z} \times_j \mathbf{F} \times_k \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_j \mathbf{G} &= \mathbf{Z} \times_j (\mathbf{G}\mathbf{F})
\end{aligned}
$$

If $\mathbf{Y} = \mathbf{Z} \times \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$, then

$$
\mathbf{Y}_{(k)} = \mathbf{A}_k \mathbf{Z}_{(k)} (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^T.
$$

The full rank multilinear Tucker product

$$\begin{aligned} y_{i,j,k} &= \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} z_{i',j',k'} a_{i',i} b_{j',j} c_{k',k} \\ \mathbf{Y} &= \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \\ &= \mathbf{Z} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \end{aligned}$$

Array-matrix multiplication: $\mathbf{Z} \times_1 \mathbf{A}$

1. Matricizize: $\mathbf{Z}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$
2. Multiply: $\mathbf{A}\mathbf{Z}_{(1)}$
3. Reform: $\mathbf{Z} \times_1 \mathbf{A} = \text{array}(\text{vec}(\mathbf{A}\mathbf{Z}_{(1)}), m_1, m_2, m_3)$

$$\begin{aligned} \mathbf{Z} \times_j (\mathbf{F} + \mathbf{G}) &= \mathbf{Z} \times_j \mathbf{F} + \mathbf{Z} \times_j \mathbf{G} \\ (\mathbf{Z} \times_j \mathbf{F}) \times_k \mathbf{G} &= (\mathbf{Z} \times_k \mathbf{G}) \times_j \mathbf{F} = \mathbf{Z} \times_j \mathbf{F} \times_k \mathbf{G} \\ (\mathbf{Z} \times_j \mathbf{F}) \times_j \mathbf{G} &= \mathbf{Z} \times_j (\mathbf{G}\mathbf{F}) \end{aligned}$$

If $\mathbf{Y} = \mathbf{Z} \times \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$, then

$$\mathbf{Y}_{(k)} = \mathbf{A}_k \mathbf{Z}_{(k)} (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^T.$$

The full rank multilinear Tucker product

$$
\begin{aligned}
y_{i,j,k} &= \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} z_{i',j',k'}\, a_{i',i}\, b_{j',j}\, c_{k',k} \\
\mathbf{Y} &= \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \\
&= \mathbf{Z} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}
\end{aligned}
$$

Array-matrix multiplication: $\mathbf{Z} \times_1 \mathbf{A}$

1. Matricizize: $\mathbf{Z}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$
2. Multiply: $\mathbf{A}\mathbf{Z}_{(1)}$
3. Reform: $\mathbf{Z} \times_1 \mathbf{A} = \text{array}(\text{vec}(\mathbf{A}\mathbf{Z}_{(1)}), m_1, m_2, m_3)$

$$
\begin{aligned}
\mathbf{Z} \times_j (\mathbf{F} + \mathbf{G}) &= \mathbf{Z} \times_j \mathbf{F} + \mathbf{Z} \times_j \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_k \mathbf{G} &= (\mathbf{Z} \times_k \mathbf{G}) \times_j \mathbf{F} = \mathbf{Z} \times_j \mathbf{F} \times_k \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_j \mathbf{G} &= \mathbf{Z} \times_j (\mathbf{G}\mathbf{F})
\end{aligned}
$$

If $\mathbf{Y} = \mathbf{Z} \times \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$, then

$$
\mathbf{Y}_{(k)} = \mathbf{A}_k \mathbf{Z}_{(k)} (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^T.
$$

The full rank multilinear Tucker product

$$
\begin{aligned}
y_{i,j,k} &= \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} z_{i',j',k'} a_{i',i} b_{j',j} c_{k',k} \\
\mathbf{Y} &= \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \\
&= \mathbf{Z} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}
\end{aligned}
$$

Array-matrix multiplication: $\mathbf{Z} \times_1 \mathbf{A}$

1. Matricizize: $\mathbf{Z}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$
2. Multiply: $\mathbf{A}\mathbf{Z}_{(1)}$
3. Reform: $\mathbf{Z} \times_1 \mathbf{A} = \text{array}(\text{vec}(\mathbf{A}\mathbf{Z}_{(1)}), m_1, m_2, m_3)$

$$
\begin{aligned}
\mathbf{Z} \times_j (\mathbf{F} + \mathbf{G}) &= \mathbf{Z} \times_j \mathbf{F} + \mathbf{Z} \times_j \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_k \mathbf{G} &= (\mathbf{Z} \times_k \mathbf{G}) \times_j \mathbf{F} = \mathbf{Z} \times_j \mathbf{F} \times_k \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_j \mathbf{G} &= \mathbf{Z} \times_j (\mathbf{G}\mathbf{F})
\end{aligned}
$$

If $\mathbf{Y} = \mathbf{Z} \times \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$, then

$$
\mathbf{Y}_{(k)} = \mathbf{A}_k \mathbf{Z}_{(k)} (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^T.
$$

The full rank multilinear Tucker product

$$
\begin{aligned}
y_{i,j,k} &= \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\sum_{k=1}^{m_3} z_{i',j',k'} a_{i',i} b_{j',j} c_{k',k} \\
\mathbf{Y} &= \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \\
&= \mathbf{Z} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}
\end{aligned}
$$

Array-matrix multiplication: $\mathbf{Z} \times_1 \mathbf{A}$

1. Matricizize: $\mathbf{Z}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$
2. Multiply: $\mathbf{A}\mathbf{Z}_{(1)}$
3. Reform: $\mathbf{Z} \times_1 \mathbf{A} = \text{array}(\text{vec}(\mathbf{A}\mathbf{Z}_{(1)}), m_1, m_2, m_3)$

$$
\begin{aligned}
\mathbf{Z} \times_j (\mathbf{F} + \mathbf{G}) &= \mathbf{Z} \times_j \mathbf{F} + \mathbf{Z} \times_j \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_k \mathbf{G} &= (\mathbf{Z} \times_k \mathbf{G}) \times_j \mathbf{F} = \mathbf{Z} \times_j \mathbf{F} \times_k \mathbf{G} \\
(\mathbf{Z} \times_j \mathbf{F}) \times_j \mathbf{G} &= \mathbf{Z} \times_j (\mathbf{G}\mathbf{F})
\end{aligned}
$$

If $\mathbf{Y} = \mathbf{Z} \times \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$, then

$$
\mathbf{Y}_{(k)} = \mathbf{A}_k \mathbf{Z}_{(k)} (\mathbf{A}_K \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^T .
$$

Separable covariance via Tucker products

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \boldsymbol{\Sigma}_2 = \mathbf{B}\mathbf{B}^T)$$

NOTE: $\mathbf{A}\mathbf{Z}\mathbf{B}^T = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}\}$

Array normal model:

$$\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \sim \text{array normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \boldsymbol{\Sigma}_2 = \mathbf{B}\mathbf{B}^T, \boldsymbol{\Sigma}_3 = \mathbf{C}\mathbf{C}^T)$$

(Hoff, 2011)

Separable covariance via Tucker products

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \boldsymbol{\Sigma}_2 = \mathbf{B}\mathbf{B}^T)$$

NOTE: $\mathbf{A}\mathbf{Z}\mathbf{B}^T = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}\}$

Array normal model:

$$\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$
$$\mathbf{Y} = \mathbf{M} + \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \sim \text{array normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \boldsymbol{\Sigma}_2 = \mathbf{B}\mathbf{B}^T, \boldsymbol{\Sigma}_3 = \mathbf{C}\mathbf{C}^T)$$

(Hoff, 2011)

Separable covariance via Tucker products

Multivariate normal model:

$$\mathbf{z} = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T)$$

Matrix normal model:

$$\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$

$$\mathbf{Y} = \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T)$$

NOTE: $\mathbf{A}\mathbf{Z}\mathbf{B}^T = \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}\}$

Array normal model:

$$\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{\text{iid}}{\sim} \text{normal}(0, 1)$$

$$\mathbf{Y} = \mathbf{M} + \mathbf{Z} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \sim \text{array normal}(\mathbf{M}, \boldsymbol{\Sigma}_1 = \mathbf{A}\mathbf{A}^T, \boldsymbol{\Sigma}_2 = \mathbf{B}\mathbf{B}^T, \boldsymbol{\Sigma}_3 = \mathbf{C}\mathbf{C}^T)$$

(Hoff, 2011)

Estimation

Given $(\boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3)$,

$$\mathbf{E} = (\mathbf{Y} - \mathbf{M}) \times \{\mathbf{I}, \boldsymbol{\Sigma}_2{}^{-1/2}, \boldsymbol{\Sigma}_3{}^{-1/2}\} \quad \sim \quad \text{array normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2}, \mathbf{I}_{m_3})$$

$\boldsymbol{\Sigma}_1$ can be estimated from $\mathbf{E}_{(1)}\mathbf{E}_{(1)}^T$

- MLE via block coordinate descent (“flip-flop” algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler

Estimation

Given $(\boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3)$,

$$\begin{aligned} \mathbf{E} = (\mathbf{Y} - \mathbf{M}) \times \{\mathbf{I}, \boldsymbol{\Sigma}_2^{-1/2}, \boldsymbol{\Sigma}_3^{-1/2}\} &\sim \text{array normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2}, \mathbf{I}_{m_3}) \\ \mathbf{E}_{(1)} &\sim \text{matrix normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2 m_3}) \end{aligned}$$

$\boldsymbol{\Sigma}_1$ can be estimated from $\mathbf{E}_{(1)}\mathbf{E}_{(1)}^T$

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler

Estimation

Given $(\boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3)$,

$$\begin{aligned} \mathbf{E} = (\mathbf{Y} - \mathbf{M}) \times \{\mathbf{I}, \boldsymbol{\Sigma}_2^{-1/2}, \boldsymbol{\Sigma}_3^{-1/2}\} &\sim \text{array normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2}, \mathbf{I}_{m_3}) \\ \mathbf{E}_{(1)} &\sim \text{matrix normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2 m_3}) \end{aligned}$$

$\boldsymbol{\Sigma}_1$ can be estimated from $\mathbf{E}_{(1)}\mathbf{E}_{(1)}^T$

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler

Estimation

Given $(\mathbf{\Sigma}_2, \mathbf{\Sigma}_3)$,

$$\mathbf{E} = (\mathbf{Y} - \mathbf{M}) \times \{\mathbf{I}, \mathbf{\Sigma}_2{}^{-1/2}, \mathbf{\Sigma}_3{}^{-1/2}\} \sim \text{array normal}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{I}_{m_2}, \mathbf{I}_{m_3})$$
$$\mathbf{E}_{(1)} \sim \text{matrix normal}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{I}_{m_2 m_3})$$

$\mathbf{\Sigma}_1$ can be estimated from $\mathbf{E}_{(1)}\mathbf{E}_{(1)}^T$

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler

Estimation

Given $(\boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3)$,

$$\begin{aligned} \mathbf{E} = (\mathbf{Y} - \mathbf{M}) \times \{\mathbf{I}, \boldsymbol{\Sigma}_2^{-1/2}, \boldsymbol{\Sigma}_3^{-1/2}\} &\sim \text{array normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2}, \mathbf{I}_{m_3}) \\ \mathbf{E}_{(1)} &\sim \text{matrix normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2 m_3}) \end{aligned}$$

$\boldsymbol{\Sigma}_1$ can be estimated from $\mathbf{E}_{(1)}\mathbf{E}_{(1)}^T$

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler

Estimation

Given $(\boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3)$,

$$\begin{aligned} \mathbf{E} = (\mathbf{Y} - \mathbf{M}) \times \{\mathbf{I}, \boldsymbol{\Sigma}_2^{-1/2}, \boldsymbol{\Sigma}_3^{-1/2}\} &\sim \text{array normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2}, \mathbf{I}_{m_3}) \\ \mathbf{E}_{(1)} &\sim \text{matrix normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \mathbf{I}_{m_2 m_3}) \end{aligned}$$

$\boldsymbol{\Sigma}_1$ can be estimated from $\mathbf{E}_{(1)}\mathbf{E}_{(1)}^T$

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim$ array normal$(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1, \dots, 30\}$ indexes exporting nation
- $j \in \{1, \dots, 30\}$ indexes importing nation
- $k \in \{1, \dots, 6\}$ indexes commodity
- $l \in \{1, \dots, 10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \boldsymbol{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim$ array normal$(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4)$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1,\ldots,30\}$ indexes exporting nation
- $j \in \{1,\ldots,30\}$ indexes importing nation
- $k \in \{1,\ldots,6\}$ indexes commodity
- $l \in \{1,\ldots,10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim$ array normal$(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times30\times6\times10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 6\times 10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim$ array normal$(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim$ array normal$(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$

International trade example

Yearly change in log exports (2000 dollars) : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30 \times 30 \times 6 \times 10}$

- $i \in \{1, \ldots, 30\}$ indexes exporting nation
- $j \in \{1, \ldots, 30\}$ indexes importing nation
- $k \in \{1, \ldots, 6\}$ indexes commodity
- $l \in \{1, \ldots, 10\}$ indexes year

Full "cell means" model:

$$y_{i,j,k,l} = \mu_{i,j,k} + e_{i,j,k,l}$$

Let $\mathbf{E} = \{e_{i,j,k,l}\}$

- iid error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{I}, \sigma^2\mathbf{I})$
- vector normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$

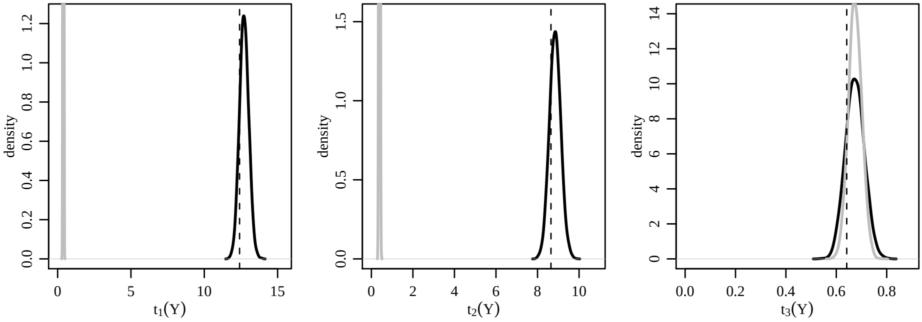
Posterior predictive comparisons

Compare $t(\mathbf{Y}_{\text{obs}})$ to $t(\mathbf{Y}_{\text{pred}})$, where $\mathbf{Y}_{\text{pred}} \sim p(\mathbf{Y}|\mathbf{Y}_{\text{obs}})$

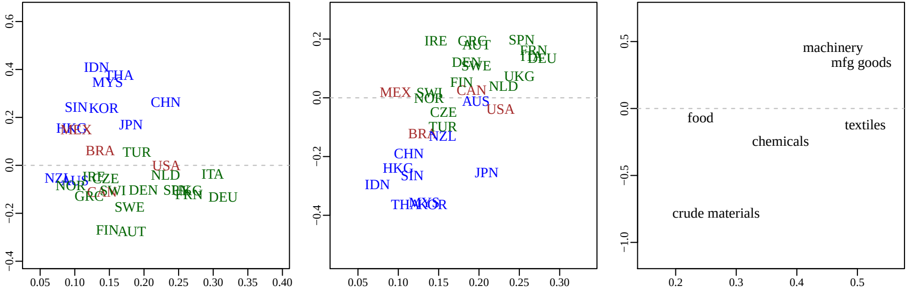
Models:

reduced: array normal$(0, \mathbf{I}, \mathbf{I}, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4)$

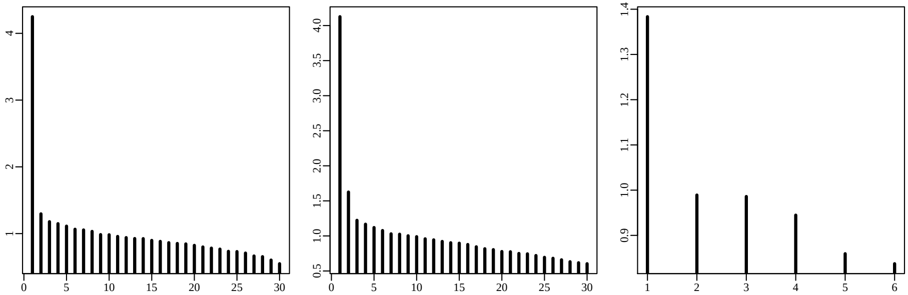
full: array normal$(0, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \boldsymbol{\Sigma}_4)$



International trade example



International trade example



Factor analysis

Vector normal factor model:

$$\begin{aligned} \mathrm{Cov}[\mathbf{y}] &= \mathbf{A}\mathbf{A}^T + \mathbf{D} \\ \mathbf{y} &\stackrel{d}{=} \mathbf{A}\mathbf{z} + \mathbf{D}^{1/2}\mathbf{e} \end{aligned}$$

where $\mathbf{A} \in \mathbb{R}^{p\times r}$ and $\mathbf{D}$ is diagonal.

Factor analysis is an alternative to likelihood penalties/priors:

An MLE of $\mathbf{A}$ and $\mathbf{D}$ exists if $n \geq r$ (Robertson and Symons, 2007)

Array normal model:

$$\begin{aligned} \mathrm{Cov}[\mathbf{Y}] &= (\mathbf{A}_1\mathbf{A}_1^T + \mathbf{D}_1) \circ \cdots \circ (\mathbf{A}_K\mathbf{A}_K^T + \mathbf{D}_K) \\ (\tilde{\mathbf{Y}}_{(1)})_i &\stackrel{d}{=} \mathbf{A}_1\mathbf{z} + \mathbf{D}_1^{1/2}\mathbf{e} \end{aligned}$$

Similarly, a FA MLE exists where the unrestricted MLE does not.

Factor analysis

Vector normal factor model:

$$\begin{aligned} \mathrm{Cov}[\mathbf{y}] &= \mathbf{A}\mathbf{A}^T + \mathbf{D} \\ \mathbf{y} &\stackrel{d}{=} \mathbf{A}\mathbf{z} + \mathbf{D}^{1/2}\mathbf{e} \end{aligned}$$

where $\mathbf{A} \in \mathbb{R}^{p \times r}$ and $\mathbf{D}$ is diagonal.

Factor analysis is an alternative to likelihood penalties/priors:

An MLE of **A** and **D** exists if $n \geq r$ (Robertson and Symons, 2007)

Array normal model:

$$\begin{aligned} \mathrm{Cov}[\mathbf{Y}] &= (\mathbf{A}_1\mathbf{A}_1^T + \mathbf{D}_1) \circ \cdots \circ (\mathbf{A}_K\mathbf{A}_K^T + \mathbf{D}_K) \\ (\tilde{\mathbf{Y}}_{(1)})_i &\stackrel{d}{=} \mathbf{A}_1\mathbf{z} + \mathbf{D}_1^{1/2}\mathbf{e} \end{aligned}$$

Similarly, a FA MLE exists where the unrestricted MLE does not.

Factor analysis

Vector normal factor model:

$$\begin{aligned} \mathrm{Cov}[\mathbf{y}] &= \mathbf{A}\mathbf{A}^T + \mathbf{D} \\ \mathbf{y} &\stackrel{d}{=} \mathbf{A}\mathbf{z} + \mathbf{D}^{1/2}\mathbf{e} \end{aligned}$$

where $\mathbf{A} \in \mathbb{R}^{p\times r}$ and $\mathbf{D}$ is diagonal.

Factor analysis is an alternative to likelihood penalties/priors:

An MLE of $\mathbf{A}$ and $\mathbf{D}$ exists if $n \geq r$ (Robertson and Symons, 2007)

Array normal model:

$$\begin{aligned} \mathrm{Cov}[\mathbf{Y}] &= (\mathbf{A}_1\mathbf{A}_1^T + \mathbf{D}_1) \circ \cdots \circ (\mathbf{A}_K\mathbf{A}_K^T + \mathbf{D}_K) \\ (\tilde{\mathbf{Y}}_{(1)})_i &\stackrel{d}{=} \mathbf{A}_1\mathbf{z} + \mathbf{D}_1^{1/2}\mathbf{e} \end{aligned}$$

Similarly, a FA MLE exists where the unrestricted MLE does not.

Mortality tables

Mean model:

$$y_{\text{age},i,j,k} = \sum_{r=0}^{4} (a_{i,r} + b_{j,r} + c_{k,r}) \times \text{age}^r + \epsilon_{\text{age},i,j,k}$$

Variance model:

$$\mathbf{E} = \{\epsilon_{\text{age},i,j,k}\} \sim \text{anorm}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4)$$
$$\mathbf{\Sigma}_k = \mathbf{A}_k \mathbf{A}_k^T + \mathbf{D}_k$$

Mortality tables

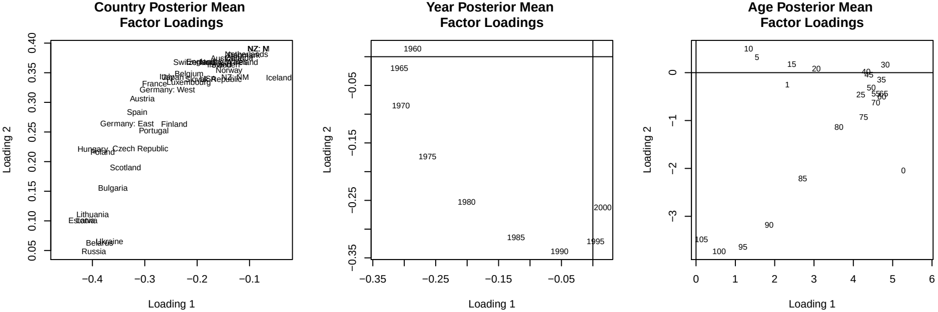
Mean model:

$$y_{\text{age},i,j,k} \quad = \quad \sum_{r=0}^{4} (a_{i,r} + b_{j,r} + c_{k,r}) \times \text{age}^r + \epsilon_{\text{age},i,j,k}$$

Variance model:

$$\begin{aligned} \mathbf{E} &= \{\epsilon_{\text{age},i,j,k}\} \sim \text{anorm}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{\Sigma}_4) \\ \mathbf{\Sigma}_k &= \mathbf{A}_k \mathbf{A}_k^T + \mathbf{D}_k \end{aligned}$$

Mortality tables



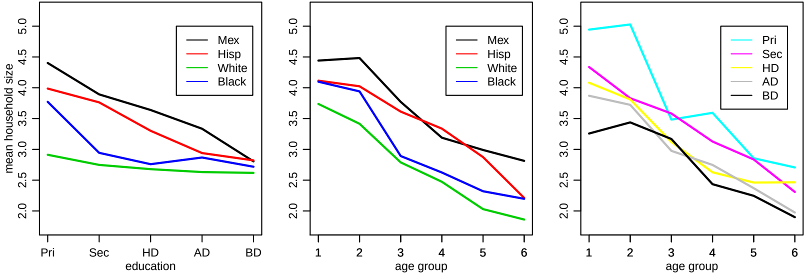
Predictive performance experiment: Predict 5% missing data

	IID	FA
mean(SSE)	273.28	3.27
sd(SSE)	20.34	0.51

Deep interaction priors

Consider the usual three-factor ANOVA decomposition model:

$$
\begin{aligned}
y_{i,j,k,l} &= \mu_{j,k,l} + \epsilon_{i,j,k,l} \\
&= \mu + \left[a_j + b_k + c_l\right] + \left[(ab)_{j,k} + (ac)_{j,l} + (bc)_{k,l}\right] + \left[(abc)_{j,k,l}\right] + \epsilon_{i,j,k,l}
\end{aligned}
$$



Array normal priors for deep interactions

main effect vectors:

$$\mathbf{a} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \boldsymbol{\Sigma}_a) \quad , \quad \mathbf{b} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \boldsymbol{\Sigma}_b) \quad , \quad \mathbf{c} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \boldsymbol{\Sigma}_c)$$

two-way interaction matrices

$$(\mathbf{ab}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b) \ , (\mathbf{ac}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_c) \ , (\mathbf{bc}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_c)$$

three-way interaction array

$$(\mathbf{abc}) \sim \text{anorm}(\mathbf{0}, \gamma_3 \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_c)$$

Array normal priors for deep interactions

main effect vectors:

$$\mathbf{a} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \mathbf{\Sigma}_a) \quad , \quad \mathbf{b} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \mathbf{\Sigma}_b) \quad , \quad \mathbf{c} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \mathbf{\Sigma}_c)$$

two-way interaction matrices

$$(\mathbf{ab}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \mathbf{\Sigma}_a, \mathbf{\Sigma}_b) \, , (\mathbf{ac}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \mathbf{\Sigma}_a, \mathbf{\Sigma}_c) \, , (\mathbf{bc}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \mathbf{\Sigma}_b, \mathbf{\Sigma}_c)$$

three-way interaction array

$$(\mathbf{abc}) \sim \text{anorm}(\mathbf{0}, \gamma_3 \mathbf{\Sigma}_a, \mathbf{\Sigma}_b, \mathbf{\Sigma}_c)$$

Array normal priors for deep interactions

main effect vectors:

$$\mathbf{a} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \boldsymbol{\Sigma}_a) \;\;,\;\; \mathbf{b} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \boldsymbol{\Sigma}_b) \;\;,\;\; \mathbf{c} \sim \text{vnorm}(\mathbf{0}, \gamma_1 \boldsymbol{\Sigma}_c)$$

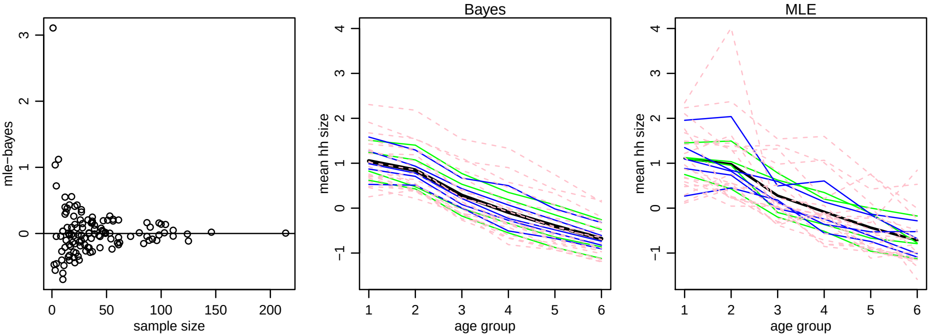
two-way interaction matrices

$$(\mathbf{ab}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b)\,,(\mathbf{ac}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_c)\,,(\mathbf{bc}) \sim \text{mnorm}(\mathbf{0}, \gamma_2 \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_c)$$

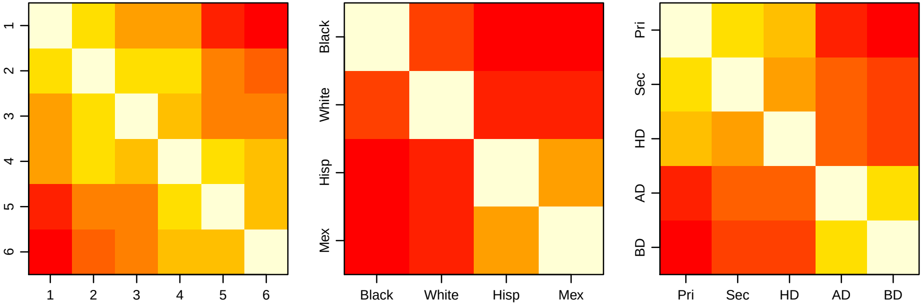
three-way interaction array

$$(\mathbf{abc}) \sim \text{anorm}(\mathbf{0}, \gamma_3 \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_c)$$

Regularization



Posterior covariance estimates



Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models

Discussion

- Data and model parameters are often in the form of a multiway array.
- Array modeling
  - Mean-modeling is reasonably well studied (ANOVA, reduced rank)
  - covariance modeling less so.
- Separable covariance models can be
  - restrictive (not a full covariance structure)
  - complex (not that parsimonious)
  - hopefully useful.
- Many interesting theoretical and methodological problems remain
  - existence and uniqueness of MLEs
  - dimension reduction and sparse solutions
  - alternatives to separable models