Statistical 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

$y_{i,j,k} =$
- $j$th measurement on $i$th 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

\[ y_{i,j,k} = \]

- \( j \)th measurement on \( i \)th 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)
$y_{i,j,k} =$

- $j$th measurement on $i$th 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

\[ y_{i,j,k} = \]

- \( j \)th measurement on \( i \)th 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

$y_{i,j,k} =$

- $j$th measurement on $i$th 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

\[ Y = \Theta + E \]

\(\Theta\) describes the “main features” (the mean),
\(E\) 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?

\(\Theta\) can be defined and estimated using

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

Can we compactly summarize deviations from \(\Theta\)?
Mean and variance structure

\[ Y = \Theta + E \]

\( \Theta \) describes the “main features” (the mean),
\( E \) 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?

\( \Theta \) can be defined and estimated using

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

Can we compactly summarize deviations from \( \Theta \)?
Mean and variance structure

\[ \mathbf{Y} = \Theta + \mathbf{E} \]

\( \Theta \) 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?

\( \Theta \) can be defined and estimated using
- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from \( \Theta \)?
Mean and variance structure

\[ Y = \Theta + E \]

\( \Theta \) describes the “main features” (the mean),

\( E \) 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?

\( \Theta \) can be defined and estimated using

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

Can we compactly summarize deviations from \( \Theta \)?
Mean and variance structure

\[ \mathbf{Y} = \Theta + \mathbf{E} \]

\( \Theta \) 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?

\( \Theta \) can be defined and estimated using

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

Can we compactly summarize deviations from \( \Theta \)?
Mean and variance structure

\[ Y = \Theta + E \]

\( \Theta \) describes the “main features” (the mean),
\( E \) 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?

\( \Theta \) can be defined and estimated using
- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from \( \Theta \)?
Mean and variance structure

\[ Y = \Theta + E \]

\[ \Theta \] describes the “main features” (the mean),
\[ E \] 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?

\[ \Theta \] can be defined and estimated using
  - sample means, given replications,
  - regression models,
  - reduced rank array representations.

Can we compactly summarize deviations from \[ \Theta \]?
Mean and variance structure

\[ Y = \Theta + E \]

\( \Theta \) describes the “main features” (the mean),

\( E \) 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?

\( \Theta \) can be defined and estimated using

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

Can we compactly summarize deviations from \( \Theta \)?
Mean and variance structure

\[ Y = \Theta + E \]

\( \Theta \) describes the “main features” (the mean),
\( E \) 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?

\( \Theta \) can be defined and estimated using
- sample means, given replications,
- regression models,
- reduced rank array representations.

Can we compactly summarize deviations from \( \Theta \)?
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: $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_{10}\}, \quad \mathbf{Y}_t = \Theta + \mathbf{E}_t$

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

- $\Theta \in \mathbb{R}^{30 \times 30 \times 6}$, 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) : \( Y = \{ y_{i,j,k,l} \} \in \mathbb{R}^{30 \times 30 \times 6 \times 10} \)

- \( i \in \{1, \ldots, 30\} \) 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: \( Y = \{ Y_1, \ldots, Y_{10} \} \), \( Y_t = \Theta + E_t \)

\[
\begin{align*}
Y_1 &= \Theta + E_1 \\
\vdots & \quad \vdots \\
Y_{10} &= \Theta + E_{10}
\end{align*}
\]

- \( \Theta \in \mathbb{R}^{30 \times 30 \times 6} \), constant over time;
- \( E_t \in \mathbb{R}^{30 \times 30 \times 6} \), changing over time.

How should the covariance among \( E = \{ E_1, \ldots, E_{10} \} \) be described?
Covariance structure of multivariate relational arrays

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

- \( i \in \{1, \ldots, 30\} \) 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: \( Y = \{Y_1, \ldots, Y_{10}\} \), \( Y_t = \Theta + E_t \)

\[
\begin{align*}
Y_1 &= \Theta + E_1 \\
\vdots & \quad \vdots \\
Y_{10} &= \Theta + E_{10}
\end{align*}
\]

- \( \Theta \in \mathbb{R}^{30 \times 30 \times 6} \), constant over time;
- \( E_t \in \mathbb{R}^{30 \times 30 \times 6} \), changing over time.

How should the covariance among \( E = \{E_1, \ldots, E_{10}\} \) be described?
Covariance structure of multivariate relational arrays

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

- $i \in \{1, \ldots, 30\}$ 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: $Y = \{Y_1, \ldots, Y_{10}\} \quad Y_t = \Theta + E_t$

- $Y_1 = \Theta + E_1$
- $\vdots$
- $Y_{10} = \Theta + E_{10}$

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

How should the covariance among $E = \{E_1, \ldots, E_{10}\}$ be described?
Covariance structure of multivariate relational arrays

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

- \( i \in \{1, \ldots, 30\} \) 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: \( Y = \{ Y_1, \ldots, Y_{10} \} \), \( Y_t = \Theta + E_t \)

\[
Y_1 = \Theta + E_1 \\
\vdots \\
Y_{10} = \Theta + E_{10}
\]

- \( \Theta \in \mathbb{R}^{30 \times 30 \times 6} \), constant over time;
- \( E_t \in \mathbb{R}^{30 \times 30 \times 6} \), changing over time.

How should the covariance among \( E = \{ E_1, \ldots, E_{10} \} \) be described?
Covariance structure of multivariate relational arrays

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

- \( i \in \{1, \ldots, 30\} \) 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: \( Y = \{Y_1, \ldots, Y_{10}\} \), \( Y_t = \Theta + E_t \)

\[
\begin{align*}
Y_1 &= \Theta + E_1 \\
\vdots & \quad \vdots \\
Y_{10} &= \Theta + E_{10}
\end{align*}
\]

- \( \Theta \in \mathbb{R}^{30 \times 30 \times 6} \), constant over time;
- \( E_t \in \mathbb{R}^{30 \times 30 \times 6} \), changing over time.

How should the covariance among \( E = \{E_1, \ldots, E_{10}\} \) be described?
Covariance structure of multivariate relational arrays

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

- \( i \in \{1, \ldots, 30\} \) 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: \( Y = \{Y_1, \ldots, Y_{10}\} , \ Y_t = \Theta + E_t \)

\[
\begin{align*}
Y_1 &= \Theta + E_1 \\
\vdots & \quad \vdots \\
Y_{10} &= \Theta + E_{10}
\end{align*}
\]

- \( \Theta \in \mathbb{R}^{30 \times 30 \times 6} \), constant over time;
- \( E_t \in \mathbb{R}^{30 \times 30 \times 6} \), changing over time.

How should the covariance among \( E = \{E_1, \ldots, E_{10}\} \) be described?
Covariance structure of multivariate relational arrays

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

- \( i \in \{1, \ldots, 30\} \) 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: \( Y = \{Y_1, \ldots, Y_{10}\} ,\ Y_t = \Theta + E_t \)

\[
\begin{align*}
Y_1 & = \Theta + E_1 \\
\vdots & \quad \vdots \\
Y_{10} & = \Theta + E_{10}
\end{align*}
\]

- \( \Theta \in \mathbb{R}^{30 \times 30 \times 6} \), constant over time;
- \( E_t \in \mathbb{R}^{30 \times 30 \times 6} \), changing over time.

How should the covariance among \( E = \{E_1, \ldots, E_{10}\} \) be described?
Longitudinal trade relations

Yearly change in log-trade averaged over commodity types

Germany  
Italy  
France  
Spain  
Thailand  
Rep. of Korea  
Malaysia  
Indonesia
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)

Log Probability of Dying within the Next Year for Males

Log Probability of Dying within the Next Year for Females

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.
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.
Examples of multiway data Separable covariance arrays Trade example Factor analysis Deep interactions Discussion

Mortality tables

Preliminary model fitting:

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

Examine the residual array \( E \in \mathbb{R}^{38 \times 23 \times 9 \times 2} \) for dependence: \( \Sigma_k \approx E(k)E_T(k) \)
Deep interaction priors

(Joint work with Alex Volfovsky)

Consider the usual three-factor “ANOVA decomposition” model:

\[ y_{i,j,k,l} = \mu_{j,k,l} + \epsilon_{i,j,k,l} \]

\[ = \mu + [a_j + b_k + c_l] + [(ab)_{j,k} + (ac)_{j,l} + (bc)_{k,l}] + [(abc)_{j,k,l}] + \epsilon_{i,j,k,l} \]

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:

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

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:

\[ y_{i,j,k,l} = \mu_{j,k,l} + \epsilon_{i,j,k,l} \]

\[ = \mu + [a_j + b_k + c_l] + [(ab)_{j,k} + (ac)_{j,l} + (bc)_{k,l}] + [(abc)_{j,k,l}] + \epsilon_{i,j,k,l} \]

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

\[ \mathbf{Y} = \Theta + \mathbf{E} \]

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

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

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance model:

\[ \text{Cov}[\mathbf{E}] = \Sigma_1 \circ \Sigma_2 \]
Separable covariance structure for matrices

\[ Y = \Theta + E \]

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

\[ \text{Cov}[E] = \{\text{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\} \]

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance model:

\[ \text{Cov}[E] = \Sigma_1 \circ \Sigma_2 \]
Separable covariance structure for matrices

\[ \mathbf{Y} = \Theta + \mathbf{E} \]

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

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

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a **separable covariance model**:

\[ \text{Cov}[\mathbf{E}] = \Sigma_1 \odot \Sigma_2 \]
**Separable covariance structure for matrices**

\[ Y = \Theta + E \]

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

\[ \text{Cov}[E] = \{\text{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\} \]

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a **separable covariance model**:

\[
\begin{align*}
\text{Cov}[E] &= \Sigma_1 \circ \Sigma_2 \\
\text{Cov}[\text{vec}(E)] &= \Sigma_2 \otimes \Sigma_1
\end{align*}
\]
Separable covariance structure for matrices

\[ Y = \Theta + E \]

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

\[
\text{Cov}[E] = \{ \text{cov}[e_{j_1,k_1}, e_{j_2,k_2}] \}
\]

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a **separable covariance model**:

\[
\begin{align*}
\text{Cov}[E] &= \Sigma_1 \circ \Sigma_2 \\
\text{Cov}[\text{vec}(E)] &= \Sigma_2 \otimes \Sigma_1 \\
E[EE^T] &= \Sigma_1 \times \text{tr}(\Sigma_2)
\end{align*}
\]
Separable covariance structure for matrices

\[ \mathbf{Y} = \Theta + \mathbf{E} \]

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

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

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a **separable covariance model**:

\[ \begin{align*}
\text{Cov}[\mathbf{E}] &= \Sigma_1 \circ \Sigma_2 \\
\text{Cov}[\text{vec}(\mathbf{E})] &= \Sigma_2 \otimes \Sigma_1 \\
\mathbb{E}[\mathbf{E}\mathbf{E}^T] &= \Sigma_1 \times \text{tr}(\Sigma_2) \\
\mathbb{E}[\mathbf{E}^T\mathbf{E}] &= \Sigma_2 \times \text{tr}(\Sigma_1)
\end{align*} \]
Separable covariance structure for matrices

\[ \mathbf{Y} = \Theta + \mathbf{E} \]
\[ \mathbf{E} \in \mathbb{R}^{m_1 \times m_2}, \text{ so } \text{Cov}[\mathbf{E}] \text{ is an } (m_1 \times m_1) \times (m_2 \times m_2) \text{ array:} \]
\[ \text{Cov}[\mathbf{E}] = \{\text{cov}[e_{j_1,k_1}, e_{j_2,k_2}]\} \]

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a **separable covariance model**:

\[
\begin{align*}
\text{Cov}[\mathbf{E}] &= \Sigma_1 \circ \Sigma_2 \\
\text{Cov}[\text{vec}(\mathbf{E})] &= \Sigma_2 \otimes \Sigma_1 \\
\mathbb{E}[\mathbf{EE}^T] &= \Sigma_1 \times \text{tr}(\Sigma_2) \\
\mathbb{E}[\mathbf{E}^T \mathbf{E}] &= \Sigma_2 \times \text{tr}(\Sigma_1) \\
\text{Cov}[e_{i,j}e_{k,l}] &= \sigma_{1,i,k}\sigma_{2,j,l}
\end{align*}
\]
Separable covariance structure for matrices

\[ Y = \Theta + E \]

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

\[ \text{Cov}[E] = \{ \text{cov}[e_{j_1,k_1}, e_{j_2,k_2}] \} \]

Usually the data are insufficient to estimate this covariance.

A parsimonious alternative is to fit a separable covariance model:

\[
\begin{align*}
\text{Cov}[E] &= \Sigma_1 \circ \Sigma_2 \\
\text{Cov}[\text{vec}(E)] &= \Sigma_2 \otimes \Sigma_1 \\
E[EE^T] &= \Sigma_1 \times \text{tr}(\Sigma_2) \\
E[E^T E] &= \Sigma_2 \times \text{tr}(\Sigma_1) \\
\text{Cov}[e_{i,j}e_{k,l}] &= \sigma_{1,i,k}\sigma_{2,j,l}
\end{align*}
\]

This is the covariance structure of the “matrix normal” model (Dawid, 1981)
Generating the matrix normal class

Multivariate normal model:

\[ z = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0, 1) \]
Generating the matrix normal class

Multivariate normal model:

\[ z = \{z_j : j = 1, \ldots, m\} \sim \text{iid normal}(0, 1) \]
\[ y = \mu + Az \sim \text{multivariate normal}(\mu, \Sigma = AA^T) \]
Generating the matrix normal class

Multivariate normal model:

\[ z = \{z_j : j = 1, \ldots, m\} \overset{\text{iid}}{\sim} \text{normal}(0, 1) \]
\[ y = \mu + Az \sim \text{multivariate normal}(\mu, \Sigma = AA^T) \]

Matrix normal model:

\[ Z = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{\text{iid}}{\sim} \text{normal}(0, 1) \]
Generating the matrix normal class

Multivariate normal model:

\[ \mathbf{z} = \{z_j : j = 1, \ldots, m\} \sim \text{iid normal}(0, 1) \]
\[ \mathbf{y} = \mathbf{\mu} + \mathbf{Az} \sim \text{multivariate normal}(\mathbf{\mu}, \Sigma = \mathbf{AA}^T) \]

Matrix normal model:

\[ \mathbf{Z} = \{z_{i,j}^{m_1,m_2}_{i=1,j=1} \sim \text{iid normal}(0, 1) \]
\[ \mathbf{Y} = \mathbf{M} + \mathbf{AZB}^T \sim \text{matrix normal}(\mathbf{M}, \Sigma_1 = \mathbf{AA}^T, \Sigma_2 = \mathbf{BB}^T) \]
Non-smooth domains

Smooth domains: For time/space data, better alternatives exist.

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

**Smooth domains:** For time/space data, better alternatives exist.

**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

**Smooth domains:** For time/space data, better alternatives exist.

**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

Smooth domains: For time/space data, better alternatives exist.

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

Smooth domains: For time/space data, better alternatives exist.

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

**Smooth domains:** For time/space data, better alternatives exist.

**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?

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

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

Let $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{mnorm}(0, \Sigma_1, \Sigma_2)$, An MLE must satisfy

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

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

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

We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$ for an MLE to exist.

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

- Penalized likelihood: Allen and Tibshirani (2011)
Data limitations

Let $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{mnorm}(0, \Sigma_1, \Sigma_2)$, An MLE must satisfy

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

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

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

We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$ for an MLE to exist.

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

- Penalized likelihood: Allen and Tibshirani (2011)
Let $Y_1, \ldots, Y_n \sim \text{mnorm}(0, \Sigma_1, \Sigma_2)$, An MLE must satisfy

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

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

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

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

We conjecture that we need $n \geq 1 + \max\{m_1/m_2, m_2/m_1\}$ for an MLE to exist.

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

- Penalized likelihood: Allen and Tibshirani (2011)
Separable covariance structure for arrays

\[ Y = \Theta + E \]

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

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

A parsimonious alternative is an “array normal” model:

\[
\text{Cov}[E] = \Sigma_1 \circ \Sigma_2 \circ \Sigma_3
\]
## Separable covariance structure for arrays

\[ \mathbf{Y} = \Theta + \mathbf{E} \]

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

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

A parsimonious alternative is an “array normal” model:

\[
\text{Cov}[\mathbf{E}] = \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3
\]
Separable covariance structure for arrays

\[ Y = \Theta + E \]

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

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

A parsimonious alternative is an “array normal” model:

\[
\begin{align*}
\text{Cov}[E] &= \Sigma_1 \circ \Sigma_2 \circ \Sigma_3 \\
\text{Cov}[\text{vec}(E)] &= \Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1
\end{align*}
\]
Separable covariance structure for arrays

\[ Y = \Theta + E \]

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

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

A parsimonious alternative is an “array normal” model:

\[
\begin{align*}
\text{Cov}[E] &= \Sigma_1 \circ \Sigma_2 \circ \Sigma_3 \\
\text{Cov}[\text{vec}(E)] &= \Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1 \\
E[E_{(k)}E_{(k)}^T] &= \Sigma_k \times \prod_{j \neq k} \text{tr}(\Sigma_j)
\end{align*}
\]
Examples of multiway data

Separable covariance arrays

Trade example

Factor analysis

Deep interactions

Discussion

Separable covariance structure for arrays

\[ \mathbf{Y} = \Theta + \mathbf{E} \]

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

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

A parsimonious alternative is an “array normal” model:

\[
\begin{align*}
\text{Cov}[\mathbf{E}] & = \Sigma_1 \circ \Sigma_2 \circ \Sigma_3 \\
\text{Cov}[\text{vec}(\mathbf{E})] & = \Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1 \\
\mathbb{E}[\mathbf{E}(k) \mathbf{E}(k)^T] & = \Sigma_k \times \prod_{j \neq k} \text{tr}(\Sigma_j) \\
\log \text{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}] & = \log \sigma_{1,j_1,j_2} + \log \sigma_{2,k_1,k_2} + \log \sigma_{3,l_1,l_2}
\end{align*}
\]
Separable covariance structure for arrays

\[ Y = \Theta + E \]

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

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

A parsimonious alternative is an “array normal” model:

\[
\begin{align*}
\text{Cov}[E] &= \Sigma_1 \circ \Sigma_2 \circ \Sigma_3 \\
\text{Cov}[	ext{vec}(E)] &= \Sigma_3 \otimes \Sigma_2 \otimes \Sigma_1 \\
E[E(k)E_T(k)] &= \Sigma_k \times \prod_{j \neq k} \text{tr}(\Sigma_j) \\
\log \text{cov}[e_{j_1,k_1,l_1}, e_{j_2,k_2,l_2}] &= \log \sigma_{1,j_1,j_2} + \log \sigma_{2,k_1,k_2} + \log \sigma_{3,l_1,l_2}
\end{align*}
\]
Generating separable covariance structure

Multivariate normal model:

\[ z = \{z_j : j = 1, \ldots, m\} \overset{iid}{\sim} \text{normal}(0, 1) \]
Generating separable covariance structure

Multivariate normal model:

\[ z = \{z_j : j = 1, \ldots, m\} \sim_{\text{iid}} \text{normal}(0, 1) \]

\[ y = \mu + Az \sim \text{multivariate normal}(\mu, \Sigma = AA^T) \]
Generating separable covariance structure

Multivariate normal model:

\[ z = \{ z_j : j = 1, \ldots, m \} \stackrel{\text{iid}}{\sim} \text{normal}(0, 1) \]
\[ y = \mu + Az \sim \text{multivariate normal}(\mu, \Sigma = AA^T) \]

Matrix normal model:

\[ Z = \{ z_{i,j} \}_{i=1,j=1}^{m_1,m_2} \stackrel{\text{iid}}{\sim} \text{normal}(0, 1) \]
Generating separable covariance structure

Multivariate normal model:

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

Matrix normal model:

\[
\begin{align*}
\mathbf{Z} &= \{z_{i.j}\}^{m_1 \times m_2}_{i=1, j=1} \overset{\text{iid}}{\sim} \text{normal}(0, 1) \\
\mathbf{Y} &= \mathbf{M} + \mathbf{A}\mathbf{Z}\mathbf{B}^T \sim \text{matrix normal}(\mathbf{M}, \Sigma_1 = \mathbf{A}\mathbf{A}^T, \Sigma_2 = \mathbf{B}\mathbf{B}^T)
\end{align*}
\]
Generating separable covariance structure

Multivariate normal model:

\[ z = \{ z_j : j = 1, \ldots, m \} \overset{iid}{\sim} \text{normal}(0, 1) \]
\[ y = \mu + Az \sim \text{multivariate normal}(\mu, \Sigma = AA^T) \]

Matrix normal model:

\[ Z = \{ z_{i,j} \}_{i=1,j=1}^{m_1,m_2} \overset{iid}{\sim} \text{normal}(0, 1) \]
\[ Y = M + AZB^T \sim \text{matrix normal}(M, \Sigma_1 = AA^T, \Sigma_2 = BB^T) \]

Array normal model:

\[ Z = \{ z_{i,j,k} \}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{iid}{\sim} \text{normal}(0, 1) \]
Generating separable covariance structure

Multivariate normal model:

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

Matrix normal model:

\[
\mathbf{Z} = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \overset{iid}{\sim} \text{normal}(0, 1)
\]
\[
\mathbf{Y} = \mathbf{M} + \mathbf{AZB}^T \overset{}{\sim} \text{matrix normal}(\mathbf{M}, \Sigma_1 = \mathbf{AA}^T, \Sigma_2 = \mathbf{BB}^T)
\]

Array normal model:

\[
\mathbf{Z} = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{iid}{\sim} \text{normal}(0, 1)
\]
\[
\mathbf{Y} = \mathbf{M} + \left[ \begin{array}{ccc} \mathbf{A} & \mathbf{Z} & \mathbf{B}^T \\ \mathbf{C} & \end{array} \right]
\]
Array decompositions and multilinear algebra


\[
Y = \sum_{r=1}^{R} \lambda_r (u_r \odot v_r \odot w_r) \quad y_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}
\]

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

\[
Y = D \times \{U, V, W\} \quad y_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}
\]

- **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)
Array decompositions and multilinear algebra


\[
Y = \sum_{r=1}^{R} \lambda_r (u_r \circ v_r \circ w_r) \\
y_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}
\]


\[
Y = D \times \{U, V, W\} \\
y_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}
\]

- 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)
Array decompositions and multilinear algebra


\[
Y = \sum_{r=1}^{R} \lambda_r (u_r \circ v_r \circ w_r) \\
y_{i,j,k} = \sum_{r=1}^{\lambda_r} u_{i,r} v_{j,r} w_{k,r}
\]

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

\[
Y = D \times \{U, V, W\} \\
y_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}
\]

- **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).
Array decompositions and multilinear algebra


\[
Y = \sum_{r=1}^{R} \lambda_r (u_r \circ v_r \circ w_r) \quad y_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}
\]


\[
Y = D \times \{U, V, W\} \quad y_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}
\]

- \( D \) 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)
Array decompositions and multilinear algebra


\[
Y = \sum_{r=1}^{R} \lambda_r (u_r \odot v_r \odot w_r) \quad y_{i,j,k} = \sum \lambda_r u_{i,r} v_{j,r} w_{k,r}
\]

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

\[
Y = D \times \{U, V, W\} \quad y_{i,j,k} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} d_{r_1,r_2,r_3} u_{i,r_1} v_{j,r_2} w_{k,r_3}
\]

- **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

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

\[ Y = Z \times \{ A, B, C \} \]

\[ = Z \times_1 A \times_2 B \times_3 C \]

**Array-matrix multiplication:** \( Z \times_1 A \)

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

\[ Z \times_j (F + G) = Z \times_j F + Z \times_j G \]

\[ (Z \times_j F) \times_k G = (Z \times_k G) \times_j F = Z \times_j F \times_k G \]

\[ (Z \times_j F) \times_j G = Z \times_j (GF) \]

If \( Y = Z \times \{ A_1, \ldots, A_K \} \), then

\[ Y_{(k)} = A_k Z_{(k)} (A_K \otimes \cdots \otimes A_{k+1} \otimes A_{k-1} \otimes \cdots \otimes A_1)^T. \]
The full rank multilinear Tucker product

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

\[ Y = Z \times \{ A, B, C \} = Z \times_1 A \times_2 B \times_3 C \]

**Array-matrix multiplication:** \( Z \times_1 A \)

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

\[ Z \times_j (F + G) = Z \times_j F + Z \times_j G \]
\[ (Z \times_j F) \times_k G = (Z \times_k G) \times_j F = Z \times_j F \times_k G \]
\[ (Z \times_j F) \times_j G = Z \times_j (GF) \]

If \( Y = Z \times \{ A_1, \ldots, A_K \} \), then

\[ Y_{(k)} = A_k Z_{(k)} (A_K \otimes \cdots \otimes A_{k+1} \otimes A_{k-1} \otimes \cdots \otimes A_1)^T. \]
The full rank multilinear Tucker product

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

\[
Y = Z \times \{A, B, C\}
\]

\[
= Z \times_1 A \times_2 B \times_3 C
\]

**Array-matrix multiplication:** \(Z \times_1 A\)

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

\[
\begin{align*}
z \times_j (F + G) &= z \times_j F + z \times_j G \\
(Z \times_j F) \times_k G &= (Z \times_k G) \times_j F = Z \times_j F \times_k G \\
(Z \times_j F) \times_j G &= Z \times_j (GF)
\end{align*}
\]

If \(Y = Z \times \{A_1, \ldots, A_K\}\), then

\[
Y_{(k)} = A_k Z_{(k)} (A_K \otimes \cdots \otimes A_{k+1} \otimes A_{k-1} \otimes \cdots \otimes A_1)^T.
\]
The full rank multilinear Tucker product

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

\[ Y = Z \times \{ A, B, C \} = Z \times_1 A \times_2 B \times_3 C \]

**Array-matrix multiplication:** \( Z \times_1 A \)

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

\[ Z \times_j (F + G) = Z \times_j F + Z \times_j G \]
\[ (Z \times_j F) \times_k G = (Z \times_k G) \times_j F = Z \times_j F \times_k G \]
\[ (Z \times_j F) \times_j G = Z \times_j (GF) \]

If \( Y = Z \times \{ A_1, \ldots, A_K \} \), then

\[ Y^{(k)} = A_k Z^{(k)} (A_K \otimes \cdots \otimes A_{k+1} \otimes A_{k-1} \otimes \cdots \otimes A_1)^T. \]
The full rank multilinear Tucker product

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

\[ Y = Z \times \{ A, B, C \} = Z \times_1 A \times_2 B \times_3 C \]

**Array-matrix multiplication:** \( Z \times_1 A \)

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

\[ Z \times_j (F + G) = Z \times_j F + Z \times_j G \]
\[ (Z \times_j F) \times_k G = (Z \times_k G) \times_j F = Z \times_j F \times_k G \]
\[ (Z \times_j F) \times_j G = Z \times_j (GF) \]

If \( Y = Z \times \{ A_1, \ldots, A_K \} \), then

\[ Y_{(k)} = A_k Z_{(k)} (A_K \otimes \cdots \otimes A_{k+1} \otimes A_{k-1} \otimes \cdots \otimes A_1)^T. \]
Examples of multiway data
Separable covariance arrays
Trade example
Factor analysis
Deep interactions
Discussion

The full rank multilinear Tucker product

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

\[ Y = Z \times \{ A, B, C \} \]
\[ = Z \times_1 A \times_2 B \times_3 C \]

Array-matrix multiplication: \( Z \times_1 A \)

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

\[ Z \times_j (F + G) = Z \times_j F + Z \times_j G \]
\[ (Z \times_j F) \times_k G = (Z \times_k G) \times_j F = Z \times_j F \times_k G \]
\[ (Z \times_j F) \times_j G = Z \times_j (GF) \]

If \( Y = Z \times \{ A_1, \ldots, A_K \} \), then

\[ Y^{(k)} = A_k Z^{(k)} (A_K \otimes \cdots \otimes A_{k+1} \otimes A_{k-1} \otimes \cdots \otimes A_1)^T. \]
Separable covariance via Tucker products

Multivariate normal model:

\[ z = \{z_j : j = 1, \ldots, m\} \sim \text{normal}(0, 1) \]
\[ y = \mu + Az \sim \text{multivariate normal}(\mu, \Sigma = AA^T) \]

Matrix normal model:

\[ Z = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \sim \text{normal}(0, 1) \]
\[ Y = M + AZB^T \sim \text{matrix normal}(M, \Sigma_1 = AA^T, \Sigma_2 = BB^T) \]

NOTE: \( AZB^T = Z \times \{A, B\} \)

Array normal model:

\[ Z = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \sim \text{normal}(0, 1) \]
\[ Y = M + Z \times \{A, B, C\} \sim \text{array normal}(M, \Sigma_1 = AA^T, \Sigma_2 = BB^T, \Sigma_3 = CC^T) \]

(Hoff, 2011)
Separable covariance via Tucker products

Multivariate normal model:

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

Matrix normal model:

\[
\begin{align*}
\mathbf{Z} &= \{ z_{i,j} \}_{i=1,j=1}^{m_1,m_2} \overset{iid}{\sim} \text{normal}(0, 1) \\
\mathbf{Y} &= \mathbf{M} + \mathbf{AZB}^T \sim \text{matrix normal}(\mathbf{M}, \Sigma_1 = \mathbf{A} \mathbf{A}^T, \Sigma_2 = \mathbf{B} \mathbf{B}^T)
\end{align*}
\]

**NOTE:** \( \mathbf{AZB}^T = \mathbf{Z} \times \{ \mathbf{A}, \mathbf{B} \} \)

Array normal model:

\[
\begin{align*}
\mathbf{Z} &= \{ z_{i,j,k} \}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \overset{iid}{\sim} \text{normal}(0, 1) \\
\mathbf{Y} &= \mathbf{M} + \mathbf{Z} \times \{ \mathbf{A}, \mathbf{B}, \mathbf{C} \} \sim \text{array normal}(\mathbf{M}, \Sigma_1 = \mathbf{A} \mathbf{A}^T, \Sigma_2 = \mathbf{B} \mathbf{B}^T, \Sigma_3 = \mathbf{C} \mathbf{C}^T)
\end{align*}
\]

(Hoff, 2011)
Separable covariance via Tucker products

Multivariate normal model:

\[ z = \{z_j : j = 1, \ldots, m\} \sim \text{normal}(0, 1) \]
\[ y = \mu + Az \sim \text{multivariate normal}(\mu, \Sigma = AA^T) \]

Matrix normal model:

\[ Z = \{z_{i,j}\}_{i=1,j=1}^{m_1,m_2} \sim \text{normal}(0, 1) \]
\[ Y = M + AZB^T \sim \text{matrix normal}(M, \Sigma_1 = AA^T, \Sigma_2 = BB^T) \]

\textbf{NOTE:} \ AZB^T = Z \times \{A, B\}

Array normal model:

\[ Z = \{z_{i,j,k}\}_{i=1,j=1,k=1}^{m_1,m_2,m_3} \sim \text{normal}(0, 1) \]
\[ Y = M + Z \times \{A, B, C\} \sim \text{array normal}(M, \Sigma_1 = AA^T, \Sigma_2 = BB^T, \Sigma_3 = CC^T) \]

(Hoff, 2011)
Estimation

Given \((\Sigma_2, \Sigma_3)\),

\[
E = (Y - M) \times \{I, \Sigma_2^{-1/2}, \Sigma_3^{-1/2}\} \sim \text{array normal}(0, \Sigma_1, I_{m_2}, I_{m_3})
\]

\(\Sigma_1\) can be estimated from \(E_{(1)}\)T \(E_{(1)}\):

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler
Given $(\Sigma_2, \Sigma_3)$,

$$E = (Y - M) \times \{ I, \Sigma_2^{-1/2}, \Sigma_3^{-1/2} \} \sim \text{array normal}(0, \Sigma_1, I_{m_2}, I_{m_3})$$

$$E_{(1)} \sim \text{matrix normal}(0, \Sigma_1, I_{m_2m_3})$$

$\Sigma_1$ can be estimated from $E_{(1)}E_{(1)}^T$

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler
Given \((\Sigma_2, \Sigma_3)\),

\[
E = (\mathbf{Y} - \mathbf{M}) \times \{ \mathbf{I}, \Sigma_2^{-1/2}, \Sigma_3^{-1/2} \} \sim \text{array normal}(\mathbf{0}, \Sigma_1, \mathbf{I}_{m_2}, \mathbf{I}_{m_3})
\]

\[
E_{(1)} \sim \text{matrix normal}(\mathbf{0}, \Sigma_1, \mathbf{I}_{m_2 m_3})
\]

\(\Sigma_1\) can be estimated from \(E_{(1)}E_{(1)}^T\)

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler
Given \((\Sigma_2, \Sigma_3)\),

\[ E = (Y - M) \times \{I, \Sigma_2^{-1/2}, \Sigma_3^{-1/2}\} \sim \text{array normal}(0, \Sigma_1, I_{m_2}, I_{m_3}) \]

\[ E_{(1)} \sim \text{matrix normal}(0, \Sigma_1, I_{m_2 m_3}) \]

\(\Sigma_1\) can be estimated from \(E_{(1)}E_{(1)}^T\)

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler
Given \((\Sigma_2, \Sigma_3)\),

\[
E = (Y - M) \times \{I, \Sigma_2^{-1/2}, \Sigma_3^{-1/2}\} \sim \text{array normal}(0, \Sigma_1, I_{m_2}, I_{m_3})
\]

\[
E_{(1)} \sim \text{matrix normal}(0, \Sigma_1, I_{m_2 m_3})
\]

\(\Sigma_1\) can be estimated from \(E_{(1)}E_{(1)}^T\)

- MLE via block coordinate descent ("flip-flop" algorithm, Dutilleul(1999))
- Equivariant Bayes estimates via Gibbs sampler
Given \((\Sigma_2, \Sigma_3)\),

\[
E = (Y - M) \times \{I, \Sigma_2^{-1/2}, \Sigma_3^{-1/2}\} \sim \text{array normal}(0, \Sigma_1, I_{m_2}, I_{m_3})
\]

\[
E_{(1)} \sim \text{matrix normal}(0, \Sigma_1, I_{m_2m_3})
\]

\(\Sigma_1\) can be estimated from \(E_{(1)}E_{(1)}^T\)

- 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: $\mathbf{E} \sim \text{array normal}(0, I, I, I, \sigma^2 I)$
- vector normal error model: $\mathbf{E} \sim \text{array normal}(0, I, I, \Sigma_3, I)$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, I, I, \Sigma_3, \Sigma_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4)$
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: \( \mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \Sigma_3, \mathbf{I}) \)
- matrix normal error model: \( \mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \Sigma_3, \Sigma_4) \)
- array normal model: \( \mathbf{E} \sim \text{array normal}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4) \)
International trade example

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

- \( i \in \{1, \ldots, 30\} \) 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 \( E = \{e_{i,j,k,l}\} \)

- iid error model: \( E \sim \text{array normal}(0, I, I, I, \sigma^2 I) \)
- vector normal error model: \( E \sim \text{array normal}(0, I, I, \Sigma_3, I) \)
- matrix normal error model: \( E \sim \text{array normal}(0, I, I, \Sigma_3, \Sigma_4) \)
- array normal model: \( E \sim \text{array normal}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4) \)
International trade example

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

- \( i \in \{1, \ldots, 30\} \) 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 \( E = \{e_{i,j,k,l}\} \)

- iid error model: \( E \sim \text{array normal}(0, I, I, I, \sigma^2 I) \)
- vector normal error model: \( E \sim \text{array normal}(0, I, I, \Sigma_3, I) \)
- matrix normal error model: \( E \sim \text{array normal}(0, I, I, \Sigma_3, \Sigma_4) \)
- array normal model: \( E \sim \text{array normal}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4) \)
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: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \Sigma_3, \mathbf{I})$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, \mathbf{I}, \mathbf{I}, \Sigma_3, \Sigma_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4)$
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: $\mathbf{E} \sim \text{array normal}(0, I, I, I, \sigma^2 I)$
- vector normal error model: $\mathbf{E} \sim \text{array normal}(0, I, I, \Sigma_3, I)$
- matrix normal error model: $\mathbf{E} \sim \text{array normal}(0, I, I, \Sigma_3, \Sigma_4)$
- array normal model: $\mathbf{E} \sim \text{array normal}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4)$
International trade example

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

- \( i \in \{ 1, \ldots, 30 \} \) 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 \( E = \{ e_{i,j,k,l} \} \)

- iid error model: \( E \sim \text{array normal}(0, I, I, I, \sigma^2 I) \)
- vector normal error model: \( E \sim \text{array normal}(0, I, I, \Sigma_3, I) \)
- matrix normal error model: \( E \sim \text{array normal}(0, I, I, \Sigma_3, \Sigma_4) \)
- array normal model: \( E \sim \text{array normal}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4) \)
Posterior predictive comparisons

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

Models:

**reduced**: array normal($0, I, I, \Sigma_3, \Sigma_4$)

**full**: array normal($0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$)
International trade example
International trade example
Factor analysis

Vector normal factor model:

\[
\begin{align*}
\text{Cov}[\mathbf{y}] &= \mathbf{A}\mathbf{A}^T + \mathbf{D} \\
\mathbf{y} &\overset{d}{=} \mathbf{A}\mathbf{z} + \mathbf{D}^{1/2}\mathbf{e}
\end{align*}
\]

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:

\[
\begin{align*}
\text{Cov}[\mathbf{Y}] &= (\mathbf{A}_1\mathbf{A}_1^T + \mathbf{D}_1) \circ \cdots \circ (\mathbf{A}_K\mathbf{A}_K^T + \mathbf{D}_K) \\
(\tilde{\mathbf{Y}}(1))_{i} &\overset{d}{=} \mathbf{A}_1\mathbf{z} + \mathbf{D}_1^{1/2}\mathbf{e}
\end{align*}
\]

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

Vector normal factor model:

\[
\text{Cov}[\mathbf{y}] = \mathbf{A}\mathbf{A}^T + \mathbf{D}
\]

\[
\mathbf{y} \overset{d}{=} \mathbf{A}\mathbf{z} + \mathbf{D}^{1/2}\mathbf{e}
\]

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:

\[
\text{Cov}[\mathbf{Y}] = (\mathbf{A}_1\mathbf{A}_1^T + \mathbf{D}_1) \circ \cdots \circ (\mathbf{A}_K\mathbf{A}_K^T + \mathbf{D}_K)
\]

\[
(\tilde{\mathbf{Y}}_{(1)})_i \overset{d}{=} \mathbf{A}_1\mathbf{z} + \mathbf{D}_1^{1/2}\mathbf{e}
\]

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

Vector normal factor model:

\[ \text{Cov}[y] = AA^T + D \]
\[ y \overset{d}{=} Az + D^{1/2}e \]

where \( A \in \mathbb{R}^{p \times r} \) and \( D \) 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:

\[ \text{Cov}[Y] = (A_1A_1^T + D_1) \circ \cdots \circ (A_KA_K^T + D_K) \]
\[ (\tilde{Y}_{(1)})_i \overset{d}{=} A_1z + D_1^{1/2}e \]

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

Mean model:

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

Variance model:

\[ \begin{align*}
E &= \{\epsilon_{age, i, j, k}\} \sim \text{anorm}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4) \\
\Sigma_k &= A_k A_k^T + D_k
\end{align*} \]
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:

\[ E = \{\epsilon_{\text{age},i,j,k}\} \sim \text{anorm}(0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4) \]
\[ \Sigma_k = A_k A_k^T + D_k \]
Mortality tables

Predictive performance experiment: Predict 5% missing data

<table>
<thead>
<tr>
<th></th>
<th>IID</th>
<th>FA</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean(SSE)</td>
<td>273.28</td>
<td>3.27</td>
</tr>
<tr>
<td>sd(SSE)</td>
<td>20.34</td>
<td>0.51</td>
</tr>
</tbody>
</table>
Consider the usual three-factor ANOVA decomposition model:

\[ y_{i,j,k,l} = \mu_{j,k,l} + \epsilon_{i,j,k,l} \]

\[ = \mu + [a_j + b_k + c_l] + [(ab)_{j,k} + (ac)_{j,l} + (bc)_{k,l}] + [(abc)_{j,k,l}] + \epsilon_{i,j,k,l} \]
Array normal priors for deep interactions

Main effect vectors:

\[ a \sim \text{vnorm}(0, \gamma_1 \Sigma_a) \quad , \quad b \sim \text{vnorm}(0, \gamma_1 \Sigma_b) \quad , \quad c \sim \text{vnorm}(0, \gamma_1 \Sigma_c) \]

Two-way interaction matrices

\[ (ab) \sim \text{mnorm}(0, \gamma_2 \Sigma_a, \Sigma_b) \quad , \quad (ac) \sim \text{mnorm}(0, \gamma_2 \Sigma_a, \Sigma_c) \quad , \quad (bc) \sim \text{mnorm}(0, \gamma_2 \Sigma_b, \Sigma_c) \]

Three-way interaction array

\[ (abc) \sim \text{anorm}(0, \gamma_3 \Sigma_a, \Sigma_b, \Sigma_c) \]
Array normal priors for deep interactions

main effect vectors:

\[ a \sim vnorm(0, \gamma_1 \Sigma_a) \quad , \quad b \sim vnorm(0, \gamma_1 \Sigma_b) \quad , \quad c \sim vnorm(0, \gamma_1 \Sigma_c) \]

two-way interaction matrices

\[ (ab) \sim mnorm(0, \gamma_2 \Sigma_a, \Sigma_b) \quad , \quad (ac) \sim mnorm(0, \gamma_2 \Sigma_a, \Sigma_c) \quad , \quad (bc) \sim mnorm(0, \gamma_2 \Sigma_b, \Sigma_c) \]

three-way interaction array

\[ (abc) \sim anorm(0, \gamma_3 \Sigma_a, \Sigma_b, \Sigma_c) \]
Array normal priors for deep interactions

main effect vectors:

\[ a \sim \text{vnorm}(0, \gamma_1 \Sigma_a) \quad , \quad b \sim \text{vnorm}(0, \gamma_1 \Sigma_b) \quad , \quad c \sim \text{vnorm}(0, \gamma_1 \Sigma_c) \]

two-way interaction matrices

\[ (ab) \sim \text{mnorm}(0, \gamma_2 \Sigma_a, \Sigma_b) \quad , \quad (ac) \sim \text{mnorm}(0, \gamma_2 \Sigma_a, \Sigma_c) \quad , \quad (bc) \sim \text{mnorm}(0, \gamma_2 \Sigma_b, \Sigma_c) \]

three-way interaction array

\[ (abc) \sim \text{anorm}(0, \gamma_3 \Sigma_a, \Sigma_b, \Sigma_c) \]
Examples of multiway data
Separable covariance arrays
Trade example
Factor analysis
Deep interactions
Discussion

Regularization

![Graphs showing sample size vs. mean household size](image)

Bayes

MLE
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
• 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
• 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