Separable covariance arrays via the Tucker product - part 2

by P. Hoff

Kean Ming Tan

May 14, 2013

International Trade data set

Yearly change in log trade value (in 2000 dollars): $\mathbf{Y} = \{y_{i,j,k,l}\}$

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

Interested in modeling the mean $M_{ijk} = \mu_{i,j,k}$ across t measurements

What can we do?

International Trade data set

Interested in the model

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

- iid error model: $\epsilon_{i,j,k,l} \sim \text{normal}(0, \sigma^2)$
- multivariate error model: $\boldsymbol{\epsilon}_{i,j,k} \sim \text{multivariate normal}(\mathbf{0}, \boldsymbol{\Sigma})$
- matrix-variate error model: $\boldsymbol{\epsilon}_{i,j} \sim \text{matrix normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$

But all four dimensions are correlated!

$$\mathbf{E} \sim ???$$

International Trade data set

Interested in the model

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

- iid error model: $\epsilon_{i,j,k,l} \sim \text{normal}(0, \sigma^2)$
- multivariate error model: $\boldsymbol{\epsilon}_{i,j,k} \sim \text{multivariate normal}(\mathbf{0}, \boldsymbol{\Sigma})$
- matrix-variate error model: $\boldsymbol{\epsilon}_{i,j} \sim \text{matrix normal}(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$

But all four dimensions are correlated!

$$\mathbf{E} \sim ???$$

Best friend in 1-dimensional case? (Normal distribution)

Goal of today's talk

Construct the Array Normal distribution for array data

$$\mathbf{Y} \sim \text{array normal}(\mathbf{M}, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_k)$$

where $\mathbf{Y} \in \mathbb{R}^{m_1 \times \ldots \times m_k}$

Goal of today's talk

Construct the Array Normal distribution for array data

$$\mathbf{Y} \sim \text{array normal}(\mathbf{M}, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_k)$$

where $\mathbf{Y} \in \mathbb{R}^{m_1 \times \ldots \times m_k}$

CAUTION!!!!!!!! It is mathematically intensive!

Definition outline

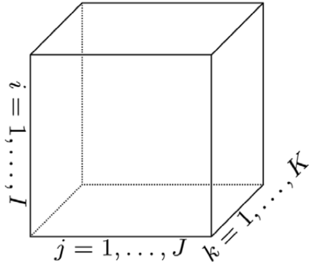
- Array
- Vectorize
- Inner product
- Matricization
- Array-matrix Product

Definition - Array

A K-array of dimension $(m_1, \ldots, m_K)$ is a set of $m_1 \times \cdots \times m_K$ points in $\mathbb{R}$

$$\mathbf{Y} = \{y_{i_1}, \ldots, y_{i_K} : i_k \in \{1, \ldots, m_k\}, k = 1, \ldots, K\}$$

An Example[1]: $\mathbf{Y} \in \mathbb{R}^{3\times4\times2}$



$$\mathbf{Y}_{[,,1]} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}$$

$$\mathbf{Y}_{[,,2]} = \begin{pmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{pmatrix}$$

[1] Kolda, T. and Bader, B. SIAM Review 2009

Definition - vec Operator

Suppose that $\mathbf{Y} \in \mathbb{R}^{2\times 2\times 2}$

$$\mathbf{Y}_{[,,1]} = \begin{pmatrix} y_{111} & y_{121} \\ y_{211} & y_{221} \end{pmatrix} \qquad \mathbf{Y}_{[,,2]} = \begin{pmatrix} y_{112} & y_{122} \\ y_{212} & y_{222} \end{pmatrix}$$

$$\text{vec}(\mathbf{Y}) = \begin{pmatrix} y_{111} \\ y_{211} \\ y_{121} \\ y_{221} \\ y_{112} \\ y_{212} \\ y_{122} \\ y_{222} \end{pmatrix}$$

Definition - vec Operator

Suppose that $\mathbf{Y} \in \mathbb{R}^{2\times2\times2}$

$$\mathbf{Y}_{[,,1]} = \begin{pmatrix} y_{111} & y_{121} \\ y_{211} & y_{221} \end{pmatrix} \qquad \mathbf{Y}_{[,,2]} = \begin{pmatrix} y_{112} & y_{122} \\ y_{212} & y_{222} \end{pmatrix}$$

$$\text{vec}(\mathbf{Y}) = \begin{pmatrix} y_{111} \\ y_{211} \\ y_{121} \\ y_{221} \\ y_{112} \\ y_{212} \\ y_{122} \\ y_{222} \end{pmatrix}$$

Linbo: If you read the indexes from the right, the numbers are always increasing as you proceed to the next row!

Definition - Inner Product

vector	$\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m_1}$	$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i$
matrix	$\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$	$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_i \sum_j X_{ij} Y_{ij}$
array	$\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$	$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_i \sum_j \sum_k X_{ijk} Y_{ijk}$

Definition - Matricization

Process of reordering elements in K-array into a matrix

Suppose $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, the 1-mode matricization is

$$\mathbf{Y}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$$

Definition - Matricization

Process of reordering elements in K-array into a matrix

Suppose $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, the 1-mode matricization is

$$\mathbf{Y}_{(1)} \in \mathbb{R}^{m_1 \times m_2 m_3}$$

Similarly,

$$\mathbf{Y}_{(2)} \in \mathbb{R}^{m_2 \times m_1 m_3}$$

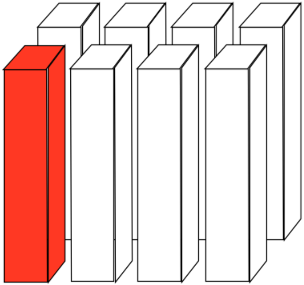
$$\mathbf{Y}_{(3)} \in \mathbb{R}^{m_3 \times m_1 m_2}$$

Definition - Matricization[3]

$\mathbf{Y} \in \mathbb{R}^{3\times4\times2}$

$$\mathbf{Y}_{[,,1]} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \qquad \mathbf{Y}_{[,,2]} = \begin{pmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{pmatrix}$$

The 1-mode matricization is[2]



$$\mathbf{Y}_{(1)} = \begin{pmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{pmatrix}$$

[2] Indebted to one of my anonymous classmates

[3] Kolda, T. and Bader, B. SIAM Review 2009

Definition - Matricization[4]

$\mathbf{Y} \in \mathbb{R}^{3\times4\times2}$

$$\mathbf{Y}_{[,,1]} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \qquad \mathbf{Y}_{[,,2]} = \begin{pmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{pmatrix}$$

The 2-mode matricization is

$$\mathbf{Y}_{(2)} = \begin{pmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 18 \\ 7 & 8 & 9 & 19 & 20 & 21 \\ 10 & 11 & 12 & 22 & 23 & 24 \end{pmatrix}$$

[4]Kolda, T. and Bader, B. SIAM Review 2009

Quiz

$$\mathbf{X} \in \mathbb{R}^{m_1 \times m_2}$$

$$\mathbf{X}_{(1)} =$$

Quiz

$$\mathbf{X} \in \mathbb{R}^{m_1 \times m_2}$$

$$\mathbf{X}_{(1)} = \mathbf{X}$$

$$\mathbf{X}_{(2)} =$$

Quiz

$$\mathbf{X} \in \mathbb{R}^{m_1 \times m_2}$$

$$\mathbf{X}_{(1)} = \mathbf{X}$$

$$\mathbf{X}_{(2)} = \mathbf{X}^T$$

Definition - k-mode product of an array

Suppose $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ and $\mathbf{A} \in \mathbb{R}^{m_1 \times m_1}$, the 1-mode product is

$$\mathbf{Z} = \mathbf{Y} \times_1 \mathbf{A} \qquad \text{if and only if} \qquad \mathbf{Z}_{(1)} = \mathbf{A}\mathbf{Y}_{(1)}$$

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

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

Definition - Array Matrix product

Suppose $\mathbf{Z} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ and $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$

$$\mathbf{Z} \times \mathbf{A} = \mathbf{Z} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3$$

Multivariate & Matrix normal model[5]

Multivariate normal model, $\mathbf{y} \in \mathbb{R}^m$:

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

[5] Adapted from Hoff's slides

Multivariate & Matrix normal model[5]

Multivariate normal model, $\mathbf{y} \in \mathbb{R}^m$:

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

Matrix-variate normal model, $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$:

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

Note that $\mathbf{Z} \times \{\mathbf{A}, \mathbf{B}\} = \mathbf{A}\mathbf{Z}\mathbf{B}^T$

[5] Adapted from Hoff's slides

Multilinear transformation

$\mathbf{Z}$ is a 3-dimensional random array with mean zero and uncorrelated covariance

$\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$ such that $\mathbf{\Sigma}_k = \mathbf{A}_k \mathbf{A}_k^T$

Let $\mathbf{Y} = \mathbf{Z} \times \mathbf{A}$

- $\text{Cov}[\mathbf{Y}] = \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3$
- $\text{Cov}[\text{vec}(\mathbf{Y})] = \mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$

Array normal model[6]

Array normal model, $\mathbf{y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$:

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

[6] Adapted from Hoff's slides

Probability density function

Multivariate normal

$$(2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp(-(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2)$$

Array normal: Note that $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3\}$

$$(2\pi)^{-m/2}(\prod_{k=1}^{3}|\boldsymbol{\Sigma}_k|^{-m/(2m_k)})\exp(-\|(\mathbf{Y}-\mathbf{M})\times\boldsymbol{\Sigma}^{-1/2}\|^2/2)$$

where $m = m_1 m_2 m_3$

Properties

If $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \overset{iid}{\sim} \text{anorm}(\mathbf{M}, \boldsymbol{\Sigma}_1 \circ \boldsymbol{\Sigma}_2 \circ \boldsymbol{\Sigma}_3)$, we have

$$\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n) \sim \text{anorm}(\mathbf{M} \circ \mathbf{1}_n, \boldsymbol{\Sigma}_1 \circ \boldsymbol{\Sigma}_2 \circ \boldsymbol{\Sigma}_3 \circ \mathbf{I}_n)$$

Estimation: Frequentist

- $\hat{\mathbf{M}} = \bar{\mathbf{Y}} = \sum \mathbf{Y}_i / n$
- Let $\mathbf{E} = \mathbf{Y} - \bar{\mathbf{Y}} \circ \mathbf{I}_n$ and repeat the following for $k = 1, 2, 3$
  - Compute $\tilde{\mathbf{E}} = \mathbf{E} \times \{\boldsymbol{\Sigma}_1^{-1/2}, \mathbf{I}_{m_k}, \boldsymbol{\Sigma}_3^{-1/2}, \mathbf{I}_n\}$ and $\mathbf{S}_k = \tilde{\mathbf{E}}_{(k)} \tilde{\mathbf{E}}_{(k)}^T$
  - $\boldsymbol{\Sigma}_k = \mathbf{S}_k / n_k$ where $n_k = n \times \prod_{j \neq k} m_j$

Analogy for univariate case $\sigma^2 = \sum (y - \bar{y})^2 / n$

Estimation: Bayesian

Assume the following prior:

- $\mathbf{M}|\mathbf{\Sigma} \sim \text{anorm}(\mathbf{M}_0, \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3/\kappa_0)$
- $\mathbf{\Sigma}_k \sim \text{inverse-Wishart}(\mathbf{S}_{0k}^{-1}, \nu_{0k})$

After A LOT of calculation, the posteriors are

- $\mathbf{M}|\mathbf{Y}, \mathbf{\Sigma} \sim \text{anorm}([\kappa_0 \mathbf{M}_0 + n\bar{\mathbf{Y}}]/[\kappa_0 + n], \mathbf{\Sigma}_1 \circ \mathbf{\Sigma}_2 \circ \mathbf{\Sigma}_3/[\kappa_0 + n])$
- $\mathbf{\Sigma}_k|\mathbf{Y}, \mathbf{\Sigma}_{-k} \sim \text{inverse-Wishart}([\mathbf{S}_{0k} + \mathbf{S}_k + \mathbf{R}_{(k)}\mathbf{R}_{(k)}^T]^{-1}, \nu_{0k} + n_k)$

where $\mathbf{R} = \sqrt{\frac{\kappa_0 n}{\kappa_0 + n}}(\bar{\mathbf{Y}} - \mathbf{M}_0) \times \{\mathbf{\Sigma}_1^{-1/2}, \mathbf{I}_{m_k}, \mathbf{\Sigma}_3^{-1/2}\}$

Gibbs Sampling

1. Sample $\mathbf{M}$ based on

$$\mathbf{M}|\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \mathbf{Y} \sim \text{anorm}(\frac{n\bar{\mathbf{Y}} + \kappa_0 \mathbf{M}_0}{n + \kappa_0})$$

2. For $k = 1, 2, 3$, sample $\boldsymbol{\Sigma}_k$ based on

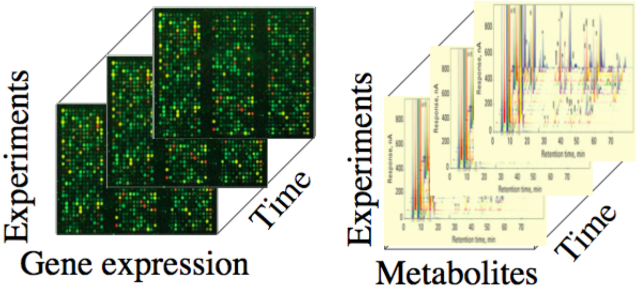
$$\boldsymbol{\Sigma}_k|\boldsymbol{\Sigma}_{-k}, \mathbf{Y} \sim \text{inverse-Wishart}([\mathbf{S}_{0k}+\mathbf{S}_k+\mathbf{R}_{(k)}\mathbf{R}_{(k)}^T]^{-1}, \nu_{0k}+n_k)$$

Next Time

- Apply to the International Trade data set

Take Home Message

You can now model CORRELATIONS of arbitrary many dimensions



No more problems to these datasets!

Questions?