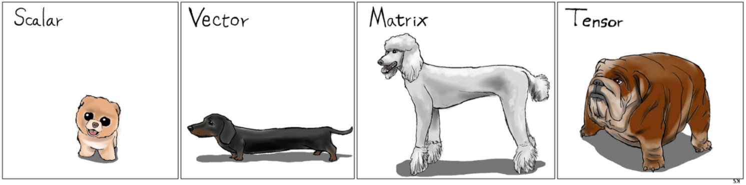
Tensor Data Analysis Overview

Carlos Llosa

STAT 680 lecture
November 08,2021

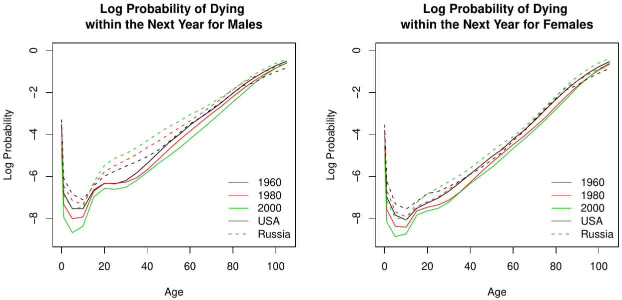


Outline

Outline

1. Tensors when the observations are scalars
2. Tensors with multivariate observations
   - Kronecker Separability for Higher Order Normality
3. Low-Rank Formats

Human Mortality Database



Tensor form along different factors:

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

$$39 \times 23 \times 9 \times 2 \text{ array!}$$

Factorial Designs

Cell means ANOVA: $y_{ijklm} = \mu_{ijkl} + e_{ijklm}$

- In STAT 500-510 deep interactions:

$$\mu_{ijkl} = \alpha_i + \beta_j + \eta_k + \gamma_l + \alpha\beta_{ij} + \alpha\eta_{ik} + \alpha\gamma_{il} + \beta\eta_{jk} + \beta\gamma_{jl} + \eta\gamma_{kl} +$$

$$(\alpha\beta\eta)_{ijk} + (\alpha\beta\gamma)_{ijl} + (\alpha\eta\gamma)_{ikl} + (\beta\eta\gamma)_{jkl} + (\alpha\beta\eta\gamma)_{ijkl}$$

- we could do $\mu_{ijkl} = \langle \mathcal{X}_{ijkl}, \mathcal{B} \rangle$, where

$$\mathcal{X}_{i^*j^*k^*l^*} = \begin{cases} 1 & if(i,j,k,l) = (i^*, j^*, k^*, l^*) \\ 0 & otherwise \end{cases}$$

  Then $\mathcal{B}$ is an array that contains all the relevant means and deep interactions!

Tensor Regression

- Simple linear regression

$$y_i = \beta x_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Multiple linear regression

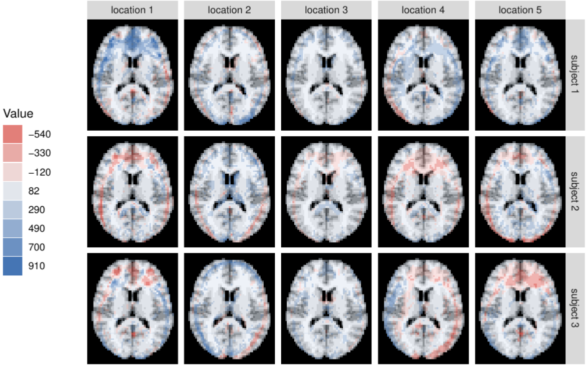
$$y_i = \boldsymbol{\beta}' \boldsymbol{x_i} + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Tensor regression

$$y_i = \langle \mathcal{B}, \mathcal{X}_i \rangle + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

Outline

Brain Imaging Data (Yamashita et.al 19)



tensor form along multiple factors and spatial dimensions!

- 9 subjects traveled to 12 imaging centers
- 3 repetitions of 240 time-steps each
- brain images of size $73 \times 73 \times 61$

$$73 \times 73 \times 61 \times 240 \times 9 \times 12 \times 3 \text{ array!}$$

Multivariate Regression

- Multivariate Multiple Linear Regression

$$\boldsymbol{y}_i = B\boldsymbol{x_i} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \stackrel{iid}{\sim} \mathcal{N}_m(0, \Sigma),$$

- Matrix-variate Regression (Ding and Cook, 2018, JRSSB)

$$Y_i = B_1 X_i B_2^{\intercal} + E_i, \quad E_i \stackrel{iid}{\sim} \mathcal{N}_{m_1,m_2}(0, \Sigma_1, \Sigma_2),$$

- Multilinear Tensor Regression (Hoff, 2015, Ann. Appl. Stat)

$$\mathcal{Y}_i = [\![\mathcal{X}_i; B_1, \ldots, B_p]\!] + \mathcal{E}_i, \quad \mathcal{E}_i \stackrel{iid}{\sim} \mathcal{N}_{m_1,\ldots,m_p}(0, \Sigma_1, \ldots, \Sigma_p),$$

The Matrix Normal Distribution

- S. N. Roy wrote its pdf in 1957
- A matrix follows a normal distribution if its vectorization follows a multivariate normal distribution
- Without any constraint on the covariance matrix, this leads to overfitting
- Kronecker separability is an intuitive constraint

Definition:

$$X \sim \mathcal{N}_{m_1,m_2}\left(M, \Sigma_1, \Sigma_2\right) \iff \mathrm{vec}(X) \sim \mathcal{N}_{m_1 \times m_2}\left(\mathrm{vec}(M), \Sigma_2 \otimes \Sigma_1\right)$$

dimensionality reduction:

$$\left[(m_1 \times m_2 + 1)(m_1 \times m_2)\right]/2 \quad \rightarrow \quad (m_1+1) \times m_1/2 + (m_2+1) \times m_2/2$$

The Tensor Normal Distribution

- A tensor follows a normal distribution if its vectorization follows a multivariate normal distribution
- The assumption of kronecker separability can be extended to higher order tensors
- Vectorization is usually defined in reverse lexicographic order to avoid an inconsistency with matrix vectorization

Definition:

$$\mathcal{X} \sim \mathcal{N}_{m_1,\ldots,m_p}\left(\mathcal{M}\ , \Sigma_1, \ldots, \Sigma_p\right)$$
$$\iff \mathrm{vec}(\mathcal{X}) \sim \mathcal{N}_{m_1 \times \ldots \times m_p}\left(\mathrm{vec}(\mathcal{M}), \bigotimes_{i=p}^{1} \Sigma_i\right)$$

Comparisons between tensor normal distributions

- Bivariate normal distribution: $\boldsymbol{x} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma), \quad \Sigma[i,j] = \sigma_{ij}$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right)$$

- Matrix normal distribution: $X \sim \mathcal{N}_{3,2}(M, \Sigma_1, \Sigma), \quad \Sigma[i,j] = \sigma_{ij}^2$

$$\operatorname{vec}(X) = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \end{bmatrix} \sim \mathcal{N}_{3\times 2}\left(\begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}, \quad \begin{bmatrix} \sigma_{11}\Sigma_1 & \sigma_{12}\Sigma_1 \\ \sigma_{21}\Sigma_1 & \sigma_{22}\Sigma_1 \end{bmatrix} \right).$$

- Third order tensor normal distribution: $\mathcal{X} \sim \mathcal{N}_{3,2,2}(\mathcal{M}, \Sigma_1, \Sigma_2, \Sigma)$

$$\begin{aligned} \operatorname{vec}(\mathcal{X}) &= \begin{bmatrix} \operatorname{vec} \mathcal{X}(:,:,1) \\ \operatorname{vec} \mathcal{X}(:,:,2) \end{bmatrix} \\ &\sim \mathcal{N}_{3\times 2\times 2}\left(\begin{bmatrix} \operatorname{vec} \mathcal{M}(:,:,1) \\ \operatorname{vec} \mathcal{M}(:,:,2) \end{bmatrix}, \begin{bmatrix} \sigma_{11}\Sigma_2 \otimes \Sigma_1 & \sigma_{12}\Sigma_2 \otimes \Sigma_1 \\ \sigma_{21}\Sigma_2 \otimes \Sigma_1 & \sigma_{22}\Sigma_2 \otimes \Sigma_1 \end{bmatrix} \right). \end{aligned}$$

Back to Matrix-Variate Regression

$$Y_i = B_1 X_i B_2^\intercal + E_i, \quad E_i \overset{iid}{\sim} \mathcal{N}_{m_1,m_2}(0, \Sigma_1, \Sigma_2),$$

- MLE under fixed (B_2, Σ_2):
  Let $H_1 = \sum_{i=1}^n X_i B_2' \Sigma_2^{-1} B_2 X_i'$, then

$$\begin{cases} \widehat{B}_1 = (\sum_{i=1}^n Y_i \Sigma_2^{-1} B_2 X_i') H_1^{-1} \\ \widehat{\Sigma}_1 = \sum_{i=1}^n Y_i \Sigma_2^{-1} Y_i' - \widehat{B}_1 H_1 \widehat{B}_1' \end{cases}$$

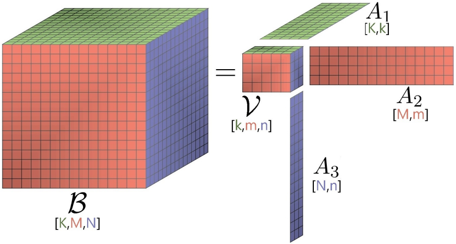
- MLE under fixed (B_1, Σ_1):
  Let $H_2 = \sum_{i=1}^n X_i' B_1' \Sigma_1^{-1} B_1 X_i$, then

$$\begin{cases} \widehat{B}_2 = (\sum_{i=1}^n Y_i' \Sigma_1^{-1} B_1 X_i) H_2^{-1} \\ \widehat{\Sigma}_2 = \sum_{i=1}^n Y_i' \Sigma_1^{-1} Y_i - \widehat{B}_2 H_2 \widehat{B}_2' \end{cases}$$

- Block relaxation algorithm!!!
- You should avoid this model → tensor on tensor regression!

Outline

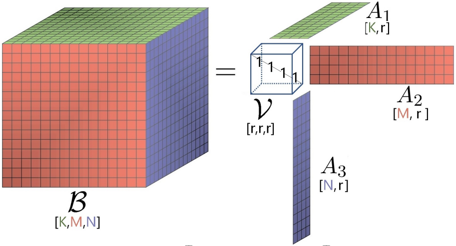
The Tucker (TK) Format



$$\mathcal{B} = [\![\mathcal{V}; A_1, A_2, A_3]\!]$$

- Unconstrained $\mathcal{B} \in \mathbb{R}^{15\times15\times15}$ has 3,375 parameters
- Constrained to a Tucker format of rank (3,4,5) leads to only 240

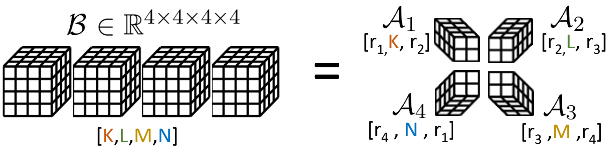
The Canonical (CP) format



$$\mathcal{B} = [\![A_1, A_2, A_3]\!]$$

- Special case when core $\mathcal{V}$ is *diagonal*
- Unconstrained $\mathcal{B}$ has 3,375 parameters
- Constrained to a CP format of rank 4 leads to only 180

The Tensor Ring (TR) Format



$$\mathcal{B} = \mathrm{tr}\left(\mathcal{A}_1 \times^1 \mathcal{A}_2 \times^1 \mathcal{A}_3 \times^1 \mathcal{A}_4\right)$$

- Referred to matrix product state (MPS) in many-body physics
- Unconstrained $\mathcal{B}$ has 256 parameters
- Constrained to a TR format of rank (2,2,2,2) leads to only 64

Tensor-on-Tensor Regression:

$$\mathcal{Y}_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + \mathcal{E}_i, \quad \mathcal{E}_i \overset{iid}{\sim} \mathcal{N}_{m_1, m_2 \dots, m_p}(0, \sigma^2 \Sigma_1, \Sigma_2, \dots, \Sigma_p),$$
$$\mathcal{X}_i \in \mathbb{R}^{h_1 \times h_2 \times \ldots \times h_l}, \quad \mathcal{Y}_i \in \mathbb{R}^{m_1 \times m_2 \times \ldots \times m_p},$$
$$\mathcal{B} \in \mathbb{R}^{h_1 \times h_2 \times \ldots \times h_l \times m_1 \times m_2 \times \ldots \times m_p}.$$

- Multilinear Tensor Regression is the case where $\mathcal{B}$ has an OP format $\mathcal{B}_{OP} = \circ[\![M_1, M_2, \dots, M_p]\!]$ because

$$\langle \mathcal{X}_i | \mathcal{B}_{OP} \rangle = [\![\mathcal{X}_i; M_1, M_2, \dots, M_p]\!].$$

- You can instead do a CP, TR or TK format!