Summary on basic time series studies

tensor data analysis with different data types

Haofan Zheng

Contents

1 High-dimentional α-PCA method **3**

1.1 Overall Summary . . . 3

1.2 Main model . . . 3

1.3 Main Statistics . . . 3

1.4 Theoretical Properties . . . 4

1.5 Simulation . . . 4

1.6 Application . . . 4

2 High-Dimensional GLM with Binary Outcomes **4**

3 Ultra-High Dimensional GFM **4**

4 Matrix-variate Logistic Regression with Measurement Error **4**

5 A Likelihood-Based Approach for Multivariate Categorical Response Regression in High Dimensions **4**

6 A likelihood-Based Approach for Semiparametric Regression with Panel Count Data **4**

7 Time Series Latent Gaussian Count **4**

8 Time Series Factor Models(tensor) **4**

1 High-dimentional α-PCA method

1.1 Overall Summary

This article considers the estimation and inference of the **low rank** components in high-dimentional matrixvariate models(tensor), and we propose an estimation method called α-PCA and it has some benefits with the high dimensions data favorably compared with other methods(traditional PCA, etc) based on the perfomance in the simulation.

1.2 Main model

The model is shown as the following:

$$\mathbf{Y}_t = \mathbf{R}\mathbf{F}_t\mathbf{C}^T + \mathbf{E}_t$$

$\mathbf{Y_t}$: $\mathbf{Y_t} \in \mathbb{R}^{p\times q}$, $1 \leq t \leq T$, observations,

$\mathbf{F_t}$: $\mathbf{F_t} \in \mathbb{R}^{k\times r}$, where $k \ll p$ and $r \ll q$ (**low rank**), latent matrix,

$\mathbf{E_t}$: $\mathbf{E_t} \in \mathbb{R}^{p\times q}$, noise matrix.

1.3 Main Statistics

An estimation procedure, namely α-PCA, aggregates the information in both first and second moments. Specifically, the two statistics are defined:

$$\widehat{\mathbf{M}}_R \triangleq \frac{1}{pq}\left((1+\alpha)\cdot\overline{\mathbf{Y}}\overline{\mathbf{Y}}^T + \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t - \overline{\mathbf{Y}})(\mathbf{Y}_t - \overline{\mathbf{Y}})^T\right)$$

$$\widehat{\mathbf{M}}_C \triangleq \frac{1}{pq}\left((1+\alpha)\cdot\overline{\mathbf{Y}}^T\overline{\mathbf{Y}} + \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t - \overline{\mathbf{Y}})^T(\mathbf{Y}_t - \overline{\mathbf{Y}})\right)$$

α : $\alpha \in [-1, +\infty)$, a hyperparameter,

$\overline{\mathbf{Y}} = \frac{1}{T}\sum_{i=1}^{T}\mathbf{Y}_t$, the sample mean.

Based on these two statistics, estimation of $\mathbf{R}$ and $\mathbf{C}$ can be obtained as $\sqrt{p}$ times the top k eigenvectors of $\widehat{\mathbf{M}}_R$ and $\sqrt{q}$ times the top q eigenvectors of $\widehat{\mathbf{M}}_C$ respectively, in descending order by corresponding eigenvalues.

1.4 Theoretical Properties

1.5 Simulation

1.6 Application

2 High-Dimensional GLM with Binary Outcomes

3 Ultra-High Dimensional GFM[1]

4 Matrix-variate Logistic Regression with Measurement Error

5 A Likelihood-Based Approach for Multivariate Categorical Response Regression in High Dimensions

6 A likelihood-Based Approach for Semiparametric Regression with Panel Count Data

7 Time Series Latent Gaussian Count

8 Time Series Factor Models(tensor)

[1] Generalized Factor Model