Tensor-variate elliptically contoured distributions with application to image learning

Carlos Llosa-Vite and Ranjan Maitra

2022 SMI conference - student paper competition

Department of Statistics
Iowa State University

Outline

The tensor-variate normal (TVN) distribution

The TVN distribution:

$$\boldsymbol{\mathcal{X}} \sim \mathcal{N}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1, \Sigma_2, \ldots, \Sigma_p) \iff \mathrm{vec}(\boldsymbol{\mathcal{X}} - \mathcal{M}) \sim \mathcal{N}_m(0, \otimes_{k=p}^{1} \Sigma_k),$$

where $\boldsymbol{m} = (m_1, m_2, \ldots, m_p)$ and $m = \prod_k m_k$.

- Defined through its vectorization.
- Kronecker-separable covariance structure.
- Also true of
  Elliptically-contoured (EC) tensor-variate (TV) distributions

Spherical tensor-variate distributions

$$\begin{aligned}\boldsymbol{\mathcal{X}} \sim \mathcal{S}_{\boldsymbol{h}}(\varphi) &\iff \psi_{\boldsymbol{\mathcal{X}}}(\mathcal{Z}) = \varphi(\langle \mathcal{Z}, \mathcal{Z} \rangle) \\ &\iff f_{\boldsymbol{\mathcal{X}}}(\mathcal{X}) = g(\langle \mathcal{X}, \mathcal{X} \rangle) \\ &\iff \text{vec}(\boldsymbol{\mathcal{X}}) \stackrel{d}{=} \Gamma \, \text{vec}(\boldsymbol{\mathcal{X}}) \ \forall \text{ orthogonal } \Gamma.\end{aligned}$$

- $\boldsymbol{h} = (h_1, h_2, \ldots, h_p)$
- φ is called the characteristic generator.
- g is called the density generator.
- $\boldsymbol{\mathcal{X}} \sim \mathcal{N}_{\boldsymbol{h}}(0, I_{h_1}, I_{h_2}, \ldots, I_{h_p})$ if $\varphi(u) = \exp(-u/2)$

Elliptically-contoured (EC) tensor-variate (TV) distributions

$$\begin{aligned}\boldsymbol{\mathcal{Y}} \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1,\ldots,\Sigma_p,\varphi) &\iff \psi_{\boldsymbol{\mathcal{Y}}}(\mathcal{Z}) = e^{i\langle \mathcal{Z},\mathcal{M}\rangle}\varphi(\langle \mathcal{Z}, [\![\mathcal{Z};\Sigma_1,\ldots,\Sigma_p]\!]\rangle) \\ &\iff f_{\boldsymbol{\mathcal{Y}}}(\mathcal{Y}) = \Big|\bigotimes_{k=p}^{1}\Sigma_k\Big|^{-1/2} g(D^2_\Sigma(\mathcal{Y},\mathcal{M})) \\ &\iff \boldsymbol{\mathcal{Y}} \overset{d}{=} \mathcal{M} + [\![\boldsymbol{\mathcal{X}}; \mathrm{Q}_1,\ldots,\mathrm{Q}_p]\!]\end{aligned}$$

- $D^2_\Sigma(\mathcal{Y},\mathcal{M})$ is the squared Mahalanobis distance

$$D^2_\Sigma(\mathcal{Y},\mathcal{M}) = \langle \mathcal{Y}-\mathcal{M}, [\![\mathcal{Y}-\mathcal{M}; \Sigma_1^{-1},\ldots,\Sigma_p^{-1}]\!]\rangle$$

- $\boldsymbol{\mathcal{X}} \sim \mathcal{S}_{\boldsymbol{h}}(\varphi)$ and $\mathrm{Q}_k\mathrm{Q}_k^\top = \Sigma_k$ is positive definite for all $k = 1,\ldots,p$.
- $\boldsymbol{\mathcal{Y}} \sim \mathcal{N}_{\boldsymbol{m}}(\mathcal{M},\Sigma_1,\ldots,\Sigma_p)$ if $\varphi(u) = \exp(-u/2)$

Derived properties

Derived properties of the EC TV family of distributions:

- Types of EC TV distributions
- Distribution of Reshapings
- Distribution of Tucker product
- Conditional distributions
- Moments
- EC TV Wishart distribution

Outline

MLE settings

Consider $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_n$ from $\mathcal{EC}_{\boldsymbol{m}}(0, \sigma^2\Sigma_1, \Sigma_2, \ldots, \Sigma_{\boldsymbol{p}}, \varphi)$

- Independent TVN sample.
- Uncorrelated but joinly EC-distributed sample.
- Tyler estimator robutst to EC type.
- Independent TVN scale mixture.
- Tensor-on-tensor regression (ToTR) under TVN scale mixture errors.

MLE setting IV: ToTR under TVN scale mixture errors

Consider for $i = 1, 2, \ldots, n$ the ToTR model

$$\boldsymbol{\mathcal{Y}}_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + Z_i^{-1/2} \boldsymbol{\mathcal{E}}_i, \quad \boldsymbol{\mathcal{E}}_i \sim \mathcal{N}_{\boldsymbol{m}}(0, \sigma^2 \Sigma_1, \ldots, \Sigma_p), \quad Z_i \sim Z.$$

AECM algorithm:

- **E-step:** obtain $\widehat{z}_i^{(t)} = \mathbb{E}_{\boldsymbol{\theta}^{(t)}}(Z_i | \boldsymbol{\mathcal{Y}}_i)\ \forall i = 1, 2, \ldots, n$. The conditional (on $\boldsymbol{\mathcal{Y}}_1, \ldots, \boldsymbol{\mathcal{Y}}_n$) expectation of the complete loglikelihood:

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = -\frac{n}{2} \log |\sigma^2 \Sigma| - \frac{1}{2\sigma^2} \sum_{i=1}^{n} D_{\Sigma}^2 (\mathcal{Y}_{w,i}^{(t)}, \langle \mathcal{X}_{w,i}^{(t)} | \mathcal{B} \rangle)$$

  where $(\mathcal{Y}_{w,i}^{(t)}, \mathcal{X}_{w,i}^{(t)}) = (\sqrt{\widehat{z}_i}^{(t)} \mathcal{Y}_i, \sqrt{\widehat{z}_i}^{(t)} \mathcal{X}_i)$.

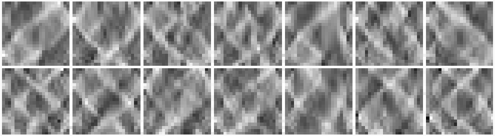
- **CM steps:** Do one iteration of the respective ToTR algorithm under TVN errors.

Outline

Predicting dogs and cats from the AFHQ database

- Animal image classification has implications in ecology.
- The animal faces HQ (AFHQ) dataset of (Choi et.al, 20) consists of 5,653 cat and 5,239 dog images, each of size 512×512.
- 500 of each animal were selected by authors for testing.
- Radon + DWT(LL) transforms were taken to each image.

$\downarrow$ Radon + DWT for 1 RGB channel $\downarrow$



Dog and cat classification: discriminant analysis

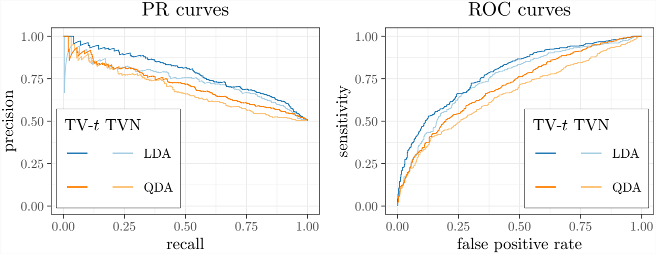
Optimal Bayes rule (under equal prior) classifies an image $\mathcal{X}$ as cat if

$$LLR(\mathcal{X}) = \log f_{cat}(\mathcal{X}) - \log f_{dog}(\mathcal{X}) > 0$$

- f_{cat} and f_{dog} are population PDFS, that can be estimated based on the training data and a statistical model.
- QDA: distinct population variances.
- LDA: equal population variances. ML fitting through TANOVA.
- We used the TVN and TV-t models.
- $\mathcal{B}$ is CP-formatted, with rank chosen using cross validation.

Dog and cat classification: PR and ROC curves

- Fitted selected models to training data: $2.5 < \widehat{\nu} < 3.6$ in all cases.
- Evaluate the LLRs at testing data, and generated PRC and ROC:



Conclusions

- We defined and characterized a family of EC TV distributions.
- We derived several properties such as moments and conditional, marginal and reshaping distributions.
- We derived ML estimation procedures under 5 scenarios, includying tensor-on-tensor regression.
- We demonstrated that considering heavier tails than Gaussian can result in better model fitting and classification performance.