Elliptically-Contoured Tensor-Variate Distributions

Carlos Llosa Vite
Joint work with Ranjan Maitra

Statistical Sciences
Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525

SAND2024-16409C

Outline

1. Elliptically-contoured TV distributions and its properties
   - Definitions and characterizations
   - Distribution of Tucker product, reshapings, conditional
   - Moments
2. Maximum likelihood estimation
3. Performance evaluations
4. Data applications
   - Labeled faces in the wild
   - Dog and cat classification

The tensor-variate normal (TVN) distribution

The TVN distribution:

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

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

- Defined through its vectorization.
- Kronecker-separable (KS) covariance structure.
- What other symmetric TV distributions can be defined through vectorization and enjoy KS covariance structure? **Elliptically-contoured tensor-variate distributions**

Spherical tensor-variate distributions

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

- $\boldsymbol{h} = (h_1, h_2, \ldots, h_p)$
- 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

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

- $D^2_{\Sigma}(\mathcal{Y}, \mathcal{M})$ 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 $\varphi(u) = \exp(-u/2)$

Types of EC TV distributions

$$\boldsymbol{\mathcal{Y}} \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1, \ldots, \Sigma_p, \varphi) \iff f_{\boldsymbol{\mathcal{Y}}}(\mathcal{Y}) = \left| \bigotimes_{k=p}^{1} \Sigma_k \right|^{-1/2} g(D^2_{\boldsymbol{\Sigma}}(\mathcal{Y}, \mathcal{M}))$$

Distribution	Additional parameters	$g(x) \propto$
Normal	–	$\exp(-x/2)$
Student's-t	$q > 0$	$(1 + q^{-1}x)^{-(q+m)/2}$
Pearson Type VII	$q > 0$	$(1 + x/q)^{-m}$
Kotz Type	$q > 0$	$x^{m-1}\exp(-qx)$
Logistic	–	$\exp(-x)/(1 + \exp(-x))^2$
Power exponential	$q > 0$	$\exp(-x^q/2)$

Scale mixture of TVN subfamily

If the DG g has the inverse Laplace transform $\lambda^{-1}[g(s)]$, then (Chu, 73)

$$\boldsymbol{\mathcal{Y}} \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1, \ldots, \Sigma_p, \varphi) \iff f_{\boldsymbol{\mathcal{Y}}}(\mathcal{Y}) = \int_0^\infty w(t) f_{\mathcal{N}_{\boldsymbol{m}}(0, t^{-1}\Sigma_1, \ldots, \Sigma_p)}(\mathcal{Y}) dt$$

- $f_{\mathcal{N}_{\boldsymbol{m}}(0, t^{-1}\Sigma_1, \ldots, \Sigma_p)}$ is the TVN PDF.
- $w(t) = (2\pi)^{m/2} t^{-m/2} \lambda^{-1}[g(2s)](t)$ is a weight function.
- If w was the PDF of a positive random variable Z, then

$$\boldsymbol{\mathcal{Y}} | (Z = z) \sim \mathcal{N}_{\boldsymbol{m}}(\mathcal{M}, z^{-1}\Sigma_1, \Sigma_2, \ldots, \Sigma_p).$$

Distribution of Reshapings

Theorem: for $m_{-k} = \prod_{q \neq k} m_q$ and $n_k = \prod_{q=1}^{k} m_q$

$$\mathcal{Y} \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1, \ldots, \Sigma_p, \varphi) \iff \text{vec}(\mathcal{Y}) \sim \mathcal{EC}_m(\text{vec}(\mathcal{M}), \bigotimes_{q=1}^{1} \Sigma_k, \varphi)$$

$$\iff \mathcal{Y}_{(k)} \sim \mathcal{EC}_{(m_k, m_{-k})}(\mathcal{M}_{(k)}, \Sigma_k, \bigotimes_{q \neq k}^{1} \Sigma_k, \varphi)$$

$$\iff \mathcal{Y}_{<k>} \sim \mathcal{EC}_{(n_k, m/n_k)}(\mathcal{M}_{<k>}, \bigotimes_{i=k}^{1} \Sigma_i, \bigotimes_{i=p}^{k+1} \Sigma_i, \varphi)$$

Distribution of Tucker product

Theorem:

$$\mathcal{Y} \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1,\ldots, \Sigma_p, \varphi) \Rightarrow$$
$$[\![\mathcal{Y}; \mathsf{A}_1,\ldots, \mathsf{A}_p]\!] \sim \mathcal{EC}_{\boldsymbol{n}}([\![\mathcal{M}; \mathsf{A}_1,\ldots, \mathsf{A}_p]\!], \mathsf{A}_1\boldsymbol{\Sigma}_1\mathsf{A}_1^\top,\ldots, \mathsf{A}_p\boldsymbol{\Sigma}_p\mathsf{A}_p^\top, \varphi)$$

- here $A_k \in \mathbb{R}^{n_k \times m_k}$ and $\boldsymbol{n} = (n_1, \ldots, n_p)$.
- Marginal distributions follow by chosing A_k appropiately

Conditional distributions

Theorem: Partition $\boldsymbol{\mathcal{Y}} \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1, \ldots, \Sigma_p, \varphi)$ and $\mathcal{M}$ over the p-th mode: $\boldsymbol{\mathcal{Y}}_1, \mathcal{M}_1 \in \mathbb{R}^{(\times_{k=1}^{p-1} m_k) \times n_1}$ and $\boldsymbol{\mathcal{Y}}_2, \mathcal{M}_2 \in \mathbb{R}^{(\times_{k=1}^{p-1} m_k) \times n_2}$ $(n_1 + n_2 = m_p)$, and let $\Sigma_{ij} \in \mathbb{R}^{n_i \times n_j}$ be the (i, j)th block of Σ_p:

$$\boldsymbol{\mathcal{Y}}_1 | (\boldsymbol{\mathcal{Y}}_2 = \mathcal{Y}_2) \sim \mathcal{EC}_{(m_1, \ldots, m_{p-1}, n_1)}\left(\bar{\mathcal{M}}, \Sigma_1, \ldots, \Sigma_{p-1}, \bar{\Sigma}_p, \bar{\varphi}\right).$$

- $\bar{\mathcal{M}} = \mathcal{M}_1 + (\mathcal{Y}_2 - \mathcal{M}_2) \times_p (\Sigma_{12}\Sigma_{22}^{-1})$
- $\bar{\Sigma}_p = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$
- $\bar{\varphi}(u) = \mathbb{E}[\ldots | \boldsymbol{\mathcal{Y}}_2 = \mathcal{Y}_2]$,
- Conditional distribution along multiple modes is possible.
- This is the TV extension of (Cambanis et.al., 81)

Moments I

Theorem: Let $\boldsymbol{\mathcal{Y}} \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \Sigma_1, \ldots, \Sigma_p, \varphi)$, then

1. $\mathbb{E}(Y_{\boldsymbol{i}}) = m_{\boldsymbol{i}}$.
2. $\mathbb{E}(Y_{\boldsymbol{i}} Y_{\boldsymbol{j}}) = m_{\boldsymbol{i}} m_{\boldsymbol{j}} - 2\varphi'(0)\sigma_{\boldsymbol{ij}}$.
3. $\mathbb{E}(Y_{\boldsymbol{i}} Y_{\boldsymbol{j}} Y_{\boldsymbol{k}}) = m_{\boldsymbol{i}} m_{\boldsymbol{j}} m_{\boldsymbol{k}} - 2\varphi'(0)(m_{\boldsymbol{i}}\sigma_{\boldsymbol{kj}} + m_{\boldsymbol{j}}\sigma_{\boldsymbol{ik}} + m_{\boldsymbol{k}}\sigma_{\boldsymbol{ij}})$.

comments:

- $\boldsymbol{i} = (i_1, \ldots, i_p)$, $\boldsymbol{j} = (j_1, \ldots, j_p)$, $\boldsymbol{k} = (k_1, \ldots, k_p)$, $\boldsymbol{l} = (l_1, \ldots, l_p)$
- $\boldsymbol{\mathcal{Y}}(i_1, \ldots, i_p) = Y_{\boldsymbol{i}}$, $\mathcal{M}(i_1, \ldots, i_p) = m_{\boldsymbol{i}}$, $\sigma_{\boldsymbol{ij}} = \prod_{q=1}^{p} \Sigma_q(i_q, j_q)$
- We describe how to derive higher moments.

Moments II

Theorem: Let $\mathcal{Y} \sim \mathcal{EC}_m(\mathcal{M}, \Sigma_1, \ldots, \Sigma_p, \varphi)$ and $D_A^2(\mathcal{X}) = \langle \mathcal{X}, [\![\mathcal{X}; \mathsf{A}_1, \ldots, \mathsf{A}_p]\!] \rangle$:

1. $\mathbb{E}(\mathcal{Y}) = \mathcal{M}$.
2. $\mathrm{Var}(\mathrm{vec}(\mathcal{Y})) = -2\varphi'(0) \bigotimes_{k=p}^{1} \Sigma_k$.
3. If $n_k = m_k$ for all $k = 1, 2, \ldots, p$, then $\mathbb{E}\left(D_A^2(\mathcal{Y})\right) = D_A^2(\mathcal{M}) - 2\varphi'(0) \prod_{k=1}^{p} \mathrm{tr}(\Sigma_k \mathsf{A}_k^\top)$.
4. If $\mathcal{V}$ is of size $n_1 \times n_2 \times \ldots \times n_p$, then
$$\mathbb{E}(\langle \mathcal{V}, [\![\mathcal{Y}; \mathsf{A}_1, \ldots, \mathsf{A}_p]\!] \rangle \mathcal{Y}) = \langle \mathcal{V}, [\![\mathcal{M}; \mathsf{A}_1, \ldots, \mathsf{A}_p]\!] \rangle \mathcal{M} - 2\varphi'(0) [\![\mathcal{V}; \Sigma_1 \mathsf{A}_1^\top, \ldots, \Sigma_p \mathsf{A}_p^\top]\!].$$
5. If $n_k = m_k$ for all $k = 1, 2, \ldots, p$, then
$$\mathbb{E}\big(D_A^2(\mathcal{Y})\mathcal{Y}\big) = D_A^2(\mathcal{M})\mathcal{M} - 2\varphi'(0)\Big[\prod_{k=1}^{p} \mathrm{tr}(\Sigma_k \mathsf{A}_k^\top)\mathcal{M} + [\![\mathcal{M}; \Sigma_1 \mathsf{A}_1, \ldots, \Sigma_p \mathsf{A}_p]\!] + [\![\mathcal{M}; \Sigma_1 \mathsf{A}_1^\top, \ldots, \Sigma_p \mathsf{A}_p^\top]\!]\Big].$$
6. If $\mathcal{M} = 0$, $\varphi^{(4)}(0) < \infty$ and $n_k = h_k = m_k$ for all $k = 1, 2, \ldots, p$, then
$$\mathbb{E}\big(D_A^2(\mathcal{Y}) D_B^2(\mathcal{Y})\big) = 4\varphi''(0)\Big[\prod_{k=1}^{p} \{\mathrm{tr}(\mathsf{A}_k \Sigma_k)\, \mathrm{tr}(\mathsf{B}_k \Sigma_k)\} + \prod_{k=1}^{p} \mathrm{tr}(\mathsf{A}_k \Sigma_k \mathsf{B}_k^\top \Sigma_k) + \prod_{k=1}^{p} \mathrm{tr}(\mathsf{A}_k \Sigma_k \mathsf{B}_k \Sigma_k)\Big].$$

Outline

MLE setting I: Independent TVN sample

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

- MLE studied by (Dutilleul 13, Hoff 11, Ohlson 13, Akdemir 11).
- Let $S_k = \sum_{i=1}^{n} \mathcal{Y}_{i(k)} \Sigma_{-k}^{-1} \mathcal{Y}'_{i(k)}$
- MLE without regard for identifiability:

$$\widehat{\Sigma}_k = S_k/(nm_{-k}\sigma^2)$$

- Constrained optimization under $\Sigma_k(1,1) = 1$:

$$\widehat{\Sigma}_k = ADJUST(nm_{-k}, \sigma^2, S_k).$$

- ADJUST procedure of Glanz and Carvalho (JMA, 18).
- $\widehat{\sigma}^2 = \mathrm{tr}(S_k \widehat{\Sigma}_k^{-1})/(nm)$

MLE setting II: Uncorrelated sample

Theorem: Let $\boldsymbol{\mathcal{Y}} \sim \mathcal{EC}_{[\boldsymbol{m},n]^\top}(\mathcal{M}_n, \sigma^2\Sigma_1, \Sigma_2, \ldots, \Sigma_p, I_n, \varphi)$. Then:

1. $\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{Y}} \times_p \boldsymbol{e}_i^m \sim \mathcal{EC}_{\boldsymbol{m}}(\mathcal{M}, \sigma^2\Sigma_1, \Sigma_2, \ldots, \Sigma_p, \varphi)$
2. $\mathbb{E}\left(\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i - \mathcal{M})\,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_j - \mathcal{M})^\top\right) = 0$

Consider an uncorrelated sample $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ from $\boldsymbol{\mathcal{Y}}$. Then the MLE of $(\mathcal{M}, \Sigma_1, \ldots, \Sigma_p)$ are the same to that under $\boldsymbol{\mathcal{Y}}_i \overset{iid}{\sim} \mathcal{N}_{\boldsymbol{m}}(\mathcal{M}, \sigma^2\Sigma_1, \ldots, \Sigma_p)$.

- Provided $h(d) = d^{nm/2} g(d)$ has a finite positive maximum d_g.
- $\widehat{\sigma}^2 = \frac{nm}{d_g}\tilde{\sigma}^2$, where $\tilde{\sigma}^2$ is the MLE under TVN.
- This is the TV extension of (Anderson et.al., 86).

MLE setting III: Independent TVN scale mixture

Consider $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_n$ iid realizations from

$$\boldsymbol{\mathcal{Y}}|(Z = z) \sim \mathcal{N}_{\boldsymbol{m}}(\mathcal{M}, \frac{\sigma^2}{z}\Sigma_1, \Sigma_2, \ldots, \Sigma_p), \quad Z \sim P$$

Conditional (on $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$) expectation of the complete loglikelihood:

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = -\frac{n}{2}\log|\sigma^2\Sigma| - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\widehat{z}_i^{(t)} D^2_{\Sigma}(\mathcal{Y}_i, \mathcal{M}),$$

AECM algorithm:

- **E-step:** obtain $\widehat{z}_i^{(t)} = \mathbb{E}_{\boldsymbol{\theta}^{(t)}}(Z_i|\boldsymbol{\mathcal{Y}}_i)\ \forall i = 1, 2, \ldots, n$.
- **CM step 1:**

$$\widehat{\mathcal{M}}^{(t+1)} = (\sum_{i=1}^{n}\widehat{z}_i^{(t)}\mathcal{Y}_i)/(\sum_{i=1}^{n}\widehat{z}_i^{(t)}).$$

- **Remaining ECM steps:** Take $\mathcal{Y}_{w,i}^{(t)} = \sqrt{\widehat{z}_i}^{(t)}(\mathcal{Y}_i - \widehat{\mathcal{M}}^{(t+1)})$ for $i = 1, 2, \ldots, n$ and do one iteration of the iterative TVN algorithm.

MLE setting IV: ToTR under TVN scale mixture errors

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

$$\mathcal{Y}_i = \langle \mathcal{X}_i | \mathcal{B} \rangle + \mathcal{E}_i,$$

where $\mathcal{E}_i$ follow a TVN scale mixture. Conditional (on $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$) expectation of the complete loglikelihood:

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

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

AECM algorithm:

- **E-step:** obtain $\widehat{z}_i^{(t)} = \mathbb{E}_{\boldsymbol{\theta}^{(t)}}(Z_i | \boldsymbol{\mathcal{Y}}_i)$ and $(\mathcal{Y}^{(t)}_{w,i}, \mathcal{X}^{(t)}_{w,i})$ $\forall i = 1, 2, \ldots, n$.
- **CM steps:** Do one iteration of the respective ToTR algorithm under TVN errors.

MLE setting V: Robust Tyler Estimator

If $\mathcal{Y} \sim \mathcal{EC}_m(0, \sigma^2\Sigma_1,\ldots,\Sigma_p, \varphi)$, then $\mathcal{Z} = \mathcal{Y}/||\mathcal{Y}||$ has PDF

$$f_{\mathcal{Z}}(\mathcal{Z}) = \frac{\Gamma(m/2)}{2\pi^{m/2}}|\Sigma|^{-1/2} D^2_{\Sigma}(\mathcal{Z}, 0)^{-m/2}.$$

Consider realizations $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ realizations, then for fixed Σ_{-k} and $S_{ik} = \mathcal{Y}_{i(k)}\Sigma_{-k}^{-1}\mathcal{Y}_{i(k)}^{\top}$, a constrained fixed-point iteration algorithm:

$$\Sigma_{k,(t+1)} = ADJUST\left(\frac{n}{m_k}, 1, \sum_{i=1}^{n} \frac{S_{ik}}{\mathrm{tr}(\Sigma_{k,(t)}^{-1} S_{ik})}\right).$$

Iterative algorithm:

(1) perform the fixed-point algorithm with lax convergence for Σ_1

$\vdots$

(p) perform the fixed-point algorithm with lax convergence for Σ_p

$(p+1)$ return to step 1 or stop if convergence is met.

Outline

Simulation I: setup

TVN model: $\boldsymbol{\mathcal{Y}}_i \sim \mathcal{N}_{\boldsymbol{m}}(\mathcal{M}, \sigma^2\Sigma_1, \ldots, \Sigma_6)$

Gamma scale mixture (GSM) model:

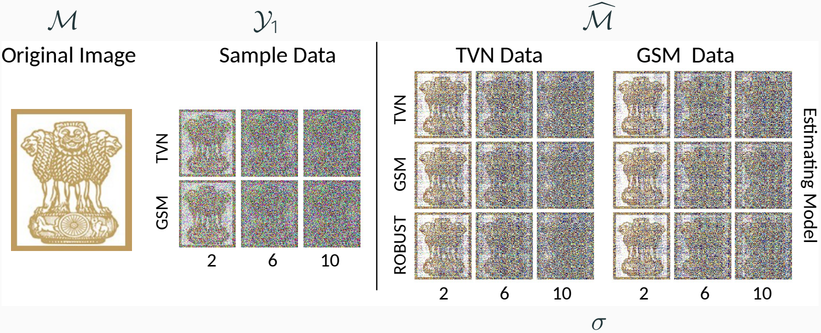
$$\boldsymbol{\mathcal{Y}}_i | (Z_i = z_i) \sim \mathcal{N}_{\boldsymbol{m}}(\mathcal{M}, \frac{a-2}{bz_i}\sigma^2\Sigma_1, \ldots, \Sigma_6), \quad Z_i \sim Gamma(a/2, b/2)$$

- $(a, b) = (3, 15)$, $n = 10$ and $\sigma = 2, 6, 10$. In E-step we need :

$$\widehat{z}_i = \mathbb{E}[Z_i | (\boldsymbol{\mathcal{Y}}_i = \mathcal{Y}_i)] = \frac{m + a}{D^2_{\sigma^2\boldsymbol{\Sigma}}(\mathcal{Y}_i, \widehat{\mathcal{M}}) + b}$$

- $\boldsymbol{m} = (7, 9, 3, 23, 7, 3)$. Σ_k taken from $\mathcal{W}_{m_k}(m_k, I_{m_k})$.
- Rearranged $\mathcal{M}$ is color image of size $(7 * 9 * 3) \times (23 * 7) \times 3$
- Generated from TVN and GSM, fitted TVN, GSM and Tyler models.

Simulation II: estimate of $\mathcal{M}$



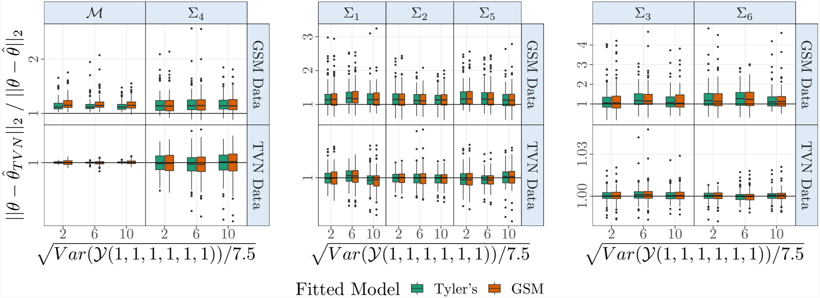
- Tyler also proposed a fixed-point algorithm for $\mathcal{M}$ that we adapt as

$$\widehat{\mathcal{M}}_{(t+1)} = \left(\sum_{i=1}^{n} \mathcal{X}_i / D_\Sigma(\mathcal{X}_i, \widehat{\mathcal{M}}_{(t)})\right) / \left(\sum_{i=1}^{n} 1 / D_\Sigma(\mathcal{X}_i, \widehat{\mathcal{M}}_{(t)})\right).$$

- Sample data and estimates are noisier with increasing σ
- $\widehat{\mathcal{M}}$ looks slightly noisier for TVN fit on GSM data

Simulation III: comparison against TVN estimate of Σ

100 distinct Σ_k from $\mathcal{W}_{m_k}(m_k, I_{m_k})$ for each setting and $k = 1, 2, \ldots, 6$.



- $\Sigma_4 \in \mathbb{R}^{23\times 23}$, $\Sigma_2 \in \mathbb{R}^{9\times 9}$, $\Sigma_1, \Sigma_5 \in \mathbb{R}^{7\times 7}$, $\Sigma_3, \Sigma_6 \in \mathbb{R}^{3\times 3}$.
- TVN data is equally fitted by all models.
- TVN model is outperformed for GSM data.

Outline

Labeled faces in the wild: Data

- We selected 605 facial images from more than 13,000.
- Three ethnic origins: African, European and Asian.
- Four cohorts: child, youth, middle-aged and senior.
- Two genders: male and female.

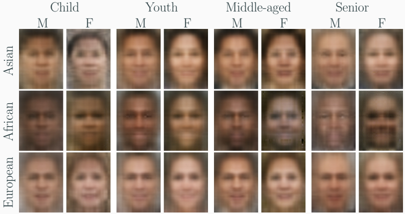
Labeled faces in the wild: model comparison

$$\mathcal{Y}_{ijkl} = \langle \mathcal{X}_{ijk} | \mathcal{B} \rangle + \mathcal{E}_{ijkl}, \quad \mathcal{E}_{ijkl} \sim t_{\boldsymbol{m}_3}(\nu; 0, \sigma^2 \Sigma_1, \Sigma_2, \Sigma_3), \tag{1}$$

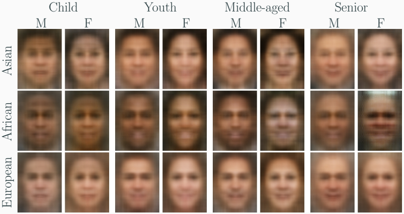
- covariate $\mathcal{X}_{ijk} \in \mathbb{R}^{2\times3\times4}$ is 0-1 and $\mathcal{B} \in \mathbb{R}^{2\times3\times4\times151\times111\times3}$.
- Response $\mathcal{Y}_{ijkl}$ and mean $\langle \mathcal{X}_{ijk} | \mathcal{B} \rangle \in \mathbb{R}^{151\times111\times3}$.
- Used the TT and CP formats, along with unformatted $\mathcal{B}$.
- compared them against the model with TVN responses of chapter 2.
- estimated DF of 0.88 (BIC of -24,065,217), so fixed at 2.01.

		Format of $\mathcal{B}$		
		TT	CP	none
Error Model	TVN	-11,190,861	-11,223,595	-4,745,807
	TV-t	-23,992,185	-24,064,833	-20,043,342

Labeled faces in the wild: $\mathcal{B}$ estimate



(a) TV-t (2.01 DF)

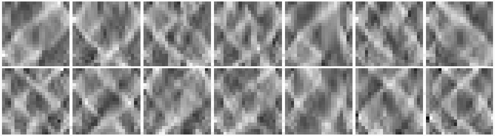


(b) TVN

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

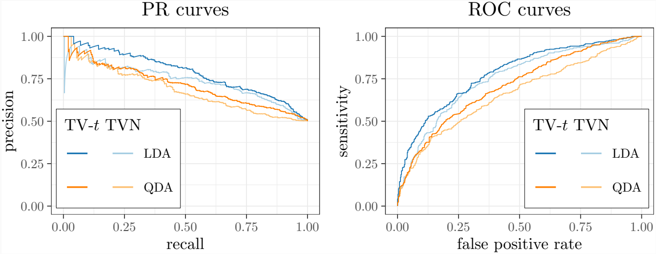
Optimal Bayes rule (under equal prior) classifies an image $\mathcal{X}$ as cat if

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

- 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 properties such as moments and conditional, marginal and reshaping distributions.
- We derived ML estimation procedures under 5 scenarios, and compared them in a performance evaluation.
- We demonstrated that considering heavier tails than Gaussian can result in better model fitting and classification performance.