# Multivariate SUN, SUNNY, and SUNSET Distributions: Properties and Applications

**Marc G. Genton**

*marc.genton@kaust.edu.sa*

*stsds.kaust.edu.sa*
*stat.kaust.edu.sa*

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

STATISTICS @KAUST

January 7, 2026

# Outline

1. THE BEGINNING

2. Selection Multivariate Distributions
   - SUN, SUNNY, SUT, SUE and properties
   - On parameterizations
   - SUNSET

3. THE END?

# 1. THE BEGINNING

- **1996**: Ph.D. from EPFL on Robustness in Spatial Statistics
- **1997**: Lecturer in Statistics at MIT in Boston
- **2000**: Seminar at University of Connecticut, met **Dipak Dey** who mentioned preprint with **Marcia Branco**
- **2001**: First paper in SPL with two students on moments of quadratic forms in SN random vectors
- **2001**: Attended VIII Latin American congress in Probability and Mathematical Statistics, Havana, Cuba, and met **Reinaldo Arellano** and **Rosangela Loschi**
- **2003**: 8a Escola de Modelos de Regressao, Conservatoria, Rio de Janeiro, Brazil, and met **Marcia Branco**
- **2004**: Edited book: Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality
- **2004**: IX CLAPEM, Punta del Este, Uruguay
- **2004**: ISBA, Vina del Mar, Chile
- **2004**: CLATSE VI, Concepcion, Chile; probably met **Victor Lachos** during this period
- **2004**: Workshop on skew distributions, Guanajuato, Mexico, and met **Arthur Pewsey** and **Nicola Loperfido**
- **2005-2006**: met **Adelchi Azzalini**, perhaps before when I was a student at EPFL; also met **Brunero Liseo** and **Bruno Scarpa**, and probably later **Antonio Canale**
- ...
- **2013**: Wrote to **Daniele Durante** for PhD fellowship

- Given two random vectors $\mathbf{U}_0 \in \mathbb{R}^m$ and $\mathbf{U}_1 \in \mathbb{R}^d$, and a subset $C \subset \mathbb{R}^m$
- Selection distribution: $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 \in C)$
- Cumulative distribution function (cdf):

$$F_{\mathbf{Z}}(\mathbf{z}) = \Pr(\mathbf{U}_1 \leq \mathbf{z} | \mathbf{U}_0 \in C) = \frac{\Pr(\mathbf{U}_0 \in C, \mathbf{U}_1 \leq \mathbf{z})}{\Pr(\mathbf{U}_0 \in C)}, \quad \mathbf{z} \in \mathbb{R}^d \tag{1}$$

- Probability density function (pdf):

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{U}_1 | \mathbf{U}_0 \in C}(\mathbf{z}) = f_{\mathbf{U}_1}(\mathbf{z}) \frac{\Pr(\mathbf{U}_0 \in C | \mathbf{U}_1 = \mathbf{z})}{\Pr(\mathbf{U}_0 \in C)}, \quad \mathbf{z} \in \mathbb{R}^d \tag{2}$$

  where $f_{\mathbf{U}_1}(\mathbf{z})$ is the pdf of $\mathbf{U}_1$

- The selection pdf (2) may also be motivated by:

$$\Pr(\mathbf{U}_0 \in C) = \mathrm{E}\{\Pr(\mathbf{U}_0 \in C | \mathbf{U}_1)\} = \int_C \Pr(\mathbf{U}_0 \in C | \mathbf{U}_1 = \mathbf{z}) f_{\mathbf{U}_1}(\mathbf{z}) \mathrm{d}\mathbf{z}$$

- More details in Arellano-Valle et al. (2006)

# SUN Distributions

- Particular selection distribution: unified skew-normal (SUN) distribution (Arellano-Valle and Azzalini, 2006)
- Selection distribution: $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$ with $\boldsymbol{\tau} \in \mathbb{R}^m$ (check: $m = 1$; $\boldsymbol{\tau} = \mathbf{0}$; $\boldsymbol{\Delta} = \mathbf{O}$)
- Joint multivariate normal distribution (with $\bar{\boldsymbol{\Omega}}^* > 0$):

$$\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{N}_{m+d}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \bar{\boldsymbol{\Omega}} \end{pmatrix} \right)$$

where $\bar{\boldsymbol{\Gamma}}$ and $\bar{\boldsymbol{\Omega}}$ are correlation matrices of $\mathbf{U}_0$ and $\mathbf{U}_1$, and $\boldsymbol{\Delta}$ is the correlation matrix between $\mathbf{U}_0$ and $\mathbf{U}_1$

- $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, with diagonal $\boldsymbol{\omega} > 0$ scale matrix and $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$, denoted by $\mathbf{Y} \sim \mathcal{SUN}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$
- Probability density function:

$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\omega}^{-1}| f_{\mathbf{Z}}\{\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})\} = \phi_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \frac{\Phi_m\left\{\boldsymbol{\tau} + \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\Delta}\right\}}{\Phi_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}})}, \quad \mathbf{y} \in \mathbb{R}^d$$

- Cumulative distribution function:

$$F_{\mathbf{Y}}(\mathbf{y}) = \Pr\{\mathbf{Z} \leq \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})\} = \frac{\Phi_{d+m}(\mathbf{y}_* - \boldsymbol{\xi}_*; \boldsymbol{\Omega}_*)}{\Phi_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}})}, \quad \mathbf{y}_* = \begin{pmatrix} \boldsymbol{\tau} \\ \mathbf{y} \end{pmatrix}, \boldsymbol{\xi}_* = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\xi} \end{pmatrix}, \boldsymbol{\Omega}_* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & -\boldsymbol{\Delta}^\top \boldsymbol{\omega} \\ -\boldsymbol{\omega}\boldsymbol{\Delta} & \boldsymbol{\Omega} \end{pmatrix}$$

# The Non-identifiability Problem of SUN Distributions (Wang et al., 2023)

- Let $\mathbf{P} \in \mathcal{P}(m) = \{\mathbf{P} \in \mathbb{R}^{m\times m} | \mathbf{P}\mathbf{P}^\top = \mathbf{P}^\top\mathbf{P} = \mathbf{I}_m \text{ and } \mathbf{P}\mathbf{1}_m = \mathbf{1}_m\}$ and $\mathbf{Z}_\mathbf{P} = (\mathbf{U}_1 | \mathbf{P}\mathbf{U}_0 + \mathbf{P}\boldsymbol{\tau} > \mathbf{0})$ with

$$\mathbf{U}_\mathbf{P} = \begin{pmatrix} \mathbf{P}\mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{N}_{m+d}(\mathbf{0}, \bar{\boldsymbol{\Omega}}^*_\mathbf{P}), \qquad \bar{\boldsymbol{\Omega}}^*_\mathbf{P} = \begin{pmatrix} \mathbf{P}\bar{\boldsymbol{\Gamma}}\mathbf{P}^\top & \mathbf{P}\boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta}\mathbf{P}^\top & \bar{\boldsymbol{\Omega}} \end{pmatrix}$$

- Then, with $\mathbf{Y}_\mathbf{P} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}_\mathbf{P}$, we have $\mathbf{Y}_\mathbf{P} \sim \mathcal{SUN}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}_\mathbf{P}, \boldsymbol{\tau}_\mathbf{P}, \bar{\boldsymbol{\Gamma}}_\mathbf{P})$ with pdf:

$$f_\mathbf{P}(\mathbf{y}) = \phi_d(\mathbf{y} - \boldsymbol{\xi}; \boldsymbol{\Omega}) \frac{\Phi_m(\boldsymbol{\tau}_\mathbf{P} + \boldsymbol{\Delta}_\mathbf{P}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \bar{\boldsymbol{\Gamma}}_\mathbf{P} - \boldsymbol{\Delta}_\mathbf{P}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}_\mathbf{P})}{\Phi_m(\boldsymbol{\tau}_\mathbf{P}; \bar{\boldsymbol{\Gamma}}_\mathbf{P})}$$

  and

$$\boldsymbol{\tau}_\mathbf{P} + \boldsymbol{\Delta}_\mathbf{P}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}) = \mathbf{P}\boldsymbol{\tau} + \mathbf{P}\boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}) = \mathbf{P}\{\boldsymbol{\tau} + \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})\},$$
$$\bar{\boldsymbol{\Gamma}}_\mathbf{P} - \boldsymbol{\Delta}_\mathbf{P}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}_\mathbf{P} = \mathbf{P}\bar{\boldsymbol{\Gamma}}\mathbf{P}^\top - \mathbf{P}\boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}\mathbf{P}^\top = \mathbf{P}(\bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta})\mathbf{P}^\top$$

- Since we only apply permutations, we have:

$$\Phi_m(\boldsymbol{\tau} + \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}) = \Phi_m(\boldsymbol{\tau}_\mathbf{P} + \boldsymbol{\Delta}_\mathbf{P}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \bar{\boldsymbol{\Gamma}}_\mathbf{P} - \boldsymbol{\Delta}_\mathbf{P}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}_\mathbf{P})$$

- So $f(\mathbf{y}) = f_\mathbf{P}(\mathbf{y})$ for all $\mathbf{y} \in \mathbb{R}^d$, which implies that $\mathcal{SUN}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}) \equiv \mathcal{SUN}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}_\mathbf{P}, \boldsymbol{\tau}_\mathbf{P}, \bar{\boldsymbol{\Gamma}}_\mathbf{P})$ for any permutation matrix $\mathbf{P}$, confirming the non-identifiability claim

# Limitations of SUN Distributions

Despite flexibility and appealing properties, the SUN has some limitations for practical use in applications:

(a) The model $\mathcal{SUN}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ has a large number of parameters, $d(d+3)/2 + dm + m(m+1)/2$, which increases as $\mathcal{O}(m^2)$ as the number $m$ of latent variables increases. In particular, the latent matrices $\boldsymbol{\Delta}$ and $\bar{\boldsymbol{\Gamma}}$ seem to contain too many parameters

(b) The SUN pdf requires the computation of $\Phi_m$, the cdf of an $m$-dimensional multivariate normal distribution, which can become problematic for large $m$

(c) The geometric interpretation of the effects of the latent parameters $\boldsymbol{\Delta}$ and $\bar{\boldsymbol{\Gamma}}$ on the resulting skewness and shape of the SUN distribution is somewhat unclear

(d) The specification of $\boldsymbol{\Delta}$ is not trivial since it must ensure that $\bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\Delta}$ is positive definite, as discussed in Arellano-Valle and Azzalini (2022)

(e) Without further constraints on the latent parameters, the SUN model is generally non-identifiable as pointed out by Wang et al. (2023)

- Particular SUN distribution: unified skew-normal with new youth (SUNNY) distribution

$$\mathcal{SUNNY}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, m) \equiv \mathcal{SUN}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \mathbf{0}_m, \bar{\boldsymbol{\Gamma}})$$

$$\text{with} \qquad \boldsymbol{\Delta} = \bar{\boldsymbol{\Omega}} \frac{\boldsymbol{\alpha}}{\sqrt{1+\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}}} \mathbf{1}_m^\top, \quad \bar{\boldsymbol{\Gamma}} = \frac{1}{1+\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}} \mathbf{I}_m + \frac{\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}}{1+\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}} \mathbf{1}_m \mathbf{1}_m^\top,$$

- Straightforward calculations lead to:

$$\boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1} = \frac{\mathbf{1}_m \boldsymbol{\alpha}^\top}{\sqrt{1+\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}}}, \qquad \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\Delta} = \frac{1}{1+\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}} \mathbf{I}_m$$

- Probability density function:

$$f(\mathbf{y}) = \frac{1}{\Phi_m(\mathbf{0}_m; \bar{\boldsymbol{\Gamma}})} \phi_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \left[ \Phi\{\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\xi})\} \right]^m, \quad \mathbf{y} \in \mathbb{R}^d$$

where $\Phi_m(\mathbf{0}_m; \bar{\boldsymbol{\Gamma}}) = K_m\left(\sqrt{\boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}}\right)$ with $K_m(\theta) = \int_{-\infty}^{+\infty} \{\Phi(\theta t)\}^m \phi(t) \mathrm{d}t = \mathrm{E}[\{\Phi(\theta X)\}^m]$ and $X \sim \mathcal{N}(0,1)$ and $K_1(\theta) = 1/2$, $K_2(\theta) = \pi^{-1}\mathrm{atan}(\sqrt{1+2\theta^2})$, $K_3(\theta) = 3/(2\pi)\mathrm{atan}(\sqrt{1+2\theta^2}) - 1/4$

- $d = 1$: Balakrishnan (2002); $d = 2$: Sharafi & Behboodian (2008); mentioned in Azzalini & Capitanio(2014)

- Particular selection distribution: unified skew-$t$ (SUT) distribution (Wang et al., 2024)
- Selection distribution: $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$ with $\boldsymbol{\tau} \in \mathbb{R}^m$
- Joint multivariate $t$ distribution (with $\bar{\boldsymbol{\Omega}}^* > 0$):

$$\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{T}_{m+d}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \bar{\boldsymbol{\Omega}} \end{pmatrix}, \nu \right)$$

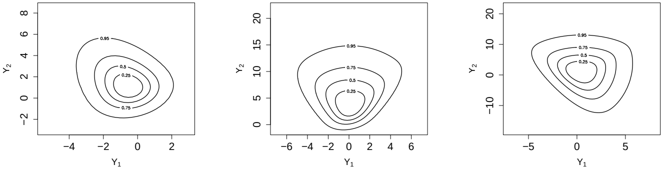
where $\bar{\boldsymbol{\Gamma}}$ and $\bar{\boldsymbol{\Omega}}$ are correlation matrices of $\mathbf{U}_0$ and $\mathbf{U}_1$, and $\boldsymbol{\Delta}$ is the correlation matrix between $\mathbf{U}_0$ and $\mathbf{U}_1$

- $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, with diagonal $\boldsymbol{\omega} > 0$ scale matrix and $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$, denoted by $\mathbf{Y} \sim \mathcal{SUT}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, \nu)$
- Probability density function ($\alpha_{\nu,Q_\mathbf{y}} = (\nu + Q_\mathbf{y})/(\nu + d)$ with $Q_\mathbf{y} = (\mathbf{y} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})$):

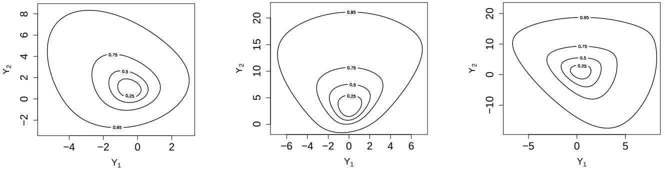
$$f_\mathbf{Y}(\mathbf{y}) = t_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) \frac{T_m\left[\alpha_{\nu,Q_\mathbf{y}}^{-1/2}\left\{\boldsymbol{\tau} + \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})\right\}; \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}, \nu + d\right]}{T_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, \nu)}, \quad \mathbf{y} \in \mathbb{R}^d$$

- Cumulative distribution function:

$$F_\mathbf{Y}(\mathbf{y}) = \frac{T_{d+m}(\mathbf{y}_* - \boldsymbol{\xi}_*; \boldsymbol{\Omega}_*, \nu)}{T_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, \nu)}, \quad \mathbf{y} \in \mathbb{R}^d$$




(a) $\mathcal{SUN}_{2,1} \equiv \mathcal{SN}_2$ (b) $\mathcal{SUN}_{2,2}$ (c) $\mathcal{SUN}_{2,3}$




(d) $\mathcal{SUT}_{2,1} \equiv \mathcal{ST}_2$ (e) $\mathcal{SUT}_{2,2}$ (f) $\mathcal{SUT}_{2,3}$

Figure 1: Contours of the pdfs of the bivariate ($d = 2$) SUN and SUT (at $\nu = 5$) distributions with the same parameter specifications with $(\boldsymbol{\Omega}, \boldsymbol{\Delta}, \bar{\boldsymbol{\Gamma}})$ specified to make the distribution skewed in the direction $(-1, 2)^\top$ ($m = 1$), $\{(-1, 2)^\top, (1, 2)^\top\}$ ($m = 2$), and $\{(-1, 2)^\top, (1, 2)^\top, (1, -6)^\top\}$ ($m = 3$)

# SUE Distributions

- Particular selection distribution: unified skew-elliptical (SUE) distribution
  see Arellano-Valle and Azzalini (2006) and, for another parameterization, Arellano-Valle and Genton (2010)
- Selection distribution: $\mathbf{Z} = (\mathbf{U}_1 | \mathbf{U}_0 + \boldsymbol{\tau} > \mathbf{0})$ with $\boldsymbol{\tau} \in \mathbb{R}^m$
- Joint multivariate elliptical distribution (with $\bar{\boldsymbol{\Omega}}^* > 0$):

$$\begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{E}_{m+d}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} \bar{\boldsymbol{\Gamma}} & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \bar{\boldsymbol{\Omega}} \end{pmatrix}, g^{(m+d)} \right)$$

- $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, with diagonal $\boldsymbol{\omega} > 0$ scale matrix and $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$, denoted $\mathbf{Y} \sim \mathcal{SUE}_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, g^{(m+d)})$
- Probability density function ($Q_{\mathbf{y}} = (\mathbf{y} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\xi})$):

$$f_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(d)}) \frac{F_m\left[\boldsymbol{\tau} + \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\omega}^{-1} (\mathbf{y} - \boldsymbol{\xi}); \bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\Delta}, g^{(m)}_{Q_{\mathbf{y}}}\right]}{F_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, g^{(m)})}, \quad \mathbf{y} \in \mathbb{R}^d$$

- Cumulative distribution function:

$$\frac{F_{d+m}(\mathbf{y}_* - \boldsymbol{\xi}_*; \boldsymbol{\Omega}_*, g^{(m+d)})}{F_m(\boldsymbol{\tau}; \bar{\boldsymbol{\Gamma}}, g^{(m)})}, \quad \mathbf{y} \in \mathbb{R}^d$$

# Properties of SUE Distributions

- Properties of SUE in another parameterization: Arellano-Valle and Genton (2010)
- Properties of SUE in this parameterization: Karling et al. (2024)

- SUE closed under linear transformation (known before in another parameterization)
- SUE closed under marginalization (known before in another parameterization)
- SUE closed under conditioning (known before in another parameterization)

- Several new SUE properties:
- If $\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim \mathcal{SUE}$ then $(\mathbf{Y}_1 | \mathbf{Y}_2 > \mathbf{0}) \sim \mathcal{SUE}$
- Let $\mathbf{\Gamma}$ be positive-definite, then $\mathcal{SUE}_{d,m}(\boldsymbol{\xi}, \mathbf{\Omega}, \mathbf{\Delta}, \boldsymbol{\tau}, \mathbf{\Gamma}, g^{(m+d)}) \stackrel{d}{=} \mathcal{SUE}_{d,m}(\boldsymbol{\xi}, \mathbf{\Omega}, \mathbf{\Delta}\boldsymbol{\gamma}^{-1}, \boldsymbol{\gamma}^{-1}\boldsymbol{\tau}, \bar{\mathbf{\Gamma}}, g^{(m+d)})$ where $\bar{\mathbf{\Gamma}}$ is a Pearson correlation matrix defined as $\bar{\mathbf{\Gamma}} = \boldsymbol{\gamma}^{-1}\mathbf{\Gamma}\boldsymbol{\gamma}^{-1}$, with $\boldsymbol{\gamma} = \text{diag}(\mathbf{\Gamma})^{1/2}$
- If $\mathbf{\Delta} = \mathbf{O}$ and $\boldsymbol{\tau} = \mathbf{0}$ then $\mathcal{SUE}$ reduces to $\mathcal{E}$

SUN case:

- Durante (2019): Probit regression
  $y \in \{0,1\}$, $\mathbf{x} \in \mathbb{R}^p$, $\mathrm{pr}(y=1|\mathbf{x},\boldsymbol{\beta}) = \Phi(\mathbf{x}^\top\boldsymbol{\beta})$ with $\boldsymbol{\beta} \in \mathbb{R}^p$
  If prior $\boldsymbol{\beta} \sim \mathcal{N}_p$ then posterior $\boldsymbol{\beta} \sim \mathcal{SUN}_{p,n}$
  If prior $\boldsymbol{\beta} \sim \mathcal{SUN}_{p,m}$ then posterior $\boldsymbol{\beta} \sim \mathcal{SUN}_{p,m+n}$

- Anceschi et al. (2023):
  - Linear regression and multivariate linear regression
  - Probit, multivariate probit and multinomial probit
  - Tobit regression
  - Extensions to skewed, nonlinear and dynamic models

Extension to SUE case:

- Zhang et al. (2023): Multivariate skew-elliptical link model for correlated binary responses

# Conjugate Bayesian Inference based on SUE: Multivariate Linear Models

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$ is the response vector, $\mathbf{X} \in \mathbb{R}^{n\times p}$ corresponds to a known design matrix, $\boldsymbol{\beta} \in \mathbb{R}^p$ denotes a vector of unknown parameters, and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is the error vector (Karling et al., 2024)

**Proposition 1.** *Assume that* $(\boldsymbol{\beta}^\top, \boldsymbol{\varepsilon}^\top)^\top \sim \mathrm{SUE}_{p+n,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, g^{(p+n+q)})$ *with parameters partitioned as*

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_\beta \\ \boldsymbol{\xi}_\varepsilon \end{bmatrix}, \qquad \boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_\beta & \boldsymbol{\Omega}_{\beta\varepsilon} \\ \boldsymbol{\Omega}_{\varepsilon\beta} & \boldsymbol{\Omega}_\varepsilon \end{bmatrix}, \qquad \boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\Delta}_\beta \\ \boldsymbol{\Delta}_\varepsilon \end{bmatrix}.$$

*Then, when* $\mathbf{y}$ *is defined as* above, *it follows that* $(\boldsymbol{\beta}^\top, \mathbf{y}^\top)^\top \sim \mathrm{SUE}_{p+n,q}(\boldsymbol{\xi}^\dagger, \boldsymbol{\Omega}^\dagger, \boldsymbol{\Delta}^\dagger, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, g^{(p+n+q)})$, *with*

$$\boldsymbol{\xi}^\dagger = \begin{bmatrix} \boldsymbol{\xi}_\beta \\ \mathbf{X}\boldsymbol{\xi}_\beta + \boldsymbol{\xi}_\varepsilon \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\xi}_\beta \\ \boldsymbol{\xi}_\mathrm{y} \end{bmatrix}, \qquad \boldsymbol{\Delta}^\dagger = \begin{bmatrix} \boldsymbol{\Delta}_\beta \\ \boldsymbol{\omega}_\mathrm{y}^{-1}(\mathbf{X}\boldsymbol{\omega}_\beta\boldsymbol{\Delta}_\beta + \boldsymbol{\omega}_\varepsilon\boldsymbol{\Delta}_\varepsilon) \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\Delta}_\beta \\ \boldsymbol{\Delta}_\mathrm{y} \end{bmatrix},$$

$$\boldsymbol{\Omega}^\dagger = \begin{bmatrix} \boldsymbol{\Omega}_\beta & \boldsymbol{\Omega}_\beta\mathbf{X}^\top + \boldsymbol{\Omega}_{\beta\varepsilon} \\ \mathbf{X}\boldsymbol{\Omega}_\beta + \boldsymbol{\Omega}_{\varepsilon\beta} & \mathbf{X}\boldsymbol{\Omega}_\beta\mathbf{X}^\top + \boldsymbol{\Omega}_{\varepsilon\beta}\mathbf{X}^\top + \mathbf{X}\boldsymbol{\Omega}_{\beta\varepsilon} + \boldsymbol{\Omega}_\varepsilon \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\Omega}_\beta & \boldsymbol{\Omega}_{\beta\mathrm{y}} \\ \boldsymbol{\Omega}_{\mathrm{y}\beta} & \boldsymbol{\Omega}_\mathrm{y} \end{bmatrix},$$

*where* $\boldsymbol{\omega}_\varepsilon = \mathrm{diag}(\boldsymbol{\Omega}_\varepsilon)^{1/2}$, $\boldsymbol{\omega}_\beta = \mathrm{diag}(\boldsymbol{\Omega}_\beta)^{1/2}$, *and* $\boldsymbol{\omega}_\mathrm{y} = \mathrm{diag}(\boldsymbol{\Omega}_\mathrm{y})^{1/2}$. *Moreover*

(a) *Prior distribution:* $\boldsymbol{\beta} \sim \mathrm{SUE}_{p,q}(\boldsymbol{\xi}_\beta, \boldsymbol{\Omega}_\beta, \boldsymbol{\Delta}_\beta, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}}, g^{(p+q)})$.

(b) *Likelihood:* $(\mathbf{y} \mid \boldsymbol{\beta}) \sim \mathrm{SUE}_{n,q}(\boldsymbol{\xi}_{\mathrm{y}|\beta}, \boldsymbol{\Omega}_{\mathrm{y}|\beta}, \boldsymbol{\Delta}_{\mathrm{y}|\beta}, \boldsymbol{\tau}_{\mathrm{y}|\beta}, \bar{\boldsymbol{\Gamma}}_{\mathrm{y}|\beta}, g^{(n+q)}_{Q_\beta(\boldsymbol{\beta})})$, *with parameters*

$$\boldsymbol{\xi}_{\mathrm{y}|\beta} = \boldsymbol{\xi}_\mathrm{y} + \boldsymbol{\Omega}_{\mathrm{y}\beta}\boldsymbol{\Omega}_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\xi}_\beta), \qquad \boldsymbol{\Omega}_{\mathrm{y}|\beta} = \boldsymbol{\Omega}_\mathrm{y} - \boldsymbol{\Omega}_{\mathrm{y}\beta}\boldsymbol{\Omega}_\beta^{-1}\boldsymbol{\Omega}_{\beta\mathrm{y}}, \qquad \boldsymbol{\Delta}_{\mathrm{y}|\beta} = \boldsymbol{\omega}_{\mathrm{y}|\beta}^{-1}(\boldsymbol{\omega}_\mathrm{y}\boldsymbol{\Delta}_\mathrm{y} - \boldsymbol{\Omega}_{\mathrm{y}\beta}\boldsymbol{\Omega}_\beta^{-1}\boldsymbol{\omega}_\beta\boldsymbol{\Delta}_\beta)\boldsymbol{\gamma}_{\mathrm{y}|\beta}^{-1},$$

$$\boldsymbol{\tau}_{\mathrm{y}|\beta} = \boldsymbol{\gamma}_{\mathrm{y}|\beta}^{-1}[\boldsymbol{\tau} + \boldsymbol{\Delta}_\beta^\top\bar{\boldsymbol{\Omega}}_\beta^{-1}\boldsymbol{\omega}_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\xi}_\beta)], \qquad \bar{\boldsymbol{\Gamma}}_{\mathrm{y}|\beta} = \boldsymbol{\gamma}_{\mathrm{y}|\beta}^{-1}(\bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}_\beta^\top\bar{\boldsymbol{\Omega}}_\beta^{-1}\boldsymbol{\Delta}_\beta)\boldsymbol{\gamma}_{\mathrm{y}|\beta}^{-1}, \qquad Q_\beta(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \boldsymbol{\xi}_\beta)^\top\boldsymbol{\Omega}_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\xi}_\beta),$$

*where* $\boldsymbol{\omega}_{\mathrm{y}|\beta} = \mathrm{diag}(\boldsymbol{\Omega}_{\mathrm{y}|\beta})^{1/2}$ *and* $\boldsymbol{\gamma}_{\mathrm{y}|\beta} = \mathrm{diag}(\bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}_\beta^\top\bar{\boldsymbol{\Omega}}_\beta^{-1}\boldsymbol{\Delta}_\beta)^{1/2}$.

(c) *Posterior distribution:* $(\boldsymbol{\beta} \mid \mathbf{y}) \sim \mathrm{SUE}_{p,q}(\boldsymbol{\xi}_{\beta|\mathrm{y}}, \boldsymbol{\Omega}_{\beta|\mathrm{y}}, \boldsymbol{\Delta}_{\beta|\mathrm{y}}, \boldsymbol{\tau}_{\beta|\mathrm{y}}, \bar{\boldsymbol{\Gamma}}_{\beta|\mathrm{y}}, g^{(p+q)}_{Q_\mathrm{y}(\mathrm{y})})$, *with parameters*

$$\boldsymbol{\xi}_{\beta|\mathrm{y}} = \boldsymbol{\xi}_\beta + \boldsymbol{\Omega}_{\beta\mathrm{y}}\boldsymbol{\Omega}_\mathrm{y}^{-1}(\mathbf{y} - \boldsymbol{\xi}_\mathrm{y}), \qquad \boldsymbol{\Omega}_{\beta|\mathrm{y}} = \boldsymbol{\Omega}_\beta - \boldsymbol{\Omega}_{\beta\mathrm{y}}\boldsymbol{\Omega}_\mathrm{y}^{-1}\boldsymbol{\Omega}_{\mathrm{y}\beta}, \qquad \boldsymbol{\Delta}_{\beta|\mathrm{y}} = \boldsymbol{\omega}_{\beta|\mathrm{y}}^{-1}(\boldsymbol{\omega}_\beta\boldsymbol{\Delta}_\beta - \boldsymbol{\Omega}_{\beta\mathrm{y}}\boldsymbol{\Omega}_\mathrm{y}^{-1}\boldsymbol{\omega}_\mathrm{y}\boldsymbol{\Delta}_\mathrm{y})\boldsymbol{\gamma}_{\beta|\mathrm{y}}^{-1},$$

$$\boldsymbol{\tau}_{\beta|\mathrm{y}} = \boldsymbol{\gamma}_{\beta|\mathrm{y}}^{-1}[\boldsymbol{\tau} + \boldsymbol{\Delta}_\mathrm{y}^\top\bar{\boldsymbol{\Omega}}_\mathrm{y}^{-1}\boldsymbol{\omega}_\mathrm{y}^{-1}(\mathbf{y} - \boldsymbol{\xi}_\mathrm{y})], \qquad \bar{\boldsymbol{\Gamma}}_{\beta|\mathrm{y}} = \boldsymbol{\gamma}_{\beta|\mathrm{y}}^{-1}(\bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}_\mathrm{y}^\top\bar{\boldsymbol{\Omega}}_\mathrm{y}^{-1}\boldsymbol{\Delta}_\mathrm{y})\boldsymbol{\gamma}_{\beta|\mathrm{y}}^{-1}, \qquad Q_\mathrm{y}(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\xi}_\mathrm{y})^\top\boldsymbol{\Omega}_\mathrm{y}^{-1}(\mathbf{y} - \boldsymbol{\xi}_\mathrm{y}),$$

*where* $\boldsymbol{\omega}_{\beta|\mathrm{y}} = \mathrm{diag}(\boldsymbol{\Omega}_{\beta|\mathrm{y}})^{1/2}$ *and* $\boldsymbol{\gamma}_{\beta|\mathrm{y}} = \mathrm{diag}(\bar{\boldsymbol{\Gamma}} - \boldsymbol{\Delta}_\mathrm{y}^\top\bar{\boldsymbol{\Omega}}_\mathrm{y}^{-1}\boldsymbol{\Delta}_\mathrm{y})^{1/2}$.

# Conjugate Bayesian Inference based on SUE (Karling et al., 2024)

Some examples:

- Multivariate linear model:

  uncorrelated $\boldsymbol{\beta} \sim \mathcal{T}_p(\boldsymbol{\xi}_{\boldsymbol{\beta}}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}, \nu)$ and $\boldsymbol{\epsilon} \sim \mathcal{SUT}_{n,q}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}, \boldsymbol{\Delta}_{\boldsymbol{\epsilon}}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu)$
  likelihood $(\mathbf{y} \mid \boldsymbol{\beta}) \sim \mathcal{SUT}_{n,q}(\mathbf{X}\boldsymbol{\beta}, \alpha_{\boldsymbol{\beta}}\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}, \boldsymbol{\Delta}_{\boldsymbol{\epsilon}}, \mathbf{0}, \bar{\boldsymbol{\Gamma}}, \nu + p)$, then posterior is $\boldsymbol{\beta} \sim \mathcal{SUT}_{p,q}$

  In particular, conjugacy of Bayesian Student's $t$ regression with:
  $\boldsymbol{\beta} \sim \mathcal{T}_p(\boldsymbol{\xi}_{\boldsymbol{\beta}}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}, \nu)$, $(\mathbf{y} \mid \boldsymbol{\beta}) \sim \mathcal{T}_n(\mathbf{X}\boldsymbol{\beta}, \alpha_{\boldsymbol{\beta}}\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}, \nu + p)$, then posterior is $\boldsymbol{\beta} \sim \mathcal{T}_p$
  with $\alpha_{\boldsymbol{\beta}} = [\nu + (\boldsymbol{\beta} - \boldsymbol{\xi}_{\boldsymbol{\beta}})^\top \boldsymbol{\Omega}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\xi}_{\boldsymbol{\beta}})]/(\nu + p)$

Extensions:

- Multivariate binary models:

  $$\mathbf{y} = [1(\bar{y}_1 > 0), \ldots, 1(\bar{y}_n > 0)]^\top, \qquad \bar{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- Multivariate censored models:

  $$\mathbf{y} = [\bar{y}_1 1(\bar{y}_1 > 0), \ldots, \bar{y}_n 1(\bar{y}_n > 0)]^\top, \qquad \bar{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Currently:

- Extension to Heckman selection models
- Computations (analytical; sampling-based; deterministic approx. (VB, EP)); robustness; spatial models

- Standard skew-normal distribution, $Z \sim \mathcal{ASN}(\alpha)$, has pdf and cf:

$$f_{\mathrm{s}}(z) \quad = \quad 2\phi(z)\Phi(\alpha z), \quad z \in \mathbb{R}, \qquad \psi_{\mathrm{s}}(t) = 2e^{-\frac{1}{2}t^2}\Phi\left(\frac{\mathrm{i}\alpha t}{\sqrt{1+\alpha^2}}\right), \quad t \in \mathbb{R}$$

- Reparameterization: $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}, \quad \alpha \in \mathbb{R} \quad \longleftrightarrow \quad \alpha = \frac{\delta}{\sqrt{1-\delta^2}}, \quad \delta \in (-1,1)$
- Standard skew-normal distribution, $Z \sim \mathcal{ASN}(\delta)$, has pdf and cf:

$$f_{\mathrm{s}}(z) \quad = \quad 2\phi(z)\Phi\left(\frac{\delta z}{\sqrt{1-\delta^2}}\right), \quad z \in \mathbb{R}, \qquad \psi_{\mathrm{s}}(t) = 2e^{-\frac{1}{2}t^2}\Phi\left(\mathrm{i}\delta t\right), \quad t \in \mathbb{R}$$

- Location-scale extension: $X = \xi + \omega Z, \quad Z \sim \mathcal{ASN}(\alpha), \quad Z \stackrel{d}{=} \frac{\alpha|U_0|+U}{\sqrt{1+\alpha^2}}, \quad U_0,\ U \stackrel{iid}{\sim} \mathcal{N}(0,1)$
- Rescaled: $\tilde{Z} = \sqrt{1+\alpha^2}\, Z, \quad \tilde{X} = \xi + \psi\tilde{Z} = \xi + \psi\alpha|U_0| + \psi U = \xi + \eta|U_0| + \psi U, \quad U_0, U \stackrel{iid}{\sim} \mathcal{N}(0,1)$

  One-to-one transformation $(\xi, \omega^2, \alpha) \longleftrightarrow (\xi, \psi^2, \eta)$ with $\psi = \frac{\omega}{\sqrt{1+\alpha^2}}$ and $\eta = \psi\alpha = \omega\delta$

$$\tilde{f}(x) \quad = \quad \frac{2}{\sqrt{\psi^2+\eta^2}}\,\phi\left(\frac{x-\xi}{\sqrt{\psi^2+\eta^2}}\right)\Phi\left\{\frac{\eta(x-\xi)}{\psi\sqrt{\psi^2+\eta^2}}\right\}, \quad x \in \mathbb{R}, \qquad \tilde{\psi}(t) = 2e^{\mathrm{i}t\xi - \frac{1}{2}\psi^2 t^2}\Phi(\mathrm{i}t\eta), \quad t \in \mathbb{R}$$

# Nonsingular/singular multivariate skew-normal distribution

$\mathbf{X} \sim \mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$ if $\mathbf{X} = \boldsymbol{\xi} + \boldsymbol{\eta} U + \mathbf{W}$ where $U = |W_0|$, $W_0 \sim \mathcal{N}(0,1)$, independent of $\mathbf{W} \sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Psi})$

- $\boldsymbol{\Psi} > 0$ (nonsingular) or $\boldsymbol{\Psi} \geq 0$ (singular)
- pdf ($\boldsymbol{\Psi} > 0$):

$$f(\mathbf{x}) = 2\phi_p(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Psi} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)\Phi\left\{\frac{\boldsymbol{\eta}^\top \boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\xi})}{\sqrt{1 + \boldsymbol{\eta}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\eta}}}\right\}, \quad \mathbf{x} \in \mathbb{R}^p$$

- cf:

$$\psi(\mathbf{t}) = 2e^{\mathrm{i}\mathbf{t}^\top \boldsymbol{\xi} - \frac{1}{2}\mathbf{t}^\top(\boldsymbol{\Psi} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)\mathbf{t}}\Phi(\mathrm{i}\mathbf{t}^\top \boldsymbol{\eta}), \quad \mathbf{t} \in \mathbb{R}^p$$

- linear transformation: $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^{q \times p}$, $\mathbf{b} \in \mathbb{R}^q$ then $\mathbf{Y} \sim \mathcal{SN}_q(\mathbf{A}\boldsymbol{\xi} + \mathbf{b}, \mathbf{A}\boldsymbol{\Psi}\mathbf{A}^\top, \mathbf{A}\boldsymbol{\eta})$
- marginals: $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ then $\mathbf{X}_j \sim \mathcal{SN}_{p_j}(\boldsymbol{\xi}_j, \boldsymbol{\Psi}_{jj}, \boldsymbol{\eta}_j)$ for $j = 1, 2$ (marginally consistent)
- link between $\mathcal{SN}_p(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\eta})$ and $\mathcal{ASN}_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$:

$$\boldsymbol{\Omega} = \boldsymbol{\Psi} + \boldsymbol{\eta}\boldsymbol{\eta}^\top, \qquad \boldsymbol{\alpha} = \frac{\boldsymbol{\omega}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}}{\sqrt{1 + \boldsymbol{\eta}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\eta}}} \qquad \boldsymbol{\omega} = \mathrm{diag}(\boldsymbol{\Omega})^{1/2}$$

$$\boldsymbol{\Psi} = \boldsymbol{\omega}(\bar{\boldsymbol{\Omega}} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)\boldsymbol{\omega}, \quad \boldsymbol{\eta} = \boldsymbol{\omega}\boldsymbol{\delta}, \quad \text{with} \quad \boldsymbol{\delta} = \frac{\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}}\boldsymbol{\alpha}}}, \quad \bar{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$$

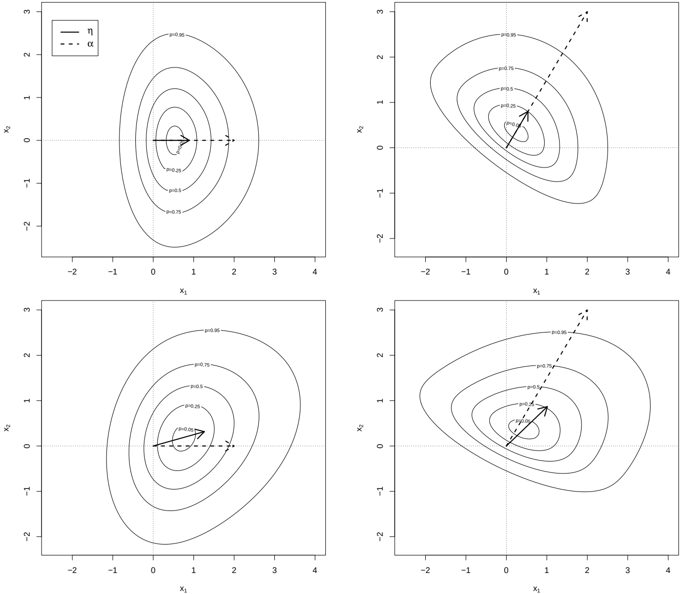
- Azzalini and Capitanio (2014) mentioned but did not explore with skewness parameter $\{\mathrm{diag}(\boldsymbol{\Psi})\}^{-1/2}\boldsymbol{\eta}$

$$\boldsymbol{\Omega}_1 = \boldsymbol{\Psi}_1 + \boldsymbol{\eta}_1\boldsymbol{\eta}_1^\top = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$\boldsymbol{\alpha}_1 = (2,0)^\top$ and $\boldsymbol{\alpha}_1 = (2,3)^\top$

$$\boldsymbol{\Omega}_2 = \boldsymbol{\Psi}_2 + \boldsymbol{\eta}_2\boldsymbol{\eta}_2^\top = \begin{pmatrix} 2 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

$\boldsymbol{\alpha}_2 = (2,0)^\top$ and $\boldsymbol{\alpha}_2 = (2,3)^\top$



# Nonsingular/singular multivariate unified skew-normal distribution

$\mathbf{X} \sim \mathcal{SUN}_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Psi}, \mathbf{H}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ if $\mathbf{X} = \boldsymbol{\xi} + \mathbf{H}\mathbf{U} + \mathbf{W}$ where $\mathbf{U} \stackrel{d}{=} (\mathbf{W}_0 \mid \mathbf{W}_0 + \boldsymbol{\tau} > \mathbf{0}_q)$, with $\mathbf{W}_0 \sim \mathcal{N}_q(\mathbf{0}_q, \bar{\boldsymbol{\Gamma}})$, is a truncated normal random vector, and $\mathbf{W} \sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Psi})$ which is independent of $\mathbf{W}_0$ and therefore of $\mathbf{U}$

- $\boldsymbol{\Psi} > 0$ (nonsingular) or $\boldsymbol{\Psi} \geq 0$ (singular)
- pdf ($\boldsymbol{\Psi} > 0$) with $\tilde{\boldsymbol{\Gamma}} = \bar{\boldsymbol{\Gamma}} - \bar{\boldsymbol{\Gamma}}\mathbf{H}^\top(\boldsymbol{\Psi} + \mathbf{H}\bar{\boldsymbol{\Gamma}}\mathbf{H}^\top)^{-1}\mathbf{H}\bar{\boldsymbol{\Gamma}} = (\bar{\boldsymbol{\Gamma}}^{-1} + \mathbf{H}^\top\boldsymbol{\Psi}^{-1}\mathbf{H})^{-1}$:

$$f(\mathbf{x}) = \frac{1}{\Phi_q(\boldsymbol{\tau}; \mathbf{0}_q, \bar{\boldsymbol{\Gamma}})}\phi_p(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Psi} + \mathbf{H}\bar{\boldsymbol{\Gamma}}\mathbf{H}^\top)\Phi_q\{\boldsymbol{\tau} + \tilde{\boldsymbol{\Gamma}}\mathbf{H}^\top\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\xi}); \mathbf{0}_q, \tilde{\boldsymbol{\Gamma}}\}, \quad \mathbf{x} \in \mathbb{R}^p$$

- cf:

$$\psi(\mathbf{t}) = e^{\mathrm{i}\mathbf{t}^\top\boldsymbol{\xi} - \frac{1}{2}\mathbf{t}^\top(\boldsymbol{\Psi} + \mathbf{H}\bar{\boldsymbol{\Gamma}}\mathbf{H}^\top)\mathbf{t}}\,\frac{\Phi_q(\boldsymbol{\tau} + \mathrm{i}\,\bar{\boldsymbol{\Gamma}}\mathbf{H}^\top\mathbf{t}; \mathbf{0}_q, \bar{\boldsymbol{\Gamma}})}{\Phi_q(\boldsymbol{\tau}; \mathbf{0}_q, \bar{\boldsymbol{\Gamma}})}, \quad \mathbf{t} \in \mathbb{R}^p$$

- linear transformation: $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^{r \times p}$, $\mathbf{b} \in \mathbb{R}^r$ then $\mathbf{Y} \sim \mathcal{SUN}_{r,q}(\mathbf{A}\boldsymbol{\xi} + \mathbf{b}, \mathbf{A}\boldsymbol{\Psi}\mathbf{A}^\top, \mathbf{A}\mathbf{H}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$
- marginals: $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ then $\mathbf{X}_j \sim \mathcal{SUN}_{p_j,q}(\boldsymbol{\xi}_j, \boldsymbol{\Psi}_{jj}, \mathbf{H}_j, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ for $j = 1, 2$ (marginally consistent $q$ fixed)
- link between $\mathcal{SUN}_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Psi}, \mathbf{H}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ and $\mathcal{AASUN}_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$:

$$\boldsymbol{\Omega} = \boldsymbol{\Psi} + \mathbf{H}\bar{\boldsymbol{\Gamma}}\mathbf{H}^\top, \quad \omega\boldsymbol{\Delta} = \mathbf{H}\bar{\boldsymbol{\Gamma}} \qquad \boldsymbol{\omega} = \mathrm{diag}(\boldsymbol{\Omega})^{1/2}$$

$$\boldsymbol{\Psi} = \boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Delta}^\top\omega, \quad \mathbf{H} = \omega\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}^{-1}$$

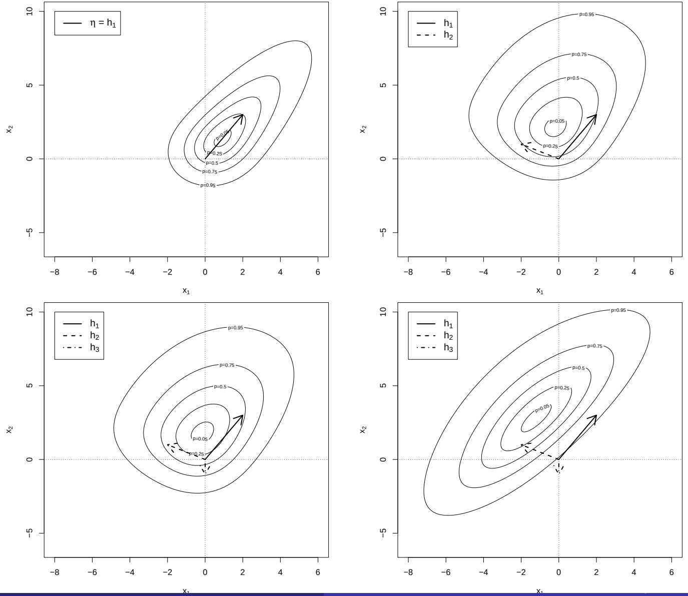
# Geometry of $\mathcal{SUN}_{p,q}$ for $p = 2$

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$\boldsymbol{\tau} = \mathbf{0}_q$, $\bar{\boldsymbol{\Gamma}} = \frac{1}{2}\mathbf{I}_q + \frac{1}{2}\mathbf{1}_q\mathbf{1}_q^\top$, $q = 1, 2, 3$

Last: $\boldsymbol{\Psi} = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$

$\boldsymbol{\tau} = (1, -1, 0)^\top$, $\bar{\boldsymbol{\Gamma}} = \frac{1}{2}\mathbf{I}_3 + \frac{1}{2}\mathbf{1}_3\mathbf{1}_3^\top$



*Multivariate unified skew-normal distribution with $S$ independent sets of latent variables,*
$\mathbf{Y} \sim \mathcal{SUNSET}_{p,q_1,\ldots,q_S}(\boldsymbol{\xi}, \boldsymbol{\Psi}, \mathbf{H}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ *if* $\mathbf{Y} \sim \mathcal{SUN}_{p,q_1+\cdots+q_S}(\boldsymbol{\xi}, \boldsymbol{\Psi}, \mathbf{H}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ *with block-diagonal* $\bar{\boldsymbol{\Gamma}}$*:*

$$\mathbf{Y} = \boldsymbol{\xi} + \sum_{s=1}^{S} \mathbf{H}_s \mathbf{U}_s + \mathbf{W}$$

- $q$-by-$q$ block-diagonal correlation matrix $\bar{\boldsymbol{\Gamma}} = \mathrm{diag}(\bar{\boldsymbol{\Gamma}}_1, \ldots, \bar{\boldsymbol{\Gamma}}_S) > 0$, $q_s$-by-$q_s$ blocks $\bar{\boldsymbol{\Gamma}}_s$, $q_1 + \cdots + q_S = q$
- $\mathbf{H} = (\mathbf{H}_1, \ldots, \mathbf{H}_S) \in \mathbb{R}^{p \times q}$, $\mathbf{H}_s \in \mathbb{R}^{p \times q_s}$, $\boldsymbol{\tau} = (\boldsymbol{\tau}_1^\top, \ldots, \boldsymbol{\tau}_S^\top)^\top \in \mathbb{R}^q$, $\boldsymbol{\tau}_s \in \mathbb{R}^{q_s}$
- $\mathbf{U}_s \overset{d}{=} (\mathbf{W}_{0s} \mid \mathbf{W}_{0s} + \boldsymbol{\tau}_s > \mathbf{0})$, with $\mathbf{W}_{0s} \sim \mathcal{N}_{q_s}(\mathbf{0}, \bar{\boldsymbol{\Gamma}}_s)$, are independent truncated normal random vectors, $s = 1, \ldots, S$, $\mathbf{W} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$ independent of $\mathbf{W}_{0s}$ and therefore of $\mathbf{U}_s$ for any $s = 1, \ldots, S$

Immediate properties:

- SUNSET naturally arises as the distribution of the sum of independent SUN random vectors
- Probability density function:

$$\phi_p(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Psi} + \mathbf{H}\bar{\boldsymbol{\Gamma}}\mathbf{H}^\top) \frac{\Phi_q\{\boldsymbol{\tau} + \tilde{\boldsymbol{\Gamma}}\mathbf{H}^\top \boldsymbol{\Psi}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \tilde{\boldsymbol{\Gamma}}\}}{\prod_{s=1}^{S} \Phi_{q_s}(\boldsymbol{\tau}_s; \bar{\boldsymbol{\Gamma}}_s)}, \quad \mathbf{y} \in \mathbb{R}^p, \quad \tilde{\boldsymbol{\Gamma}} = (\bar{\boldsymbol{\Gamma}}^{-1} + \mathbf{H}^\top \boldsymbol{\Psi}^{-1}\mathbf{H})^{-1}$$

- Fewer parameters than full SUN distributions, even more so for equicorrelated $\bar{\boldsymbol{\Gamma}}_s$; simplified marginals

# Beautiful SUNSET

*Multivariate beautiful unified skew-normal distribution with $S$ independent sets of latent variables,* $\mathbf{Y} \sim \mathcal{BSUNSET}_{p,q_1,\ldots,q_S}(\boldsymbol{\xi}, \boldsymbol{\Psi}, \mathbf{h}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ *is* $\mathcal{SUNSET}_{p,q_1,\ldots,q_S}(\boldsymbol{\xi}, \boldsymbol{\Psi}, \mathbf{H}, \boldsymbol{\tau}, \bar{\boldsymbol{\Gamma}})$ *with:*

- $\boldsymbol{\Psi} = \sum_{i=1}^{p} \lambda_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^\top$
- $\mathbf{H} = (h_1\boldsymbol{\psi}_1, \ldots, h_q\boldsymbol{\psi}_q)$, where $\mathbf{h} = (h_1, \ldots, h_q)^\top$ controls the magnitude of the skewness in the directions given by the first $q$ eigenvectors of $\boldsymbol{\Psi}$, hence providing a clear geometrical interpretation ($q \leq p$)
- If $q > p$, then we use all the $p$ eigenvectors of $\boldsymbol{\Psi}$ and add $q - p$ vectors to form a basis of $\mathbb{R}^q$

Immediate properties:

- BSUNSET has better geometrical interpretation (structure of $\mathbf{H}$ can also be used for SUN distributions)
- Probability density function:

$$\phi_p(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Psi} + \mathbf{H}\bar{\boldsymbol{\Gamma}}\mathbf{H}^\top) \prod_{s=1}^{S} \frac{\Phi_{q_s}\{\boldsymbol{\tau}_s + (\bar{\boldsymbol{\Gamma}}_s + \mathbf{D}_s)\mathbf{H}_s^\top \boldsymbol{\Psi}^{-1}(\mathbf{y} - \boldsymbol{\xi}); (\bar{\boldsymbol{\Gamma}}_s^{-1} + \mathbf{D}_s)^{-1}\}}{\Phi_{q_s}(\boldsymbol{\tau}_s; \bar{\boldsymbol{\Gamma}}_s)}, \quad \mathbf{y} \in \mathbb{R}^p$$

  where $\mathbf{D}_s = \mathrm{diag}(h_i^2/\lambda_i) \in \mathbb{R}^{q_s \times q_s}$
- Fewer parameters than SUNSET distributions

# BSUNSET Advantages

The BSUNSET distribution addresses the aforementioned limitations of the SUN distribution:

(a) The BSUNSET model has $p(p+3)/2 + 2q + \sum_{s=1}^{S} q_s(q_s-1)/2$ parameters, a smaller number of parameters than the SUN model. In particular, the latent matrices $\mathbf{H}$ and $\bar{\boldsymbol{\Gamma}}$ have few parameters since the columns of $\mathbf{H}_s$ are proportional to the eigenvectors of $\boldsymbol{\Psi}$ and $\bar{\boldsymbol{\Gamma}}$ is block-diagonal

(b) The BSUNSET pdf requires the computation of $\Phi_{q_s}$ with $q_s < q$, thus it is easier to evaluate

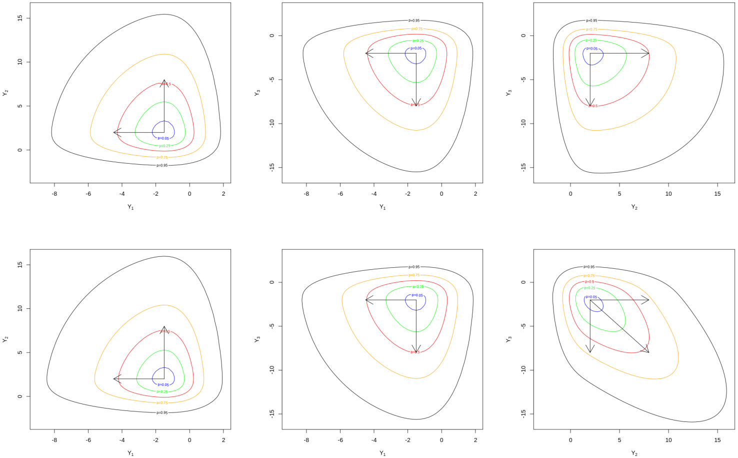
(c) The geometric interpretation of the effects of the latent parameters $\mathbf{H}$ and $\bar{\boldsymbol{\Gamma}}$ on the resulting skewness and shape of the BSUNSET distribution is clear. In particular, $\mathbf{h}$ in the construction of $\mathbf{H}$ represents the intensity of the skewness in the directions given by the first $q$ eigenvectors of $\boldsymbol{\Psi}$, and $\bar{\boldsymbol{\Gamma}}_s$ dictates the correlation between a sub-group of latent variables

(d) With the setting of $\boldsymbol{\Sigma}_s$, an inverse of a positive definite matrix $\bar{\boldsymbol{\Gamma}}_s^{-1}$ plus a diagonal matrix with non-negative entries $\mathbf{H}_s^\top \boldsymbol{\Psi}^{-1}\mathbf{H}_s$, the covariance matrix $\boldsymbol{\Sigma}_s$ is guaranteed to be positive definite

(e) With the structure of $\mathbf{H}_s$, the results in Wang et al. (2023) imply that the corresponding BSUNSET model is identifiable because the columns of $\mathbf{H}$ are strictly ordered by the decreasing eigenvalues of $\boldsymbol{\Psi}$ and hence leave no flexibility for potential permutations (this also applies for SUN)

(f) A parsimonious choice of $\bar{\boldsymbol{\Gamma}}_s$ is $\bar{\boldsymbol{\Gamma}}_s = (1-\rho_s)\mathbf{I}_{q_s} + \rho_s \mathbf{1}_{q_s}\mathbf{1}_{q_s}^\top$, with $-\frac{1}{q_s-1} < \rho_s < 1$

# Geometry of BSUNSET Distributions



$$\bar{\mathbf{\Gamma}} = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 1 & \rho \\ 0 & \rho & 1 \end{array}\right)$$

$$\mathbf{H}_1 = \begin{pmatrix} h_1 \\ 0 \\ 0 \end{pmatrix}, \ \mathbf{H}_2 = \begin{pmatrix} 0 & 0 \\ h_2 & 0 \\ 0 & h_3 \end{pmatrix}$$

SUNSET/BSUNSET
for Bayesian conjugacy?

Figure 2: Pairwise contour plots of the pdf of $\mathbf{Y} \sim \mathcal{SUNSET}_{3,1,2}(\mathbf{0}, \mathbf{I}_3, \mathbf{h}, \mathbf{0}, \bar{\mathbf{\Gamma}})$ with $\rho = 0.2$ (first row) and $\rho = 0.8$ (second row) for $h_1 = -3$, $h_2 = 6$ and $h_3 = -6$.

# Selected list of publications related to this presentation


Anceschi, N., Fasano, A., Durante, D., and Zanella, G. (2023). Bayesian conjugacy in probit, tobit, multinomial probit and extensions: a review and new results. *J. Amer. Statist. Assoc.*, 118(542):1451–1469.

Arellano-Valle, R. and Azzalini, A. (2022). Some properties of the unified skew-normal distribution. *Statistical Papers*, 63:461–487.

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33(3):561–574.

Arellano-Valle, R. B., Branco, M. D., and Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics*, 34(4):581–601.

Arellano-Valle, R. B. and Genton, M. G. (2010). Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, 1(1):17–33.

Durante, D. (2019). Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika*, 106(4):765–779.

Karling, M. J., Durante, D., and Genton, M. G. (2024). Conjugacy properties of multivariate unified skew-elliptical distributions. *J. Multivariate Anal.*, 204:105357.

Wang, K., Arellano-Valle, R. B., Azzalini, A., and Genton, M. G. (2023). On the non-identifiability of unified skew-normal distributions. *Stat*, 12:e597.

Wang, K., Karling, M. J., Arellano-Valle, R. B., and Genton, M. G. (2024). Multivariate unified skew-$t$ distributions and their properties. *J. Multivariate Anal.*, 203:105322.

Zhang, Z., Arellano-Valle, R. B., Genton, M. G., and Huser, R. (2023). Tractable Bayes of skew-elliptical link models for correlated binary data. *Biometrics*, 79(3):1788–1800.

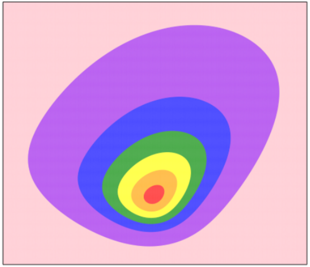
# 3. THE END? (the book will survive for a while...)

**Multivariate Statistics Beyond Normality**

Reinaldo B. Arellano-Valle
&
Marc G. Genton

Pontificia Universidad Católica de Chile
&
King Abdullah University of Science and Technology

2026



- CRC/Chapman & Hall
- 10 chapters
- $\sim$ 350 pages
- over 130 exercises
- selected solutions on website
- R code on website
- over 30 open research problems

# Summary

1. THE BEGINNING
2. Selection Multivariate Distributions
   - SUN, SUNNY, SUT, SUE and properties
   - On parameterizations
   - SUNSET
3. THE END?

Thank You! Questions?