Univariate and bivariate truncated von Mises distributions

Pablo Fernandez-Gonzalez, Concha Bielza, Pedro Larrañaga

Department of Artificial Intelligence, Universidad Politécnica de Madrid

Abstract: In this article we study the univariate and bivariate truncated von Mises distribution, as a generalization of the von Mises distribution (Jupp and Mardia (1989)), (Mardia and Jupp (2000)). This implies the addition of two or four new truncation parameters in the univariate and, bivariate cases, respectively. The results include the definition, properties of the distribution and maximum likelihood estimators for the univariate and bivariate cases. Additionally, the analysis of the bivariate case shows how the conditional distribution is a truncated von Mises distribution, whereas the marginal distribution that generalizes the distribution introduced in Singh (2002). From the viewpoint of applications, we test the distribution with simulated data, as well as with data regarding leaf inclination angles (Bowyer and Danson. (2005)) and dihedral angles in protein chains (Murzin AG (1995)). This research aims to assert this probability distribution as a potential option for modelling or simulating any kind of phenomena where circular distributions are applicable.

Key words and phrases: Angular probability distributions, Directional statistics, von Mises distribution, Truncated probability distributions.

1 Introduction

The von Mises distribution has received undisputed attention in the field of directional statistics (Jupp and Mardia (1989)) and in other areas like supervised classification (Lopez-Cruz et al. (2013)). Thanks to desirable properties such as its symmetry, mathematical tractability and convergence to the wrapped normal distribution (Mardia and Jupp (2000)) for high concentrations, it is a viable option for many statistical analyses. However, angular phenomena may present constraints on the outcomes that are not properly accounted forby the density function of the von Mises probability distribution. Thus, a truncated distribution with the capabilities of the von Mises distribution is strongly suggested. Additionally, this direction has need of development, since there is hardly any literature, and to the best of our knowledge, only one paper (Bistrian and Iakob (2008)), proposes a definition of the truncated von Mises distribution.

In this article, we propose a truncated probability distribution for angular values whose parent distribution is the von Mises distribution. The univariate and bivariate cases of this distribution are explicitly developed.

Section 2 introduces the definition for the univariate case and derives some properties of the distribution, calculates the maximum likelihood estimators of the parameters and studies the distribution moments. Section 3 addresses the definition of the bivariate truncated von Mises, maximum likelihood estimation of the parameters and the definition and study of the conditional and marginal truncated distributions. Section 4 reports the simulation studies applying the above developments and experimentally testing the observed behaviors. Section 5 reports experiments on real datasets containing leaf inclination angles in the univariate case and dihedral angles in protein chains in the bivariate case. Finally, Section 6 discusses the summary and conclusions.

The proofs of all results can be found in the supplementary material.

2 Univariate truncated von Mises distribution

Definition 2.1 *The truncated von Mises distribution is presented as a four-parameter generalization of the non-truncated case for truncation parameters a, b as*

$$f_{tvM}(\theta; \mu, \kappa, a, b) = \begin{cases} \frac{e^{\kappa\cos(\theta-\mu)}}{\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta} & if \quad \theta \in \mathbb{O}_{a,b} \\ \\ 0 & if \; \theta \in \mathbb{O}_{b,a} \end{cases} \tag{1}$$

where $\mu \in \mathbb{O}$ is the location parameter, $\kappa > 0$ the concentration parameter, $\mathbb{O}$ is the circular set of points ($\mathbb{O} : (x, y)$ such that $x^2 + y^2 = 1$), $\mathbb{O}_{a,b} \subset \mathbb{O}$ is obtained by selecting the points in the circular path from $a \in \mathbb{O}$ to $b \in \mathbb{O}$ in the preferred direction (counterclockwise) and $\mathbb{O}_{b,a}$ is its counterpart w.r.t. $\mathbb{O}$. Our proposed definition differs from Bistrian and Iakob (2008) in the circular definition of the truncation parameters. In their article, the truncation parameters were bounded to a linear definition involving the location parameter. Thanks to this difference, our distribution only needs truncation parameter values contained in $[0, 2\pi]$ in order to represent all possible distributions. This also affects the calculations involving the truncation parameters.

Truncation parameters have a big influence on the shape of the truncated distribution (Figure 1).

Lemma 2.1 *$\exists a, b, \mu \in \mathbb{O}$ such that in $\mathbb{O}_{a,b}$:*

1. *$f_{tvM}(\theta; \mu, \kappa, a, b)$ is a strictly decreasing function.*
2. *$f_{tvM}(\theta; \mu, \kappa, a, b)$ is a strictly increasing function.*
3. *$f_{tvM}(\theta; \mu, \kappa, a, b)$ increases and decreases reaching a single critical point that is a maximum.*
4. *$f_{tvM}(\theta; \mu, \kappa, a, b)$ increases and decreases reaching a single critical point that is a minimum.*
5. *$f_{tvM}(\theta; \mu, \kappa, a, b)$ increases and decreases reaching two critical points, a maximum and a minimum.*

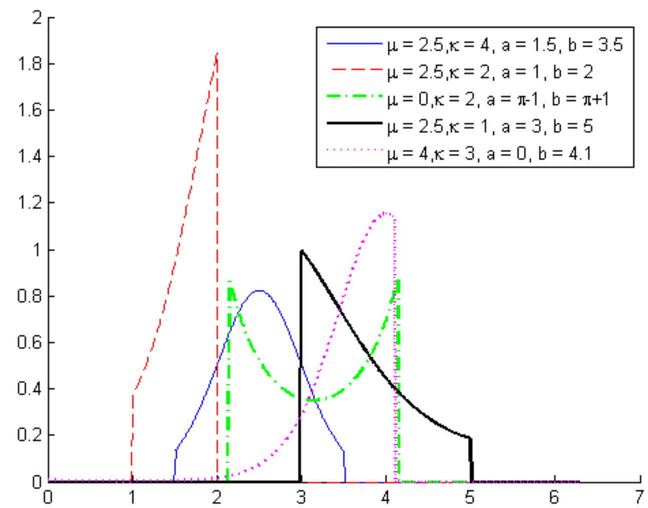

Figure 1: Several truncated von Mises distributions that include all cases of Lemma 2.1. Symmetrical function with maxima not at the extrema (thin continuous line), strictly increasing function (dashed line), strictly decreasing function (thick continuous line), unique critical point that is a minimum (dash-dot line) and two critical points, a maximum and a minimum (dotted line).

Looking at the normalization constant, we find that given $f_{vM}(\theta;\mu,\kappa)$ and $f_{tvM}(\theta;\mu,\kappa,a,b)$ such that $\mathbb{O}_{b,a} \neq \varnothing$ then $f_{vM}(\theta) < f_{tvM}(\theta), \quad \forall\theta \in \mathbb{O}_{a,b}$, as would be expected from a truncated distribution. Note then that while truncation parameters are circular quantities, the values for the integration coefficients are linear. Therefore, we use $b + 2\pi$ if $b < a$ as integration coefficients and a, b if $a \leq b$ (Unless otherwise stated, this will be assumed and not written explicitly throughout).

It is a well-known result (Abramowitz and Stegun (1964)) that $2\pi I_0(\kappa) = \int_0^{2\pi} e^{\kappa\cos(\theta-\mu)}d\theta$, where $I_0(\kappa)$ is the modified Bessel function of the first kind and order 0, that is,

$$I_0(\kappa) = \sum_{m=0}^{\infty} \frac{x^{2m}}{(m!)^2 2^m}.$$

The above expression suffices for truncation parameters a, b such that $\mathbb{O}_{a,b} = \mathbb{O}$. However, it is necessary to calculate the general case for non-restricted truncation parameters. Taking $w = \lfloor \frac{n}{2} \rfloor + \operatorname{mod}\frac{n}{2} - 1$, we obtain:

Lemma 2.2 $\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta = I(b;\mu,\kappa) - I(a;\mu,\kappa)$, *where*

$$I(\theta;\mu,\kappa) = \sum_{n=0}^{\infty}\frac{\kappa^n}{n!}\left(\sin(\theta-\mu)\sum_{i=0}^{w}\left(\cos^{n-2i-1}(\theta-\mu)\prod_{j=0}^{2i}(n-j)^{-(-1)^j}\right)\right.$$
$$\left.+\frac{((-1)^n+1)\prod_{j=0}^{w}(n-j)^{-(-1)^j}(\theta-\mu)}{2}\right). \tag{2}$$

$I(\theta;\mu,\kappa)$ is the distribution function of the positive support of the truncated von Mises density.

2.1 Maximum likelihood estimation

Provided we have a sample of observations $\theta_1,\theta_2,\ldots,\theta_n$ from a truncated von Mises distribution (1), we obtain:

$$\begin{aligned}\ln L(\mu,\kappa,a,b;\theta_1,\theta_2,\ldots,\theta_n) &= \sum_{i=1}^{n}\ln\left(\frac{e^{\kappa\cos(\theta_i-\mu)}}{\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta}\right)\\ &= \sum_{i=1}^{n}\kappa\cos(\theta_i-\mu) - n\ln\left(\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta\right)\end{aligned} \tag{3}$$

where $\ln L(\mu,\kappa,a,b;\theta_1,\theta_2,\ldots,\theta_n)$ is the log-likelihood function for the truncated von Mises distribution.

We now seek to solve the system of four log-likelihood equations created by the four parameters of the distribution. For parameters μ,κ, we have

$$\frac{\partial\ln L}{\partial\mu} = 0$$
$$\frac{\partial\ln L}{\partial\kappa} = 0.$$

As parameters a,b, define the region of the greater-than-zero density, we find that all $\theta_1,\ldots,\theta_n$ observations necessarily lie within the subset $\mathbb{O}_{a,b}$. Thanks to this consideration, together with the $-n\ln\left(\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta\right)$ sub term of (3), i.e., the integral of a solely positive function as the argument of the strictly increasing logarithmic function and $n\in\mathbb{N}$, we can isolate the estimators

$$\mathbb{O}_{\hat{a},\hat{b}} = \underset{a,b}{\operatorname{argmax}}(\max(\{A(\mathbb{O}_{\theta'_1,\theta'_2}),\ldots,A(\mathbb{O}_{\theta'_{n-1},\theta'_n}),A(\mathbb{O}_{\theta'_n,\theta'_1})\})), \tag{4}$$

where $A(\mathbb{O}_{\theta_1,\theta_2})$ is the angle between the point with angle θ_1 w.r.t. 0 and the point with angle θ_2 w.r.t. 0, and $\{\theta'_1,\ldots,\theta'_n\}$ is the sample sorted in ascending order of value. Intuitively, the truncation parameters are separated by the largest angle and are contiguous in a sorted finite circular sample. Notice that we input the angle between the last and the first element in the sample in order to complete the circle. Consequently, every truncated distribution with truncation parameters whose positive support does not include 0° maximizes the integral subterm of (3) by means of this inclusion in (4).

From this result, we can say that the truncation parameters of the truncated von Mises distribution have existing and population-only dependent maximum likelihood estimators. For parameters μ and κ, interdependency is a consequence of the possibly non-symmetrical shape of the distribution. If we observe the expression of the partial derivatives

$$\frac{1}{n}\sum_{i=1}^{n}\sin(\theta_i-\mu) - \frac{e^{\kappa\cos(a-\mu)} - e^{\kappa\cos(b-\mu)}}{\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta} = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\cos(\theta_i-\mu) - \frac{\int_a^b \cos(\theta-\mu)e^{\kappa\cos(\theta-\mu)}d\theta}{\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta} = 0,$$

$e^{\kappa\cos(a-\mu)} - e^{\kappa\cos(b-\mu)} = 0$ holds jf a,b are symmetrical w.r.t. μ, reducing the location parameter estimator to that of the non-truncated case (Mardia and Jupp (2000)), the circular sample mean $\hat{\mu}$. As no population-only dependent expressions of the parameters μ and κ were found, we use optimization techniques to maximize the log-likelihood function for those parameters in our study. To be precise, we regard the optimization of μ and κ as a non-linear programming problem that we can solve as a system of Karush-Kuhn-Tucker conditions.

2.2 Moments

The moments in circular statistics are particular values of the characteristic function. The r-th moment about a direction d can be written as

$$m_{r_{tvM}} = \mathbb{E}[e^{ir(X-d)}].$$

The first moment about the 0 direction for the truncated von Mises is calculated as

$$m_{1_{tvM}} = \frac{\int_a^b \cos(\theta) e^{\kappa \cos(\theta-\mu)} d\theta}{\int_a^b e^{\kappa \cos(\theta-\mu)} d\theta} + \frac{i \int_a^b \sin(\theta) e^{\kappa \cos(\theta-\mu)} d\theta}{\int_a^b e^{\kappa \cos(\theta-\mu)} d\theta}, \tag{5}$$

and we can relate (5) to the first moment about the μ direction, denoted as $m'_{1_{tvM}}$ as

$$m_{1_{tvM}} = e^{i\mu} m'_{1_{tvM}}. \tag{6}$$

Notice that if $\cos(a-\mu) = \cos(b-\mu)$, then $m'_{1_{tvM}} = \frac{\int_a^b \cos(x-\mu) e^{\kappa \cos(x-\mu)} d\theta}{\int_a^b e^{\kappa \cos(x-\mu)} d\theta} = R$, the mean resultant length of μ and thus $m_{1_{tvM}} = e^{i\mu} R$.

An alternative expression for $m_{1_{tvM}}$ can be found by considering equations $\mathbb{E}[\cos(x)] = R' \cos(\mu')$ and $\mathbb{E}[\sin(x)] = R' \sin(\mu')$, where R' and μ' are the sample mean resultant length and sample mean, respectively. We can then state

$$m_{1_{tvM}} = \mathbb{E}[\cos(x)] + i\mathbb{E}[\sin(x)] = R' \cos(\mu') + iR' \sin(\mu') = R' e^{i\mu'}. \tag{7}$$

Thus, merging Equations (6) and (7), we obtain

$$e^{i(\mu'-\mu)} R' = m'_{1_{tvM}},$$

which can be seen as a valuable expression as it contains the sample mean (μ') and the location parameter of the distribution (μ) .

3 Bivariate truncated von Mises distribution

The non-truncated bivariate von Mises distribution was first proposed by Singh (2002) and extended and developed in Mardia et al. (2008) and Mardia and Voss

(2011). It is a unimodal/bi-modal function on the torus $f_{btvM} : \mathbb{O} \times \mathbb{O} \to \mathbb{R}$ obtained by replacing the quadratic and linear terms of the normal bivariate distribution with their circular analogues. This distribution is known as the "sin variant bivariate von Mises distribution" and is defined for dependent pairs of angular variables. It is expressed for variables θ_1 and θ_2, as

$$f(\theta_1, \theta_2) = Ce^{\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)},$$

where $\kappa_1, \kappa_2 \geq 0, \lambda \in \mathbb{R}, \mu_1, \mu_2 \in \mathbb{O}$ and C is the normalization constant.
We propose the density function for the truncated case as a nine-parameter function with density defined as follows:

Definition 3.1 *We wite the density function for the truncated case as a nine-parameter function with density*

$$f_{btvM}(\theta_1, \theta_2; \boldsymbol{W}) = \begin{cases} \frac{f_{ubvM}(\theta_1, \theta_2; \boldsymbol{W})}{\int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{ubvM}(\theta_1, \theta_2; \boldsymbol{W}) d\theta_2 d\theta_1} & \text{if } \theta_1 \in \mathbb{O}_{a_1,b_1}, \theta_2 \in \mathbb{O}_{a_2,b_2}, \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where $\boldsymbol{W} = \{\lambda, \mu_1, \mu_2, \kappa_1, \kappa_2, a_1, b_1, a_2, b_2\}$ is the parameter vector and $f_{ubvM}(\theta_1, \theta_2; \boldsymbol{W}) = e^{\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)}$ is the unnormalized bivariate von Mises distribution. Parameters μ_1, μ_2 and κ_1, κ_2 are analogous to parameters μ and κ, respectively, in the univariate truncated case. Truncation parameters a_1, b_1, a_2 and b_2 are similar to the univariate truncation parameters. The $\lambda \in \mathbb{R}$ parameter accounts for the dependency between the variable components (Figure 2). If $\lambda = 0$, then θ_1 and θ_2 are independent and each is distributed as a univariate von Mises distribution. Also, if θ_1, θ_2 are independent, then $\lambda = 0$.

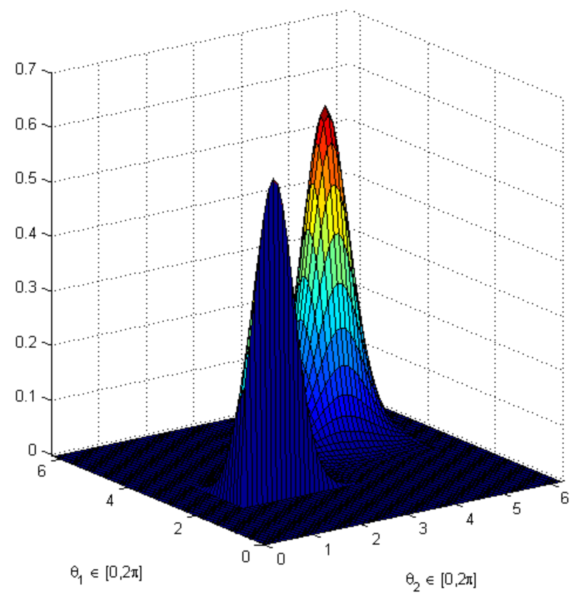

Figure 2: Example of the bi-dimensional von Mises distribution with parameters $\lambda = 6, \mu_1 = \pi, \mu_2 = 3, \kappa_1 = 2, \kappa_2 = 1, a_1 = 2, b_1 = 6, a_2 = 1, b_2 = 5.5$, showing truncated bi-modality.

A desirable property of a joint distribution is that it should have closed distributions under marginalization and conditioning, i.e., the marginal and conditional distributions should also follow the univariate distribution. Particularizing for the von Mises family, the bivariate von Mises distribution presents closed distributions only under conditioning as shown by Singh (Singh (2002)). We want to find out whether this also holds for the truncated case.

3.1 Maximum Likelihood Estimation

The maximum likelihood estimator for the bivariate distribution takes data of the form $\{(\theta_{1i}, \theta_{2i})\}$ $i = 1, \ldots, n$. The resulting log-likelihood function is

$$
\begin{aligned}
&\ln L(\boldsymbol{W};(\theta_{11},\theta_{21}),\ldots,(\theta_{1n},\theta_{2n}))\\
= &\sum_{i=1}^{n}\ln\left(\frac{e^{\kappa_1\cos(\theta_{1i}-\mu_1)+\kappa_2\cos(\theta_{2i}-\mu_2)+\lambda\sin(\theta_{1i}-\mu_1)\sin(\theta_{2i}-\mu_2)}}{\int_{a_1}^{b_1}\int_{a_2}^{b_2}e^{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)}d\theta_2 d\theta_1}\right)\\
= &\sum_{i=1}^{n}(\kappa_1\cos(\theta_{1i}-\mu_1)+\kappa_2\cos(\theta_{2i}-\mu_2)+\lambda\sin(\theta_{1i}-\mu_1)\sin(\theta_{2i}-\mu_2))\\
&-n\ln\left(\int_{a_1}^{b_1}\int_{a_2}^{b_2}e^{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)}d\theta_2 d\theta_1\right).
\end{aligned}
$$

Thus we have

$$\frac{\partial}{\partial\mu_1}\ln L(\boldsymbol{W};(\theta_{11},\theta_{21}),\ldots,(\theta_{1n},\theta_{2n}))=0,$$

that is,

$$\sum_{i=1}^{n}\kappa_1\sin(\theta_{1i}-\mu_1)-\lambda\cos(\theta_{1i}-\mu_1)\sin(\theta_{2i}-\mu_2)-\frac{n\left(\int_{a_2}^{b_2}f_{ubvM}(a_1,\theta_2)-f_{ubvM}(b_1,\theta_2)d\theta_2\right)}{\int_{a_1}^{b_1}\int_{a_2}^{b_2}f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}=0,$$

where $f_{ubvM}(\theta_1,\theta_2)$ is the following unnormalized bivariate truncated von Mises function

$$f_{ubvM}(\theta_1,\theta_2)=e^{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)}.$$

Similarly, the partial derivate w.r.t. μ_2 gives

$$\sum_{i=1}^{n}\kappa_2\sin(\theta_{2i}-\mu_2)-\lambda\cos(\theta_{2i}-\mu_2)\sin(\theta_{1i}-\mu_1)-\frac{n\left(\int_{a_1}^{b_1}f_{ubvM}(\theta_1,a_2)-f_{ubvM}(\theta_1,b_2)d\theta_1\right)}{\int_{a_1}^{b_1}\int_{a_2}^{b_2}f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}=0.$$

For κ_1 we have

$$\frac{\partial}{\partial\kappa_1}\ln L(\boldsymbol{W};(\theta_{11},\theta_{21}),\ldots,(\theta_{1n},\theta_{2n}))=0,$$

that is,

$$\frac{1}{n}\sum_{i=1}^{n}\cos(\theta_{1i}-\mu_1)-\frac{\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cos(\theta_1-\mu_1)f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}{\int_{a_1}^{b_1}\int_{a_2}^{b_2}f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}=0. \qquad (9)$$

Similarly, the partial derivate w.r.t. κ_2 gives

$$\frac{1}{n}\sum_{i=1}^{n}\cos(\theta_{2i}-\mu_2)-\frac{\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cos(\theta_2-\mu_2)f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}{\int_{a_1}^{b_1}\int_{a_2}^{b_2}f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}=0. \quad (10)$$

At this point, we can see that both equations (9) and (10), involving κ_1, κ_2 parameters, respectively, preserve their analogy with the univariate case. Their second addend corresponds to the definition of the estimators of $\mathbb{E}[\cos(\theta_1-\mu_1)]$ and $\mathbb{E}[\cos(\theta_2-\mu_2)]$, respectively.

For the parameter λ we obtain

$$\frac{\partial}{\partial\lambda}\ln L(\boldsymbol{W};(\theta_{11},\theta_{21}),\ldots,(\theta_{1n},\theta_{2n}))=0,$$

that is,

$$\frac{1}{n}\sum_{i=1}^{n}\sin(\theta_{1i}-\mu_1)\sin(\theta_{2i}-\mu_2)-\frac{\int_{a_1}^{b_1}\int_{a_2}^{b_2}\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}{\int_{a_1}^{b_1}\int_{a_2}^{b_2}f_{ubvM}(\theta_1,\theta_2)d\theta_2 d\theta_1}=0,$$

which analogously corresponds to the estimator of $\mathbb{E}[\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)]$. As in the univariate case, the truncation parameters has the following isolated estimators

$$\begin{aligned}\mathbb{O}_{\hat{a}_1,\hat{b}_1} &= \underset{a_1,b_1}{\operatorname{argmax}}(\max(\{A(\mathbb{O}_{\theta'_{11},\theta'_{12}}),\ldots,A(\mathbb{O}_{\theta'_{1n-1},\theta'_{1n}}),A(\mathbb{O}_{\theta'_{1n},\theta'_{11}})\}))\\ \mathbb{O}_{\hat{a}_2,\hat{b}_2} &= \underset{a_2,b_2}{\operatorname{argmax}}(\max(\{A(\mathbb{O}_{\theta'_{21},\theta'_{22}}),\ldots,A(\mathbb{O}_{\theta'_{2n-1},\theta'_{2n}}),A(\mathbb{O}_{\theta'_{2n},\theta'_{21}})\})),\end{aligned}$$

while as yielded by the above calculations, the expressions regarding the non-truncation parameters enhibit interdependency. We optimize them as a non-linear programming problem in the form of a system of Karush-Kuhn-Tucker conditions, as we did in univariate case.

3.2 Conditional truncated von Mises distribution

The density of the conditional truncated von Mises distribution is defined as:

Definition 3.2 *The conditional truncated von Mises distribution has density*

$$f_{ctvM}(\theta_2|\theta_1;\lambda,\mu_1,\mu_2,\kappa 2,a_2,b_2)=\begin{cases}\frac{e^{\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)}}{\int_{a_2}^{b_2}e^{\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)}d\theta_2} & \text{if } \theta_2\in\mathbb{O}_{a_2,b_2}.\\ 0 & \text{otherwise}\end{cases} \quad (11)$$

It is a six-parameter distribution where the parameters hold the same meaning as in the bivariate case, with the simplification of parameters κ_1, a_1, b_1 for $f_{ctvM}(\theta_2|\theta_1)$ (or κ_2, a_2, b_2 for $f_{ctvM}(\theta_1|\theta_2)$). Worthy of note, however, is that $\theta_1 \in \mathbb{O}_{a_1,b_1}$ in $f_{ctvM}(\theta_2|\theta_1)$ since otherwise, by the definition of the conditional distribution $f_{ctvM}(\theta_2|\theta_1) = \frac{f_{btvM}(\theta_2,\theta_1)}{f_{tvM}(\theta_1)}$, $f_{ctvM}(\theta_2|\theta_1)$ is not defined.

Theorem 3.1 *A conditional truncated von Mises distribution corresponds to a univariate truncated von Mises distribution through*

$$f_{ctvM}(\theta_2|\theta_1;\lambda,\mu_1,\mu_2,\kappa_2,a_2,b_2) =$$
$$f_{tvM}\left(\theta_2;\mu_2+\arctan\left(\frac{\lambda\sin(\theta_1-\mu_1)}{\kappa_2}\right),\sqrt{\kappa_2^2+(\lambda\sin(\theta_1-\mu_1))^2},a_2,b_2\right),$$

which completely specifies the behavior and properties of the conditional distribution and is analogous to the non-truncated conditional case (Singh (2002)).

3.3 Marginal truncated von Mises distribution

We can define the density function of the marginal truncated von Mises distribution as:

Definition 3.3 *The density function of the marginal truncated von Mises distribution can be written as*

$$f_{mtvM}(\theta_1;\boldsymbol{W}) = \begin{cases} \frac{\int_{a_2}^{b_2} e^{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)}d\theta_2}{\int_{a_1}^{b_1}\int_{a_2}^{b_2} e^{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)}d\theta_2 d\theta_1} & \text{if } \theta_1\in\mathbb{O}_{a_1,b_1} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

It is a nine-parameter distribution that shares all the parameters with the bivariate truncated von Mises distribution. In the original publication, Singh (2002) studied the distribution and reported the "frontiers" of bi-modality (for $\mu = 0$) as

$$\frac{I_1(\kappa_2)}{I_0(\kappa_2)} = \frac{\kappa_1\kappa_2}{\lambda^2}$$

where the distribution is unimodal if $\frac{I_1(\kappa_2)}{I_0(\kappa_2)} \geq \frac{\kappa_1\kappa_2}{\lambda^2}$, and bimodal with two equal maxima otherwise. Additionally, the modes were calculated to be symmetrical w.r.t μ_1 and at the distance value θ_1^* that solves the equation (for $\mu_1 = 0$):

$$\frac{A\left(\sqrt{\kappa_2 + \lambda^2 \sin^2(\theta_1^*)}\right)}{\sqrt{\kappa_2 + \lambda^2 \sin^2(\theta_1^*)}} \cos(\theta_1^*) = \frac{\kappa_1}{\lambda^2},$$

where $A(x) = \frac{I_1(x)}{I_0(x)}$. In order to generalize this analysis to cover the truncated case in (12), we need to account for the contribution made by the parameters μ_2, a_2 and b_2 to the shape of the distribution. μ_2 is not necessarily simplified by the symmetry of a_2 and b_2 in the non-truncated case. Additionally, a_2 and b_2 do not behave as truncation parameters, as they do not address the argument of the function and appear in both the numerator of the expression and the normalization term, actively influencing the resultant shape of the proposed distribution. Contrary to the non-truncated case, a truncated marginal distribution that exhibits two maxima may have only one global maximum; and. even if it has only one maximum, the distribution is not necessarily centered around the mean (Figure 3). Therefore, our analysis determines the different parameter configurations that produce the whole range of behaviors, with a special focus on bi-modality/unimodality results. Parameters a_1 and b_1 behave as the truncation parameters studied for the univariate case.

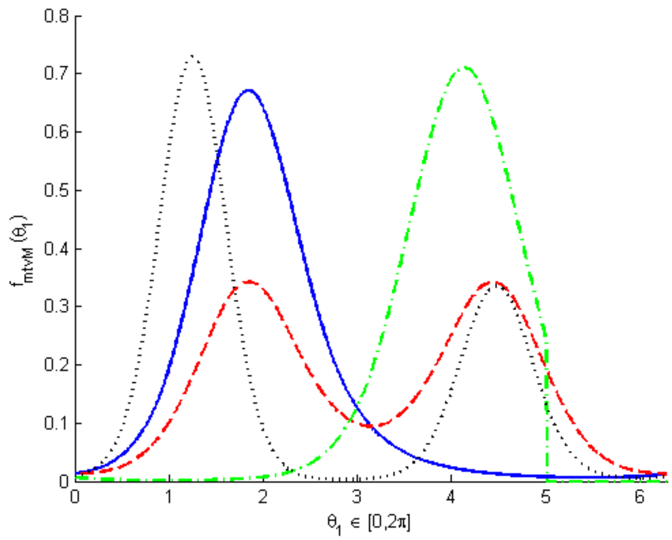

Figure 3: Several truncated marginal distributions showing unimodality (continuous line) with parameters $\lambda = 5, \mu_1 = \pi, \mu_2 = 0, \kappa_1 = 1, \kappa_2 = 4, a_1 = 0, b_1 = 2\pi, a_2 = \pi - 0.2, b_2 = 2\pi$, two equal maxima (dashed line) with parameters $\lambda = 5, \mu_1 = \pi, \mu_2 = 0, \kappa_1 = 1, \kappa_2 = 4, a_1 = 0, b_1 = 2\pi, a_2 = 0, b_2 = 2\pi$, truncated unimodality (dash-dot line) with parameters $\lambda = 1, \mu_1 = 4, \mu_2 = 2, \kappa_1 = 3, \kappa_2 = 4, a_1 = 0, b_1 = 5, a_2 = 2, b_2 = 2\pi$ and two distinct maxima (dotted line) with parameters $\lambda = 10, \mu_1 = 6, \mu_2 = 1, \kappa_1 = 0.3, \kappa_2 = 6, a_1 = 0, b_1 = 2\pi, a_2 = 0, b_2 = 5$ respectively.

If, without loss of generality, we take $\theta_{1'} = \theta_1 - \mu_1$, we can postulate the following theorem:

Theorem 3.2 *All different behaviors w.r.t. the unimodality/bi-modality of the marginal truncated von Mises distribution can be accounted for as follows*

1. $f_{mtvM}(\theta_{1'})$ *is unimodal with mode (maximum) in* μ_1*, if and only if* $T(\lambda, \mu_2, \kappa_1, \kappa_2, a_2, b_2) < 0$ *and* $\cos(b_2 - \mu_2) = \cos(a_2 - \mu_2)$.

2. $f_{mtvM}(\theta_{1'})$ *is bi-modal with equal maxima, if and only if* $T(\lambda, \mu_2, \kappa_1, \kappa_2, a_2, b_2) > 0$ *and* $\cos(b_2 - \mu_2) = \cos(a_2 - \mu_2)$*. Also in this case, a minimum is found at* $\theta_{1'} = 0$.

3. $f_{mtvM}(\theta_{1'})$ *presents two differentiated maxima if and only if one of the two following cases applies:*

(a) $\cos(b_2 - \mu_2) < \cos(a_2 - \mu_2)$ *and* $f'_{umtvM}(\theta_{1'}; \lambda, \mu_1, \mu_2, \kappa_1, \kappa_2, \mu_2, a_2, b_2)$ *has exactly two zero points in* $\theta_{1'} \in [-\frac{\pi}{2}, 0]$

(b) $\cos(b_2 - \mu_2) > \cos(a_2 - \mu_2)$ *and* $f'_{umtvM}(\theta_{1'}; \lambda, \mu_1, \mu_2, \kappa_1, \kappa_2, \mu_2, a_2, b_2)$ *has exactly two zero points in* $\theta_{1'} \in [0, \frac{\pi}{2}]$

4. $f_{mtvM}(\theta_{1'})$ *is unimodal with mode not at* μ_1 *if the parameters do not match any of the above cases,*

where $T(\lambda, \mu_2, \kappa_1, \kappa_2, a_2, b_2)$ is the test function and is defined as

$$T(\lambda, \mu_2, \kappa_1, \kappa_2, a_2, b_2) = -\frac{\kappa_1}{\lambda^2} + \frac{\int_{a_2}^{b_2} \sin^2(\theta_2 - \mu_2) e^{\kappa_2 \cos(\theta_2 - \mu_2)} d\theta_2}{\int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2 - \mu_2)} d\theta_2}, \tag{13}$$

and $f'_{umtvM}(\theta_{1'}; \lambda, \mu_1, \mu_2, \kappa_1, \kappa_2, \mu_2, a_2, b_2)$ is the unnormalized truncated marginal von Mises derivative function.

4 Simulation

In this section we experimentally test the behavior and properties of the univariate and bivariate truncated distributions. We have implemented an acceptance-rejection algorithm for simulation and maximum likelihood estimation by optimization over the likelihood function for both cases. We will confirm the expected effect of manipulating the parameters of the distribution by conducting different samplings and estimation operations. In the bivariate case, our studies are focus on the effect of the λ and truncation parameters.

4.1 Univariate case simulation

For the univariate case:

1. In our first simulation, we sampled 20000 points from the univariate distribution $f_{tvM}(\theta; \pi, 2, 1, 5)$ (Figure 4A). This resulted in a nearly symmetrical distribution, where the sample mean $\hat{\theta} = 3.1299 \approx \pi$. In this case, $\mathbb{O}_{a,b}$ selects most of the area of the original von Mises distribution. Maximizing the log-likelihood function (3) yielded values $\hat{\mu} = 3.1412, \hat{\kappa} = 1.9932, \hat{a} = 1.0014$ and $\hat{b} = 4.9940$. These values are indicative of a successful estimation of the parameters with errors roughly around 10^{-3}.

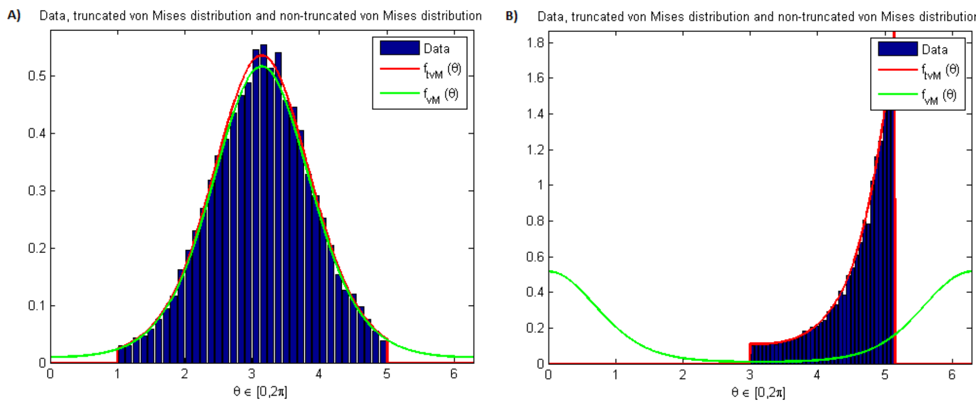

Figure 4: Simulation of two truncated von Mises distributions. The data were grouped in same-length value intervals in order to ilustrate its relative frequency.

2. We sampled 20000 points from the univariate distribution $f_{tvM}(\theta; 0, 2, 3, \pi + 2)$ (Figure 4B). Its shape is highly asymmetrical with truncation parameters that satisfy $\mu + \pi \in \mathbb{O}_{a,b}$ and $\mu \notin \mathbb{O}_{a,b}$. We chose this case because it was apparently troublesome to estimate and identify as a truncated von Mises distribution, i.e., the differences w.r.t. the original von Mises distribution are very noticeable. We find that the sample mean $\hat{\theta} = 4.6335$ clearly differs from the location parameter. Maximizing the likelihood function yielded parameter values $\hat{\mu} = 6.3044$ (which, given the periodicity of the function, could be also considered $\mu = 6.3044 - 2\pi \approx 0$), $\hat{\kappa} = 2.0446$, $\hat{a} = 3.0004$, $\hat{b} = 5.1415$.

3. We tested the distribution for relatively high concentrations by sampling 20000 points from the univariate distribution $f_{tvM}(\theta; \pi, 15, 0.5, \pi)$ (Figure 5). We can see how the truncation parameters retain slightly less than half of the area of the non-truncated distribution, this in turn causes the density to be slightly greater than twice the non-truncated density in its possitive support. Maximizing the log-likelihood function yielded parameter values $\hat{\mu} = 3.1386, \hat{\kappa} = 15.2597, \hat{a} = 2.0877, \hat{b} = 3.1416$.

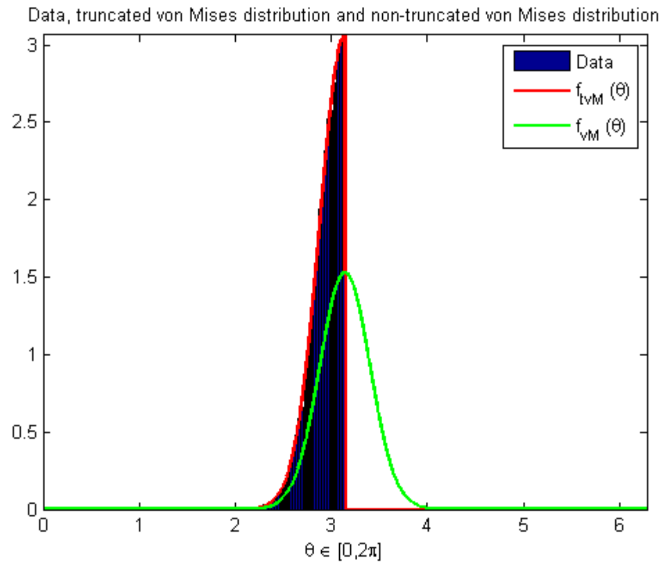

Figure 5: Simulation of the truncated von Mises distribution of the third study.

4.2 Bivariate case simulation

For the bivariate case:

1. We sampled 20000 points from the bivariate distribution $f_{btvM}(\theta_1, \theta_2; 1, 2, \pi, 3, 1, 0, 4, 3, 6)$ (Figure 6). From a visual inspection, this has a low lambda parameter and shows unimodality. The truncation parameters for this case more than halve the volume by selecting the positive support in the $\mathbb{O}_{a_1,b_1} \times \mathbb{O}_{a_2,b_2}$ region. Thus we find, for example, that it is possible to build a truncated bivariate von Mises distribution that shows only one maximum whereas its associated non-truncated distribution produces a bi-maximal distribution. The maximization of the log-likelihood function yielded parameter values $\hat{\lambda} = 0.7579, \hat{\mu_1} = 2.0540, \hat{\mu_2} = 3.2657, \hat{\kappa}_1 = 2.9753, \hat{\kappa}_2 = 1.0516, \hat{a}_1 = 0.0353, \hat{b}_1 = 3.9846, \hat{a}_2 = 3.0006, \hat{b}_2 = 5.9910$. This is a lowery quality approximation with an error of around 10^{-1}.

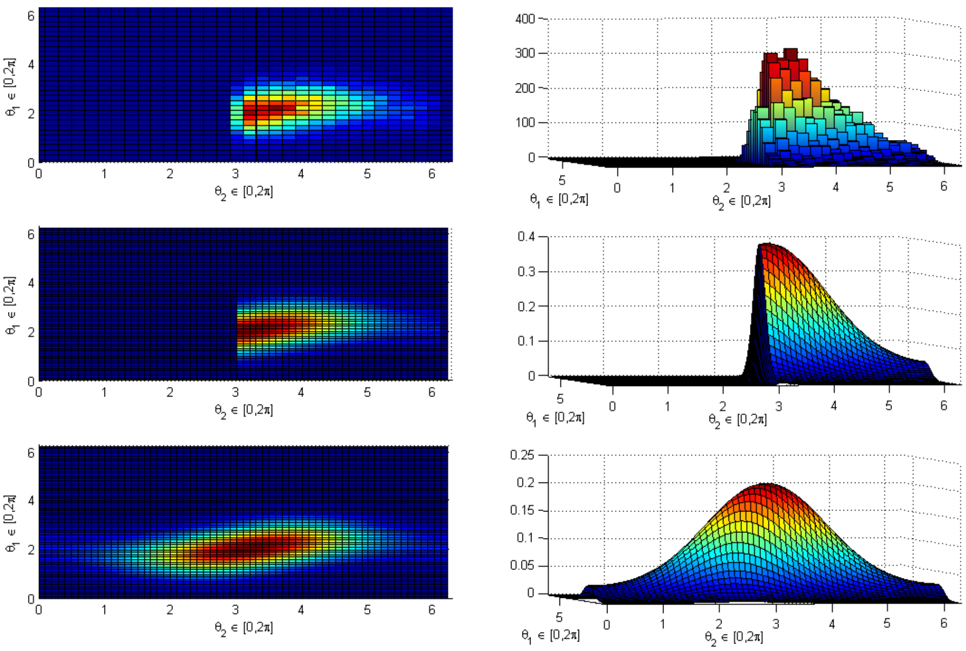

Figure 6: Simulation of a bivariate truncated von Mises distribution with parameters $\lambda = 1, \mu_1 = 2, \mu_2 = \pi, \kappa_1 = 3, \kappa_2 = 1, a_1 = 0, b_1 = 4, a_2 = 3, b_2 = 6$ from two different perspectives. The top plot corresponds to the data grouped by equal-length×equal-length value square areas in order to clearly ilustrate its relative frequency, the middle plot corresponds to the truncated bivariate von Mises distribution and the bottom plot corresponds to the associated bivariate distribution.

2. The next study was similar but reformulated the parameters to accont for the case when truncation omits one of the maxima of an otherwise bimodal distribution. To be precise, we studied 20000 points sampled from $f_{btvM}(\theta_1, \theta_2; 5, 2, \pi, 4, 2, 2.5, 4, 3, 6)$ (Figure 7), where the truncation parameters a_1 and a_2 were altered in order to select only part of one of the maxima. Notice also that in this case a significantly higher λ parameter was used in order to achieve bimodality. Maximizing the log-likelihood function for this case yielded parameter values $\hat{\lambda} = 4.9199, \hat{\mu}_1 = 2.0093, \hat{\mu}_2 = 3.0476, \hat{\kappa}_1 = 4.0643, \hat{\kappa}_2 = 1.48, \hat{a}_1 = 2.5008, \hat{b}_1 = 3.9788, \hat{a}_2 = 3.0014, \hat{b}_2 = 5.9947$ with a similar error of around 10^{-1}.

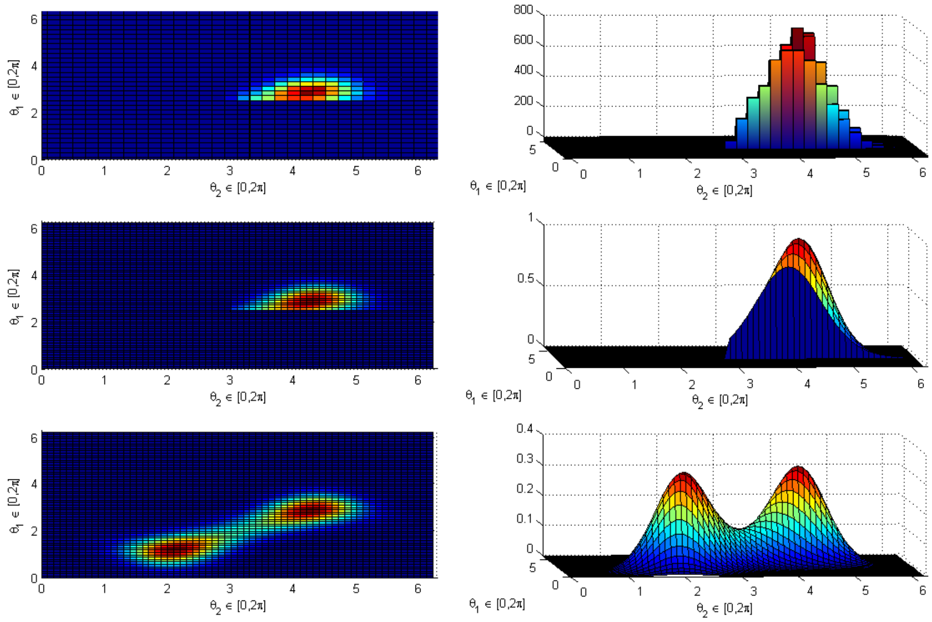

Figure 7: Simulation of a bivariate truncated von Mises distribution with parameters $\lambda = 5, \mu_1 = 2, \mu_2 = \pi, \kappa_1 = 4, \kappa_2 = 2, a_1 = 2.5, b_1 = 4, a_2 = 3, b_2 = 6$ from two different perspectives.

5 Real data applications

We further illustrate the truncated case by analyzing the data obtained by Bowyer and Danson. (2005) for the univariate case and Murzin AG (1995) for the bivariate case.

5.1 Leaf angle inclination

The data in Bowyer and Danson. (2005) was collected during a safari along the Kalahari Transect, southwest Botswana in 2001. It contains measurements of leaf inclination angles of four different woody plant species (*Acacia erioloba*, *Grewia flava*, *Acacia leuderitzii* and *Acacia mellifera*) across three different regions (Mabuasehube, Tsabong and Tshane). The measurements were taken using a clinometer.

In order to formally test the goodness-of-fit of the estimated distributions, we transform the data by means of the random variable $U = 2\pi\frac{[I(\theta,\mu,\kappa)-I(a,\mu,\kappa)]}{\int_a^b e^{\kappa\cos(\theta-\mu)}d\theta}$ mod 2π that is applied over the sorted sample $\theta_1,\ldots,\theta_n$. If the data distribute according to the truncated von Mises distribution, then the above random variable has a uniform distribution. As shown in Mardia and Jupp (2000), the modified Rayleigh statistic $S^* = (1-\frac{1}{2n})2nR^2+\frac{nR^4}{2}$, where n is the sample size and R the mean resultant length, distributes as a χ^2_2 distribution.

1. For the first study, the whole dataset containing a total of 741 samples was observed without further regard for region or type of plant (Table 1, Figure 8). A visual inspection of the plot clearly shows that the truncated von Mises distribution performs better. Formally, for the truncated case, $S^* = 2.8887$, which corresponds to a significance level of between 0.2 and 0.3 and acceptance of this distribution hypothesis. For the non-truncated case, $S^* = 25.5028$, with is a clear rejection with a significance level of less than 0.001. From these results we conclude that the truncated distribution is significantly better for these data. Truncation parameters conform the circular interval $\mathbb{O}_{0,\frac{\pi}{2}}$, which indicates no angle greater than 90° was measured in this study.

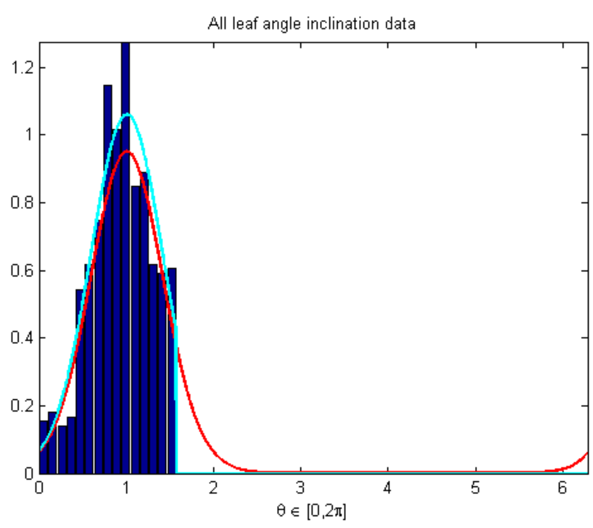

Figure 8: The study distribution and data representation of the entire dataset. The estimated truncated von Mises distribution (lighter line) clearly has higher density values than its associated von Mises distribution (darker line). The data are grouped by value intervals in order to observe its relative frequency (bars).

Table 1: Parameter values obtained after conducting the first study

	μ	κ	a	b	No.Samples
All data	1.0063	5.9602	0	1.5708	741

2. For the second study, we grouped the data by plant types without regards for region. This yielded four different distributions. A visual inspection shows that the univariate distributions are clearly better than the non-truncated von Mises distribution at describing the resulting data (Table 3, Figure 9), except for the case of *A. erioloba*. The goodness-of-fit tests (Table 2) revealed that the non-truncated distribution is rejected in all cases but in *A.erioloba*, whereas the truncated distribution hypothesis was more strongly accepted than that of the non-truncated distribution in all cases. Thus we can conclude that, for this study, the truncated distribution models the data better.

Table 2: Modified Rayleigh statistic values for the second study

	Truncated von Mises S^*	Non-truncated von Mises S^*
A. erioloba	3.014	3.5534
Grewia flava	0.0038	20.6273
A. leuderitzii	2.6073	10.1990
A. mellifera	1.3157	7.3046

Truncation parameters were consistently found to be in $\mathbb{O}_{0,\frac{\pi}{2}}$ except for *A.erioloba*, which also presented a significantly higher concentration parameter than in any of the other estimations. The irregularities in *A.erioloba* could partially be explained by the small sample size, which causes the estimations to be less reliable. More data would be needed to clarify the current results. On the whole, the remaining studies show few variations in the location-concentration parameters, which closely resemble the ones obtained in the first study.

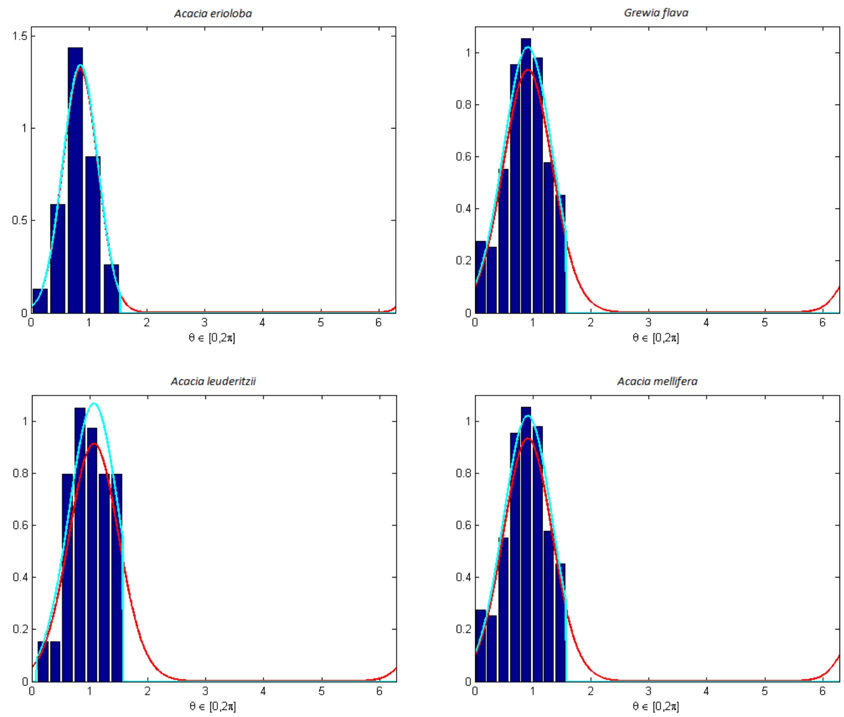

Figure 9: Studies of each type of plant.

Table 3: Parameter values yielded after conducting the second study

	μ	κ	a	b	No.Samples
A. Erioloba	0.8516	11.1894	0	1.5359	100
Grewia flava	1.1261	5.2668	0	1.5708	254
A. Leuderitzii	1.0706	5.5138	0	1.5708	184
A. Mellifera	0.9125	5.7396	0	1.5708	203

3. For the third study, we separately fitted univariate truncated distributions to the data for each plant in each region. Since not all plants were measured in all regions, this procedure produced eight different univariate truncated von Mises estimations. The distributions are generally observed to clearly differ from their associated non-truncated von Mises distribution, except in the first of the eight plots (Table 5, Figure 10). The goodness-of-fit tests (Table 4) are also consistent with previous studies. All truncated von Mises hypotheses were accepted, while around half of the non-truncated

distributions were rejected. Also, acceptance is stronger for the truncated case in all cases where both tested distributions were accepted. Thus, there is a strong suggestion that the truncated von Mises distribution properly models the underlying behavior that yielded the data.

Table 4: Parameter values yielded after conducting the third study

	Truncated von Mises S^*	Non-truncated von Mises S^*
A. erioloba, Mabuasehube	3.014	3.5534
Grewia flava, Mabuasehube	1.1543	8.9599
A. leuderitzii, Tsabong	2.0981	7.3115
Grewia flava, Tsabong	0.2050	3.8702
A. mellifera, Tsabong	0.1199	4.2131
Grewia flava(2), Tsabong	0.1165	9.7290
A. leuderitzii, Tshane	0.7002	2.8717
A. mellifera, Tshane	1.0525	10.2656

For this study, each distribution was estimated from a relatively small sample size ranging from 50 to 104 samples, which may have caused estimations to be less precise than desired. The concentration parameter shows the highest variability across the different cases (from 4.4078 to 11.1894 across the whole study or even from 4.8340 to 7.4245 in the case of *A. leuderitzii*). With more data it might be possible to distinguish if the variations in the concentration parameter are clearly influenced by the region of the plant species or the small sample size. Regarding the location parameter, there are few variations in the parameter value on the whole, *A. mellifera* being the species that experienced the highest variations w.r.t. one of the measurements in the first study. Truncation parameters remained consistently within the $\mathbb{O}_{0,\frac{\pi}{2}}$ interval.

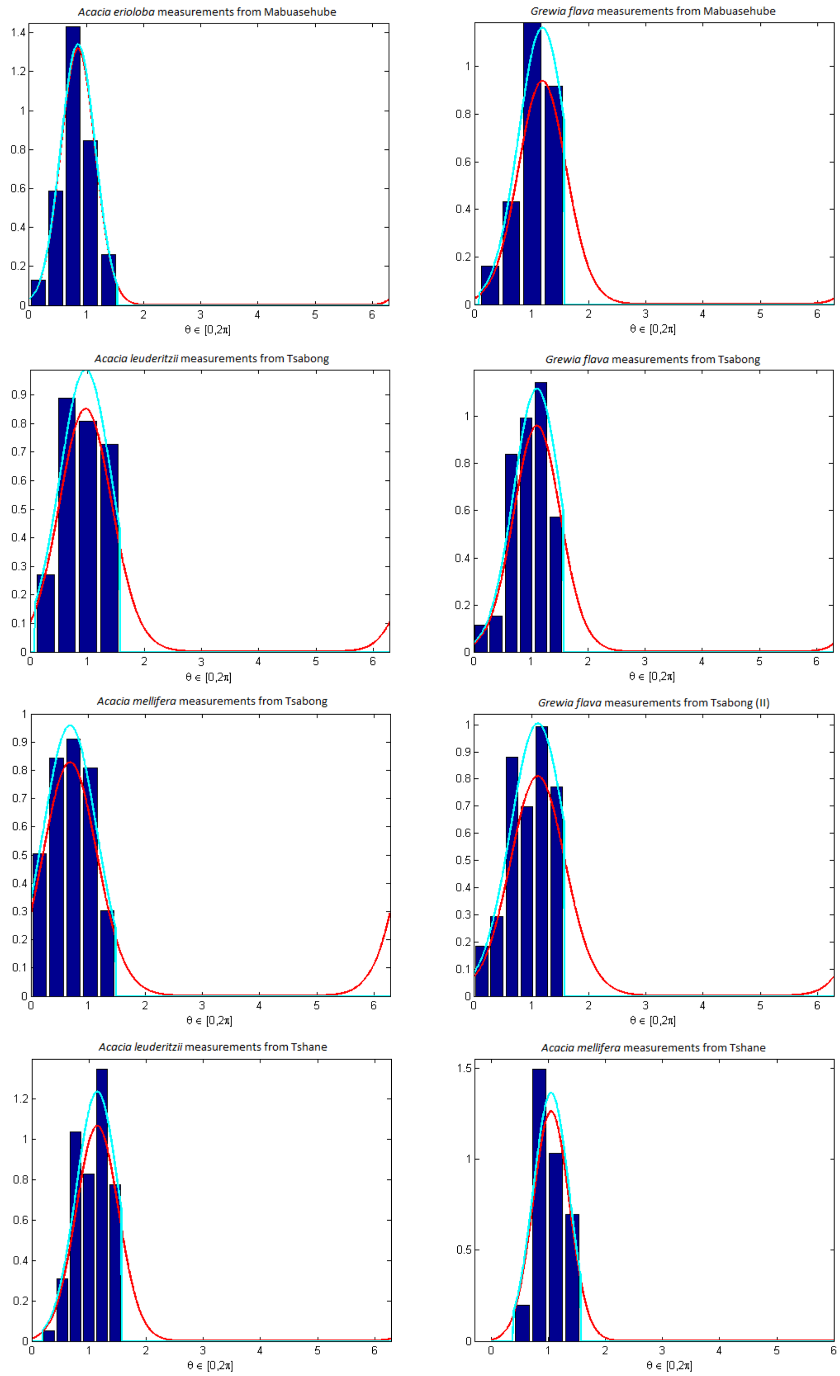

Figure 10: Studies of each type of plant in each region.

Table 5: Parameter values yielded after conducting the third study

	μ	κ	a	b	No.Samples
A. erioloba, Mabuasehube	0.8516	11.1894	0	1.5359	100
Grewia flava, Mabuasehube	1.1882	5.8142	0.0873	1.5708	50
A. leuderitzii, Tsabong	0.9712	4.8340	0.0873	1.5708	100
Grewia flava, Tsabong	1.1082	6.0832	0	1.5708	100
A. mellifera, Tsabong	0.6844	4.5884	0	1.4835	100
Grewia flava (2), Tsabong	1.1091	4.4078	0	1.5708	104
A. leuderitzii, Tshane	1.1474	7.4245	0.1920	1.5708	84
A. mellifera, Tshane	1.0525	10.2656	0.4014	1.5708	103

5.2 Pairs of dihedral angles in protein chains

This study consists of a single estimation for the bivariate case. The dataset corresponds to the spatial coordinates of the β-class atom in the SCOP-ASTRAL database (Murzin AG (1995)). Specifically, we extract the data on the position of the alpha-carbon atoms within a protein chain. The data are properly transformed from a set of coordinates to a set of pairs of angles (θ, ϕ) that account for the relative angular deviation of the next node in the protein chain has the previous part of the chain. After applying this procedure, the resulting sample contained 288610 observations.

The data have two clearly distinguishable modes that are separated by an area of low or zero density. This area is known to be empty in protein chains, which our model will assume by fixing the truncation parameters at $\mathbb{O}_{4.8,2.2}$. For this dataset, the bivariate truncated distribution is capable of capturing bimodality and defining an area of empty density as desired, and the parameter estimation by the maximum likelihood method provided a model that clearly exhibited theseh features (Figure 11).

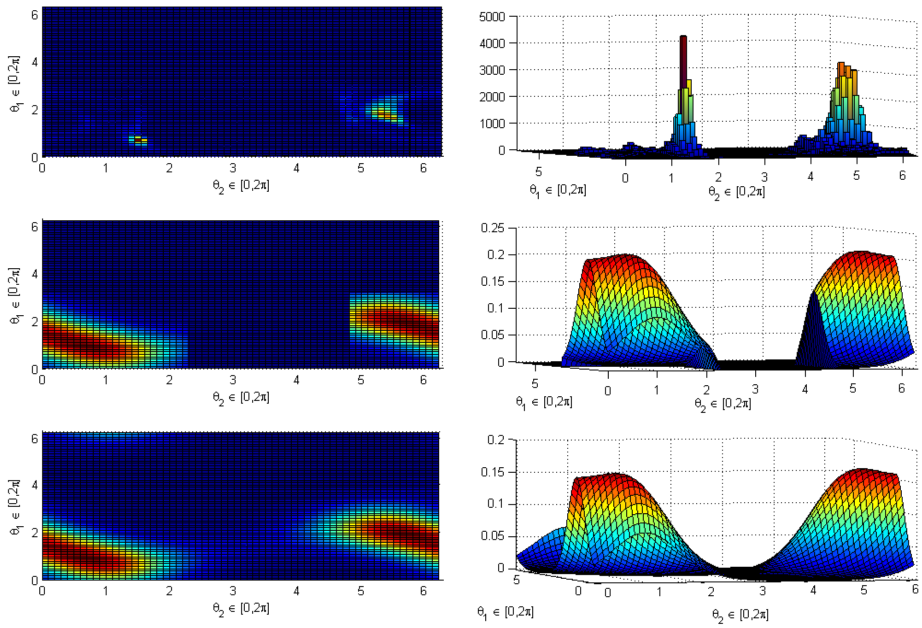

Figure 11: Estimated truncated von Mises distribution for the entire dataset. This distribution corresponds to the parameter values in Table 4. The top plot corresponds to a data frequency plot, the middle plot corresponds to the estimated bivariate truncated von Mises distribution and the bottom plot corresponds to its associated non-truncated distribution.

Table 6: Parameter values yielded after conducting the study on the protein chain dataset

λ	μ_1	μ_2	κ_1	κ_2	a_1	b_1	No.Samples
-1.93333	1.3859	0	2.1477	1.1213	0.0023	3.1363	288,610

In this case (Table 6), there seems to be a consistent inverse correlation between the pairs of angles as shown by the value of the lambda parameter. This plays a role in creating the observed bi-modality.

We conclude that the truncated von Mises distribution can be used to effectively model, simulate and summarize data about real-world phenomena and may be applied in any experiment whose outcome can be expressed in angular values.

6 Summary and conclusions

In this article we developed the theoretical framework of the univariate and the bivariate truncated von Mises distribution. To do this, we gave

1. The definition of a truncated von Mises distribution in the circle $\mathbb{O}$. The circular distribution is defined by means of the $\mathbb{O}$ subset, as the periodicity and properties of the circlehave to be naturally acknowledged for. If the a, b parameters were to not include 0° (our reference point) or $f_{tvM}(0; \mu, \kappa, a, b) = 0$, then linear definitions for the truncation parameters could be used. Therefore, $\mathbb{O}_{a,b}$ becomes $[a, b]$ with the restriction of $a < b$ and maximum likelihood estimators $\hat{a}, \hat{b}$ become $\hat{a} = \min\{\theta_1, \ldots, \theta_n\}$, $\hat{b} = \max\{\theta_1, \ldots, \theta_n\}$ respectively. The extension for the bivariate case is trivial.

2. The successfully determined expressions of the maximum likelihood estimators. For both univariate and bivariate cases, solely sample-dependent maximum likelihood estimators of the truncation parameters were found, while the other parameters showed interdependency.

3. The resulting moments of the univariate case and existing interrelationships.

4. The properties of both bivariate and univariate cases, especially the results concerning the additional manipulability and possible shapes of the distribution when modifying the truncation parameters, that is, a distribution can be made to be strictly increasing or strictly decreasing symmetrical or non-symmetrical function, or to concentrate its positive support in an arbitrarily short sub-interval.

5. The bivariate case and studies of the shape and behavior of marginal and conditional distributions. We determined that every conditional truncated von Mises distribution is a univariate truncated von Mises distribution. For the case of the marginal distribution, we concluded that only for parameter $\lambda = 0$ does the distribution behave like a truncated univariate von Mises distribution. When $\lambda \neq 0$, the resultant marginal distribution is a potentially bi-maximal not a von Mises distribution. We obtained four different cases in order to characterize all the different truncated marginal von Mises

behaviors. To be precise, when the marginal distribution is bi-maximal, it exhibits either one of two or two global maxima and the minimum value is not at μ_1 if truncation parameters a_2 and b_2 are not symmetrical w.r.t. μ_2, and it is if they are. When the marginal distribution is unimodal, the maximum value is not at μ_1 if truncation parameters a_2 and b_2 are not symmetrical w.r.t. μ_2, and it is if they are. This covers all possible shapes or behaviors.

SUPPLEMENTARY MATERIAL

Appendix All proofs of introduced lemmas and theorems (.pdf file).

References

Abramowitz, M., and Stegun, I. (1964), *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Applied Mathematics Series Dover Publications.

Bistrian, D. A., and Iakob, M. (2008), "One-dimensional truncated von Mises distribution in data modeling," *Annals of Faculty of Engineering Hunedoara, tome VI (year 2008),fascicule 3*, .

Bowyer, P., N. M. T., and Danson., F. M. (2005), "SAFARI 2000 Canopy Structural Measurements, Kalahari Transect, Wet Season 2001. Data set.,".

Jupp, P. E., and Mardia, K. V. (1989), "A unified view of the theory of directional statistics, 1975-1988," *International Statistical Review*, 57(3), 261–294.

Lopez-Cruz, P., Bielza, C., and Larrañaga, P. (2013), "Directional naive Bayes classifiers," *Pattern Analysis and Applications*, pp. 1–22.

Mardia, K., and Jupp, P. (2000), *Directional Statistics*, Wiley Series in Probability and Statistics.

Mardia, K. V., Hughes, G., Taylor, C. C., and Singh, H. (2008), "A multivariate von Mises distribution with applications to bioinformatics," *Canadian Journal of Statistics*, 36, 99–109.

Mardia, K. V., and Voss, J. (2011), "Some fundamental properties of a multivariate von Mises distribution," *ArXiv e-prints*, .

Murzin AG, Brenner SE, H. T. C. C. (1995), "SCOP: a structural classification of proteins database for the investigation of sequences and structures,".

Singh, H. (2002), "Probabilistic model for two dependent circular variables," *Biometrika*, 89(3), 719–723.

Computational Intelligence Group, Department of Artificial Intelligence, Universidad Politécnica de Madrid

E-mail: pablo.fernandezgonz@fi.upm.es

Department of Artificial Intelligence, Universidad Politécnica de Madrid

E-mail: mcbielza@fi.upm.es

Department of Artificial Intelligence, Universidad Politécnica de Madrid

E-mail: plarranaga@fi.upm.es

Suplementary Documents

Pablo Fernandez-Gonzalez, Concha Bielza, Pedro Larrañaga

Department of Artificial Intelligence, Universidad Politécnica de Madrid

1 APPENDIX. Proofs of the results

1.1 Proof of Lemma 1

1. If $f_{tvM}(\theta;\mu,\kappa,a,b)$ is strictly decreasing $\Leftrightarrow$ a,b satisfy $a,b \in \mathbb{O}_{\mu,\mu+\pi}$.The proof follows simply from noting that the von Mises distribution decreases from its maximum to its minimum value in $\mathbb{O}_{\mu,\mu+\pi}$. If $\mathbb{O}_{a,b} \subset \mathbb{O}_{\mu,\mu+\pi}$, the resulting truncated von Mises distribution exhibits a monotonic decreasing behavior.

2. Analogously, if $f_{tvM}(\theta;\mu,\kappa,a,b)$ is strictly increasing $\Leftrightarrow$ truncation parameters a,b satisfy $\mathbb{O}_{a,b} \subset \mathbb{O}_{\mu-\pi,\mu}$.

3. If $f_{tvM}(\theta;\mu,\kappa,a,b)$ increases and decreases reaching a single maximum $\Leftrightarrow$ truncation parameters a,b satisfy $\mu \in \mathbb{O}_{a,b}$ and $\mu+\pi, \notin \mathbb{O}_{a,b}$

4. If $f_{tvM}(\theta;\mu,\kappa,a,b)$ increases and decreases reaching a single minimum $\Leftrightarrow$ truncation parameters a,b satisfy $\mu+\pi \in \mathbb{O}_{a,b}$ and $\mu,\mu+2\pi \notin \mathbb{O}_{a,b}$.

5. If $f_{tvM}(\theta;\mu,\kappa,a,b)$ increases and decreases with both single maximum and single minimum $\Leftrightarrow$ the truncation parameters a,b satisfy either $\mu,\mu+\pi \in \mathbb{O}_{a,b}$ or $\mu,\mu-\pi \in \mathbb{O}_{a,b}$

1.2 Proof of Lemma 2.

We have, by means of the power series expansion of the $e^{(\cdot)}$ function,

$$I(\theta;\mu,\kappa) = \int f_{uvM}(\theta;\mu,\kappa)d\theta = \int e^{\kappa\cos(\theta-\mu)}d\theta = \int \sum_{n=0}^{\infty} \frac{(\kappa\cos(\theta-\mu))^n}{n!}d\theta,$$

where $f_{uvM}(\theta;\mu,\kappa)$ is the unnormalized von Mises distribution, and $I(\theta;\mu,\kappa)$ is its distribution function. Therefore, $\int_a^b f_{uvM}(\theta;\mu,\kappa) = I(b;\mu,\kappa) - I(a;\mu,\kappa)$.

Considering that $\sum_{n=0}^{\infty} \frac{|(\kappa\cos(\theta-\mu))^n|}{n!}$ is a solely positive continuous bounded function in $[1,e^{\kappa}]$, and, therefore, for any finite integral coefficients $i_1,i_2 \in \mathbb{R}$, it satisfies $\int_{i_1}^{i_2}\sum_{n=0}^{\infty}\frac{(\kappa\cos(\theta-\mu))^n}{n!}d\theta <$

∞, we can conclude that it satisfies the Fubini-Tonelli theorem conditions for integral summation exchange.

We then follow with the procedure for the indefinite integral:

$$\begin{aligned} I(\theta;\mu,\kappa) &= \int \sum_{n=0}^{\infty} \frac{(\kappa\cos(\theta-\mu))^n}{n!} d\theta \\ &= \sum_{n=0}^{\infty} \int \frac{(\kappa\cos(\theta-\mu))^n}{n!} d\theta \\ &= \sum_{n=0}^{\infty} \frac{\kappa^n}{n!} \int \cos^n(\theta-\mu) d\theta. \end{aligned} \tag{1}$$

The above integral is defined in a recursive way as

$$\int \cos^n(\theta-\mu)d\theta = \frac{\sin(\theta-\mu)\cos^{n-1}(\theta-\mu)}{n} + \frac{n-1}{n}\int \cos^{n-2}(\theta-\mu)d\theta.$$

And it can be calculated by the procedure of integration by parts. In this appendix, however, we give a non-recursive expression:

$$\int \cos^n(\theta-\mu)d\theta = \sin(\theta-\mu)\left(\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor + \bmod \frac{n}{2} - 1} \left(\cos^{n-2i-1}(\theta-\mu)\frac{\prod_{j=0}^{2i}(n-j)}{\prod_{j=0}^{i}(n-2j)^2}\right)\right) \ \forall n \text{ such that } n = 2m+1$$

with $m \in \mathbb{N}$. This materializes out of the observation of the numerical regularities that appear when "unfolding" the recursive expression:

$$\begin{aligned} \int \cos^n(\theta-\mu)d\theta &= \frac{\sin(\theta-\mu)\cos^{n-1}(\theta-\mu)}{n} + \frac{n-1}{n}\int \cos^{n-2}(\theta-\mu)d\theta \\ &= \frac{\sin(\theta-\mu)\cos^{n-1}(\theta-\mu)}{n} + \frac{n-1}{n}\left(\frac{\sin(\theta-\mu)\cos^{n-3}(\theta-\mu)}{n-2} + \frac{n-3}{n-2}\int \cos^{n-4}(\theta-\mu)d\theta\right) \\ &= \frac{1}{n}\sin(\theta-\mu)\cos^{n-1}(\theta-\mu) + \frac{n-1}{n(n-2)}\sin(\theta-\mu)\cos^{n-3}(\theta-\mu) \\ &\quad + \frac{(n-1)(n-3)}{n(n-2)(n-4)}\sin(\theta-\mu)\cos^{n-5}(\theta-\mu) + \frac{(n-1)(n-3)(n-5)}{n(n-2)(n-4)}\int \cos^{n-6}(\theta-\mu)d\theta \end{aligned}$$

They can be primary generalized using the expression

$$\sin(\theta-\mu)\left(\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor + \bmod \frac{n}{2} - 1} \left(\cos^{n-2i-1}(\theta-\mu)\frac{\prod_{j=0}^{2i}(n-j)}{\prod_{j=0}^{i}(n-2j)^2}\right)\right)$$

However, while this first expression does suffice for odd n, an extra term appears if n is even as we reach the point at which the term $\int \cos^0(\theta-\mu)d\theta$ is computed. This can be reflected properly by adding an addend that takes into account the parity of the formula. In our case, it has the form:

$$g(n,x) = \frac{(-1)^n h(x) + h(x)}{2} = \frac{((-1)^n + 1)h(x)}{2},$$

where $\forall n \in \mathbb{N}$ such that $n = 2m$ and $m \in \mathbb{N}, g(n,x) = h(x)$ and 0 otherwise.

In a shorter notation and adding the parity term, the expression becomes

$$\int \cos^n(\theta-\mu)d\theta = \sin(\theta-\mu)\left(\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor + \bmod \frac{n}{2} - 1}\left(\cos^{n-2i-1}(\theta-\mu)\prod_{j=0}^{2i}(n-j)^{-(-1)^j}\right) + \frac{((-1)^n+1)\prod_{j=0}^{\lfloor \frac{n}{2} \rfloor + \bmod \frac{n}{2} - 1}(n-j)^{-(-1)^j}(\theta-\mu)}{2}\right).$$

Thus, substituting in (A.1) we obtain the final expression for $\int e^{\kappa\cos(\theta-\mu)}d\theta$.

1.3 Proof of Theorem 1

The theorem is entirely derived by means of the trigonometrical equality:

$$\begin{aligned} &\kappa_2\cos(x) + c_2\sin(x) \\ &= \left[\kappa_2\cos\left(\arctan\left(\frac{c_2}{\kappa_2}\right)\right) + c_2\sin\left(\arctan\left(\frac{c_2}{\kappa_2}\right)\right)\right]\cos\left(x - \arctan\left(\frac{c_2}{\kappa_2}\right)\right). \end{aligned} \tag{2}$$

From the equality we can express the exponent of the conditional distribution in (11) using a formula of the type $\kappa'\cos(x-\mu')$. Now if we consider that

$$\kappa_2\cos\left(\arctan\left(\frac{c_2}{\kappa_2}\right)\right) + c_2\sin\left(\arctan\left(\frac{c_2}{\kappa_2}\right)\right) = \frac{\kappa_2 + \frac{c_2^2}{\kappa_2}}{\sqrt{1+\left(\frac{c_2}{\kappa_2}\right)^2}} = \sqrt{\kappa_2^2 + c_2^2},$$

then (A.1) becomes

$$\kappa_2\cos(x) + c_2\sin(x) = \sqrt{\kappa_2^2 + c_2^2}\cos\left(x - \arctan\left(\frac{c_2}{\kappa_2}\right)\right). \tag{3}$$

Thus, we can adapt the truncated conditional distribution to the univariate truncated von Mises exponent by properly selecting:

$$\begin{aligned} \kappa' &= \sqrt{\kappa_2^2 + c_2^2} \\ \mu' &= \mu_2 + \arctan\left(\frac{c_2}{\kappa_2}\right), \end{aligned}$$

where $c_2 = \lambda\sin(\theta_1 - \mu_1)$.

1.4 Proof of Theorem 2

We consider

$$f_{umtvM}(\theta_{1'}) = e^{\kappa_1 \cos(\theta_{1'})} \int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2-\mu_2)+\lambda \sin(\theta_{1'})\sin(\theta_2-\mu_2)} d\theta_2 \tag{4}$$

to be the unnormalized marginal truncated von Mises distribution. For simplicity's sake, the proof is developed in a linear context (using classical intervals [x,y], with their associated constraints, instead of circular intervals $\mathbb{O}_{x,y}$), whose extension to the circle is deemed as known and trivial at this point. Also, unless otherwise specified, $\lambda > 0$ is assumed and a_2, b_2 truncation parameters are referred to simply as the truncation parameters. The proof is as follows:

(a) Determination of the derivative expression and the $T(\cdot,\cdot,\cdot,\cdot,\cdot,\cdot)$ function

(b) Analysis of the marginal expression with focus on the case of symmetrical truncation parameters in order to prove cases 1 and 2

(c) Further analysis for the case of non-symmetrical truncation parameters, determining all distinctive behaviors of the integral subterm of the marginal expression

(d) Monotony study divided by cases of the circular distance of the truncation parameters w.r.t. μ_2 and subintervals of the $\theta_{1'} \in [-\pi, \pi]$ interval in order to prove case 3. Case 4 is proven by ruling out every other possible outcome.

In (a), $T(\cdot,\cdot,\cdot,\cdot,\cdot,\cdot)$ is derived from a particularization of the second derivative of the marginal function. The meaning of the value of the $T(\cdot,\cdot,\cdot,\cdot,\cdot,\cdot)$ function is clarified for the symmetrical truncation parameters. In (b) and (c), the analysis aims to characterize the behavior of the integral term of the marginal distribution. In (b), the analysis will first observe the particularities of the integral term, especially, how $\theta_{1'}$ modifies the location and concentration parameters of the von Mises distribution inside the integral, and then derive from it some properties and insights will also be used for the proof of case 3. We then prove how thsese variations affect the area under the curve and their relationships to the truncation parameters. Finally, partial and total analyses of the derivate of the integral term are performed, concluding the proof of the first two cases of the theorem. In (c), an analysis of the derivate of the integral term for non-symmetrical truncation parameters w.r.t. μ_2 is performed. Using the previous insights, the analysis first determines the cases where, according to the truncation parameter values, the marginal integral term follows a unimodal distribution. The analysis then focuses on the remaining cases in order to prove that the global maximum of the integral term necessarily appears at the associated point of the truncation parameter ($-\frac{\pi}{2}$ for a_2 and $\frac{\pi}{2}$ for b_2), which has the largest circular distance w.r.t. μ_2. Also, in the bi-modal case for non-symmetrical truncation parameters, we analyze how the minimum comprehended between the modes appears in the $\frac{\pi}{2}$−length interval with 0 as an extrema associated with the truncation parameter that has the smallest circular distance w.r.t. μ_2 ($[-\frac{\pi}{2}, 0]$ for a_2 and $[0, \frac{\pi}{2}]$ for b_2), and its relationship with the minimum that appears in $[-\pi, -\frac{\pi}{2}]$ for the associated interval $[-\frac{\pi}{2}, 0]$ or in $[\frac{\pi}{2}, \pi]$ for the associated interval $[0, \frac{\pi}{2}]$. In (d), the monotony study identifies all different behaviors and the subinterval in which more than one critical point can occur, thus enabling us to detect bi-modality with different valued maxima with the proposed criteria.

(a) By differentiating $f_{umtvM}(\theta_{1'})$ w.r.t. θ_1 we obtain:

$$\begin{aligned} f'_{umtvM}(\theta_{1'}) &= -\kappa_1 \sin(\theta_{1'}) e^{\kappa_1 \cos(\theta_{1'})} \int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2-\mu_2)+\lambda \sin(\theta_{1'}) \sin(\theta_2-\mu_2)} d\theta_2 \\ &\quad +\lambda \cos(\theta_{1'}) e^{\kappa_1 \cos(\theta_{1'})} \int_{a_2}^{b_2} \sin(\theta_2-\mu_2) e^{\kappa_2 \cos(\theta_2-\mu_2)+\lambda \sin(\theta_{1'}) \sin(\theta_2-\mu_2)} d\theta_2 \\ &= e^{\kappa_1 \cos(\theta_{1'})} \Big(-\kappa_1 \sin(\theta_{1'}) \int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2-\mu_2)+\lambda \sin(\theta_{1'}) \sin(\theta_2-\mu_2)} d\theta_2 \\ &\quad +\lambda \cos(\theta_{1'}) \int_{a_2}^{b_2} \sin(\theta_2-\mu_2) e^{\kappa_2 \cos(\theta_2-\mu_2)+\lambda \sin(\theta_{1'}) \sin(\theta_2-\mu_2)} d\theta_2 \Big). \end{aligned} \tag{5}$$

We observe that

$$\begin{aligned} f'_{umtvM}(0) &= \lambda e^{\kappa_1} \left(\int_{a_2}^{b_2} \sin(\theta_2-\mu_2) e^{\kappa_2 \cos(\theta_2-\mu_2)} d\theta_2 \right) \\ &= \frac{\lambda}{\kappa_2} e^{\kappa_1} \left(e^{\kappa_2 \cos(a_2-\mu_2)} - e^{\kappa_2 \cos(b_2-\mu_2)} \right). \end{aligned} \tag{6}$$

If and only if $\cos(b_2-\mu_2) = \cos(a_2-\mu_2)$, it follows that $f_{umtvM}(\theta_{1'})$ has a critical point at μ_1.

Solving and assessing the equation $f''_{umtvM}(\theta_{1'}) = 0$ in order to obtain information about the curvature for $\theta_{1'} = 0$ results in

$$-\frac{\kappa_1}{\lambda^2} + \frac{\int_{a_2}^{b_2} \sin^2(\theta_2-\mu_2) e^{\kappa_2 \cos(\theta_2-\mu_2)} d\theta_2}{\int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2-\mu_2)} d\theta_2} = 0,$$

from which we can define the $T(\cdot,\cdot,\cdot,\cdot,\cdot,\cdot)$ function as

$$T(\lambda, \mu_2, \kappa_1, \kappa_2, a_2, b_2) = -\frac{\kappa_1}{\lambda^2} + \frac{\int_{a_2}^{b_2} \sin^2(\theta_2-\mu_2) e^{\kappa_2 \cos(\theta_2-\mu_2)} d\theta_2}{\int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2-\mu_2)} d\theta_2}. \tag{7}$$

However, we still need to understand whether Equation (A.7) is sufficient to distinguish between cases 1 and 2 established in the theorem.

(b) In order to understand the truncated marginal behavior, if we rewrite the integral term in $f_{umtvM}(\theta_{1'})$ by means of Equation (A.3) we have

$$f_{umtvM}(\theta_{1'}) = e^{\kappa_1 \cos(\theta_{1'})} \int_{a_2}^{b_2} e^{\sqrt{\kappa_2^2+(\lambda \sin(\theta_{1'}))^2} \cos\left(x_2-\mu_2-\arctan\left(\frac{\lambda \sin(\theta_{1'})}{\kappa_2}\right)\right)} d\theta_2.$$

It is apparent that the integral term computes the area of location-concentration varying von Mises distributions as $\int_{a_2}^{b_2} f_{tvM}\left(\theta_2; \mu_2 + \arctan\left(\frac{\lambda \sin(\theta_{1'})}{\kappa_2}\right), \sqrt{\kappa_2^2 + (\lambda \sin(\theta_{1'}))^2}\right) d\theta_2$. If we consider the location variations over $[-\pi, \pi]$ by means of the $\sin(\theta_{1'})$ function, the distribution in the integrand is displaced over the interval $[-\arctan\left(\frac{\lambda}{\kappa_2}\right), 0]$ when $\sin(\theta_{1'}) < 0$ (from displacement 0 to displacement $-\arctan\left(\frac{\lambda}{\kappa_2}\right)$ when $\theta_{1'} \in [-\pi, -\frac{\pi}{2}]$ and from displacement $-\arctan\left(\frac{\lambda}{\kappa_2}\right)$ to displacement 0 when $\theta_{1'} \in [-\frac{\pi}{2}, 0]$), and over the interval $[0, \arctan\left(\frac{\lambda}{\kappa_2}\right)]$ when $\sin(\theta_{1'}) > 0$

(similary for $\theta_{1'} \in [0, \frac{\pi}{2}]$ and $\theta_{1'} \in [\frac{\pi}{2}, \pi]$). If we consider concentration variations, we can regard the source of bi-modality of the integral term as the $\sqrt{\kappa_2^2 + (\lambda \sin(\theta_{1'}))^2}$ subterm, given that $\sin^2(\theta_{1'})$ is a π−periodic solely positive function. Additionally, from $\theta_{1'} = 0$ to $\theta_{1'} = \frac{\pi}{2}$ and from $\theta_{1'} = -\pi$ to $\theta_{1'} = -\frac{\pi}{2}$, the concentration parameter grows from its minimum value κ_2 to its maximum value $\sqrt{\kappa_2^2 + \lambda^2}$, while it decreases from its maximum to its minimum value in the cases of $\theta_{1'}$ from $-\frac{\pi}{2}$ to 0 and from $\frac{\pi}{2}$ to π.

The proof then follows trivially by noting that, truncation parameters aside, the function's behavior in $[\mu_2 - \pi, \mu_2]$ can be considered symmetrical w.r.t. μ_2 to the function's behavior in $[\mu_2, \mu_2 + \pi]$. The symmetry w.r.t. μ_2 in the truncation parameters selects two subintervals of symmetrical behavior w.r.t. μ_1, thus producing a function that is symmetrical w.r.t. μ_1.

Further analyzing the integral term we look to determine the critical points and understand how the selection of truncation parameters affects the integral term behaviour. We take

$$\begin{aligned} v_1(\theta_{1'}) &= \int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_{1'}) \sin(\theta_2 - \mu_2)} d\theta_2 \\ v_2(\theta_{1'}) &= \int_{a_2}^{b_2} \sin(\theta_2 - \mu_2) e^{\kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_{1'}) \sin(\theta_2 - \mu_2)} d\theta_2, \end{aligned}$$

where

$$\lambda \cos(\theta_{1'}) v_2(\theta_{1'}) = v_1'(\theta_{1'}).$$

We now want to analyze $v_2(\theta_{1'})$ as it is part of the derivate expression of $v_1(\theta_{1'})$. Taking the integrand of $v_2(\theta_{1'})$ to be

$$f_{v_2}(\theta_2; \theta_{1'}) = \sin(\theta_2 - \mu_2) e^{\kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_{1'}) \sin(\theta_2 - \mu_2)}$$

Note that, in $f_{v_2}(\theta_2; \theta_{1'})$, the argument is θ_2 since it creates the area that is to be computed in $v_2(\theta_{1'})$. $\theta_{1'}$ can be considered here as a modifying parameter. The $f_{v_2}(\theta_2; \theta_{1'})$ function comprises the product of a strictly positive function $e^{(\cdot)}$ and a $\sin(\cdot)$ function. Therefore, the sign of $f_{v_2}(\theta_2; \theta_{1'})$ is solely determined by the sign of the $\sin(\cdot)$ function. To be precise, if $\theta_2 \in [\mu_2 - \pi, \mu_2]$ then $f_{v_2}(\theta_2; \theta_{1'}) \leq 0$ and if $\theta_2 \in [\mu_2, \mu_2 + \pi]$ then $f_{v_2}(\theta_2; \theta_{1'}) \geq 0$. Therefore, we can subdivide $v_2(\theta_{1'})$ as

$$v_2(\theta_{1'}) = \int_{a_2}^{\mu_2} f_{v_2}(\theta_2; \theta_{1'}) d\theta_2 + \int_{\mu_2}^{b_2} f_{v_2}(\theta_2; \theta_{1'}) d\theta_2,$$

where the first addend is a solely negative term and the second addend is a solely positive term provided that $\mu_2 \in (a_2, b_2)$. In the symmetry case, if $\theta_{1'} = 0$ we have

$$-\int_{a_2}^{\mu_2} f_{v_2}(\theta_2; 0) d\theta_2 = \int_{\mu_2}^{b_2} f_{v_2}(\theta_2; 0) d\theta_2; \tag{8}$$

for $\theta_{1'} \in (0, \pi)$ we have

$$-\int_{a_2}^{\mu_2} f_{v_2}(\theta_2; \theta_{1'}) d\theta_2 < \int_{\mu_2}^{b_2} f_{v_2}(\theta_2; \theta_{1'}) d\theta_2; \tag{9}$$

and for $\theta_{1'} \in (-\pi, 0)$ we have

$$-\int_{a_2}^{\mu_2} f_{v_2}(\theta_2;\theta_{1'})d\theta_2 > \int_{\mu_2}^{b_2} f_{v_2}(\theta_2;\theta_{1'})d\theta_2 \tag{10}$$

Intuitively, the displaced exponential w.r.t. the μ_2 term increases all the values of either the negative or the positive curve of the $\sin(\theta_2 - \mu_2)$ function and reduces the curve of the opposite sign in less amount, therefore defining the sign and the value of $v_2(\theta_{1'})$. Formally, this to hold, we need to prove that $\forall\theta_{1'} \in (-\pi, 0)$ $f_{v_2}(\theta_2; 0) - f_{v_2}(\theta_2;\theta_{1'}) > 0$ if $\theta_2 \in (\mu_2 - \pi, \mu_2)$ and $\forall\theta_{1'} \in (-\pi, 0)$ $f_{v_2}(\theta_2; 0) - f_{v_2}(\theta_2;\theta_{1'}) < 0$ if $\theta_2 \in (\mu_2, \mu_2 + \pi)$ for the negative displacement, and an analogous statement for $\theta_{1'} \in (0, \pi)$ positive displacement. For the negative displacement case, it follows that

$$\begin{aligned}
\sin(\theta_2 - \mu_2)e^{\kappa_2\cos(\theta_2-\mu_2)} - \sin(\theta_2 - \mu_2)e^{\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_{1'})\sin(\theta_2-\mu_2)} &> 0\\
\sin(\theta_2 - \mu_2)\left(e^{\kappa_2\cos(\theta_2-\mu_2)} - e^{\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_{1'})\sin(\theta_2-\mu_2)}\right) &> 0.
\end{aligned}$$

As $\sin(\theta_2 - \mu_2) < 0$ in $\theta_2 \in [\mu_2 - \pi, \mu_2]$ it suffices if

$$e^{\kappa_2\cos(\theta_2-\mu_2)} - e^{\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_{1'})\sin(\theta_2-\mu_2)} < 0$$

in $\theta_2 \in [\mu_2 - \pi, \mu_2]$. We proceed as follows:

$$\begin{aligned}
e^{\kappa_2\cos(\theta_2-\mu_2)} - e^{\kappa_2\cos(\theta_2-\mu_2)+\lambda\sin(\theta_{1'})\sin(\theta_2-\mu_2)} &< 0\\
e^{-\lambda\sin(\theta_{1'})\sin(\theta_2-\mu_2)} &< 1\\
-\lambda\sin(\theta_{1'})\sin(\theta_2 - \mu_2) &< 0
\end{aligned}$$

and, since we have specified $\theta_{1'} \in (-\pi, 0)$ and then $\sin(\theta_{1'}) < 0$, we have $-\lambda\sin(\theta_{1'}) > 0$. Therefore, the sign of $-\lambda\sin(\theta_{1'})\sin(\theta_2 - \mu_2)$ follows from that of $\sin(\theta_2 - \mu_2)$. This proves the statement for both θ_2 intervals in the case of negative displacement. The proof for positive displacement is analogous.

This result implies that the selection of truncation parameters that are symmetrical w.r.t. μ_2 does not change the monotony of $v_1(\theta_{1'})$. More generally, this result implies that no selection of truncation parameters changes the monotonicity of $v_2(\theta_{1'})$, that is, increasing in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and decreasing otherwise.

Since (A.8), (A.9) and (A.10) hold, we can now perform the sign and critical points analysis of $\lambda\cos(\theta_{1'})v_2(\theta_{1'})$ to obtain that $v_1(\theta_{1'})$ follows the monotony of $\sin^2(\theta_{1'})$ for any a_2, b_2 such that $\cos(b_2 - \mu_2) = \cos(a_2 - \mu_2)$, with critical points $\{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$. Therefore, in Equation (A.4), unimodal/bimodal observed distributions are "decided" for this case by the product of $v_1(\theta_{1'})$ with $e^{\kappa_1\cos(\theta_{1'})}$.

Therefore, if $T(\lambda, \mu_2, \kappa_1, \kappa_2, a_2, b_2) > 0$ then $f_{umtvM}(\theta_{1'})$ presents a minimum critical point at μ_1 and the distribution has two equal symmetrical maxima in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ (the maxima location interval can be proven as a result of monotony and sign comparisons between $v_1(\theta_{1'})$ and $e^{\kappa_1\cos(\theta_{1'})}$). Respectively, if $T(\lambda, \mu_2, \kappa_1, \kappa_2, a_2, b_2) < 0$ then $f_{umtvM}(\theta_{1'})$ presents a maximum critical point and the distribution is unimodal. This result generalizes the outcome for the non-truncated case to symmetrical parameters other than a_2, b_2 such that $b_2 - a_2 = 2\pi$ (Singh (2002)). This suffices to prove cases 1 and 2 of the theorem.

(c) For case 3 we want to observe the behavior of the marginal distribution for different cases of circular distances of a_2, b_2 truncation parameters w.r.t. μ_2. Thus, we need knowledge about the subterm $v_2(\theta_{1'})$ when a_2, b_2 truncation parameters are not symmetrical w.r.t. μ_2 in order to reach useful results. We will address this point first.

If we now observe $\lambda \cos(\theta_{1'}) v_2(\theta_{1'}) = 0$ for non-symmetrical parameters we can as before, isolate two critical points:

$$\theta_{1'} = -\frac{\pi}{2}, \qquad \theta_{1'} = \frac{\pi}{2}$$

and a third critical point at some $\theta_{1'}$ such that $-\int_{a_2}^{\mu_2} f_{v_2}(\theta_2;\theta_{1'}) + \int_{\mu_2}^{b_2} f_{v_2}(\theta_2;\theta_{1'}) = 0$ if a_2, b_2 are not truncation parameters that satisfy any of the following conditions:

(i) $a_2, b_2 \in [\mu_2, \mu_2 + \pi]$ as then $v_2(\theta_{1'}) > 0\ \forall \theta_{1'} \in [-\pi, \pi]$

(ii) $a_2, b_2 \in [\mu_2 - \pi, \mu_2]$ as then $v_2(\theta_{1'}) < 0\ \forall \theta_{1'} \in [-\pi, \pi]$

(iii) $\mu_2 \in (a_2, b_2)$ such as $-\int_{a_2}^{\mu_2} f_{v_{2'}}(\theta_2; -\frac{\pi}{2}) d\theta_2 \le \int_{\mu_2}^{b_2} f_{v_{2'}}(\theta_2; -\frac{\pi}{2}) d\theta_2$ as then $v_2(\theta_{1'}) > 0\ \forall \theta_{1'} \in [-\pi, \pi]$

(iv) $\mu_2 \in (a_2, b_2)$ such as $\int_{\mu_2}^{b_2} f_{v_{2'}}(\theta_2; \frac{\pi}{2}) d\theta_2 \le -\int_{a_2}^{\mu_2} f_{v_{2'}}(\theta_2; \frac{\pi}{2}) d\theta_2$ as then $v_2(\theta_{1'}) < 0\ \forall \theta_{1'} \in [-\pi, \pi]$.

Notice that from the viewpoint of truncation parameters, cases (iii) and (iv) can be considered opposite. Also, as highlighted by the previous analysis, it is clear that case (iii) implies $\cos(b_2 - \mu_2) < \cos(a_2 - \mu_2)$ (more intuitively, $\cos(b_2 - \mu_2) \ll \cos(a_2 - \mu_2)$) and case (iv) $\cos(b_2 - \mu_2) > \cos(a_2 - \mu_2)$ (more intuitively, $\cos(b_2 - \mu_2) \gg \cos(a_2 - \mu_2)$). We will refer to cases (iii) and (iv) as the strong lower parameter cases.

Therefore, by manipulating a_2, b_2 truncation parameters, it is possible to reshape $v_1(\theta_{1'})$ to exhibit a minimum in $-\frac{\pi}{2}$ and a maximum in $\frac{\pi}{2}$ if case (i) or (iii) applies or to exhibit a maximum in $-\frac{\pi}{2}$ and a minimum in $\frac{\pi}{2}$ if case (ii) or (iv) applies. In these cases, $v_1(\theta_{1'})$ is an integral term with unimodal behavior.

It follows that any other case for non-symmetrical truncation parameters implies $\mu_2 \in (a_2, b_2)$, and $v_1(\theta_{1'})$ exhibits two differentiated maxima in $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Also, $v_2(-\frac{\pi}{2}) < 0$ and $v_2(\frac{\pi}{2}) > 0$. If we examine the case of $\theta_{1'} = 0$ for truncation parameters a_2, b_2 such that $\cos(b_2 - \mu_2) > \cos(a_2 - \mu_2)$ then $-\int_{a_2}^{\mu_2} f_{v_2}(\theta_2; 0) d\theta_2 > \int_{\mu_2}^{b_2} f_{v_2}(\theta_2; 0) d\theta_2$ and therefore $v_2(\theta_{1'}) = 0$ for some $\theta^*_{1'} \in [0, \frac{\pi}{2}]$ such that $v_2(\theta_{1'}) < 0$ if $\theta_{1'} \in [0, \theta^*_{1'})$ and $v_2(\theta_{1'}) > 0$ if $\theta_{1'} \in (\theta^*_{1'}, \frac{\pi}{2}]$. It follows that this also impliesthe existence of another minimum in $[\frac{\pi}{2}, \pi]$ as $v_2(\theta_{1'}) > 0\ \forall \theta_{1'} \in [\frac{\pi}{2}, \pi - \theta^*_{1'})$ and $v_2(\theta_{1'}) < 0\ \forall \theta_{1'} \in (\pi - \theta^*_{1'}, \pi]$. Similarly, if $\cos(b_2 - \mu_2) < \cos(a_2 - \mu_2)$ then $-\int_{a_2}^{\mu_2} f_{v_2}(\theta_2; 0) d\theta_2 < \int_{\mu_2}^{b_2} f_{v_2}(\theta_2; 0) d\theta_2$ and therefore $v_2(\theta_{1'}) = 0$ for some $\theta^*_{1'} \in [-\frac{\pi}{2}, 0]$ and $-\pi - \theta^*_{1'} \in [-\pi, -\frac{\pi}{2}]$, that is, the minimum of $v_1(\theta_{1'})$ that appears in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ is more precisely located in the $\frac{\pi}{2}$–length interval associated with the truncation parameter that presents the smallest circular distance w.r.t. μ_2 and imples an additional minimum located in the contiguous $\frac{\pi}{2}$–length interval more distant from $\theta_{1'} = 0$.

Additionally, the global maximum of the two differentiated maxima is that of the $\frac{\pi}{2}$–length interval associated with the truncation parameter that has the largest circular distance w.r.t. μ_2. We can prove this by comparing both maxima as follows:

$$v_1\left(-\frac{\pi}{2}\right) - v_1\left(\frac{\pi}{2}\right) > 0 \text{ if } \cos(b_2 - \mu_2) > \cos(a_2 - \mu_2).$$

Thus if we take $\kappa' = \sqrt{\kappa_2^2 + (\lambda)^2}$ we have

$$\int_{a_2}^{b_2} e^{\kappa' \cos\left(\theta_2 - \mu_2 - \arctan\left(-\frac{\lambda}{\kappa_2}\right)\right)} d\theta_2 - \int_{a_2}^{b_2} e^{\kappa' \cos\left(\theta_2 - \mu_2 - \arctan\left(\frac{\lambda}{\kappa_2}\right)\right)} d\theta_2 > 0.$$

Expressing this by means of the distribution function we obtain

$$\left[I(\theta, -\mu_2 - \arctan\left(-\frac{\lambda}{\kappa_2}\right), \kappa')\right]_{a_2}^{b_2} - \left[I(\theta, -\mu_2 - \arctan\left(\frac{\lambda}{\kappa_2}\right), \kappa')\right]_{a_2}^{b_2} > 0. \tag{11}$$

Clearly, $I(\theta, \mu, \kappa)$ is strictly increasing and $e^{\kappa' \cos\left(\theta_2 - \mu_2 - \arctan\left(-\frac{\lambda}{\kappa_2}\right)\right)}$ is symmetrical to $e^{\kappa' \cos\left(\theta_2 - \mu_2 - \arctan\left(\frac{\lambda}{\kappa_2}\right)\right)}$ w.r.t. μ_2. Therefore

1. $$\left[I(\theta, -\mu_2 - \arctan\left(\frac{-\lambda}{\kappa_2}\right), \kappa')\right]_{2\mu_2 - b_2}^{\mu_2} = \left[I(\theta, -\mu_2 - \arctan\left(\frac{\lambda}{\kappa_2}\right), \kappa')\right]_{\mu_2}^{b_2}$$

2. $$\left[I(\theta, -\mu_2 - \arctan\left(\frac{-\lambda}{\kappa_2}\right), \kappa')\right]_{\mu_2}^{2\mu_2 - a_2} = \left[I(\theta, -\mu_2 - \arctan\left(\frac{\lambda}{\kappa_2}\right), \kappa')\right]_{a_2}^{\mu_2}$$

taking

$$\left[I(\theta, -\mu_2 - \arctan\left(\frac{-\lambda}{\kappa_2}\right), \kappa')\right] = Ie_1(\theta)$$
$$\left[I(\theta, -\mu_2 - \arctan\left(\frac{\lambda}{\kappa_2}\right), \kappa')\right] = Ie_2(\theta),$$

we can rewrite inequation (A.11) as

$$[Ie_1(\theta)]_{a_2}^{\mu_2} + [Ie_1(\theta)]_{\mu_2}^{b_2} - [Ie_2(\theta)]_{a_2}^{\mu_2} - [Ie_2(\theta)]_{\mu_2}^{b_2} > 0,$$

substituting,

$$\begin{aligned}
[Ie_1(\theta)]_{\mu_2}^{a_2} + [Ie_1(\theta)]_{b_2}^{\mu_2} - [Ie_1(\theta)]_{\mu_2}^{2\mu_2 - a_2} - [Ie_1(\theta)]_{2\mu_2 - b_2}^{\mu_2} &> 0 \\
-Ie_1(a_2) + Ie_1(b_2) - Ie_1(2\mu_2 - a_2) + Ie_1(2\mu_2 - b_2) &> 0 \\
[Ie_1(\theta)]_{a_2}^{2\mu_2 - b_2} - [Ie_1(\theta)]_{b_2}^{2\mu_2 - a_2} &> 0,
\end{aligned}$$

that is, the inequation reduces to the comparison between the area in two subintervals of equal length that are symmetrical w.r.t. μ_2. By this symmetry and by the fact that the mode is in $(-\frac{\pi}{2}, 0)$ and the anti-mode in $(\frac{\pi}{2}, \pi)$ in $e^{\kappa' \cos\left(\theta_2 - \mu_2 - \arctan\left(-\frac{\lambda}{\kappa_2}\right)\right)}$, we can safely conclude that the inequation holds thus proving the statement. Therefore, for any marginal truncated distribution, the global maximum in the integral term is located in $\theta_{1'} = \frac{\pi}{2}$ if $\cos(a_2 - \mu_2) > \cos(b_2 - \mu_2)$ and in $\theta_{1'} = -\frac{\pi}{2}$ if $\cos(a_2 - \mu_2) < \cos(b_2 - \mu_2)$.

At this point all behaviors for critical points and monotony of $v_1(\theta_{1'})$ have been characterized.

Analogously to the non-truncated case, the effect of the $e^{\kappa_1 \cos(\theta_{1'})}$ subterm has to be taken into consideration in order to determine the shape of the distribution. To do this, we perform a monotony study that incorporates all previous developments.

(d) After conducting the study on $v_2(\theta_{1'})$ and $v_1(\theta_{1'})$, we proceed by equating function (A.5) to zero, resulting in

$$-\kappa_1 \sin(\theta_{1'}) v_1(\theta_{1'}) + \lambda \cos(\theta_{1'}) v_2(\theta_{1'}) = 0.$$

If we consider the cases where $a_2, b_2 \in [\mu_2, \mu_2 + \pi]$ or a_2 is a strong lower parameter w.r.t b_2 we have:

1. $v_2(\theta_{1'}) > 0 \ \forall \theta_{1'} \in [-\pi, \pi]$.

2. If $\theta_{1'} \in [-\pi, -\frac{\pi}{2}]$, then $\sin(\theta_{1'}) \leq 0$ and $\cos(\theta_{1'}) \leq 0$. In this case, at least a minimum and a critical point of $f_{umtvM}(\theta_{1'})$ can be found in the examined interval as shown by:

$$\begin{aligned} f'_{umtvM}(-\pi) &= e^{-\kappa_1} \left(-\lambda \int_{a_2}^{b_2} \sin(\theta_2 - \mu_2) e^{\kappa_2 \cos(\theta_2 - \mu_2)} d\theta_2 \right) \\ f'_{umtvM}\left(-\frac{\pi}{2}\right) &= \kappa_1 \int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2 - \mu_2) - \lambda \sin(\theta_2 - \mu_2)} d\theta_2 > 0, \end{aligned}$$

   where $f'_{umtvM}(-\pi) < 0$. Notice that if $a_2, b_2 \in [\mu_2, \mu_2 + \pi]$ the critical point neccesarily exists regardless of the effect of the other parameters.

3. If $\theta_{1'} \in [-\frac{\pi}{2}, 0]$, then $\sin(\theta_{1'}) \leq 0$ and $\cos(\theta_{1'}) \geq 0$. $f_{umtvM}(\theta_{1'})$ exhibits a monotonic increasing behavior, as all terms involved in the expression are positive.

4. If $\theta_{1'} \in [0, \frac{\pi}{2}]$, then $\sin(\theta_{1'}) \geq 0$ and $\cos(\theta_{1'}) \geq 0$. Here, at least a maximum and a critical point can be found in the interval by considering Equation (A.6), where $f'_{umtvM}(0) > 0$, and

$$f'_{umtvM}\left(\frac{\pi}{2}\right) = -\kappa_1 \int_{a_2}^{b_2} e^{\kappa_2 \cos(\theta_2 - \mu_2) - \lambda \sin(\theta_2 - \mu_2)} d\theta_2 < 0.$$

5. If $\theta_{1'} \in [\frac{\pi}{2}, \pi]$, then $\sin(\theta_{1'}) \geq 0$ and $\cos(\theta_{1'}) \leq 0$. $f_{umtvM}(\theta_{1'})$ exhibits a monotonic decreasing behavior, as all terms involved in the expression are negative.

Therefore, for this case, the distribution exhibits critical points in two non-contiguous intervals. By the previous developments, such a distribution of critical points would only correspond to the unimodal case and also, as the contribution of $e^{\kappa_1 \cos(\theta_{1'})}$ is symmetrical w.r.t. μ_1 or $\theta_{1'} = 0$, the marginal function could only have one global maximum in $\theta_{1'} \in [0, \frac{\pi}{2}]$ interval and one global minimum in $\theta_{1'} \in [-\pi, -\frac{\pi}{2}]$.

The case where $a_2, b_2 \in [\mu_2 - \pi, \mu_2]$ or b_2 is a strong lower parameter w.r.t a_2 can be understood as "symmetric behavior w.r.t μ_1", since the results for $\theta_{1'} \in [-\pi, -\frac{\pi}{2}]$ now hold for $\theta_{1'} \in [\frac{\pi}{2}, \pi]$ and the results for $\theta_{1'} \in [-\frac{\pi}{2}, 0]$ now hold for $\theta_{1'} \in [0, \frac{\pi}{2}]$. This property, general to the $[-\pi, \pi]$ interval, guarantees that in our case, it suffices to determine the behavior for one of the two remaining cases to completely determine the behavior of the marginal function.

We now consider the remaining parameter configurations that satisfy $\cos(b_2 - \mu_2) > \cos(a_2 - \mu_2)$.

1. If $\theta_{1'} \in [-\pi, -\frac{\pi}{2}]$, then $v_2(\theta_{1'}) < 0$, thus resulting in $f_{umtvM}(\theta_{1'})$, which exhibits a strictly increasing behavior, as all terms involved in the expression are now positive.

2. If $\theta_{1'} \in [-\frac{\pi}{2}, 0]$, then $v_2(\theta_{1'}) < 0$. In this case, after performing sign comparisons on the extrema, there is at least one critical point and one maximum in the interval.

3. If $\theta_{1'} \in [0, \frac{\pi}{2}]$, $v_2(\theta_{1'}) < 0 \ \forall \theta_{1'} \in [0, \theta^*_{1'})$ and $v_2(\theta_{1'}) > 0 \ \forall \theta_{1'} \in [\theta^*_{1'}, \frac{\pi}{2})$. Therefore, no critical point exists in $[0, \theta^*_{1'})$, since $f_{umtvM}(\theta_{1'})$ exhibits a decreasing behavior and all terms involved in the expression are negative. In $[\theta^*_{1'}, \frac{\pi}{2})$, no, one or two critical points can occur as both sign and monotony comparisons were not conclusive.

4. If $\theta_{1'} \in [\frac{\pi}{2}, \pi]$, then $v_2(\theta_{1'}) > 0 \ \forall \theta_{1'} \in [\frac{\pi}{2}, \pi - \theta^*_{1'})$ and $v_2(\theta_{1'}) < 0 \ \forall \theta_{1'} \in (\pi - \theta^*_{1'}, \pi]$. Therefore, no critical point exists in $[\frac{\pi}{2}, \pi - \theta^*_{1'})$ since $f_{umtvM}(\theta_{1'})$ exhibits a decreasing behavior as all terms involved in the expression are negative. In $(\pi - \theta^*_{1'}, \pi]$, after performing sign comparisons on the extrema, at least one critical point can occur. Therefore, for this case, the distribution has three contiguous intervals containing critical points. Since clearly no more than two critical points are allowed in a $\frac{\pi}{2}$–length interval, the case with two possible critical points in $[\theta^*_{1'}, \frac{\pi}{2})$ is the case of bi-maximality (differentiated maxima) with a minimum and a maximum in $\theta_{1'} \in [\theta^*_{1'}, \frac{\pi}{2})$ and a maximum in $\theta_{1'} \in [-\frac{\pi}{2}, 0]$. Complementarily, this distribution of critical points "corresponds" to the bi-maximal (differentiated maxima) behavior of $v_1(\theta_{1'})$, and, therefore, the critical point in $\theta_{1'} \in [-\frac{\pi}{2}, 0]$ is necessarily a maximum, and the critical point in $[\frac{\pi}{2}, \pi]$ is necessarily a minimum. Thus, it can be concuded that in the case of bimodality, the interval associated with the truncation parameter that has the shortest circular distance w.r.t. μ_2 contains the two critical points, whereas the interval associated with the truncation parameter that has the largest circular distance w.r.t. μ_2 contains the global maximum.

If $\lambda < 0$, the proof follows trivially by noting that the displacement caused by the $\sin(\cdot)$ function in the exponent that appears in the $v_1(\theta_{1'})$ subterm is the opposite. This in turn causes the distribution to have an opposite symmetrical behaviour w.r.t. μ_1. This suffices to prove case 3 of the theorem. Case 4 can also be proven with the developed theory. However, it can additionally be proven by ruling out any other possible outcome, considering the three previously developed cases.