

Tree-Structured Bayesian Networks for Wrapped Cauchy Directional Distributions

Ignacio Leguey^(✉), Concha Bielza, and Pedro Larrañaga

Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid,
Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain
ig.leguey@upm.es, {mcbielza,pedro.larranaga}@fi.upm.es
<http://cig.fi.upm.es>

Abstract. Modelling the relationship between directional variables is a nearly unexplored field. The bivariate wrapped Cauchy distribution has recently emerged as the first closed family of bivariate directional distributions (marginals and conditionals belong to the same family). In this paper, we introduce a tree-structured Bayesian network suitable for modelling directional data with bivariate wrapped Cauchy distributions. We describe the structure learning algorithm used to learn the Bayesian network. We also report some simulation studies to illustrate the algorithms including a comparison with the Gaussian structure learning algorithm and an empirical experiment on real morphological data from juvenile rat somatosensory cortex cells.

1 Introduction

Directional distributions are widely used in many areas such as geography, geology, geophysics, medicine, meteorology, oceanography or biology [1]. Traditional statistics methods are sometimes unequal to the task of dealing with directional data because data periodicity has to be taken into account. For example, whereas 0 and 360 are different points in non-directional data, dealing with angles, 0° and 360° are considered as the same point. Hence, the analysis of directional data is different and more challenging than non-directional data. There is substantial literature on directional data [6, 8, 18]. The von Mises distribution [20] is the best-known directional model and can be considered the directional analogue of the univariate normal distribution. Mardia [15, 16] introduced a bivariate von Mises distribution and its extension to the multivariate case [17]. He showed that the conditional distributions are also von Mises distributions. However, the marginal distributions are either unimodal or bimodal, and only the unimodal case could be approximated to a von Mises distribution when the concentration parameter was large. The wrapped Cauchy distribution is another popular symmetric distribution on the circle introduced by Lévy [14]. It was studied

by Wintner [25]. Further, it was shown that it could be obtained by mapping Cauchy distributions [19] onto the circle. Recently Kato and Pewsey [10] proposed a five-parameter bivariate wrapped Cauchy distribution for toroidal data, whose marginals and conditionals follow univariate wrapped Cauchy distributions. Their family is therefore said to be closed.

Probabilistic graphical models are widely used for non-directional data. These models, as a marriage between graph theory and probability theory, have some interesting properties that make them a useful and interesting tool for data modelling. For example they are easy interpretable, handle missing data nicely, treat inference and learning together or focus on conditional independence and computational issues. These models have been successfully applied in several different areas such as medicine, education or neuroscience [12], but, no directional probabilistic graphical models have yet been developed. One exception is Boomsma et al. [3], who modelled a specific part of the geometry of proteins called $C\alpha$ using a hidden Markov model that outputs amino acid symbols, secondary structure symbols and unit vectors. They used the directional five-parameter Fisher-Bingham distribution [11] on the unit sphere to represent the unit vectors. Also Razavian et al. [22] developed an undirected von Mises [20] graphical model using L_1 regularization for structure learning and Gibbs sampling for performing inference. They also compared their model with a Gaussian graphical model, demonstrating that the von Mises graphical model achieves higher accuracy than the Gaussian graphical model.

Our aim is to find a tree-structured model that reveals the relationship between several directional variables. A family distribution that is closed under conditionality and marginalization (marginals and conditionals belong to the same family) is necessary to make the modelling phase easier and simplify the subsequent inference process. In this paper we introduce a tree-structured Bayesian network for a five-parameter bivariate wrapped Cauchy, learning the structure from data using a directional mutual information criteria. The remainder of this paper is organized as follows. Section 2 reviews the univariate and bivariate wrapped Cauchy distributions. In Sect. 3 we explain the Bayesian network structure learning from data. Section 4 presents the simulation process for this model. Section 5 addresses a real data example using the proposed tree-structured Bayesian network. The paper ends in Sect. 6 with some concluding remarks and some proposals for future work.

2 Univariate and Bivariate Wrapped Cauchy

One of the best-known directional distributions is the wrapped Cauchy distribution [14]. A random variable Θ that follows a wrapped Cauchy distribution $wC(\mu, \varepsilon)$ has a density function

$$f(\theta) = \frac{1}{2\pi} \frac{1 - \varepsilon^2}{1 + \varepsilon^2 - 2\varepsilon \cos(\theta - \mu)}, \quad \theta, \mu \in (-\pi, \pi], \varepsilon \in [0, 1) \quad (1)$$

where μ is the mean angle and ε the concentration parameter. Equation (1) is unimodal and symmetric about μ , unless $\varepsilon = 0$ which yields the directional

uniform distribution. Data simulation from a $wC(\mu, \varepsilon)$ is already implemented in the ‘‘Circular’’ R package [21].

Kato and Pewsey [10] proposed a five-parameter bivariate wrapped Cauchy distribution, which is unimodal, pointwise symmetric around the mean and has a closed-form expression for the mode. A dependence parameter controls the correlation from total independence to perfect correlation. This distribution materializes as an appealing submodel of a six-parameter distribution obtained by applying a restricted version of the Möbius circle transformation to a random vector from a bivariate directional distribution previously proposed by Kato [9].

A random vector (Θ_1, Θ_2) is said to follow a bivariate wrapped Cauchy distribution $bwC(\mu_1, \mu_2, \varepsilon_1, \varepsilon_2, \rho)$ if its density function is given by

$$f(\theta_1, \theta_2) = c[c_0 - c_1 \cos(\theta_1 - \mu_1) - c_2 \cos(\theta_2 - \mu_2) - c_3 \cos(\theta_1 - \mu_1) \cos(\theta_2 - \mu_2) - c_4 \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)]^{-1}, \quad \theta_1, \theta_2 \in (-\pi, \pi] \quad (2)$$

where

$c = (1 - \rho^2)(1 - \varepsilon_1^2)(1 - \varepsilon_2^2)/4\pi^2$, $c_0 = (1 + \rho^2)(1 + \varepsilon_1^2)(1 + \varepsilon_2^2) - 8|\rho|\varepsilon_1\varepsilon_2$, $c_1 = 2(1 + \rho^2)\varepsilon_1(1 + \varepsilon_2^2) - 4|\rho|(1 + \varepsilon_1^2)\varepsilon_2$, $c_2 = 2(1 + \rho^2)(1 + \varepsilon_1^2)\varepsilon_2 - 4|\rho|\varepsilon_1(1 + \varepsilon_2^2)$, $c_3 = -4(1 + \rho^2)\varepsilon_1\varepsilon_2 + 2|\rho|(1 + \varepsilon_1^2)(1 + \varepsilon_2^2)$, $c_4 = 2\rho(1 - \varepsilon_1^2)(1 - \varepsilon_2^2)$, $\mu_1, \mu_2 \in (-\pi, \pi]$, $\varepsilon_1, \varepsilon_2 \in [0, 1]$, $\rho \in (-1, 1)$, and with ε_1 and ε_2 regulating the concentration of the marginal distributions and ρ being the correlation coefficient between Θ_1 and Θ_2 . When $\varepsilon_1, \varepsilon_2 > 0$ Eq. (2) is unimodal and pointwise symmetric about (μ_1, μ_2) .

As explained by McCullagh [19], computations are simplified in many wrapped Cauchy models by representing them in complex form. Let $Z = e^{i\Theta}$, where Θ is distributed as in Eq. (1), then the density function of Z is

$$f(z; \lambda) = \frac{1}{2\pi} \frac{|1 - |\lambda|^2|}{|z - \lambda|^2}, \quad z \in \Omega, \lambda \in \hat{\mathbb{C}} \setminus \Omega \quad (3)$$

where $\lambda = \varepsilon e^{i\mu}$, $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ and $\Omega = \{z \in \mathbb{C} : |z| = 1\}$. We use the notation $Z \sim C^*(\lambda)$ to denote that Z is distributed as in Eq. (3).

Similarly to Eq. (3), by representing random variables in complex form, let $(Z_1, Z_2) = (e^{i\Theta_1}, e^{i\Theta_2})$, where (Θ_1, Θ_2) is distributed as in Eq. (2), then the density of (Z_1, Z_2) is:

$$f(z_1, z_2) = \frac{(4\pi^2)^{-1}(1 - \rho^2)(1 - \varepsilon_1^2)(1 - \varepsilon_2^2)}{|a_{11}(\bar{z}_1\eta_1)^q z_2 \bar{\eta}_2 + a_{12}(\bar{z}_1\eta_1)^q + a_{21}z_2\bar{\eta}_2 + a_{22}|^2}, \quad z_1, z_2 \in \Omega \quad (4)$$

where q is the sign of ρ , $\eta_k = e^{i\mu_k}$ with $k \in \{1, 2\}$, \bar{z}_n is the complex conjugate of z_n , $a_{11} = \varepsilon_1\varepsilon_2 - |\rho|$, $a_{12} = |\rho|\varepsilon_2 - \varepsilon_1$, $a_{21} = |\rho|\varepsilon_1 - \varepsilon_2$, $a_{22} = 1 - |\rho|\varepsilon_1\varepsilon_2$, $\varepsilon_1, \varepsilon_2 \in [0, 1]$, $\rho \in (-1, 1)$ and $\eta_1, \eta_2 \in \Omega$.

Following the complex notation, we denote $(Z_1, Z_2) \sim bC^*(\eta_1, \eta_2, \varepsilon_1, \varepsilon_2, \rho)$ if (Z_1, Z_2) is distributed as in Eq. (4). This complex representation of a five-parameter bivariate wrapped Cauchy verifies the following result.

Theorem 1. (Kato and Pewsey [10]) A random vector (Z_1, Z_2) with density given by Eq. (4) has marginals $Z_1 \sim C^*(\varepsilon_1 \eta_1)$ and $Z_2 \sim C^*(\varepsilon_2 \eta_2)$, and conditionals $Z_1|Z_2 = z_2 \sim C^*(-\eta_1[A \circ (z_2 \overline{\eta_2})^q])$ and $Z_2|Z_1 = z_1 \sim C^*(-\eta_2[A^T \circ (z_1 \overline{\eta_1})^q])$, where A is defined in Eq. (5), A^T is the transpose of A , and

$$A \circ z = \frac{a_{11}z + a_{12}}{a_{21}z + a_{22}}.$$

$$A: \begin{bmatrix} a_{11} = \varepsilon_1 \varepsilon_2 - |\rho| & a_{12} = |\rho| \varepsilon_2 - \varepsilon_1 \\ a_{21} = |\rho| \varepsilon_1 - \varepsilon_2 & a_{22} = 1 - |\rho| \varepsilon_1 \varepsilon_2 \end{bmatrix}. \quad (5)$$

To the best of our knowledge, this is the first bivariate directional distribution for which marginal and conditional distributions are well-known, mathematically tractable and from the same family. Thus, we consider the wrapped Cauchy distribution to be suitable for developing our Bayesian network model, as we are using a tree structure that needs only bivariate and conditional densities.

2.1 Parameter Estimation

Working with the density given by Eq. (2), there is no closed-form expression for the maximum likelihood estimates, and numerical optimization methods must be used to find them. Although maximum likelihood estimation is the most common parameter estimation method, Kato and Pewsey [10] showed that method of moments is more efficient for our purpose. We use the method of moments, where all formulas for the estimates can be expressed in a closed form, as it is easier to implement and is computationally very fast.

Let $\{(\theta_{1j}, \theta_{2j}), j = 1, \dots, n, \}$ be a random sample from a $bwC(\mu_1, \mu_2, \varepsilon_1, \varepsilon_2, \rho)$ as stated in Eq. (2). Then the method of moments estimators [4] of $\mu_1, \mu_2, \varepsilon_1, \varepsilon_2$ and ρ from [10] are

$$\begin{aligned} \tilde{\mu}_1 &= \text{Argument}(\bar{R}_1) \quad \text{with} \quad \bar{R}_1 = \frac{1}{n} \sum_{j=1}^n e^{i\theta_{1j}}, \\ \tilde{\mu}_2 &= \text{Argument}(\bar{R}_2) \quad \text{with} \quad \bar{R}_2 = \frac{1}{n} \sum_{j=1}^n e^{i\theta_{2j}}, \\ \tilde{\varepsilon}_1 &= |\bar{R}_1|, \quad \tilde{\varepsilon}_2 = |\bar{R}_2|, \\ \tilde{\rho} &= \frac{1}{n} \left(\left| \sum_{j=1}^n e^{i(\Phi_{1j} - \Phi_{2j})} \right| - \left| \sum_{j=1}^n e^{i(\Phi_{1j} + \Phi_{2j})} \right| \right) \\ &\text{with} \quad \Phi_{rj} = 2 \arctan \left(\frac{1 + \tilde{\varepsilon}_r}{1 - \tilde{\varepsilon}_r} \tan \left(\frac{\theta_{rj} - \tilde{\mu}_r}{2} \right) \right), \quad r = 1, 2 \end{aligned} \quad (6)$$

3 Wrapped Cauchy Tree-Structured Bayesian Network: Structure Learning

Assuming that the topology of the graph that we want to represent is unknown and not given, we must learn the structure of the model from our random sample. The so-called Bayesian networks structure learning has been studied at length for linear data. However, it is far from straightforward to adapt the structure learning algorithms to the directional domain. There are two different kinds of algorithms for structure learning in Bayesian networks. One kind of algorithm

learns a network structure by capturing the conditional independences between the different triplets of variables used for the model. The best-known method for this kind of structure learning is the PC algorithm [23]. The other kind of algorithm, which we’ve chose for our structure learning problem, is based on score and search. It tackles the problem of structure learning as an optimization problem.

We use a maximum weight spanning tree structure learning algorithm. This is a variant of the algorithm introduced by Chow and Liu [5]. Instead of the mutual information measure, traditionally used for running this algorithm, for the first time we introduce circular mutual information (CMI) for directional variables.

Let Θ_i, Θ_j be wrapped Cauchy random variables, the CMI between a pair of variables is

$$CMI(\Theta_i, \Theta_j) = \int_0^{2\pi} \int_0^{2\pi} f(\theta_i, \theta_j) \log\left(\frac{f(\theta_i, \theta_j)}{f(\theta_i)f(\theta_j)}\right) d\theta_j d\theta_i. \quad (7)$$

where their marginal density functions $f(\theta_i), f(\theta_j)$ and the joint density function $f(\theta_i, \theta_j)$ have been previously estimated.

Since the resulting integral is intractable, it has to be approximated using numerical methods. We use the Cubature method [2, 24]. Cubature is an adaptive multidimensional integration algorithm over hypercubes. It is the best method in terms of approximation error and computational cost. This method is also best suited for dimensions lower than seven. This applies to the tree structure, whose maximum dimension is two.

Like the traditional mutual information measure for linear variables, the CMI represents the weight of the edge that links Θ_i and Θ_j . To determine the parent of each node (only one parent per node), we select the root node and follow the structure learned by our algorithm (Algorithm 1) (i.e., in Fig. 1a, the selected root node is Θ_1 , which is the parent of Θ_2 and Θ_3 , Θ_2 is the parent of Θ_4 and Θ_3 is the parent of Θ_5 and Θ_6). Given a tree structure with N nodes, there are N possible resulting trees depending on the selected root node (see Fig. 1a).

Algorithm 1. Adaptation of Chow Liu algorithm

- 1: Given $\Theta_1, \Theta_2, \dots, \Theta_N$ directional random variables, compute the joint distribution $f(\theta_i, \theta_j)$ for all variable pairs.
 - 2: Using the pairwise distributions, compute all $N(N - 1)/2$ edge weights and order them by CMI value.
 - 3: Assign the largest two edges to the tree to be represented.
 - 4: Examine the next-largest edge, and add it to the tree unless it forms a loop, in which case discard it and examine the next largest edge.
 - 5: Repeat step 4 until $N - 1$ edges have been selected (so the spanning tree is finished).
-

We use this algorithm for several reasons: (a) It uses only second-order statistics, which are easily and reliably measured from data, (b) we avoid the use

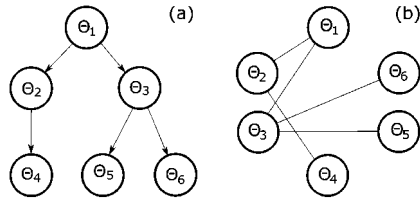


Fig. 1. (a) Wrapped Cauchy tree-structured network representation with $N = 6$. Each node $\Theta_i (i = 1, \dots, 6)$ represents a wC random variable. The root node selected is Θ_1 . (b) Wrapped Cauchy non-directed tree-structured network learned by Algorithm 1 from the structure of (a) created by simulation.

of conditional independence tests, which are often computationally expensive and only available by simulation [10] from wrapped Cauchy and (c) is a way of ensuring a tree-structure.

4 Simulation

In order to demonstrate the accuracy of the learning algorithm, we report the results of several simulation studies and the comparison with the Gaussian structure learning algorithm [7]. Note that we present only some selected results from a broader simulation study. For each simulation, we generated a tree-structured Bayesian network (i.e., the six-node network in Fig. 1a), assigning random parameters to the nodes ($0 < \mu < 2\pi, 0 < \varepsilon < 1$) and enforcing some dependence between parent and child nodes ($0.5 < |\rho| < 1$). From each node, we simulated n wrapped Cauchy samples using the “Circular” R package from [21]. Once we had completed the data simulation, we tested the algorithm by constructing the network from the created dataset (i.e., Fig. 1b) which we compared with the original network. In order to measure the accuracy of the method, we counted the number of edges that are misplaced in the created network with respect to the arcs of the original graph, therefore it wouldn’t be necessary to direct the edges in the simulated network for measure accuracy. We simulated several networks changing the number of nodes from 3,6,10 (Fig. 3), 20 and 30 and simulating

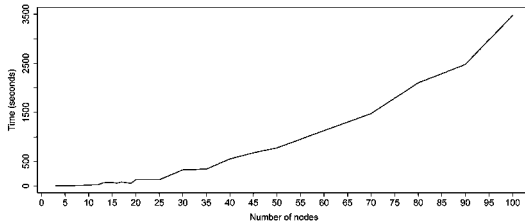


Fig. 2. Line chart that represents the increment of the computational time of our algorithm with the increment of the number of nodes in the network.

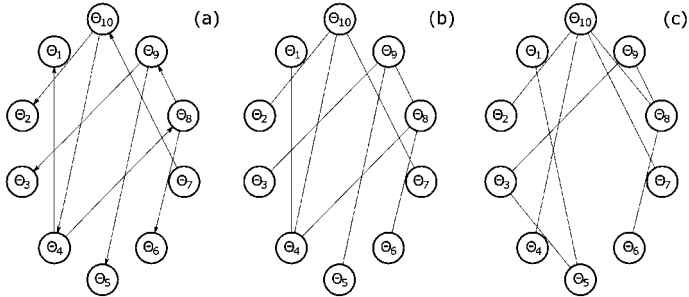


Fig. 3. Simulation output: wrapped Cauchy tree-structured network comparison in a 10-node network. (a) The original network. (b) Non-directed network, learned from data simulated from (a) using our algorithm. Accuracy is 9/9 (100%). (c) Non-directed network, learned from the same dataset using the Gaussian structure learning algorithm. Accuracy is 6/9 (67%).

100 observations per node for the dataset. The number of simulations differs for each network type due to the high computational cost of the biggest networks. For this reason, the maximum number of nodes used for simulation is 30. We have scaled our algorithm (Fig. 2) in order to show the computation complexity, as our algorithm has to perform $N * (N - 1)/2$ iterations.

The results (Table 1) show that of our method is highly accurate and the constructed graph is the same as the initial network in most cases, and outperforms the Gaussian method in terms of structure learning. Any cases where one or more than one edge is misplaced could probably be due to the randomization of the given parameters.

Table 1. Simulation results. Accuracy values are given by the mean of the non-misplaced edges in the total number of simulations.

			Wrapped Cauchy	Gaussian
Number of nodes	Simulations	Number of edges	Accuracy	Accuracy
3	1794	2	1.87	1.46
6	1764	5	4.37	2.61
10	1666	9	7.65	4.09
20	674	19	15.76	8.61
30	200	29	24.09	11.71

5 Real Data Example

Since our main application interest is in neuroscience, we applied our proposal to a real-world dataset of 3027 combinations of dendritic bifurcation angles composed by 288 3D pyramidal neurons from six different layers of the 14-day-old

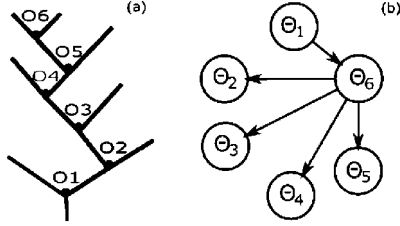


Fig. 4. (a) Angles of different branch orders measured between sibling segments in a dendritic arbor showing orders from 1 to 6. (b) Tree-structured Bayesian network for dendritic bifurcation angles from orders 1–6. Root node is Θ_1 .

(P14) rat hind limb somatosensory (S1HL) neocortex recently published in [13]. Dendritic bifurcation angles are an important part of the geometry of pyramidal cell basal arbors. The comprehension and capability of modelling these angles is crucial for advances in neuroscience in order to replicate brain structure and functioning. Our purpose is to model the conditional relationship between the bifurcation angles of the whole neuron dataset in a tree-structured Bayesian network for wrapped Cauchy distributions. Each angle is generated by two sibling segments originating from the bifurcation of basal dendritic trees. Using the same notation as in [13], we denote the first bifurcation that takes place in a dendritic arbor path that starts from the soma and ends at the angle as O1, the second bifurcation would be O2, etc. (Fig. 4a). Bifurcation angles from orders higher than six are not included in the model because they are relatively few in number.

We fitted a wrapped Cauchy distribution for every set of angles of the same bifurcation order (Table 2). The resulting tree-structured Bayesian network learned from the dataset is the six-node directed graph shown in Fig. 4b. Looking at Fig. 4b, it is evident that the Θ_6 node has the highest CMI values for every pair of nodes. We have selected Θ_1 as the root node in order to finish the structure of our network and transform the undirected graph into a directed graph.

Table 2. Characteristics of the 6 different branch orders shown in Fig. 4a. Circular mean (μ) is measure in radians.

Bifurcation order	Variable	Number of angles	μ (radians)	ε
O1	Θ_1	1607	1.02	0.90
O2	Θ_2	2072	0.9	0.91
O3	Θ_3	1773	0.82	0.92
O4	Θ_4	998	0.78	0.92
O5	Θ_5	382	0.77	0.92
O6	Θ_6	106	0.81	0.92

6 Conclusions and Future Work

The main objective of this paper is to introduce the first Bayesian network that deals with directional wrapped Cauchy variables. We introduced the methods for learning and representing the model. The proposed model applied to real and simulated data is capable of representing the relationship between directional data nodes that fit wrapped Cauchy distributions.

In future work we intend to explore different distributions in order to extend this model to other directional families. This task is tough because the known directional family distributions are not closed. Another interesting possibility is to develop a supervised classification algorithm using wrapped Cauchy variables. An analogue of the tree-augmented naive Bayes would be rather straightforward, since we already developed the inference process.

This preliminary model is confined to tree structures. We hope to extend this model to a more general Bayesian network case, capable of accounting for more than one parent per node.

Acknowledgments. This work has been partially supported by the Spanish Ministry of Economy and Competitiveness through the TIN2013-41592-P and Cajal Blue Brain (C080020-09), by the Regional Government of Madrid through the S2013/ICE-2845-CASI-CAM-CM project. I.L. is supported by the Spanish Ministry of Education, Culture and Sport Fellowship (FPU13/01941). The authors thankfully acknowledge the Cortical Circuits Laboratory (CSIC-UPM) for the neurons dataset.

References

1. Batschelet, E.: *Circular Statistics in Biology*. Academic Press, London (1981)
2. Berntsen, J., Espelid, T.O., Genz, A.: An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Trans. Math. Softw.* **17**(4), 437–451 (1991)
3. Boomsma, W., Kent, J.T., Mardia, K.V., Taylor, C.C., Hamelryck, T.: Graphical models and directional statistics capture protein structure. *Interdisc. Stat. Bioinform.* **25**, 91–94 (2006)
4. Bowman, K., Shenton, L.: Methods of moments. *Encycl. Stat. Sci.* **5**, 467–473 (1985)
5. Chow, C., Liu, C.: Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **14**(3), 462–467 (1968)
6. Fisher, N.I.: *Statistical Analysis of Circular Data*. Cambridge University, Cambridge (1995)
7. Geiger, D., Heckerman, D.: Learning gaussian networks. In: *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, pp. 235–243. Morgan Kaufmann Publishers Inc. (1994)
8. Jammalamadaka, S.R., Sengupta, A.: *Topics in Circular Statistics*. World Scientific, River Edge (2001)
9. Kato, S.: A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli* **15**(3), 898–921 (2009)
10. Kato, S., Pewsey, A.: A Möbius transformation-induced distribution on the torus. *Biometrika* **102**(2), 359–370 (2015)

11. Kent, J.T.: The Fisher-Bingham distribution on the sphere. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **44**(1), 71–80 (1982)
12. Koller, D., Friedman, N.: *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge (2009)
13. Leguey, I., Bielza, C., Larrañaga, P., Kastanauskaitė, A., Rojo, C., Benavides-Piccione, R., DeFelipe, J.: Dendritic branching angles of pyramidal cells across layers of the juvenile rat somatosensory cortex. *J. Comp. Neurol.* **524**(13), 2567–2576 (2016)
14. Lévy, P.: L’addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société Mathématique de France* **67**, 1–41 (1939)
15. Mardia, K.V.: Statistics of directional data. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **37**, 349–393 (1975)
16. Mardia, K.V.: Bayesian analysis for bivariate von Mises distributions. *J. Appl. Stat.* **37**(3), 515–528 (2010)
17. Mardia, K.V., Hughes, G., Taylor, C.C., Singh, H.: A multivariate von Mises distribution with applications to bioinformatics. *Can. J. Stat.* **36**(1), 99–109 (2008)
18. Mardia, K.V., Jupp, P.E.: *Directional Statistics*. Wiley, Hoboken (2009)
19. McCullagh, P.: Möbius transformation and Cauchy parameter estimation. *Ann. Stat.* **24**(2), 787–808 (1996)
20. von Mises, R.: Über die Ganzzahligkeit der Atomgewichte und verwandte Fragen. *Zeitschrift für Physik* **19**, 490–500 (1918)
21. R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2008). ISBN 3-900051-07-0, <http://www.R-project.org>
22. Razavian, N., Kamisetty, H., Langmead, C.J.: The von Mises graphical model: regularized structure and parameter learning. Technical report CMU-CS-11-108. Carnegie Mellon University, Department of Computer Science (2011)
23. Spirtes, P., Glymour, C.N., Scheines, R.: *Causation, Prediction, and Search*. MIT Press, Cambridge (2000)
24. Van Dooren, P., de Ridder, L.: An adaptive algorithm for numerical integration over an N -dimensional cube. *J. Comput. Appl. Math.* **2**(3), 207–217 (1976)
25. Wintner, A.: On the shape of the angular case of Cauchy’s distribution curves. *Ann. Math. Stat.* **18**(4), 589–593 (1947)