Directional Data: theoretical issues and environmental applications

Giovanna Jona Lasinio [1] & Gianluca Mastrantonio [2] & Antonello Maruotti [3]

[1] *Sapienza Università di Roma, P.le Aldo Moro 5, 00185 Rome, ITALY, giovanna.jonalasinio@uniroma1.it*
[2] *Roma Tre University, Via Silvio D'Amico 77, Rome, 00145, Italy , gianluca.mastrantonio@uniroma3.it*
[3] *LUMSA University, Borgo Sant'Angelo, 13, 00193 Roma, Italy, a.maruotti@lumsa.it*

Abstract. Directional data when lacking information on the orientation and initial direction should be treated and analyzed with caution as statistical inference may strongly depend on the chosen reference system. This is particularly relevant in environmental applications where finding patterns in data is required and this leads to relevant policy decisions. In this work we give the definition of two invariance properties and provide necessary and sufficient conditions to avoid inferential problems and misinterpretation of parameter estimates for any circular distributions.

Keywords. Circular data, wind directions, Invariance properties.

1 Introducing the problem and sketching the solution

There is increasing interest in analyzing directional data collected overspace, time and jointly over space and time. Examples arise, for instance, in oceanography (wave directions), meteorology (wind directions), biology (study of animal movement). Standard techniques cannot be used to analyze circular data, mainly due to the circular geometry of the sample space (for details, see Mastrantonio *et al.* (2015)). Many *ad-hoc* methods and statistical techniques have been developed to analyze and understand circular data (Mardia, 1972; Fisher, 1996; Mardia and Jupp, 1999; Jammalamadaka and SenGupta, 2001; Lee, 2010; Pewsey *et al.*, 2013), leading to important probability distribution theory and inferential results.

Probability distributions for circular data are often argued to have a general structure stemming from that of considering the unit circle as support and having a closed form density with relatively simple parameters estimation procedure. However, circular data have many specific features that should be taken into account in any analyses. Indeed, on this scale, there is no designed zero (i.e. initial direction), and neither end; moreover, in contrast to a linear scale, the designation of the natural orientation is arbitrary. Although having good fitting characteristics and tractable forms, the use of well-known circular distributions may lead to misleading inferential results as long as issues related to the initial direction and the orientation system are not accounted for.

The two properties The two key properties that a circular distribution must have are formalized as:

Definition 1 *We say that f_Θ is invariant under change of initial direction (ICID) if $\forall \xi \in \mathbb{D}$, $\forall \theta \in \mathbb{D}$ and $\forall \boldsymbol{\psi} \in \boldsymbol{\Psi}$, where $\boldsymbol{\psi}$ is the vector of parameters, there exists $\boldsymbol{\psi}^* \in \boldsymbol{\Psi}$ such that $f_\Theta(\theta|\boldsymbol{\psi}) = f_\Theta(\theta - \xi|\boldsymbol{\psi}^*)$.*

Definition 2 *We say that f_θ is invariant under changes of the reference system orientation (ICO) if $\forall \xi \in \mathbb{D}$, $\forall \theta \in \mathbb{D}$ and $\forall \boldsymbol{\psi} \in \boldsymbol{\Psi}$ there exists $\boldsymbol{\psi}^* \in \boldsymbol{\Psi}$ such that : $f_\Theta(\xi - \theta|\boldsymbol{\psi}) = f_\Theta(\xi + \theta|\boldsymbol{\psi}^*)$.*

To clarify the issue consider Figure 1 where in (a)-(c), a wrapped skew normal density (Pewsey, 2000) and in (d)-(f) a wrapped exponential are plotted with changing reference systems. The origin (initial direction) is set to zero radiant and the orientation is anticlockwise in (a) and (d). By changing the initial direction (Figure 1 (b)), or the system orientation (Figure 1 (c)), we obtain wrapped skew normal pdfs with shapes exactly as the one in Figure 1 (a). Then the wrapped skew normal distribution is an invariant distribution with respect to any change in the reference system, i.e. it is both ICI and ICO. Looking at Figure 1(d) to (f) we can see that the shape of the wrapped exponential changes substantially when the reference system changes, defining a not invariant distribution.

2 Data application

We show, with a simple data application, the misleading inference that can be obtained under different reference systems. We have data about the wind direction on January 2000 at the monitoring station of Capo Palinuro. The monitoring station of Capo Palinuro (WMO code 16310) is one of the coastal stations managed by the Meteorological Service of the Italian Air Force. The station is located on the rocky cape of Capo Palinuro, in the town of Centola in the province of Salerno, South Italy. Wind directions are monitored and routinely collected by several environmental agencies. Analyzed data come from reports prepared at the station and provided by the National Center of Aeronautical Meteorology and Climatology (C.N.M.C.A.), special office of the Meteorological Service of the Italian Air Force. The database includes date and time of registration, direction of the wind in degrees, with eight daily measurements (every three hours) in the month of January 2000, i.e. we have 240 observations. The measuring instrument, anemometer, is placed away from obstacles and at an height of 10 meters above ground. A relevant issue with this recordings is that the measurement instrument asses wind directions on a discrete scale dividing the circle into ten-degree intervals ($l = 36$).

The most common distribution for a discrete circular random variable Θ is the wrapped Poisson (WP), that, in this study, has density

$$\sum_{k=0}^{\infty} \frac{\lambda^{\theta \frac{36}{2\pi} + k36} e^{-\lambda}}{(\theta \frac{36}{2\pi} + k36)!}, \ \lambda \in \mathbb{R}$$

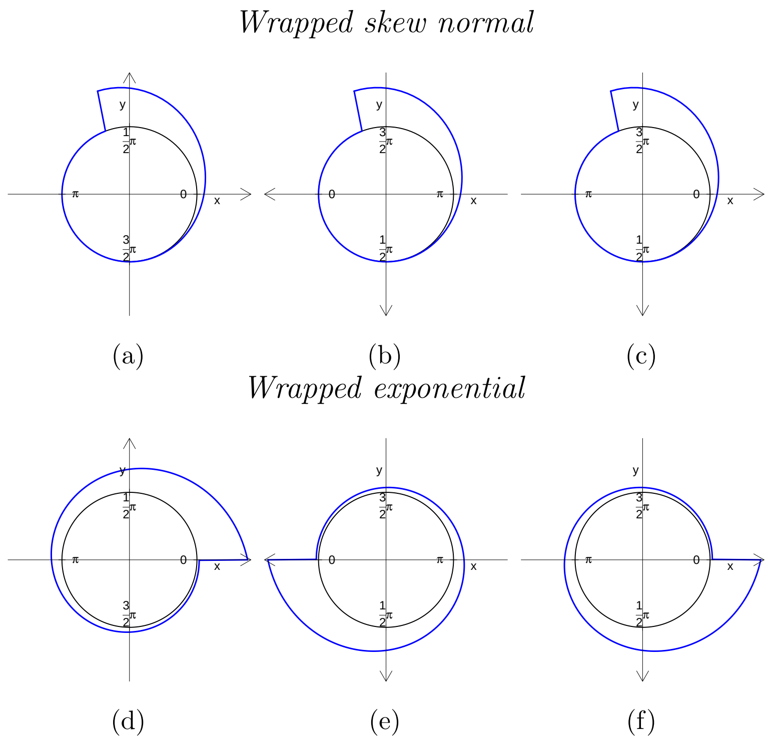


Figure 1: Probability density functions of a wrapped skew normal (a-c) and a wrapped exponentil (d-f) under different initial directions and orientations. The arrows indicate the axis orientation

It can be shown that the WP is a not invariant density. Then we proposed (see Mastrantonio *et al.*, 2015, for details) a generalization of the WP, called the *invariant wrapped Poisson* (IWP) which introduces two new parameters $\delta \in \{-1, 1\}$ and $\xi \in \{j2\pi/36\}_{j=0}^{35}$ that can take into account changes in the reference system, with density

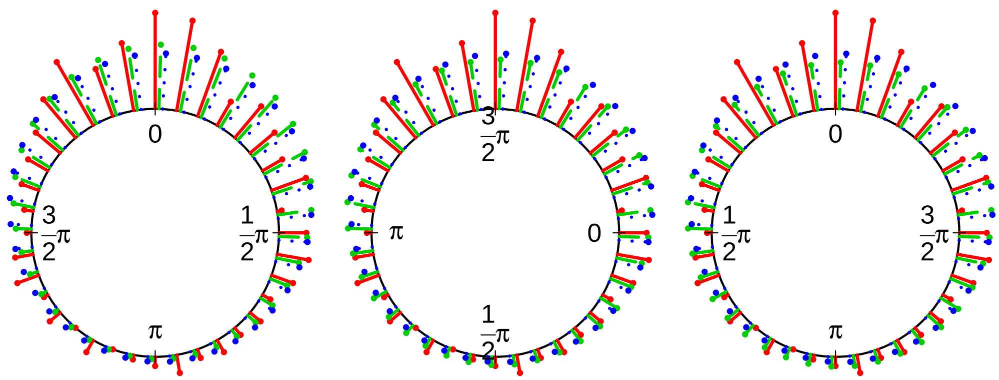
$$\sum_{k=0}^{\infty} \frac{\lambda^{(\delta\theta-\xi)\frac{36}{2\pi}+k36} e^{-\lambda}}{\left((\delta\theta-\xi)\frac{36}{2\pi}+k36\right)!}.$$

Note that with $\delta = 1$ and $\xi = 0$ the IWP reduces to the WP.

Results We use three reference systems: the first fixes North as zero direction and chooses a clockwise orientation (RS1), in the second reference system (RS2) we move the zero direction to East, while we obtain the third reference system (RS3) by changing the orientation of RS1. In each reference system, we find the MLEs of the invariant distribution parameters, (IWP), and the MLE of the corresponding non invariant distribution parameters, (i.e. the WP), together with their circular means and concentrations.
We indicate with $\hat{\lambda}_i$, $\hat{\mu}_{1,i}$ and $\hat{c}_{1,i}$ the MLEs of, respectively, the non invariant distribution

	WP			IWP				
	$\hat{\lambda}_i$	$\hat{\mu}_{1,i}$	$\hat{c}_{1,i}$	$\hat{\lambda}_i^*$	$\hat{\delta}_i^*$	$\hat{\xi}_i^*$	$\hat{\mu}_{1,i}^*$	$\hat{c}_{1,i}^*$
RS1	37.3434	0.2014	0.567	50.1254	1	4.0143	0.1520	0.4657
RS2	64.1283	4.8526	0.3775	50.1254	1	2.4435	4.8644	0.4657
RS3	71.3427	12.3885	0.3383	50.1254	-1	4.0143	6.1312	0.4657

Table 1: MLE of the WP and IWP parameters, circular mean and circular concentration



(a) RS1 (b) RS2 (c) RS3

Figure 2: Density estimate of observed data (solid line), wrapped Poisson (dashed line) and invariant wrapped Poisson (dotted line) density computed using the MLE of the parameters in the three reference systems

parameter, circular mean and concentration in the i^{th} reference system, while $\hat{\lambda}_i^*$, $\hat{\delta}_i^*$, $\hat{\xi}_i^*$, $\hat{\mu}_{1,i}^*$ and $\hat{c}_{1,i}^*$ are the parameters of the invariant distribution, circular mean and concentration in the i^{th} reference system. The MLEs are reported in Table 1 while the barplot of the observed data and the WP and IWP densities, in the three reference systems obtained with the MLEs, are shown in Figure 2.

3 General comments and concluding remarks

In this example, see Table 1, we can appreciate how the MLEs of the invariant density parameters are coherent in moving among the three reference systems. More precisely $\hat{\lambda}_1^* = \hat{\lambda}_2^* = \hat{\lambda}_3^*$, i.e. the MLE of λ^* is unaffected by changes of orientation or zero direction, we have $(\hat{\delta}_2^*, \hat{\xi}_2^*) = \left(\hat{\delta}_1^*, \hat{\xi}_1^* - \frac{\pi}{2}\right)$ since in the RS2 we change the zero direction to $\frac{\pi}{2}$, and in the RS3, where we modify the sense of orientation, we have $(\hat{\delta}_3^*, \hat{\xi}_3^*) = \left(-\hat{\delta}_1^*, \hat{\xi}_1^*\right)$.

Moreover the circular concentration remains the same in the three reference system, i.e. $\hat{c}_{1,1} = \hat{c}_{1,2} = \hat{c}_{1,3}$, while the circular mean changes according to the reference system, i.e. $\hat{\mu}_{1,2} = \hat{\mu}_{1,2} - \frac{\pi}{2}$ and $\hat{\mu}_{1,3} = -\hat{\mu}_{1,2}$. For the non invariant density, even the circular concentration changes with the reference system as well as the MLE of the parameter λ and the circular mean. The shape of the invariant density remains the same in the three reference systems, as shown in Figure 2.
We provided in Mastrantonio *et al.* (2015) a general framework to build invariant circular distributions according to the definitions above, here we show a simple example of the ability of this new methodological tool in an environmental example, however its usefulness is totally general and can be used whenever circular/directional data are involved.

References

[1] Fisher, N. I. (1996). *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge.

[2] Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific, Singapore.

[3] Lee, A. (2010). Circular Data. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(4), 477–486.

[4] Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press, London.

[5] Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. John Wiley and Sons, Chichester.

[6] Mastrantonio, G., Jona Lasinio, G., Maruotti, A., and Calise, G. (2015). On initial direction, orientation and discreteness in the analysis of circular variables. *ArXiv e-prints*.

[7] Pewsey, A. (2000). The wrapped skew-normal distribution on the circle. *Communications in Statistics - Theory and Methods*, **29**(11), 2459–2472.

[8] Pewsey, A., Neuhäuser, M., and Ruxton, G. D. (2013). *Circular Statistics in R*. Oxford University Press, Croydon.