Econometrics and Statistics 000 (2016) 1–14

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

A tractable, parsimonious and flexible model for cylindrical data, with applications

Toshihiro Abe[a], Christophe Ley[b,*]

[a] Department of Information Systems and Mathematical Sciences, Nanzan University, 18 Yamazato-cho, Showa-ku, Nagoya, 446-8673, Japan
[b] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281, S9, Campus Sterre, 9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 31 December 2015
Revised 6 March 2016
Accepted 26 April 2016
Available online xxx

Keywords:
Circular–linear data
Circular–linear regression
Distributions on the cylinder
Sine-skewed von Mises distribution
Weibull distribution

ABSTRACT

New cylindrical distributions are proposed by combining the sine-skewed von Mises distribution (circular part) with the Weibull distribution (linear part). This new model, the WeiSSVM, enjoys numerous advantages: simple normalizing constant and hence very tractable density, parameter-parsimony and interpretability, good circular–linear dependence structure, easy random number generation thanks to known marginal/conditional distributions, and flexibility illustrated via excellent fitting abilities. Inferential issues, such as independence testing, circular–linear respectively linear–circular regression, can easily be tackled with the new model, which is applied on two real data sets.

1. Introduction

Cylindrical data are observations that consist of a directional part (a set of angles), which is often of a circular nature (a single angle), and a linear part (mostly a positive real number). This explains the alternative terminology of directional–linear or circular–linear data. Such data occur frequently in natural sciences; typical examples are wind direction and another climatological variable such as wind speed or air temperature, the direction an animal moves and the distance moved, or wave direction and wave height. Recent studies of cylindrical data include the exploration of wind direction and SO_2 concentration (García-Portugués et al., 2013), the analysis of Japanese earthquakes (Wang et al., 2013), the link between wildfire orientation and burnt area (García-Portugués et al., 2014), and space-time modeling of sea currents in the Adriatic Sea (Lagona et al., 2015b; Wang et al., 2015).

A non-trivial yet fundamental problem is the joint modeling of the directional/circular and linear variables via the construction of cylindrical probability distributions. The best known examples stem from Mardia and Sutton (1978), conditioning from a trivariate normal distribution, and Johnson and Wehrly (1978), invoking maximum entropy principles. The latter also provide in their paper a general way, based on copulas, to construct circular–linear distributions with specified marginals.

What desirable properties should a “good” cylindrical distribution possess? It should be able to model diverse shapes, in other words present good fitting aptitudes, yet it should ideally remain of a tractable form (this is crucial for stochastic properties, estimation purposes, and circular–linear regression) and be parsimonious in terms of parameters at play. The

* Corresponding author.
E-mail addresses: abetosh@ss.nanzan-u.ac.jp (T. Abe), christophe.ley@ugent.be (C. Ley).

http://dx.doi.org/10.1016/j.ecosta.2016.04.001
2452-3062/

marginal and conditional distributions should optimally be well-known and flexible (e.g., there is no reason for the circular component to be always symmetric), whilst the dependence structure has to take care of a reasonable joint behavior. Indeed, numerous examples of cylindrical data require that the circular concentration tends to increase with the linear component, as identified in the seminal paper Fisher and Lee (1992).

All these conditions are well fulfilled by the new model we propose in the present paper. Its probability density function (pdf) is of the form

$$(\theta, x) \mapsto \frac{\alpha\beta^{\alpha}}{2\pi\cosh(\kappa)}(1+\lambda\sin(\theta-\mu))x^{\alpha-1}\exp[-(\beta x)^{\alpha}(1-\tanh(\kappa)\cos(\theta-\mu))], \qquad (1)$$

where $(x, \theta) \in [0, \infty) \times [-\pi, \pi)$, α, $\beta > 0$, $-\pi \leq \mu < \pi$, $\kappa \geq 0$ and $-1 \leq \lambda \leq 1$. The roles of the distinct parameters will be explained in Section 2, as well as the construction underpinning (1). Stochastic properties such as marginal and conditional distributions, random number generation, moment and correlation calculations are presented in Section 3. We will in particular stress the capacity of our new density to model cylindrical data with length-increasing circular concentration. Maximum likelihood estimation and the ensuing efficient likelihood ratio tests (including tests for circular–linear independence) are discussed in Section 4, as well as circular–linear and linear–circular regression. A Monte Carlo simulation study (Section 5) reveals a good behavior of maximum likelihood estimates. In order to give the reader a better idea of the strengths of our new model, we review in Section 6 the main competitor cylindrical distributions from the literature, and compare them to our model on basis of objective criteria. The excellent modeling capacities of our new model are illustrated by means of two real data sets in Section 7. We conclude the paper by some final comments in Section 8, where we refer to two very recent papers that have already used our model as important building block.

2. A new model for circular–linear data: the WeiSSVM

Johnson and Wehrly proposed in Johnson and Wehrly (1978) a very simple distribution able to fit cylindrical data where the circular concentration increases with the length of the linear part. Their density reads

$$(\theta, x) \mapsto \frac{\beta}{2\pi\cosh(\kappa)}\exp[-\beta x(1-\tanh(\kappa)\cos(\theta-\mu))], \qquad (2)$$

with $-\pi \leq \mu < \pi$, $\beta > 0$ and $\kappa \geq 0$. The linear conditional density is the (negative) exponential, while the circular conditional density given $X = x$ is of the form

$$\theta \mapsto \frac{1}{2\pi I_0(x\beta\tanh(\kappa))}\exp[\beta x\tanh(\kappa)\cos(\theta-\mu)], \qquad (3)$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero. The mapping (3) is the popular von Mises density with location μ and concentration $\beta x \tanh(\kappa)$, often considered as the circular analogue of the normal distribution. We attract the reader's attention to the fact that we have slightly reparameterized the original Johnson–Wehrly parameterization which would correspond to using β and $\kappa_1 = \beta\tanh(\kappa)$ instead of β and κ, and hence adding the condition that $\kappa_1 < \beta$ in view of $\beta/\cosh(\kappa) = (\beta^2 - \kappa_1^2)^{1/2}$. With our parameterization we avoid this condition, which is an advantage for numerical maximization methods.

A drawback of the Johnson–Wehrly model (2) is its lack of flexibility. Both its conditional and marginal circular densities are symmetric (see Section 3.2 for details), when $\kappa = 0$ the circular contribution in (2) boils down to the uniform law on $[-\pi, \pi)$, and the circular concentration varies linearly with x (see (3)). In order to overcome these limitations, we have applied two separate transformations to the Johnson–Wehrly density: a power transformation $x \mapsto x^{1/\alpha}$ for $\alpha > 0$ to the linear part, and a perturbation of the circular part via multiplication with $\theta \mapsto (1 + \lambda\sin(\theta - \mu))$ for $\lambda \in [-1, 1]$. The former is the classical way to turn an exponential distribution on $\mathbb{R}^+$ to the Weibull distribution with pdf $x \mapsto \alpha\beta x^{\alpha-1}\exp[-\beta x^{\alpha}]$, which is a very popular distribution to model diverse natural phenomena. The effect of the perturbation is known in circular statistics as "sine-skewing" a reflectively symmetric distribution, see Abe and Pewsey (2011). Whenever $\lambda \neq 0$, the resulting density becomes skewed, whereas symmetry is retrieved for $\lambda = 0$; moreover, the perturbation leaves the normalizing constant untouched. The combined effect of both transformations (plus the change from β to β^{α} mainly for aesthetic reasons) thus yields the pdf

$$\frac{\alpha\beta^{\alpha}}{2\pi\cosh(\kappa)}(1+\lambda\sin(\theta-\mu))x^{\alpha-1}\exp[-(\beta x)^{\alpha}(1-\tanh(\kappa)\cos(\theta-\mu))],$$

which we term *WeiSSVM* for the interplay between the linear Weibull part and the circular sine-skewed von Mises part. 2D contour plots of the density (1) are given in Fig. 1 and show the versatility of our new model.

Parameter interpretation becomes now clear: μ and λ respectively endorse the role of circular location and skewness parameters, while β and α are linear scale and shape parameters. The parameter κ plays, as in the original Johnson–Wehrly model, the role of circular concentration and circular–linear dependence parameter. Independence is attained when $\kappa = 0$, in which case the density (1) becomes the product of the linear Weibull and the circular cardioid distribution with location $\mu + \pi/2$ and concentration λ, see the first row of Fig. 1.

Please cite this article as: T. Abe, C. Ley, A tractable, parsimonious and flexible model for cylindrical data, with applications, Econometrics and Statistics (2016), http://dx.doi.org/10.1016/j.ecosta.2016.04.001