Computational Statistics and Data Analysis 71 (2014) 506–519

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

# A Bayesian model for longitudinal circular data based on the projected normal distribution

Gabriel Nuñez-Antonio [a,b,*], Eduardo Gutiérrez-Peña [c]

[a] Department of Statistics, Universidad Carlos III de Madrid, Spain
[b] Department of Mathematics, UAM-I, Mexico
[c] Department of Probability and Statistics, IIMAS-UNAM, Mexico

ARTICLE INFO

*Article history:*
Received 10 October 2011
Received in revised form 25 July 2012
Accepted 26 July 2012
Available online 3 August 2012

*Keywords:*
Circular data
Gibbs sampler
Latent variables
Longitudinal data
Mixed-effects linear models
Projected normal distribution

ABSTRACT

The analysis of short longitudinal series of circular data may be problematic and to some extent has not been fully developed. A Bayesian analysis of a new model for such data is presented. The model is based on a radial projection onto the circle of a particular bivariate normal distribution. Inference about the parameters of the model is based on samples from the corresponding joint posterior density, which are obtained using a Metropolis-within-Gibbs scheme after the introduction of suitable latent variables. The procedure is illustrated using both simulated data sets and a real data set previously analyzed in the literature.


## 1. Introduction

Several approaches have been proposed to analyze longitudinal data. See, for example, Diggle et al. (2002), Fitzmaurice et al. (2004), Gelman and Hill (2007), and Hedeker and Gibbons (2006). All these books discuss models for longitudinal 'scalar' (i.e. linear) responses. In contrast, methodological proposals to describe relationships within repeated measurements of directional data are rather limited. This may be due to the difficulties in working with probability distributions commonly associated with directional data and to the intrinsic dependency inherent to longitudinal structures.

Circular data are a particular case of directional data. Specifically, circular data represent directions in two dimensions. For a survey of this and related topics, we refer the reader to Fisher (1993), Fisher et al. (1987), Jammalamadaka and SenGupta (2001) and Mardia and Jupp (2000). More recent contributions include Abe and Pewsey (2011), Oliveira et al. (2012), Pewsey (2008) and Qin et al. (2011). See also Arnold and SenGupta (2006) for an overview of applications of circular data analysis in ecological and environmental sciences. In many of these applications, the observations on the variable of interest are longitudinal in nature. For example, in studies concerning the orientation mechanism of birds, it is relevant to analyze the angular differences between a bird's position at consecutive times after release (Artes and Jørgensen, 2000; Artes et al., 2000). In motion studies of small animals, the aim is often to describe the effect of covariates on the directional behavior of those animals (D'Elia, 2001; D'Elia et al., 2001). On the other hand, in the analysis of cell-cycle gene expression data, a problem of interest is to estimate the phase angles using information regarding the order among them (Rueda et al., 2009).

* Correspondence to: Departamento de Estadística, Universidad Carlos III de Madrid, C/Madrid 126, 28903, Getafe (Madrid), Spain. Tel.: +34 91 624 5713; fax: +34 91 624 9849.
*E-mail address:* gab.nunezantonio@gmail.com (G. Nuñez-Antonio).

0167-9473/$ – see front matter 
doi:10.1016/j.csda.2012.07.025

All of these situations illustrate the need for models that allow us to analyze longitudinal structures where the response variable is circular.

As pointed out above, from a methodological point of view there does not seem to be a general framework for the analysis of longitudinal directional data. Longitudinal data where the response variable is circular have been analyzed using quasi-likelihood methods, such as the generalized estimating equations (GEE) originally proposed by Liang and Zeger (1986) to analyze linear data. Specifically, Artes et al. (2000) derive estimating equations for the parameters of a family of circular distributions and obtain asymptotic inference for the parameters of a mixed-effects model. In turn, Artes and Jørgensen (2000) extend GEE methods to deal with Jørgensen's dispersion models (Jørgensen, 1997a,b) and employ this approach to model longitudinal circular data. They also present a simulation study for a simple model which only involves the mean direction and a single covariate. They note that in some situations their proposal may have troubles with convergence, and point out that their method requires either high correlations between the longitudinal observations or large samples in order to achieve satisfactory performance. On the other hand, D'Elia et al. (2001) propose a generalized linear model to study the directional behavior of sandhoppers under natural conditions in repeated trials. In the same vein, D'Elia (2001) assumes a variance components model to describe the orientation mechanism and uses a simulated maximum likelihood approach. She points out that the use of this approach may raise several problems. Recently, Song (2007) has used a generalized linear model approach where the random component belongs to the family of dispersion models. He suggests using penalized pseudo-likelihood and restricted maximum likelihood estimation to bypass the analytical difficulties arising from the nonlinearity of the corresponding score functions. Nevertheless, in some cases it is not possible to make inferences about all the parameters involved in the proposed models.

We feel that most of the procedures currently available for analyzing longitudinal data with circular responses suffer from certain flaws that render some of the required inferences unfeasible in general settings. These limitations include troubles for fitting, model comparison and prediction, as well as convergence problems of some of the iterative methods employed. Song (2007) argues that different approaches may be required to analyze series of repeated measurements, depending on the length of the series. In the case of (several) short time series, modeling typically focuses on the relationship between the response variable and the corresponding covariates. In such situations the correlations are treated as nuisance parameters, as opposed to the case of a long time series where the correlations are usually modeled explicitly via a stochastic process.

In this paper, we introduce a new model to describe short series of longitudinal data where the response variable is circular. The model considers linear covariates and is based on a version of the projected bivariate normal distribution. In our proposal, each of the components of the model is specified by a mixed-effects linear model. In addition, we present a Bayesian analysis that allows us to make inference about any of the parameters of interest.

The paper is organized as follows. In the next section, we introduce the projected circular longitudinal model (henceforth called the *PCL model*) and describe some of its properties, including the longitudinal structures that can be obtained from it. In Section 3, we discuss a Bayesian analysis of the model and derive all the full conditionals needed for a Gibbs sampler. In Section 4, we present some illustrative examples. Finally, Section 5 contains some concluding remarks.

## 2. The PCL model

### 2.1. Description of the model

We start this section by briefly reviewing the projected normal distribution. For further details, we refer the reader to Mardia and Jupp (2000), Nuñez-Antonio and Gutiérrez-Peña (2005), Presnell et al. (1998), Presnell and Rumcheva (2008), and the references therein.

There are several ways of generating probability distributions for circular data. One relatively straightforward way is to radially project on the unit circle probability distributions originally defined on the plane. In the general case, let $\boldsymbol{Y}$ be a $q$-dimensional random vector such that $\Pr(\boldsymbol{Y} = \boldsymbol{0}) = 0$. Then $\boldsymbol{U} = \|\boldsymbol{Y}\|^{-1}\boldsymbol{Y}$ is a random point on the $q$-dimensional unit sphere. Its *mean direction* is the unit vector $\boldsymbol{\eta} = E(\boldsymbol{U})/\rho$, where $\rho = \|E(\boldsymbol{U})\|$, $0 \le \rho \le 1$; here $E(\cdot)$ represents the usual expectation for random vectors, and $\|\cdot\|$ denotes the usual Euclidean norm. The parameter $\rho$ is called the *mean resultant length* and can be regarded as a measure of concentration.

An important instance of this class of distributions is that in which $\boldsymbol{Y}$ has a $q$-variate normal distribution, $N_q(\cdot|\boldsymbol{\mu}, \boldsymbol{\Lambda})$, with mean vector $\boldsymbol{\mu} = E(\boldsymbol{Y})$ and precision matrix $\boldsymbol{\Lambda} = \mathrm{Var}(\boldsymbol{Y})^{-1}$. In this case $\boldsymbol{U}$ is said to have a $q$-dimensional *projected normal distribution*, here denoted by $PN(\cdot|\boldsymbol{\mu}, \boldsymbol{\Lambda})$. In the circular case, $q = 2$, $\boldsymbol{U}$ is a two-dimensional unit vector and so it can be alternatively specified by means of a single angle $\Theta$, say. A version of the projected normal linear model for the circular case has been analyzed by Presnell et al. (1998) using a maximum likelihood approach. Nuñez-Antonio et al. (2011) present and discuss a Bayesian analysis of the same model. See also Nuñez-Antonio and Gutiérrez-Peña (2005).

The aim of this work is to introduce a model to describe short series of longitudinal data, where the response is a circular variable $\Theta$, in terms of one or more explanatory variables or covariates $\boldsymbol{x} = (x_1, \ldots, x_m)^t$. Even though the results presented here can be extended to other cases, only linear covariates will be considered.

Suppose that, on each of a number of occasions $j$ ($j = 1, \ldots, n_i$), measurements are taken on the $i$th individual in the study ($i = 1, \ldots, N$) and arranged in an $n_i \times 1$ vector of responses $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{in_i})^t$. Thus, we have a design with $N$ individuals and $n_i$ angular observations, $\theta_{ij}$, on each individual.