Journal of the Korean Data & Information Science Society
2011, **22**(6), 1217–1222

한국데이터정보과학회지

Exponential family of circular distributions

Sungsu Kim[1]

[1]Department of Statistics, Kyungpook National University

Received 9 September 2011, revised 30 September 2011, accepted 13 October 2011

Abstract

In this paper, we show that any circular density can be closely approximated by an exponential family of distributions. Therefore we propose an exponential family of distributions as a new family of circular distributions, which is absolutely suitable to model any shape of circular distributions. In this family of circular distributions, the trigonometric moments are found to be the uniformly minimum variance unbiased estimators (UMVUEs) of the parameters of distribution. Simulation result and goodness of fit test using an asymmetric real data set show usefulness of the novel circular distribution.

Keywords: Approximation, circular distribution, trigonometric polynomial, uniformly minimum variance unbiased estimator.

1. Introduction

Circular random variables are found in various areas of research such as biology, medicine, just to name a few. Because of the periodic nature of a circular variable, it is necessary to use a circular distribution to model a circular variable. Up to date, there are so many circular distributions available in literatures and books. Some of them are flexible enough to model asymmetric or multimodal circular distributions. For various types of circular distribution, including von Mises (VM) or circular normal distribution, the readers can refer to Jammalamadaka and SenGupta (2001).

The circular normal distribution, which is symmetric, has been mainly used to model a circular random variable. However, circular distributions are rarely symmetric, i.e. they are usually asymmetric and even multi-modal. Therefore, the VM distribution is not suitable to model such a data set. In fact, this is also the case in linear statistical analysis (Arnold and Beaver, 2000; Azzalini, 1985) that the normal distribution is often not suitable. One way to model an asymmetric and multimodal distribution is using a mixture of von Mises distributions (Batschelet, 1981). Another model suitable for an aymmetric and/or multimodal circular distribution is based on nonnegative trigonometric sums (Fernandez-Duran, 2004). Other existing asymmetric and/or bimodal circular distributions are appeared in Jammadamalaka and Kozubowski (2004), Getto and Jammalamadaka (2007), and Umbach and Jammalamadaka (2009).

[1] Assistant professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dr.sungsu@gmail.com

In this paper, we present a family of circular distributions that is an exponential family of distributions with a trigonometric polynomial inside the exponent. The novel family of distributions is useful to model any shape of circular distribution, and possess a particular advantage of having sample trigonometric moments as the UMVUEs of the parameters.

2. Methodology

2.1. Approximation technique

Proposition 2.1 Every circular density can be closely approximated in the density form of a full exponential family of distributions, using the trigonometric polynomial of order s, where s=1, ...

Proof: [Sketch of proof] Let $f(\theta)$ denote a density of circular distribution, and the Fourier coefficients of $\log(f(\theta))$ are given as

$$c_i = \int_0^{2\pi} \cos(i\theta)\log(f(\theta))d\theta, \quad i = 1, \ldots, \qquad s_j = \int_0^{2\pi} \sin(j\theta)\log(f(\theta))d\theta, \quad j = 1, \ldots.$$

Then, we approximate the logarithm of the density using the trigonometric polynomial of order 3, i.e.

$$f(\theta) = \exp(\log(f(\theta))) \cong \exp(c_1 \cos\theta + s_1 \sin\theta + c_2 \cos 2\theta). \tag{2.1}$$

In (2.1), one may replace $\cos 2\theta$ with $\sin 2\theta$. Then, the density of approximated 3-parameter family of circular distributions is given by

$$f(\theta) \cong \frac{\exp(c_1 \cos\theta + s_1 \sin\theta + c_2 \cos 2\theta)}{\int_0^{2\pi} \exp(c_1 \cos\theta + s_1 \sin\theta + c_2 \cos 2\theta)d\theta}, \tag{2.2}$$

where the integral in the denominator is introduced as the normalizing constant. The density of approximated 2-parameter family of circular distributions is given by those members of (2.2) for which $c_2 = 0$, which becomes the density of a von Mises distribution. The 4-parameter approximated circular density is, after adding another trigonometric term inside the exponent of (2.2), given by

$$f(\theta) \cong \frac{\exp(c_1 \cos\theta + s_1 \sin\theta + c_2 \cos 2\theta + s_2 \sin 2\theta)}{\int_0^{2\pi} \exp(c_1 \cos\theta + s_1 \sin\theta + c_2 \cos 2\theta + s_2 \sin 2\theta)d\theta}.$$

In this manner, a density of $s(= k + m)$-parameters family of circular distributions can be built as shown below, by adding trigonometric terms successively inside the exponent of (2.3),

$$f(\theta) \cong \frac{\exp(\sum_{i=1}^{k} c_i \cos i\theta + \sum_{j=1}^{m} s_j \sin j\theta)}{\int_0^{2\pi} \exp(\sum_{i=1}^{k} c_i \cos i\theta + \sum_{j=1}^{m} s_j \sin j\theta)d\theta}. \tag{2.3}$$

□

Proposition 2.2 The sample trigonometric moments are given as the complete sufficient statistics for the parameters in (2.3), therefore they are the uniformly minimum variance unbiased estimators of the corresponding parameters.

Proof: Since (2.3) belongs to a full exponential family of distributions, the sample trigonometric moments are given as the complete sufficient statistics for the parameters, c_i's and s_j's (Casella and Berger, 2001), where $i = 1, \ldots, k$ and $j = 1, \ldots, m$. □

2.2. Error of approximation

Suppose we approximate a circular density with $k + m$-parameters family of distributions as shown in (2.3). The approximation error of $\log f(\theta)$ is given by the sum of all the left out trigonometric terms in the fourier series expansion, i.e.

$$\sum_{i=k+1}^{\infty} c_i \cos i\theta + \sum_{j=m+1}^{\infty} s_j \sin j\theta.$$

Then the root mean square error, squaring $\log f(\theta)$ and integrating over [0,2π), is given by

$$\text{rms} = \sqrt{\sum_{i=k+1,j=m+1}^{\infty} (c_i^2 + s_j^2)},$$

according to the Parseval's theorem (Rudin, 1976). The root mean square error for (2.3) is given by exp(rms), which is smaller than rms since $f(\theta)$ ranges from 0 to 1.

The convergence rate of the trigonometric polynomial approximation partially depends on the smoothness and the dimension of the original density. The more smoothness and the lower in dimension the original density has, the faster the rate of convergence is (Rudin, 1976), where a circular density is a smooth function of a circular variable and the case of circular distributions in this paper is univariate. It is well known that trigonometric polynomial approximations of smooth density curves are very close even with the first 2 or 3 terms.

3. Goodness of fit

In this section the proposed family of circular distributions is fitted to an asymmetric data set. The data set refers to the directions chosen by 100 ants in response to an evenly illuminated black target and they are randomly selected values by Fisher (1993, p.243) taken from Jander (1957). Figure 3.1 represents the raw density plot.

From this it is clear that the data set is left-skewed about π, the direction in which a black target has been placed. The estimation of the parameters was made by using the maximum likelihood method. The maximum likelihood parameter estimates were obtained by using the optimization routine function called 'nlm' in R.

This data set has been analyzed by Fisher (1993) with the aim of fitting a von Mises distribution using the maximum likelihood estimation method. After checking Q-Q plot and applying a formal test of goodness of fit, Fisher concludes that the von Mises distribution is not a suitable model for this data set. To check the goodness of fit of the new model, we use the Watson's test (Watson, 1961). The Watson's test tests $H_0 : F(\alpha) = F_0(\alpha)$ against $H_1 : F(\alpha) \neq F_0(\alpha)$, where $\alpha_1, \ldots, \alpha_n$ constitutes a random sample from a continuous distribution and $F_0(\cdot)$ is a specified distribution function. Since the observed test statistic value has the value of 0.001, using its asymptotic distribution found in Jammadamalaka and

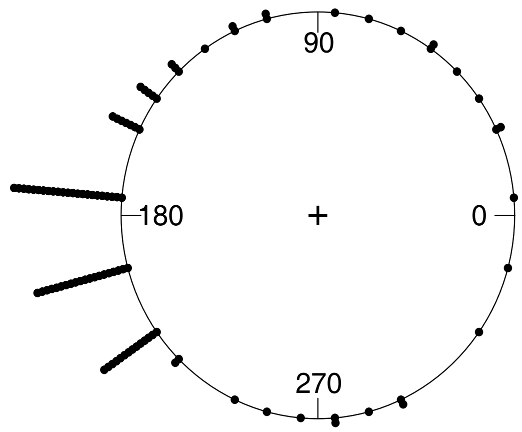

Figure 3.1 Row plot of response directions of 100 ants

SenGupta (2001), we conclude that the fit is good for the new model. Critical values for the Watson's statistic are available in Lockhart and Stephens (1985).

4. Simulation

The goodness of fit for approximating a (5 parameter) truncated bivariate circular distribution given in (4.1) is illustrated in Figure 4.1.

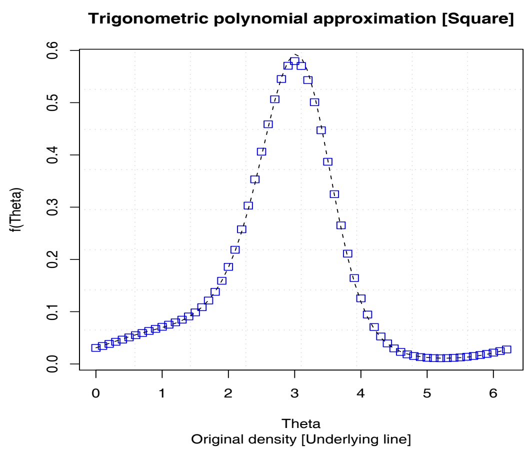

Figure 4.1 Trigonometric polynomial approximation using the first five terms

$$f(\theta)=\frac{\int_{1.5}^{2\pi}\exp(\cos\theta+\cos\phi+5\cos\theta\cos\phi+1.2\sin\theta\sin\phi)d\phi}{\int_0^{2\pi}\int_0^{2\pi}\exp(\cos\theta+\cos\phi+5\cos\theta\cos\phi+1.2\sin\theta\sin\phi)d\theta d\phi}. \tag{4.1}$$

Since the distribution whose density is given in (4.1) is not in a form of a familiar univariate circular distribution, the acceptance- rejection sampling method is employed to generate values of Θ, using R. The envelope function used is the product of M and the uniform density, where M denotes the maximum of (4.1). A 3-parameter approximated density of (4.1), which is overlayed in Figure 4.1 using squares, is shown below:

$$f(\theta)\cong\frac{\exp(1.1\cos\theta-2.6\sin\theta+3.1\cos 2\theta)}{\int_0^{2\pi}\exp(1.1\cos\theta-2.6\sin\theta+3.1\cos 2\theta)d\theta}.$$

5. Discussion

It is known that the usual goodness of fit tests based on the χ^2 statistic are not immediately applicable to circular data since it depends on how the cells are chosen, which in turn will depend on the choice of origin (Jammalamadaka and SenGupta, 2001). Invariant versions of such chi-square tests are considered in Ajne (1968) and Rao (1972). Similarly the class of tests based on the empirical distributions, like the Kolmogorov-Smirnov or the Cramer-von Mises tests, are also not directly applicable to circular data since their values again depend on the choice of origin. In this paper, we used the invariant version of such tests which are due to Kuiper (1960) and Watson (1961). The nearly equal performance of Kuiper's, Watson's and Ajne/s tests was noted by Stephens (1969) by a simulation.

Alternatively, those tests based on the gaps between successive points, called Spacing tests, can be directly applied to circular data. Indeed, spacings form the maximal invariant statistic under changes in origin so that every rotationally invariant statistic that is useful for the circular context, can be expressed in terms of spacings (Jammalamadaka and Sen-Gupta, 2001). We like to refer our readers to Jammalamadaka (1984) for a survey article on nonparametric methods for directional data.

In many practical situations, well-known circular models like the von Mises or wrapped stable densities may not provide an adequate description of the data. In this paper, a method of approximating a circular density is introduced, which leads to a family of circular distributions that are robust for a large class of possible models. Since any circular density can be approximated using an exponential family of distributions with a trigonometric polynomial inside the exponent, we propose an exponential family of distributions with a trigonometric polynomial inside the exponent as a new family of circular distributions.

Approximation technique was demonstrated using a simulation and an asymmetric real data set. Theoretical result of error of approximation was provided, however, the numerical result was not suitable to provide since it involves calculation of the sum of infinitely many trigonometric terms of the sine and the cosine. The plot of approximated density and the original density shows that they are almost the same density. The Watson's goodness of fit test using Jander's ants' data (Jander, 1957) shows that the fit is excellent, while it was shown in Fisher (1993) that a von Mises distribution does not provide a good fit for the data set. The novel family of circular distributions has the UMVUEs of the parameters as the sample trigonometric moments, which is a particularly attracting aspect of the family of distributions.

References

Ajne, B. (1968). A simple test for uniformity for a circular distribution. *Biometrika,* **55**, 343-354.

Arnold, B. and Beaver, R. (2000). Hidden truncation models. *Sankhya,* **62**, 23-35.

Azzalini, A. (2006). A class of distributions which includes the normal ones. *Scandinavian Journal of Statististics,* **12**, 171-178.

Batschelet, E. (1981). *Circular statistics in Biology,* Academic Press, Inc., London-New York.

Casella, G. and Berger, R. (2001). *Statistical inference,* Duxbury Advanced Series, Pacific Grove, CA.

Fernandez-Duran, J. J. (2004) Circular distributions based on nonnegative trigonometric sums. *Biometrics,* **60**, 499-503.

Fisher, N. I. (1993). *Statistical analysis of circular data,* Cambridge University Press, Cambridge.

Gatto, R. and Jammalamadaka, S. R. (2007). The generalized von Mises distribution. *Statistical Methodology,* **4**, 341-353.

Jammalamadaka, S. R. (1984). New families of wrapped distributions for modeling skew circular data. *Communication in Statistics-Theory and Methods,* **33**, 2059-2074.

Jammalamadaka, S. R. and Kozubowski, T. (2004). New families of wrapped distributions for modeling skew circular data. *Communication in Statistics-Theory and Methods,* **33**, 2059-2074.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in circular statistics,* Scientific Publishing CO., New York.

Jander, R. (1957). Die optische richtungsorientierung der roten waldermeise (Formica Rufa L.). *Zeitschrift fur Vergleichende Physiologie,* **40**, 162-238.

Kuiper, N. H. (1960). Tests concerning random points on a circle. *Nederlandse Akademie van Wetenschappen Proceedings,* **63**, 38-47.

Lockhart, R. A. and Stephens, M. A. (1985). Tests of fit for the von Mises distribution. *Biometrika,* **72**, 647-652.

Rao, J. S. (1972). Some variants of chi-square for testing uniformity on the circle. *Zeitschrift fur Wahrscheinlichkeitsthorie and Vervandt Gebiete,* **22**, 33-44.

Rudin, (1976). *Principal of mathematical analysis*, McGraw-Hill Science/Engineering/Math, New York.

Stephens, M. A. (1969). A goodness of fit statistic for the circle, with some comparison. *Biometrika,* **56**, 161-168.

Umbach, D. and Jammalamadaka, S. (2009). Building asymmetry into circular distribution. *Statistics and Probability Letters,* **79**, 659-663.

Watson, G. S. (1961). Goodness of fit tests on the circle. *Biometrika*, **48**, 109-114.