A unifying generator for cardioid modelling in circular statistics

Delene Martine van Wyk (16099754)

Mini-dissertation submitted in partial fulfillment of the degree

MSc Advanced Data Analytics

Supervisor: Prof. A. Bekker

Co-supervisors: Prof. M. Arashi and Prof. J.T. Ferreira

Department of Statistics, University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Abstract

One of the most important distributions in the circular world is the cardioid distribution. This distribution, however, suffers limitations for multimodal or asymmetric occurrences in circular data. This study addresses this shortfall by developing a flexible cardioid generated class that can capture inherent skewness and multimodality that data might present. This class is obtained by expressing the density in a general form by making use of a measurable function termed as generator. Some of the existing extensions of the cardioid distribution is shown to be special cases of this class of cardioid-generator distribution, and flexible behaviour of the proposed class in terms of skewness and multimodality is graphically demonstrated. Special cases of new and existing models of the newly proposed class are compared by making use of simulation studies and data sets. The newly proposed cardioid-t distribution proved a worthy addition to the class of symmetric models, outperforming well known circular distributions in two of the real data applications presented here. Introducing a more flexible cardioid generated class opens new insights for the cardioid model and equips the practitioner with a modified set of tools towards uncovering the intricate world of directional statistics.

Acknowledgements

This work would not have been possible without the help of the National Research Foundation and South African Statistical Association, whose financial contribution towards my studies have aided me greatly. I would furthermore not have been able to complete this work without the insightful and meticulous guidance of my research supervisors, Prof. A. Bekker, Prof. M. Arashi and Dr. J.T. Ferreira. As teachers and mentors they have taught me more than I can ever give them credit for here. I am also grateful to all those who have helped me throughout this journey and with whom I have had the pleasure to work with - in particular Prof. N.N. Rad for her assistance and Mr. G. de Ridder for general reinforcement and support. I am thankful to the people who have shared my goals and equipped me with the tools to achieve this final product.

Contents

List of Figures

List of Tables

Nomenclature

CU	Circular uniform distribution	EC	Exponentiated cardioid distribution
AEC	Asymmetrical extended cardiod distribution	vM	von Mises distribution
GvM	Generalized von Mises distribution	WC	Wrapped Cauchy distribution
cdf	Cumulative distribution function	GT	Generalized t-distribution

Chapter 1

Background and Motivation

1.1 Introduction

Circular statistics contribute greatly in uncovering and describing meaningful results in many topics of research where direction is involved. An example of circular statistics used in biology can be found in the study of Tremblay and Castro [29], where the growth tendencies of an endangered epiphytic orchid in Puerto Rico was investigated via circular statistics. In meteorology, Jones, Daniels and Bach [18] studied the diffusion of downwind air pollutants from a certain source with the desire to expand the resolution of wind directions from the widely used 16 main points of the compass. Circular statistics are also applied in the geological world. Beck [6] studied the value of inclination and paleolatitude. Circular distributions are also applicable in many medical branches, where Kim, SenGupta and Arnold [20] used a multivariate circular distribution in approaching protein structure prediction.

Various scientific areas use measurements in the form of directions noted in two or three dimensions [16]. Sets of directional measurements are known as directional data. If the data is two-dimensional, it is referred to as circular data which can be displayed as angles measured with respect to a suitable zero-direction starting point and rotation in either a clockwise or anti-clockwise fashion [16]. Directions are represented as points on the circumference of a unit circle centered at the origin [16]. For three dimensions, the data is referred to as spherical data which can be represented as points on the surface of a unit sphere or as unit vectors in three dimensions [16]. Directional measurements can be generalized to any higher number of dimensions to be visualized on the surface of a p-dimensional hypersphere. Developing suitable data analysis strategies when the observed variables are directions, presents new challenges as it intuitively differs from the formalizations of multivariate data [22].

Based on the fact that most parametric inferential approaches for directional data are obtained from only a handful of proposed models, there are still ongoing discussions surrounding the development of approaches that

will increase the robustness of most existing directional models [22]. Methods for deciphering directional data have been developed towards characterizing population distributions from which circular data may arise and most of the tools that are generally implemented, rely heavily on the von Mises and Fisher distributions [22]. The effect is that many of the existing methods appropriate for directional statistics are parametric in nature and reliant on restricted distributional assumptions, such as having a unimodal and symmetric spread. It has been suggested by Jammalamadaka [15] that a general non-parametric method that is based on randomization might be useful in the analysis of directional data over the broader spectrum of statistical challenges. In an attempt to address multimodality and skewness, a number of distributions have been proposed during the last few decades in which directional statistics have received a renewed interest. Some of these models, specifically related to the cardioid distribution, include the exponentiated cardioid [25], Papakonstantinou's extension of cardioid distributions [1] and the asymmetrical cardioid distribution, that has been developed through transformation of argument [2]. These distributions were all constructed with an increased ability to address skewness and multimodality where sharper or more flat-topped modes are concerned [2].

More robust procedures are not as easily affected by anomalies in data and can be generalized to a broader range of data analysis situations. In light of the general lack of robustness in classical directional methods, advances are continuously being made towards finding distributions that are more flexible. The cardioid distribution is a relevant and tested model for circular data, but lacks the ability to appropriately address multimodality and skewness [25]. In an attempt to address these shortcomings, some generator functions have been proposed for representing circular distributions in a different form [25]. A simple generator for representing circular and directional data is the perturbation procedure introduced by Jeffreys [17]. This approach is based on combining a circular density and some chosen function by multiplication so that the result is again a circular density [25]. Moreover, generators can be defined by transforming the argument of an existing density to a function of the specific argument, as seen in the development of the asymmetrical cardioid and extensions of the Jones-Pewsey family of distributions proposed by Abe, Pewsey and Shimizu [2]. Siew, Kato and Shimizu [28] also suggested a generalized t-distribution on the circle, based on a combination of the Jones-Pewsey and the generalized von Mises distributions.

1.1.1 Aims and Objectives

A cardioid-generator class is proposed with the aim of addressing multimodality and skewness, as there is an increasing need for circular distributions to capture these features in a broad range of scientific fields.

The objectives of this study are to:

1. Obtain a thorough understanding of the cardioid distribution and its existing generalizations by investigating existing methods.

2. Propose a new generator class of distributions and consider some special cases thereof (in particular the

proposed generalized t-distribution and the cardioid-von Mises distribution).

3. Investigate the structural properties of special cases of the generator class.

4. Graphically demonstrate the behaviour of the proposed class and to investigate the effect of a change in parameters for the different members.

5. Investigate the maximum likelihood estimation of the generalized t- and cardioid-von Mises distributions through nonparametric methods when closed forms do not exist.

6. Conduct a simulation study of the generalized t- and cardioid-von Mises distributions as special cases of the cardioid-generator class of distributions.

7. Determine and evaluate the performance of some cardioid-generator distributions (specifically the von Mises, cardioid, cardioid-t, wrapped Cauchy and cardioid-von Mises) by making use of different information criteria (such as the Akaike information criteria (AIC) and Bayesion information criteria (BIC)) by means of application to real data sets.

1.1.2 Outline of the Study

In the Chapter 1 of this study, the cardioid distribution and some generalizations thereof will be investigated. An overview is essential in order to identify shortcomings and gain insight in the existing techniques. Substantial literature reviews accompany this process to ensure novelty and intellectual property.

Mathematical and statistical methods will be applied towards determining a cardioid-generator class of distributions in Chapter 2. These methods will include, but are not limited to, embedding the kernel of the cardioid density within a measurable function, termed as a generator. Through Maclaurin series expansion, the normalizing constant can be obtained for specific choices of the generator so that existing as well as new special cases can be obtained [7]. This will then be followed by determining the characteristics of the class and visualisation of these characteristics.

Simulation studies are carried outin Chapter 3 to evaluate the performance of the proposed cardioid-generator distributions. A synopsis of the estimation procedure utilized in the study, namely maximum likelihood estimation, is presented. Finally, in Chapter 4, real data applications are conducted towards validating the practical use of the proposed generator class and special cases thereof. Concluding remarks and suggestions for future studies are mentioned in Chapter 5.

1.2 Background

In an attempt to obtain a set of distributions that are able to model multimodality and skewness, Fernández-Durán [10] suggests that a family of circular distributions can be defined based on non-negative trigonometric

sums. Resulting from a theorem developed by Fejér [10] that is stated in the Appendix as Theorem 1, Fernández-Durán [10] defines a family of circular distributions to be

$$f(\theta; n) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{n}\{a_k cos(k\theta) + b_k \sin(k\theta)\}, \tag{1.1}$$

where (1.1) relates to Theorem 1 in the sense that $\sum_{k=0}^{n} |c_k|^2 = \frac{1}{2\pi}$ and $a_k - ib_k = 2\pi \sum_{v=0}^{n-k} c_{v+k}\bar{c}_v$ for complex numbers $c_k = c_{rk} + ic_{ck}$ for $k = 0, 1, ..., n$. In order to make the parameters of the model identifiable, certain constraints are implemented so that $c_{r0} > 0$ and $c_{c0} = 0$ [10].

In the following section certain circular distributions will be briefly revisited as background for the development of the cardioid generator class in Chapter 2.

1.2.1 Circular uniform distribution

The circular uniform (CU) distribution is obtained when total probability has a uniform spread on the circumference of a circle [16]. The effect is that all directions on the circle have the same likelihood of being obtained in a random sample. For a discreet CU distribution on m points, the probability mass function is described by Pewsey, Neuhäuser and Ruxton [27] to be of the form

$$f(\theta) = \frac{1}{m}.$$

The discrete case can be useful in determining, for example, the position of a ball on an unbiased roulette wheel and a Poisson distribution can be fitted on the circle in this manner [22].

The continuous CU distribution has a constant density given by

$$f(\theta) = \frac{1}{2\pi}, \qquad 0 \leq \theta < 2\pi. \tag{1.2}$$

For the continuous CU distribution, no specific direction experiences concentration about it [22].

Following from (1.2) it is important to note that the continuous CU distribution is a special case of (1.1) where $a_k = b_k = 0$, for $k = 1, ..., n$.

In Figure 1.1, the density plots for the continuous CU are displayed. It is clear from the density in (1.2) that the CU distribution is spread out symmetrically around the circumference of the circle.

1.2.2 Cardioid distribution and extensions

An important model for circular data is the cardioid distribution, denoted by $C(\mu, \rho)$. The cardioid distribution is mainly used as a small-concentration approximation to the von Mises (vM) distribution. The two-parameter cardioid distribution was introduced by Jeffreys in 1948 [17] and the density of the cardioid distribution can

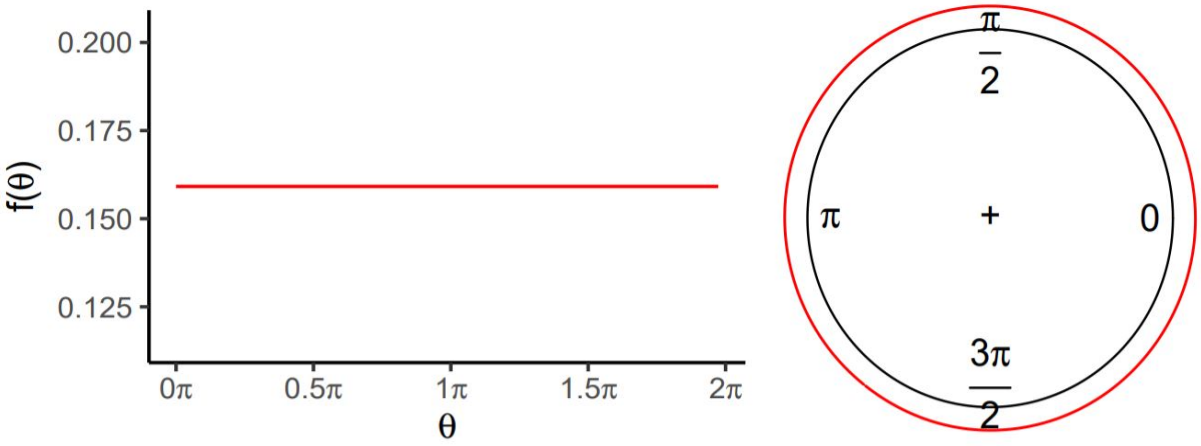


Figure 1.1: Density plots of a continuous CU distribution (1.2).

be expressed as:

$$f(\theta;\mu,\rho) = \frac{1}{2\pi}\{1 + 2\rho\cos(\theta-\mu)\}, \qquad |\rho| < \frac{1}{2}. \tag{1.3}$$

The two-parameter cardioid distribution's cumulative distribution function (cdf) is given by

$$F(\theta;\mu,\rho) = \frac{\theta}{2\pi} + \frac{\rho}{\pi}(\sin(\theta-\mu) + sin(\mu)). \tag{1.4}$$

Mardia and Jupp (2000) states that when θ is distributed as $C(\mu,\rho)$ then ρ represents the mean resultant length of $C(\mu,\rho)$, while μ represents the mean direction whenever $\rho > 0$, and $(1-\rho)$ is then the circular variance [22]. When $\rho = 0$ the distribution reduces to the circular uniform distribution as displayed in (1.2). The mode of this unimodal and symmetric distribution is found at μ.

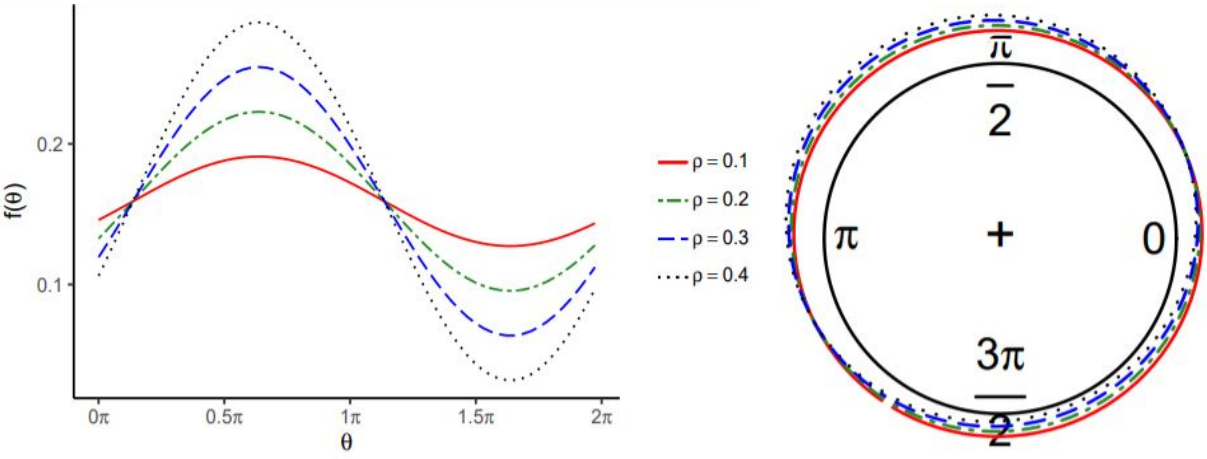


Figure 1.2: Density plots of the cardioid distribution for $\mu = 2$ (1.3).

In Figure 1.2 and Figure 1.3 the densities for the cardioid distribution are displayed. In Figure 1.2 the effect of a change in ρ is examined. It can be seen from the plots how the overall effect of parameter ρ is found in the change of the concentration of points for varying choices of ρ. The images in Figure 1.3 describe the

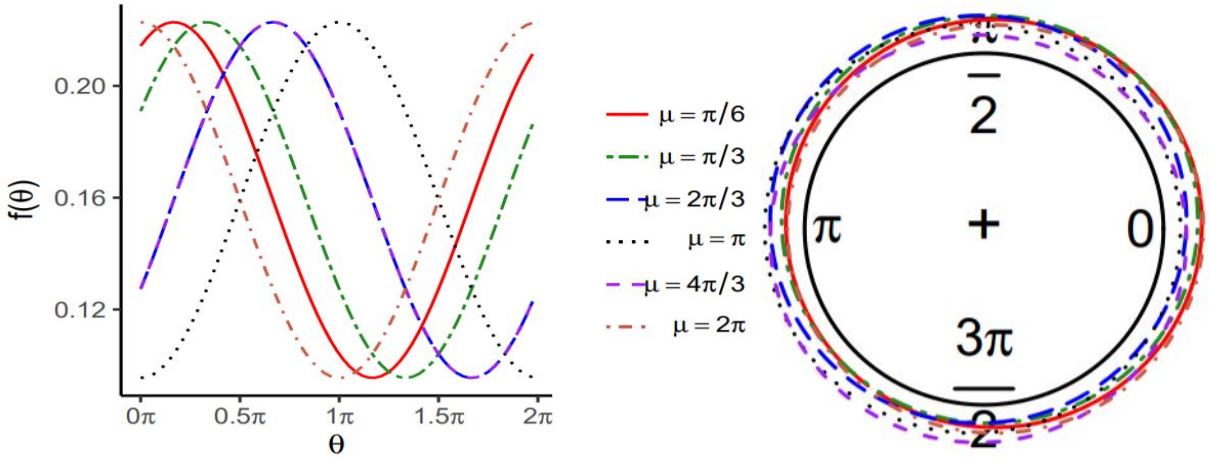


Figure 1.3: Density plots of the cardioid distribution for $\rho = 0.2$ (1.3).

effect of a change in μ. Here it is evident from the density plots that the value of μ affects the position of the mode.

Extensions of the cardioid distribution are, however, required for asymmetric and multimodal approaches. The cardioid distribution is in essence a symmetric distribution. Extended models have been introduced by developing distributions with one or more parameters added to a standard cardioid distribution [25] that have proven themselves to be useful in contributing towards a more flexible family of distributions for disciplines like actuarial science and survival analysis [22]. Some of these extensions have, however, presented difficult analytic formulas for their densities, as will briefly be discussed.

A family of symmetric circular distributions were described by Batschelet [5] to have the following density:

$$h(\theta;\mu,\rho,\nu) = \frac{1}{2\pi\{1-\rho I_1(\nu)\}}\left[1+\rho\cos\{(\theta-\mu)+\nu\sin(\theta-\mu)\}\right], \qquad -\pi<\theta<\pi. \tag{1.5}$$

In this equation $-\pi < \mu < \pi$ indicates the mean direction and $-1 \leq \rho \leq 1$ is a concentration parameter. Furthermore, $-1 < \nu < 1$ is a shape parameter and $I_n(\aleph)$ denotes a Bessel function of the first kind and nth order [1]. The bounds for the shape parameter, ν, ensures that (1.5) is unimodal.

It is then assumed by Abe, Pewsey and Shimizu [1] that $0 \leq \rho \leq 1$ and $-\infty < \nu < \infty$, where ν controls both the modality and the shape of the distribution in (1.5). When $\nu = 0$ in (1.5), the cardioid distribution is obtained. This family of distributions is an extension of the cardioid distribution and are referred to as Papakonstantinou's family. Abe, Pewsey and Shimizu [1] reported that for $|\nu| \leq 1$, the distribution is unimodal and for $|\nu| > 1$ the distribution is multimodal. Furthermore, when $-1 \leq \nu < 0$ the distribution obtained is more flattened at the top than a cardioid distribution and when $0 < \nu \leq 1$, the top is more peaked [1].

An advantage of Papakonstantinou's family is that the trigonometric moments thereof can be expressed by means of Bessel functions [1]. This has the effect of standard measures used as summaries of the circular distributions te be expressed in simple formats. The class of unimodal Papakonstantinou distributions, such

as the cardioid, are flexible models for data distributed symmetrically over the whole of the unit circle [1].

Papakonstantinou's asymmetrical extended cardioid distribution (AEC) was derived by Abe, Pewsey and Shimizu [2] based on the approaches followed by Papakonstantinou in the symmetric extended cardioid and Batschelet who extended the shape parameter for the von Mises distribution. Both the densities of the von Mises and the cardioid distributions make use of the $\cos(\theta - \mu)$ transformation. In addition to this transformation, the following expression can be used to extend the shape characteristics of circular distributions:

$$\cos(\theta - \mu + \eta\cos(\theta - \mu)), \qquad -\infty < \eta < \infty. \tag{1.6}$$

By substituting θ with $\phi - \frac{\pi}{2}$ in (1.6) and making use of the trigonometric cofunction identity

$$\cos(\phi - \frac{\pi}{2}) = \sin(\phi),$$

we have that

$$\cos((\phi - \tfrac{\pi}{2}) - \mu + \eta\cos((\phi - \tfrac{\pi}{2}) - \mu)) \quad = \sin(\phi - \mu + \eta\sin(\phi - \mu)), \tag{1.7}$$

for $-\infty < \eta < \infty$. When (1.7) is substituted into the cardioid density (1.3) and we let $\kappa = \rho$ here, it leads to the following density:

$$f(\theta; \eta, \rho, \mu) = \frac{1}{2\pi}(1 + \kappa\sin((\theta - \mu) - \eta\sin(\theta - \mu))), \tag{1.8}$$

where $0 \leq \kappa \leq 1$ is the concentration parameter. This density is known as Papakonstantinou's *asymmetric* extended cardioid (AEC) density and is favourable in light of the closed form of its normalizing constant [2]. This density is symmetric about $\frac{\pi}{2}$ if $\eta = 0$, but otherwise it is asymmetric [2]. For any $\phi = \theta - \frac{\pi}{2}$, we have that when $\mu = 0$, $\sin(\theta - \eta\sin(\theta)) = \cos(\phi + \eta\cos(\phi))$ for any $-\frac{3\pi}{2} \leq \phi < \frac{\pi}{2}$ [2]. This density is unimodal with mode at $\theta = 0$ and antimode at $\theta = -\pi$ as long as $|\eta| \leq 1$, but more modes are obtainable when $|\eta| > 1$.

In Figure 1.4 it can be seen that a change in the value of η, while keeping the other parameters constant at $\mu = \pi$ and $\kappa = 0.2$, leads to the density forming more than one mode as the value of η is increased.

In Figure 1.5 the effect of a change in μ is investigated, while keeping the other parameters constant at $\eta = 2$ and $\kappa = 0.2$. The density plots show that the mode shifts to the right or left depending on the chosen value of μ.

When setting $\mu = \pi$ and $\eta = 2$, it can be seen in Figure 1.6 that a change in the shape parameter κ leads to densities that expand further away from the mean as the values of κ are increased.

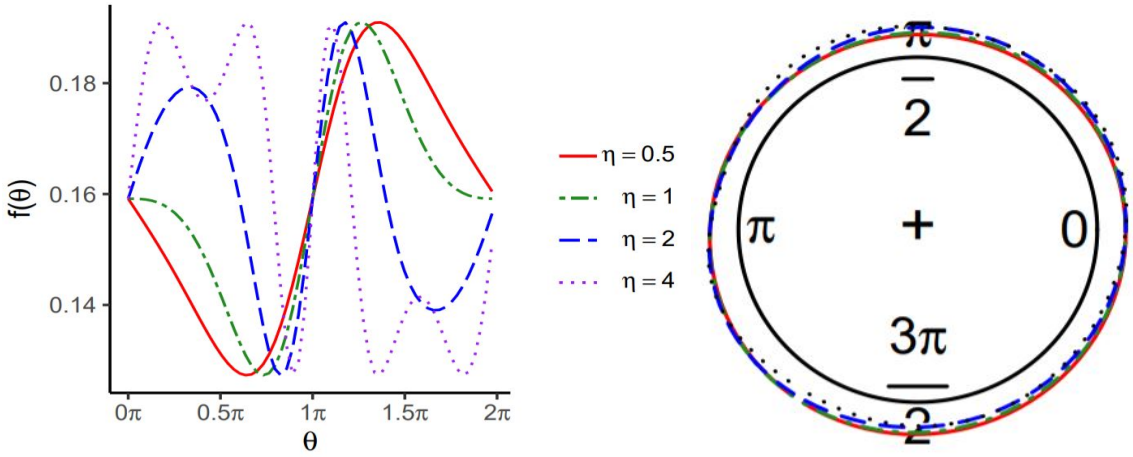


Figure 1.4: Density plots of the AEC distribution for $\kappa = 0.2$ and $\mu = \pi$ (1.8).

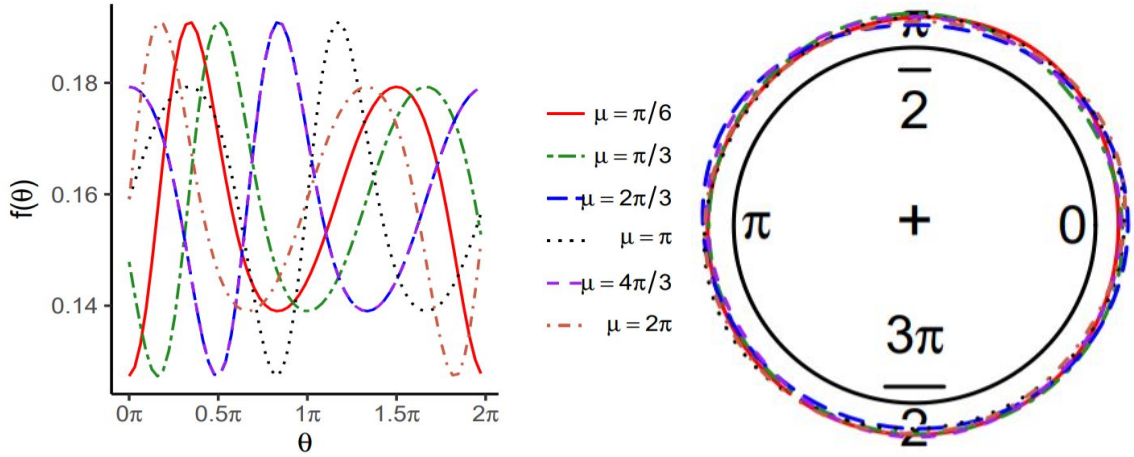


Figure 1.5: Density plots of the AEC distribution for $\eta = 2$ and $\kappa = 0.2$ (1.8).

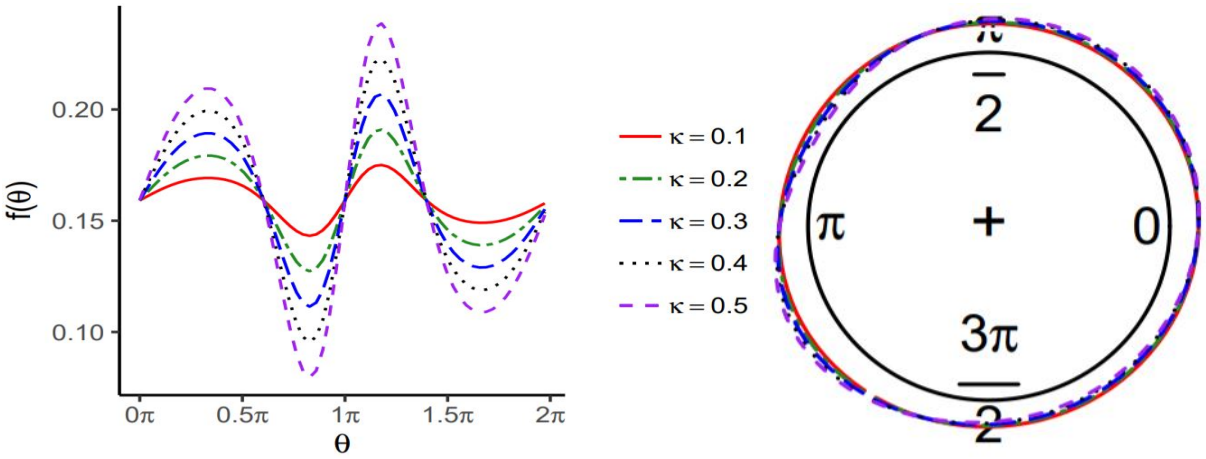


Figure 1.6: Density plots of the AEC distribution for $\mu = \pi$ and $\eta = 2$ (1.8).

1.2.3 Von Mises distribution and extensions

The von Mises (vM) distribution is useful in its wide application to circular data and was introduced by von Mises towards studying deviations of measured atomic weights from integral values in 1918 [22]. The density

of the vM distribution can be given by

$$f_{vM}(\theta) = [2\pi I_0(\kappa)]^{-1} \exp\left(\kappa \cos(\theta - \mu)\right), \qquad 0 \leq \theta < 2\pi \tag{1.9}$$

for $0 \leq \mu < 2\pi$ and $\kappa \geq 0$ and $I_0(\kappa)$ a modified Bessel function of the first kind of order zero [25]. The parameter κ is the concentration parameter and the μ indicates the mean direction [22]. The vM distribution is symmetric about its mean at $\theta = \mu$ and unimodal with a mode at $\theta = \mu$ and antimode at $\theta = \mu + \pi$ [22]. If $\kappa = 0$ in (10), then the probability distribution simplifies to the uniform distribution in equation (1.2). It is also noteworthy that the vM distribution can be approximated by die cardioid distribution if the concentration parameter κ is small and $\exp(x) \simeq 1 + x$ [22].

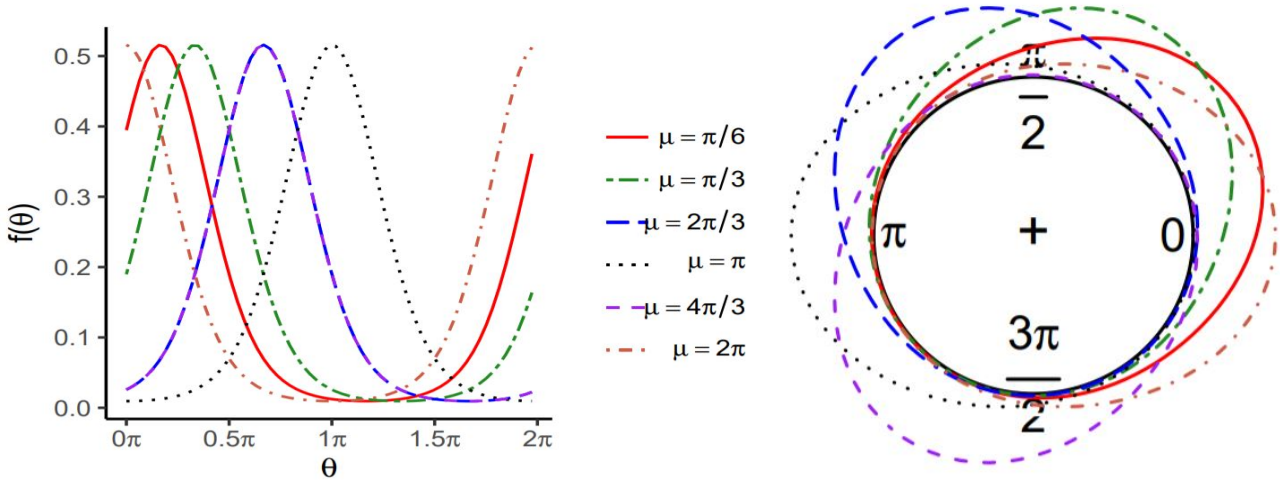


Figure 1.7: Density plots of the von Mises distribution for $\kappa = 2$ (1.9).

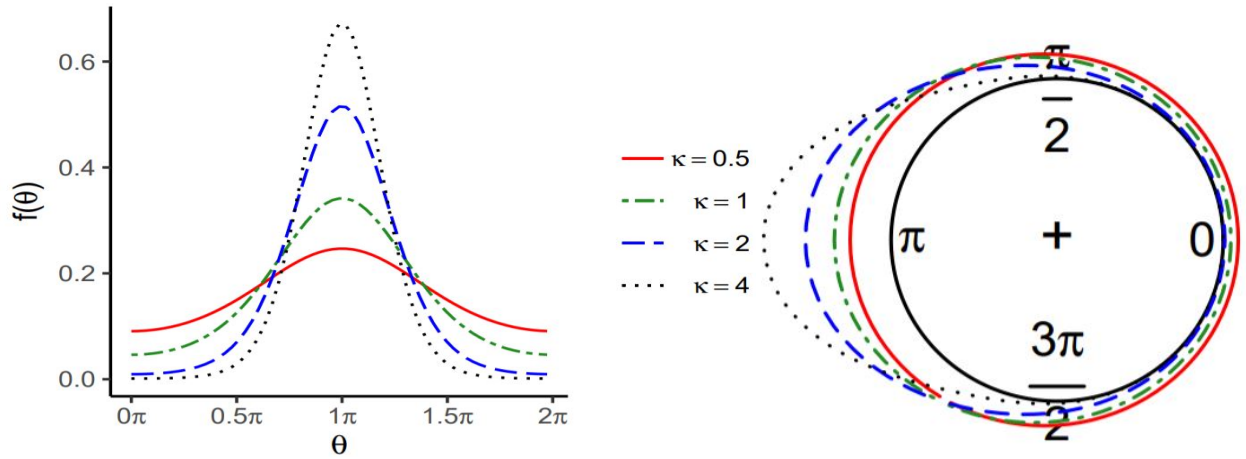


Figure 1.8: Density plots of the von Mises distribution for $\mu = \pi$ (1.9).

In Figure 1.7 and Figure 1.8 densities are displayed for the vM distribution. In Figure 1.7 the effect of a change in μ is examined. It is clear from the plots how a change in μ affects the densities in terms of the

direction in which most of the data is found. There are notable shifts in the densities distorting the symmetric shape to some extent as μ is increased. Turning to Figure 1.8, the images describe the effect of a change in κ. Here it is evident that the size of the concentration parameter κ plays an important part in the densities, with smaller κ values leading to a more evenly spread contentration about $\mu = \pi$. Larger values of κ lead the data points to become more concentrated around $\mu = \pi$.

A generalized von Mises (GvM) distribution was also introduced by Gatto and Jammalamadaka [12]. The GvM distribution is an extension of the vM and its density can be expressed as:

$$f(\theta|\mu_1, \mu_2, \rho_1, \rho_2) = \frac{1}{2\pi G_0(\delta, \rho_1, \rho_2)} \exp(\rho_1 \cos(\theta - \mu_1) + \rho_2 \cos(2(\theta - \mu_2))),$$

for $\theta\epsilon[0, 2\pi)$, $\mu_1\epsilon[0, 2\pi)$, $\mu_2\epsilon[0, \pi)$, $\delta = (\mu_1 - \mu_2)$ mod π, $\rho_1, \rho_2 > 0$ with normalizing constant

$$G_0(\delta, \rho_1, \rho_2) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\rho_1 \cos\theta + \rho_2 \cos(2(\theta + \delta)))d\theta.$$

The GvM was found to display much greater flexibility than the vM distribution and was applied successfully on data from meteorology [12]. While being more robust, the GvM distribution retains valuable characteristics such as its relationship with the normal distribution and forming part of the exponential family [12].

1.2.4 Wrapped Cauchy

Wrapped distributions originate from the concept that a distribution on a line can be wrapped around the circumference of a circle with unit radius [22]. One important wrapped distribution in the world of circular statistics, is the wrapped Cauchy (WC) distribution. This distribution is both unimodal and symmetric about μ and is closely connected to the projected normal distribution [22]. When considered on the real line, the density of the Cauchy distribution is given by

$$f(x; \mu, b) = \frac{1}{\pi} \frac{b}{b^2 + (x - \mu)^2} \qquad -\infty < \mu < \infty, \quad b > 0.$$

When wrapped around the circumference of the unit radius circle, this density becomes

$$f_{wc}(\theta; \mu, \rho) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{k=1}^{\infty} \rho^k \cos(k(\theta - \mu)) \right\} \qquad \text{for } \rho = e^{-b}, \tag{1.10}$$

while its distribution function is found to be

$$F_{WC}(\theta) - F_{WC}(\mu) = \frac{1}{2\pi} \arccos \frac{(1 + \rho^2)\cos(\theta - \mu) - 2\rho}{1 + \rho^2 - 2\rho\cos(\theta - \mu)}, \qquad \mu \leq \theta \leq \mu + \pi. \tag{1.11}$$

It can be shown that the real part of the geometric series $\sum_{k=1}^{\infty} \rho^k e^{-ik(\theta-\mu)}$ causes (1.10) to reduce to

$$wc(\theta;\mu,\rho) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}. \tag{1.12}$$

The mean direction is $\mu(\mathrm{mod}2\pi)$, with mean resultant length ρ [22]. When $\rho \to 0$, the WC tends to the uniform distribution in (1.2) and when $\rho \to 1$, the WC becomes concentrated at μ.

Attempting to address skewness in the generally accepted symmetric environment of the WC distribution, Kato and Jones [19] introduced a four-parameter extended family of distributions related to the WC distribution. This family of distributions has a closed form and can be asymmetric depending on the choice of parameters. The derived trigonometric moments exist in closed forms and is therefore more mathematically tractable than many other circular distributions that need to be estimated with numerical methods [19].

In Figure 1.9 it can be seen that varying μ over a constant $\rho = 0.2$, leads to the plot of the WC density to clearly display the modes centred at the proposed mean direction. It can also be seen from Figure 1.10 that keeping the mean direction constant at $\mu = \pi$ and varying ρ, causes the circular density to clearly exhibit a stronger concentration around μ when ρ is closer to 1.

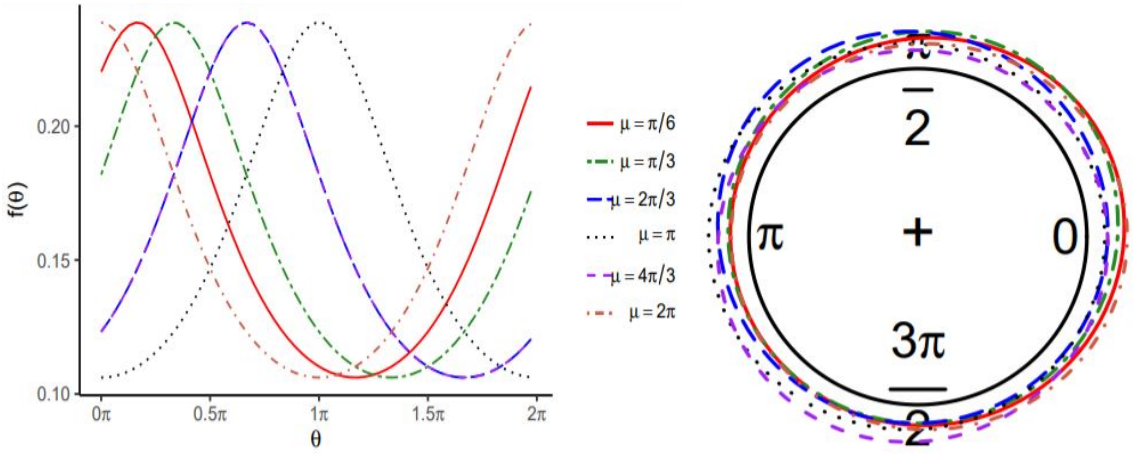


Figure 1.9: Density plots of the WC distribution for $\rho = 0.2$ (1.10).

1.2.5 Generalized t

Siew, Kato and Shimizu [28] introduced an extended version of the t-distribution on the unit circle by conditioning a normal mixture model. The generalized t-distribution (GT) accommodates multimodality and asymmetry and special cases thereof, such as the generalized von Mises (when $\psi \to 0$), generalized cardioid (when $\psi = 1$) and generalized wrapped Cauchy (when $\psi = -1$), can be obtained by selecting the parameters in a certain way [28].

The density of the GT distribution on the circle, is given by

$$f(\theta) = c_{gt}\left\{1 + \tanh(\kappa_1\psi)\cos(\theta-\mu) + \tanh(\kappa_2\psi)\cos(\theta-\nu)\right\}^{1/\psi}, \tag{1.13}$$

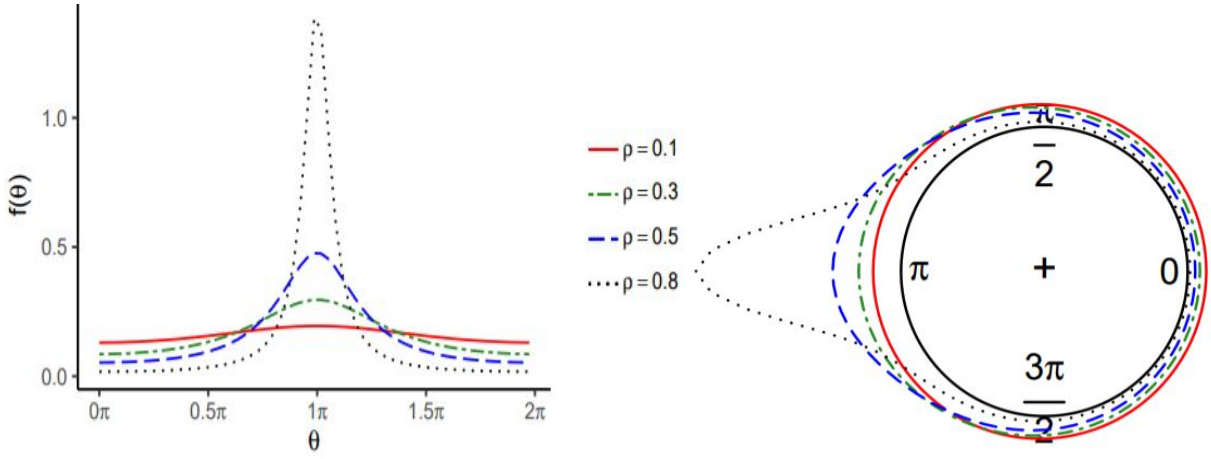


Figure 1.10: Density plots of the WC distribution for $\mu = \pi$ (1.10).

where $\psi \neq 0$, $\kappa_1, \kappa_2 \geq 0$, $0 \leq \mu < 2\pi$ and $0 \leq \nu < \pi$. The normalizing constant, c_{gt}, is expressed as

$$c_{gt} = \left[\int_0^{2\pi} \{1 + \tanh(\kappa_1 \psi) \cos(\theta - \mu) + \tanh(\kappa_2 \psi) \cos(\theta - \nu)\}^{1/\psi} \right]^{-1}. \tag{1.14}$$

The additional parameters in this distribution results in the distribution being more flexible towards multimodality and asymmetry than some of the other distributions discussed in earlier sections.

In Figures 1.11 to 1.15 the density plots for various combinations of parameters were obtained.

- In Figure 1.11 the values of ψ is varied while all other parameters remain constant. Smaller values of ψ lead to a sharper peak in the direction of $\mu = \nu = \pi$.
- In Figure 1.12 the values of κ_1 is varied while all other parameters remain constant. Larger values of κ_1 lead to a higher concentration of data points around $\mu = \nu = \pi$.
- In Figure 1.13 the values of κ_2 is varied while all other parameters remain constant. Larger values of κ_2 lead to more prominent multimodal outcomes around the chosen $\mu = \nu = \pi$ as well as the opposite direction at 0.
- In Figure 1.14 the values of μ is varied while all other parameters remain constant. It is clear that the modes shift according to the value of μ, but the concentration is not altered.
- In Figure 1.15 the values of ν is varied while all other parameters remain constant. It can be seen from the density plots that the influence of $\mu = 0$ is strong and results in a shift in the modes. The effect of a change in ν is less prominent, but adds to the distribution being able to address asymmetry.

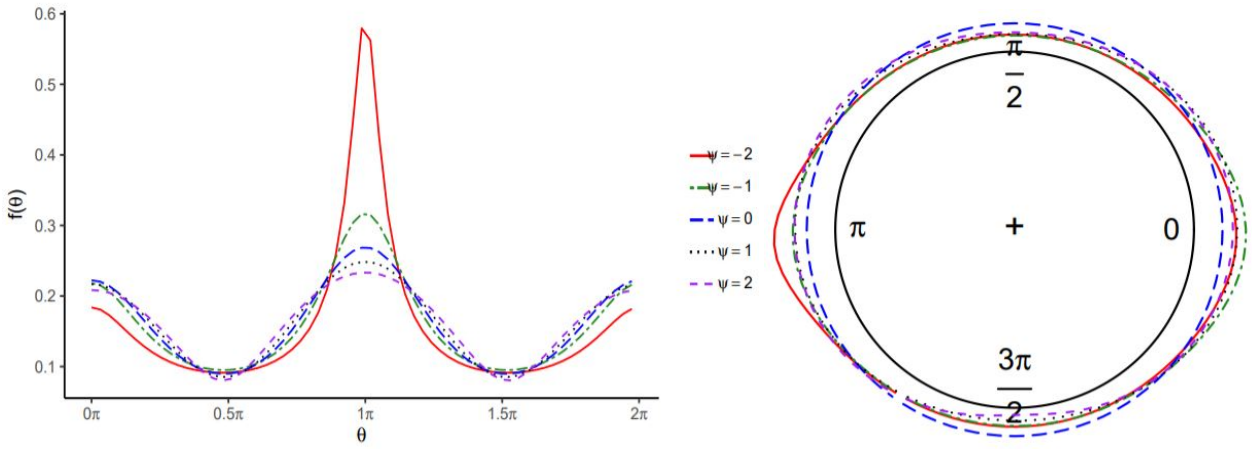


Figure 1.11: Density plots of the GT distribution for $\mu = \nu = \pi$, $\kappa_1 = 0.1$ and $\kappa_2 = 0.5$ (1.13).

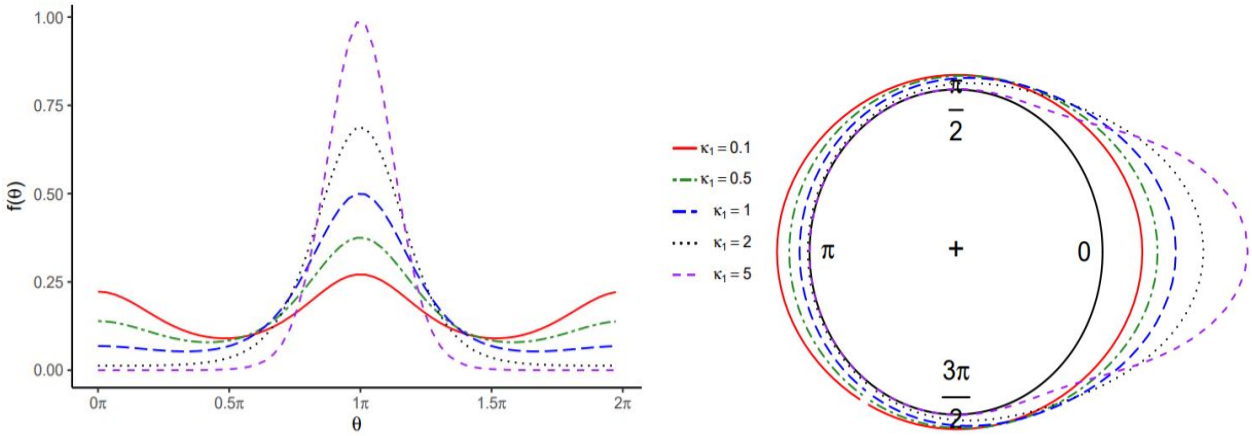


Figure 1.12: Density plots of the GT distribution for $\mu = \nu = \pi$, $\psi = 0.01$ and $\kappa_2 = 0.5$ (1.13).

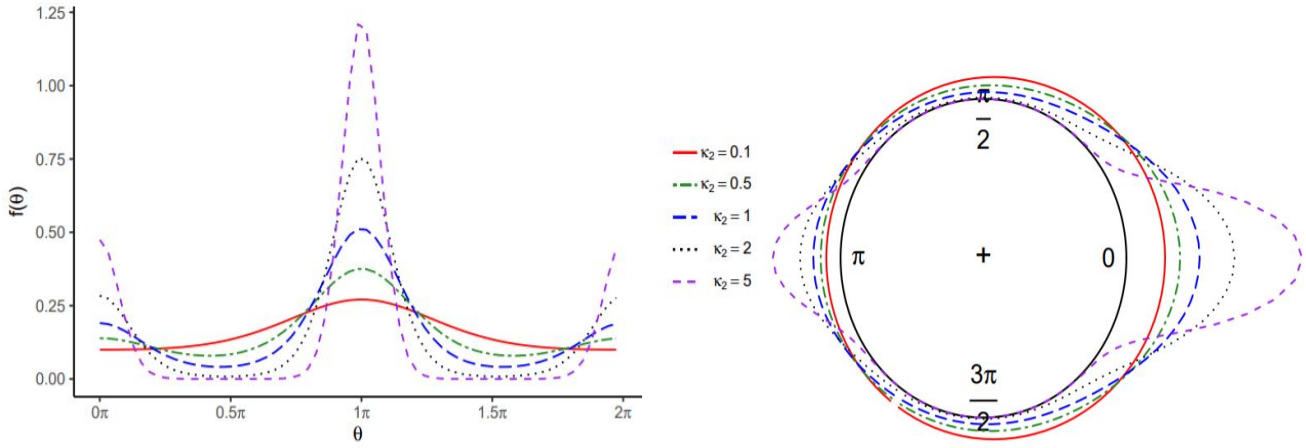


Figure 1.13: Density plots of the GT distribution for $\mu = \nu = \pi$, $\kappa_1 = 0.5$ and $\psi = 0.01$ (1.13).

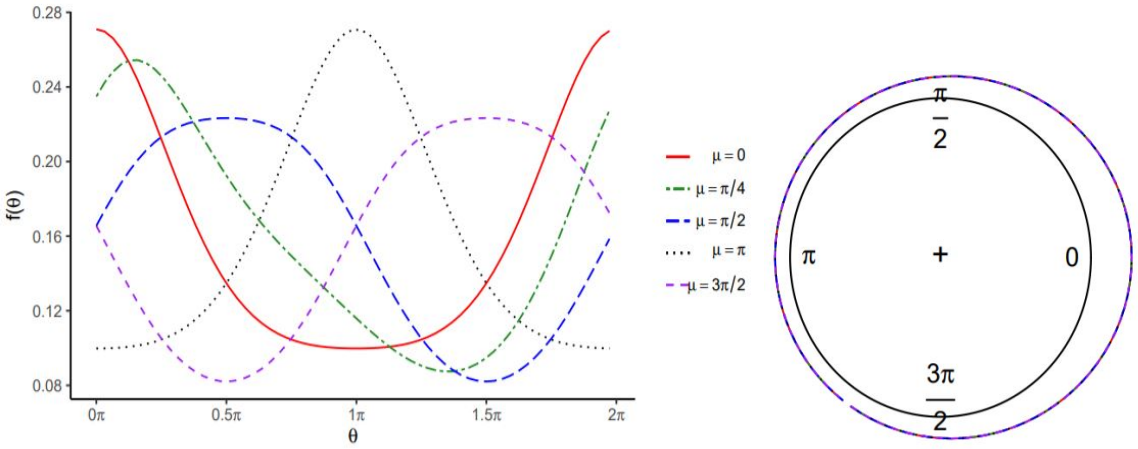


Figure 1.14: Density plots of the GT distribution for $\psi = 0.01$, $\nu = \pi$, $\kappa_1 = 0.5$ and $\kappa_2 = 0.1$ (1.13).

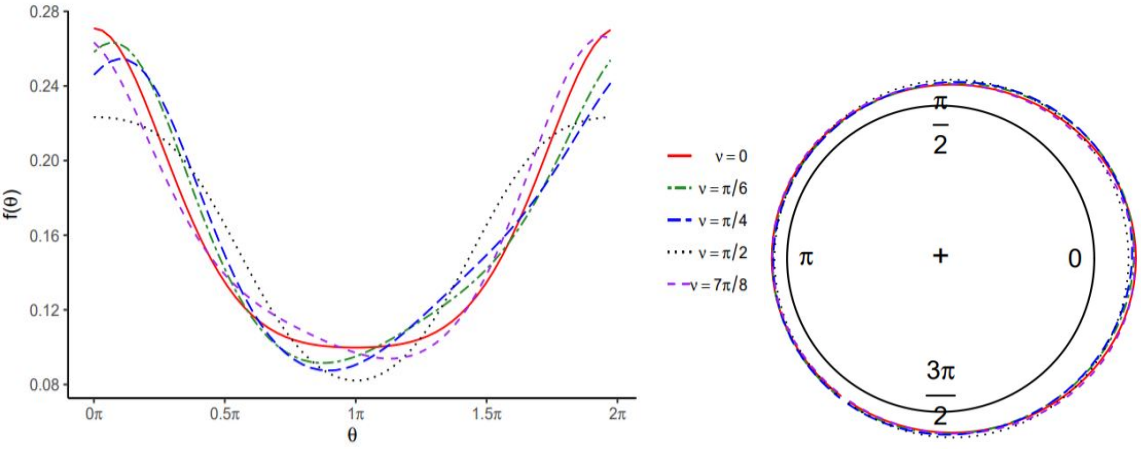


Figure 1.15: Density plots of the GT distribution for $\mu = 0$, $\psi = 0.01$, $\kappa_1 = 0.5$ and $\kappa_2 = 0.1$ (1.13).

Chapter 2

Generalized cardioid-generator class

A more flexible generalized class of distributions for circular data is proposed of the following form:

$$f(\theta) = c.g(\cos(\theta - \mu)). \tag{2.1}$$

$g(\cdot)$ is a Borel measurable function which admits Taylor series expansion and $g \neq 1$. $g(\cdot)$ is continuous and can be referred to as a generator function, and c is a normalizing constant. In this case, it is clear that the generator function, $g(\cos(\theta - \mu))$, relates to the cardioid distribution in (1.3).

2.1 Deriving the generalized cardioid-generator class

Subsequently a general formula for the appropriate normalizing constant for (2.1) is obtained. The proposed generalized formula below only requires the derivatives of the function $g(cos(\theta - \mu))$ evaluated at 0 in each case. In order to arrive at such a formula, the definition of a probability distribution of an absolutely continuous distribution [22] needs to be used. Since

$$\int_0^{2\pi} f(\theta)\mathrm{d}\theta = 1, \tag{2.2}$$

it follows that

$$c = \left(\int_0^{2\pi} g(\cos(\theta - \mu))\mathrm{d}\theta\right)^{-1}. \tag{2.3}$$

It will be assumed that $a = 0$ in equation (6.2) in the Appendix, so that the Taylor series reduces to the Maclaurin series.

For the function $g(\cos(\theta - \mu))$, the Maclaurin series can be written as

$$g(\cos(\theta-\mu)) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} \cos^k(\theta-\mu), \tag{2.4}$$

where $g^{(k)}(0)$ represente the k^{th} derivative of $g(\cdot)$ around zero. Note that this series can also be represented as

$$\begin{aligned} g(\cos(\theta-\mu)) &= \tfrac{g^{(0)}(0)}{0!}\cos^0(\theta-\mu) + \textstyle\sum_{k=1}^{\infty} \frac{g^{(k)}(0)}{k!}\cos^k(\theta-\mu) \\ &= g(0) + \textstyle\sum_{k=1}^{\infty} \frac{g^{(k)}(0)}{k!}\cos^k(\theta-\mu), \end{aligned} \tag{2.5}$$

which will be referred to for a specific distribution under consideration, in (2.26).

Substituting (2.4) into (2.3), it follows that

$$\begin{aligned} c &= \left(\textstyle\int_0^{2\pi}\left(\sum_{k=0}^{\infty}\frac{g^{(k)}(0)}{k!}\cos^k(\theta-\mu)\right)\mathrm{d}\theta\right)^{-1} \\ &= \left(\textstyle\sum_{k=0}^{\infty}\frac{g^{(k)}(0)}{k!}\int_0^{2\pi}\cos^k(\theta-\mu)\mathrm{d}\theta\right)^{-1}. \end{aligned} \tag{2.6}$$

Below three approaches are described and derived to determine the value of the normalizing constant c.

Approach 1:

Using Gradshteyn and Ryzhik [13] for even values of k and $(k-1)!! = (k-1)(k-3)(k-5)...1$, the integral in (2.6) can be evaluated at 0 and 2π to solve for

$$\begin{aligned} \int\cos^k(\theta-\mu)\mathrm{d}\theta &= \tfrac{\sin(\theta-\mu)}{k}\left(\cos^{k-1}(\theta-\mu) + \textstyle\sum_{b=1}^{\frac{k}{2}-1}\frac{(k-1)(k-3)...(k-2b+1)}{2^b(\frac{k}{2}-1)(\frac{k}{2}-2)...(\frac{k}{2}-b)}\cos^{(k-2b-)1}(\theta-\mu)\right) \\ &\quad + \tfrac{(k-1)!!}{2^{\frac{k}{2}}\frac{k}{2}!}(\theta-\mu). \end{aligned} \tag{2.7}$$

For odd values of k, the integral the integral in (2.6) can be evaluated at 0 and 2π to solve for

$$\int\cos^k(\theta-\mu)\mathrm{d}\theta = \tfrac{\sin(\theta-\mu)}{k}\left(\cos^{k-1}(\theta-\mu) + \textstyle\sum_{b=0}^{\frac{k-1}{2}-1}\frac{2^{b+1}(\frac{k-1}{2})(\frac{k-1}{2}-1)...(\frac{k-1}{2}-b)}{(k-2)(k-4)...((k-1)-2b-1)}\cos^{((k-1)-2b-2)}(\theta-\mu)\right). \tag{2.8}$$

Approach 2:

Another approach to obtain c in (2.6) could be to make use of Euler's formula, equation (6.3) in the Appendix. Instead of substituting $\cos^k(\theta-\mu)$ in (2.4), the equation in (6.3) can be included in the integral to find

$$\begin{aligned} \textstyle\int_0^{2\pi} f(\theta)\mathrm{d}\theta &= c\textstyle\sum_{k=0}^{\infty}\frac{g^{(k)}(0)}{k!}\int_0^{2\pi}\left(\frac{\exp(i(\theta-\mu))+\exp(-i(\theta-\mu))}{2}\right)^k\mathrm{d}\theta \\ &= \tfrac{c}{2^k}\textstyle\sum_{k=0}^{\infty}\frac{g^{(k)}(0)}{k!}\int_0^{2\pi}\left(\exp(i(\theta-\mu))+\exp(-i(\theta-\mu))\right)^k\mathrm{d}\theta \\ &= \tfrac{c}{2^k}\textstyle\sum_{k=0}^{\infty}\frac{g^{(k)}(0)}{k!}\int_0^{2\pi}\exp(-ik(\theta-\mu))(\exp(2i(\theta-\mu))+1)^k\mathrm{d}\theta \end{aligned} \tag{2.9}$$

Approach 3:

In this approach, we solve (2.6) by applying a transformation. Let $x = \cos(\theta - \mu)$ (with bounds $-1 \leq x = \cos(\theta - \mu) \leq 1$). We find $\theta = \arccos(x) + \mu$ and $d\theta = -\arcsin(x)dx$. Therefore, from Gradshteyn and Ryzhik [13],

$$\int_0^{2\pi} \cos^k(\theta - \mu)d\theta = \int_{-1}^{1} x^k \arcsin(x)\mathrm{d}x = \frac{\frac{\pi}{2}(1+(-1)^{k+1})}{k+1} - \frac{1}{k+1}\int_{-1}^{1} \frac{x^{k+1}}{\sqrt{1-x^2}}\mathrm{d}x. \tag{2.10}$$

Furthermore, if we let $x = \sin(u)$, we find that $u = \arcsin(x)$ and $dx = \cos(u)\mathrm{d}u$. Based on the identity $1 - \sin^2(u) = cos^2(u)$, we can use substitution to solve the remaining integral:

$$\begin{aligned} \int_{-1}^{1} \frac{x^{k+1}}{\sqrt{1-x^2}}\mathrm{d}x &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^{k+1}(u)\cos(u)}{\sqrt{1-sin^2(u)}}\mathrm{d}u \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+1}(u)\mathrm{d}u \end{aligned} \tag{2.11}$$

The resulting normalizing constant is thus found, from the same definition as in (2.3), to be

$$\begin{aligned} c &= \left(\sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} \int_{-1}^{1} x^k \arcsin(x)\mathrm{d}x\right)^{-1} \\ &= \left(\sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} \left(\frac{\frac{\pi}{2}(1+(-1)^{k+1})}{k+1} - \frac{1}{k+1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+1}(u)\mathrm{d}u\right)\right)^{-1} \end{aligned} \tag{2.12}$$

In this simplified form, we can then solve the integral for even powers of the form $k+1$ in the following manner:

$$\int \sin^{k+1}(u)\mathrm{d}u = \frac{1}{2^{k+1}}\binom{k+1}{\frac{k+1}{2}}u + \frac{1}{2^k}(-1)^{\frac{k+1}{2}} \sum_{p=0}^{\frac{k-1}{2}} (-1)^p \binom{k+1}{p} \frac{\sin(k+1-2p)u}{k+1-2p}. \tag{2.13}$$

For odd powers of the form $k+1$, the integral solves to

$$\int \sin^{k+1}(u)\mathrm{d}u = \frac{1}{2^k}(-1)^{\frac{k+2}{2}} \sum_{p=0}^{\frac{k}{2}} (-1)^p \binom{k+1}{p} \frac{\cos(k+1-2p)u}{k+1-2p}. \tag{2.14}$$

2.2 Special cases of the generalized cardioid-generator class

By letting $x = \cos(\theta - \mu)$, we can obtain the generator functions of Table 2.1 through transformations.

2.2.1 Cardioid-von Mises combination

The following generator function is proposed for the expression $\cos(\theta - \mu)$:

$$g(\cos(\theta - \mu)) = (1 + 2\alpha\cos(\theta - \mu))e^{\alpha\cos(\theta-\mu)}, \tag{2.15}$$

so that the appropriate density, as described in (2.1), can be given as

Table 2.1: Special cases of the generator function.

Distribution	**g(x)**	**c**
Cardioid [22]	$(1+2\rho x)$	$(2\pi)^{-1}$
Von Mises [22]	$e^{\rho x}$	$(2\pi I_0(\rho))^{-1}$
Wrapped Cauchy [22]	$(1+\rho^2)(1+\rho^2-2\rho x)^{-1}$	$(2\pi)^{-1}$
Exponentiated cardioid [25]	$\left\{2\rho\sin(\mu)+(\arccos(x)+\mu)+2\rho\sqrt{1-x^2}\right\}^{\beta-1}\beta(1+2\rho x)$	$(2\pi)^{-\beta}$
Asymmetrical extended cardioid [2]	$1+\rho(w\sin(\nu\sqrt{1-x^2})+\sqrt{1-x^2}\cos(\nu\sqrt{1-x^2}))$	$(2\pi)^{-1}$
Cardioid-von Mises [14]	$(1+2\alpha x)e^{\alpha x}$	c_{cvm} (2.17)
Cardioid-t	$\left(1+\frac{\kappa}{n}x\right)^{-n}$	c_{tc}(2.33)

$$f_{cvm}(\theta) = c_{cvm}(1+2\alpha\cos(\theta-\mu))e^{\alpha\cos(\theta-\mu)}, \tag{2.16}$$

for $0 \leq \alpha \leq \frac{1}{2}$ [14]. The normalizing constant, c_{cvm}, is found as

$$c_{cvm} = \left[\int_0^{2\pi}(1+2\alpha\cos(\theta-\mu))e^{\alpha\cos(\theta-\mu)}\mathrm{d}\theta\right]^{-1} = \frac{1}{2\pi I_0(\alpha)+4\alpha\pi I_1(\alpha)}, \tag{2.17}$$

where $I_r(z)$ is the modified Bessel function of the first kind and of order r. In terms of (2.12), the normalizing constant can be expressed as

$$c*_{cvm} = \left(\sum_{k=0}^{\infty}\frac{(1+2k)\alpha^k}{k!}\left\{\frac{\frac{\pi}{2}(1-(-1)^{k+1})}{k+1}-\frac{1}{k+1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin^{k+1}(u)\mathrm{d}u\right\}\right)^{-1} \tag{2.18}$$

where

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin^{k+1}(u)\mathrm{d}u = \begin{cases}\left[\frac{1}{2^{k+1}}\binom{k+1}{\frac{k+1}{2}}u+\frac{(-1)^{\frac{k+1}{2}}}{2^k}\sum_{s=0}^{\frac{k-1}{2}}(-1)^s\binom{k+1}{s}\frac{\sin(k+1-2s)u}{k+1-2s}\right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} & \text{if } k \text{ is odd}\\ 0 & \text{if } k \text{ is even.}\end{cases} \tag{2.19}$$

and

$$\frac{\frac{\pi}{2}(1-(-1)^{k+1})}{k+1} = \begin{cases}0 & \text{if } k \text{ is odd}\\ \frac{\pi}{k+1} & \text{if } k \text{ is even.}\end{cases} \tag{2.20}$$

Based on the equation in (2.17),Table 2.2 is constructed for the normalizing constant that were obtained by numerical integration for various values of the parameter α. Note that the normalizing constant is independent of μ. Trigonomically this makes sense in the circular world and it can also be deduced from the final derivation in (2.17).

Based on Table 2.2 and the normalizing constants that were found through numerical integration, the

Table 2.2: Finding the normalizing constant, c_{cvm}, through numerical integration.

α	c_{cvm}
0.1	0.15719
0.2	0.15154
0.3	0.14291
0.4	0.13223
0.5	0.12045

density plots in Figures 2.1 and 2.2 were constructed. In Figure 2.1 it can be seen that a change in parameter μ affects the central position of the plots and shifts the plot left or right with regard to the origin. In Figure 2.2 the effect of different α-values over a constant parameter $\mu = 0$ can be seen.

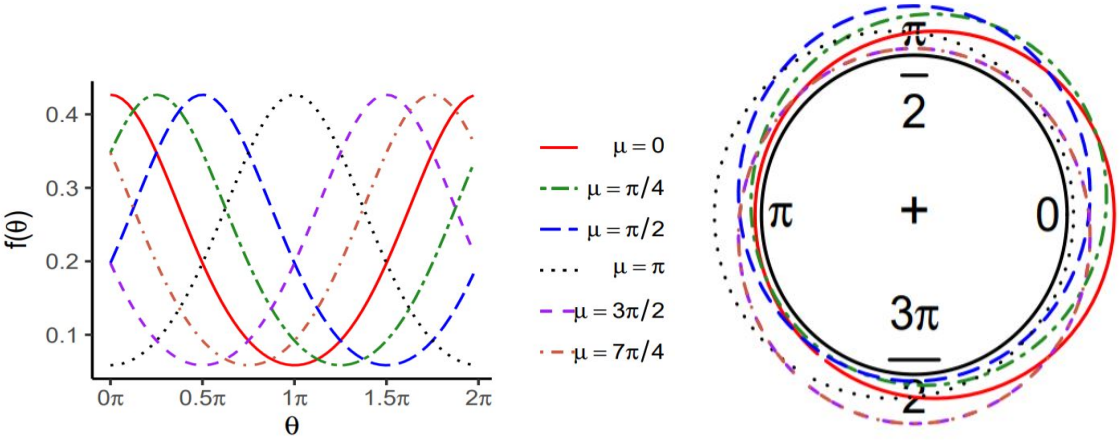


Figure 2.1: Density plots of the cardioid-von Mises distribution for $\alpha = 0.3$, varying over μ.

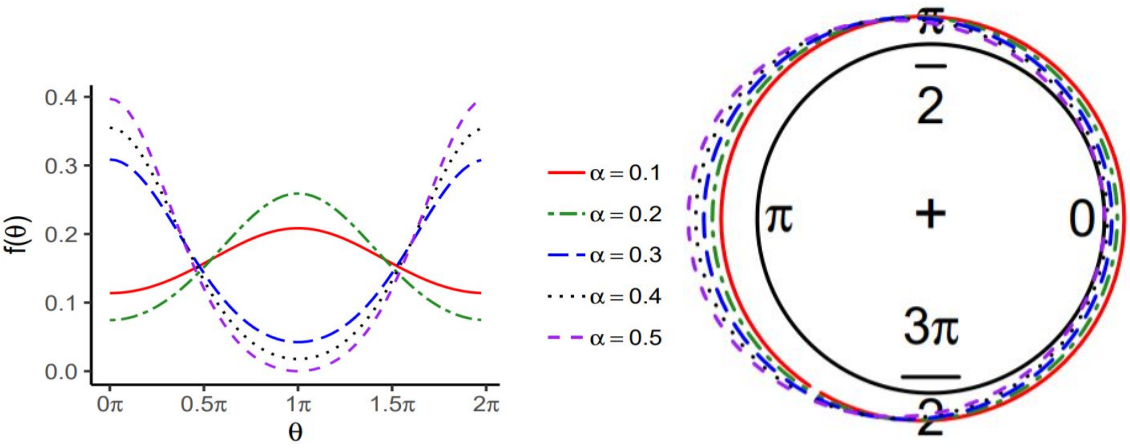


Figure 2.2: Density plots of the cardioid-von Mises distribution for $\mu = \pi$, varying over α.

Hatami and Alamatsaz [14] proposed a transformation of circular random variables with special focus on the cardioid circular distribution function. The cardioid-von Mises distribution that was proposed has the following density:

$$f^*_{CvM}(\theta) = \frac{1+\lambda\cos(\theta-\mu)}{2\pi I_0(\kappa)} e^{\kappa\cos(\theta-\mu+\lambda\sin(\theta-\mu)-\tau)}, \tag{2.21}$$

where $-\pi \leq \theta < \pi$, $\kappa \geq 0$, $0 \leq \lambda \leq 1$, $-\pi < \tau \leq \pi$ and $I_r(z)$ is the modified Bessel function of the first kind and of order r. This particular distribution was found to be useful in modelling both asymmetric and symmetric, unimodal or bimodal data with a wide range of skewness and kurtosis [14].

2.2.2 Cardioid-t distribution

In this section a $g(\cdot)$ function, stemming from a t-distribution, will be proposed as generator.

Exploration stage

In the first case, based on (2.4) and the transformation $x = \cos(\theta-\mu)$, the following $g(\cdot)$ function is proposed:

$$g_{ct}(x) = \left(1 + \frac{x^2}{2n-1}\right)^{-n}, \tag{2.22}$$

from the density of the Student's t-distribution in the form

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{2.23}$$

from where it was determined that $\nu = 2n-1$ and $n = \frac{\nu+1}{2}$.

In this case, the density would then be given as

$$f_{ct}(x) = c_{ct}\left(1 + \frac{x^2}{2n-1}\right)^{-n} \arcsin x. \tag{2.24}$$

For this generator class and proposed density, the normalizing constant can be found based on the Mclaurin series used in (2.4), together with the expression in (2.22). The first 20 derivatives of (2.22), evaluated at $x = 0$, is given by:

- $k = 1$: 0
- $k = 2$: $-2n(2n-1)^{-1}$
- $k = 3$: 0
- $k = 4$: $12n(n+1)(2n-1)^{-2}$
- $k = 5$: 0
- $k = 6$: $-120n(n+1)(n+2)(2n-1)^{-3}$
- $k = 7$: 0

- $k = 8$: $1680n(n+1)(n+2)(n+3)(2n-1)^{-4}$
- $k = 9$: 0
- $k = 10$: $-30240n(n+1)(n+2)(n+3)(n+4)(2n-1)^{-5}$
- $k = 11$: 0
- $k = 12$: $665280n(n+1)(n+2)(n+3)(n+4)(n+5)(2n-1)^{-6}$
- $k = 13$: 0
- $k = 14$: $-17297280n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(2n-1)^{-7}$
- $k = 15$: 0
- $k = 16$: $518918400n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(n+7)(2n-1)^{-8}$
- $k = 17$: 0
- $k = 18$: $-17643225600n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(n+7)(n+8)(2n-1)^{-9}$
- $k = 19$: 0
- $k = 20$: $670442572800n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(n+7)(n+8)(n+9)(2n-1)^{-10}$

This pattern holds for all values of k and can be confirmed with mathematical software such as Mathematica. From the fact that the 0-values, determined when k is odd in the above list, will directly cause the entire expression in (2.12) to be 0, these values are excluded from the pattern for the sake of simplicity.

The effect is that we re-label these k-terms as

- $k = 1$: $-2n(2n-1)^{-1}$
- $k = 2$: $12n(n+1)(2n-1)^{-2}$
- $k = 3$: $-120n(n+1)(n+2)(2n-1)^{-3}$
- $k = 4$: $1680n(n+1)(n+2)(n+3)(2n-1)^{-4}$
- $k = 5$: $-30240n(n+1)(n+2)(n+3)(n+4)(2n-1)^{-5}$
- $k = 6$: $665280n(n+1)(n+2)(n+3)(n+4)(n+5)(2n-1)^{-6}$
- $k = 7$: $-17297280n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(2n-1)^{-7}$
- $k = 8$: $518918400n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(n+7)(2n-1)^{-8}$
- $k = 9$: $-17643225600n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(n+7)(n+8)(2n-1)^{-9}$

- $k = 10$: $670442572800n(n+1)(n+2)(n+3)(n+4)(n+5)(n+6)(n+7)(n+8)(n+9)(2n-1)^{-10}$

From here, it is necessary to determine a simplified pattern for the terms so that it may be expressed in a condensed form that can easily be implemented in mathematical workings and coding procedures. The pattern of coefficients (not considering the signs at this point) is given by

- $k = 1$: 2
- $k = 2$: 12 (2×6)
- $k = 3$: 120 (12×10)
- $k = 4$: 1 680 (120×14)
- $k = 5$: 30 240 (1 680×18)
- $k = 6$: 665 280 (30 240×22)
- $k = 7$: 17 297 280 (665 280×26)
- $k = 8$: 518 918 400 (17 297 280×30)
- $k = 9$: 17 643 225 600 (518 918 400×34)
- $k = 10$: 670 442 572 800 (17 643 225 600×38).

It was found that this pattern has a common factor in the form of a linear series, determined from the following values:

6 , 10 , 14 , 18 , 22 , 26 , 30 , 34 , 38 , ...

This linear factor can be expressed as (2+4p). The effect of this observation is that the p^{th} term of this sequence can be obtained by multiplying the $(p-1)^{th}$ term by (2+4$(p-1)$), expressed symbolically as $T_p = T_{p-1} \times (2 + 4(p-1))$.

Based on the above conclusion, the k^{th} term of the derivative of the function g, as in (2.22), evaluated at 0, can be found to be

$$k^{th} \text{ term} = \left(\frac{-1}{2n-1}\right)^k \left\{\prod_{p=0}^{k-1}(2+4p)(n+p)\right\} \qquad \text{for} \quad k \geq 1, \tag{2.25}$$

when the signs and n-terms are also considered. When(2.22) itself, or in other words when the $k = 0^{th}$ derivative of the generator function, is evaluated at 0, we find the value to be equal to 1. Then, as explained in the alternative form of the Maclaurin series in (2.5), we determine the normalizing constant by substituting (2.25) into the formula in (2.12), together with the value of 1 found when $k = 0$:

$$
\begin{aligned}
c_{ct} &= \left(\sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} \left\{ \frac{\frac{\pi}{2}(1-(-1)^{k+1})}{k+1} - \frac{1}{k+1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+1}(u) \mathrm{d}u \right\}\right)^{-1} \\
&= \\
&= \left(\left\{ 1 + \sum_{k=1}^{\infty} \frac{\left(\frac{-1}{2n-1}\right)^k \left\{\prod_{p=0}^{k-1}(2+4p)(n+p)\right\}}{k!} \right\} \left\{ \frac{\frac{\pi}{2}(1-(-1)^{k+1})}{k+1} - \frac{1}{k+1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+1}(u) \mathrm{d}u \right\}\right)^{-1} \\
&= \left(\pi + \sum_{k=1}^{\infty} \frac{\left(\frac{-1}{2n-1}\right)^k \left\{\prod_{p=0}^{k-1}(2+4p)(n+p)\right\}}{k!} \left\{ \frac{\frac{\pi}{2}(1-(-1)^{k+1})}{k+1} - \frac{1}{k+1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+1}(u) \mathrm{d}u \right\}\right)^{-1}
\end{aligned}
\tag{2.26}
$$

under the same conditions described in (2.19) and (2.20).

Numerical outcomes associated with the first k derivatives evaluated at zero were calculated individually for different fixed values of the parameter n at $n = 2, 5, 10$ and 50, with the purpose of calculating the normalizing constant c based on (2.26). This is displayed in Table 2.3.

Table 2.3: Single theoretical outcomes for k derivatives evaluated at zero, calculated for different n-parameters.

k	$n = 2$	$n = 5$	$n = 10$	$n = 50$
1	0.571	0.476	0.451	0.433
2	4.189	2.327	1.915	1.635
3	1.732	0.561	0.375	0.266
4	65.159	11.262	5.791	3.218
5	25.554	2.208	0.837	0.344
6	2866.994	117.983	31.764	9.194
7	919.792	17.349	3.218	0.626
8	248472.829	2082.915	259.276	32.437
9	61635.819	223.897	18.279	1.414
10	35669655.055	54970.255	2885.037	133.022
11	6651702.042	4271.205	141.580	3.758
12	7657085951.878	2017142.135	41569.608	615.199
13	1056147520.196	112615.247	1422.654	11.387
14	2297125785563.401	97975475.119	749141.255	3152.26
15	231953311412.0118	3916025.777	17915.892	38.582
Normalizing constant	$c_{ct} = 0.3304 \times 10^{-20}$	$c_{ct} = 0.9608 \times 10^{-8}$	$c_{ct} = 0.1229 \times 10^{-5}$	$c_{ct} = 0.2496 \times 10^{-3}$

Some form of oscillation between the even and odd values of k was detected, as can be found on closer inspection of Table 2.5. The inverse of the sum of the values in each colomn is represented as the normalizing constant for the chosen values of parameter n. It can be seen that for all chosen values of parameter n, the calculated normalizing constant is almost zero. This is true for all values of n. In light of this discovery based on the theoretical approach through the Maclaurin series expansion, it was thought to alternatively investigate a more robust approach in determining the normalizing constant through numerical integration.

For the numerical integration approach, it is necessary to focus on the original expression of the generator function, $g(\cos(\theta - \mu))$. In this case we have

$$
g(\cos(\theta - \mu)) = \left(1 + \frac{\cos^2(\theta - \mu)}{2n - 1}\right)^{-n}, \tag{2.27}
$$

so that, with (2.27) substituted into the density:

$$1 = \int_0^{2\pi} c.\left(1 + \frac{\cos^2(\theta - \mu)}{2n - 1}\right)^{-n} \mathrm{d}\theta. \tag{2.28}$$

From (2.28) it can be determined that

$$\frac{1}{c} = \int_0^{2\pi} \left(1 + \frac{\cos^2(\theta - \mu)}{2n - 1}\right)^{-n} \mathrm{d}\theta. \tag{2.29}$$

To start the numerical integration, it is important to see how the generator function behaves on the proposed range of 0 to 2π. For the expression of the generator function in (2.27), the size of parameter n is found to be crucial to the shape of the function. For small values of n, the "corners" of the cusps are more rounded than when n becomes large. The y-values obtained in the produced graphs are always positive due the fact that the cosinus function is squared for all x-values between 0 and 2π. It was found that singularities arise at $\frac{\pi}{2} + \mu$ and $\frac{3\pi}{2} + \mu$. When we incorporate this into the numerical integration technique in Mathematica under a local adaptive method where the regions of the function is subdivided until all local errors are sufficiently small, Table 2.4 could be constructed for different values of n and μ towards discovering the value of the normalizing constant, c.

Based on these graphs, it was clear that singularities arise at $\frac{\pi}{2} + \mu$ and $\frac{3\pi}{2} + \mu$. We incorporate this into the numerical integration approach in Mathematica under a local adaptive method where the regions of the function is subdivided until all local errors are sufficiently small. Table 2.4 could be constructed for different values of n and μ towards discovering the value of the normalizing constant, c.

Table 2.4: Finding the normalizing constant, c, through numerical integration.

Number	n	μ	c^{-1}
1	0.5005	0	324.171
2	1	0	3.70433×10^{17}
3	2	0	1.74692×10^{49}
4	5	0	1.26792×10^{151}
5	10	0	7.926131×10^{356}
6	0.5005	π	210.659
7	1	π	3.56985×10^{18}
8	2	π	6.52272×10^{51}
9	5	π	3.35434×10^{150}
10	10	π	2.901272×10^{347}
11	0.5005	$\frac{3\pi}{4}$	6.7897×10^{8}
12	1	$\frac{3\pi}{4}$	1.068×10^{25}
13	2	$\frac{3\pi}{4}$	1.42423×10^{57}
14	5	$\frac{3\pi}{4}$	2.76692×10^{152}
15	10	$\frac{3\pi}{4}$	7.83569×10^{303}

From Table 2.4, it is clear that the values of the normalizing constant, c, once again tend to 0 in all presented cases. When n is very small, we get the best possible results of c, even under a robust approach such as numerical integration.

In light of this observation, another approach had to be considered in finding a meaningful result, based on the work of Siew, Kato and Shimizu [28], where the following density was introduced:

$$f(\theta) = C_\nu \left\{1 + \tanh(\kappa_1\nu)\cos(\theta - \mu) + \tanh(\kappa_2\nu)\cos 2(\theta - \phi)\right\}^{1/\nu} \tag{2.30}$$

where $\nu \neq 0, \kappa_1, \kappa_2 \geq 0, 0 \leq \mu \leq 2\pi$ and $0 \leq \phi < \pi$.

Final model

The proposed generator function is then as follows:

$$g(\cos(\theta - \mu)) = \left(1 + \frac{\kappa}{n}\cos(\theta - \mu)\right)^{-n}, \tag{2.31}$$

so that the appropriate density, as described in (2.1), can be given as

$$f_{tc}(\theta) = c_{tc}\left(1 + \frac{\kappa}{n}\cos(\theta - \mu)\right)^{-n} \tag{2.32}$$

for $0 < \kappa \leq n$ and n a non-negative integer.

The normalizing constant, c_{tc}, can then be found as

$$c_{tc} = \left[\int_0^{2\pi}\left(1 + \frac{\kappa}{n}\cos(\theta - \mu)\right)^{-n} \mathrm{d}\theta\right]^{-1} \tag{2.33}$$

or in terms of equation (2.12), under the conditions described in (2.19) and (2.20), as

$$\begin{aligned} c_{tc} &= \left(\sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!}\left\{\frac{\pi}{2}\frac{(1-(-1)^{k+1})}{k+1} - \frac{1}{k+1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+1}(u)\mathrm{d}u\right\}\right)^{-1} \\ &= \left(\sum_{k=0}^{\infty} \frac{(-1)^k\kappa^k(n+(k-1))!}{n^{k-1}n!k!}\left\{\frac{\pi}{2}\frac{(1-(-1)^{k+1})}{k+1} - \frac{1}{k+1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k+1}(u)\mathrm{d}u\right\}\right)^{-1}, \end{aligned} \tag{2.34}$$

following a similar approach in determining the appropriate equation to substitute into the Maclaurin series, as described in the exploration stage.

Based on the equation in (2.33), the following results for the normalizing constant were obtained in Table 2.5 by numerical integration for various values of the parameters.

Based on Table 2.5 and the normalizing constants that were found through numerical integration, the density plots, based on (2.32), in Figures 2.3, 2.4 and 2.5 were constructed. In Figure 2.3 it can be seen that a change in parameter μ, while keeping the other parameters constant at $n = 100$ and $\kappa = 2$, that the center of the different plots shift as μ, the location parameter, changes. In the circular plot, the density seems to remain symmetric around 0.

In Figure 2.4 the effect of different values of n is considered, while keeping $\kappa = 2$ and $\mu = 0$. Here it can be seen that the bell shape is centered at π and that n influences the peak of where the density curve finally

Table 2.5: Finding the normalizing constant, c, through numerical integration.

n	κ	μ	c_{tc}^{-1}
5	2	0	20.5113
25	2	0	15.1320
50	2	0	14.7111
100	2	0	14.5133
1 000	2	0	14.3417
10 000	2	0	14.3249
100	0.05	0	6.28715
100	0.5	0	6.68637
100	1	0	7.97708
100	2	0	14.5133
100	5	0	190.3580
100	10	0	28761.4
100	2	0	14.5133
100	2	$\pi/4$	14.5133
100	2	$\pi/2$	14.5133
100	2	π	14.5133
100	2	$3\pi/2$	14.5133

settles. Once again, in the circular plot, we see a symmetric shape around the chosen μ.

Lastly, in Figure 2.5, the influence of particular values of the concentration parameter κ is considered. Higher values of κ, while keeping all other parameters constant, lead to more concentrated peaks in the plots than when the value of κ is closer to zero. In the circular plot it can also be seen that some amounts of skewness occur when κ becomes large. For smaller values of κ, the shape is more symmetric around the chosen value of μ.

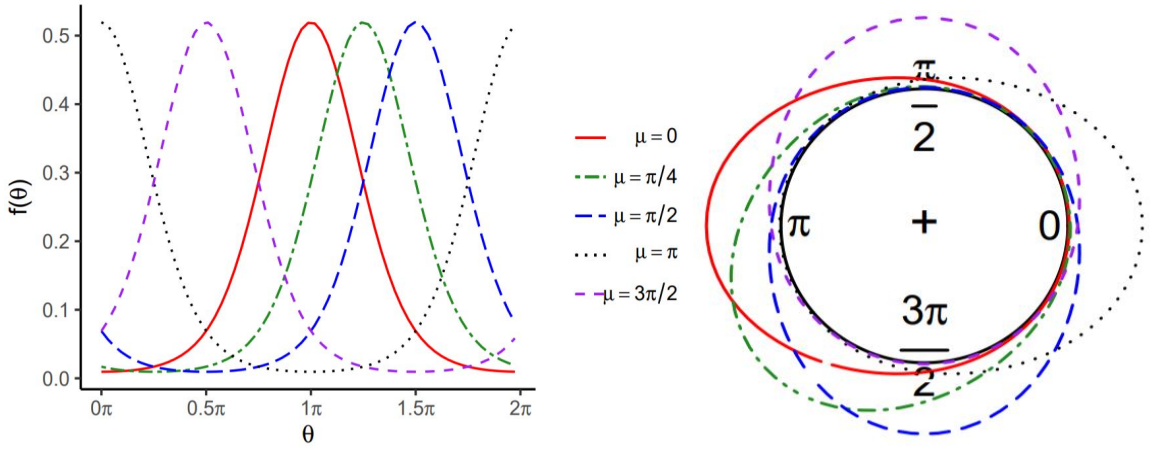


Figure 2.3: Density plots of the cardioid-t distribution for $n = 100$ and $\kappa = 2$, varying over μ.

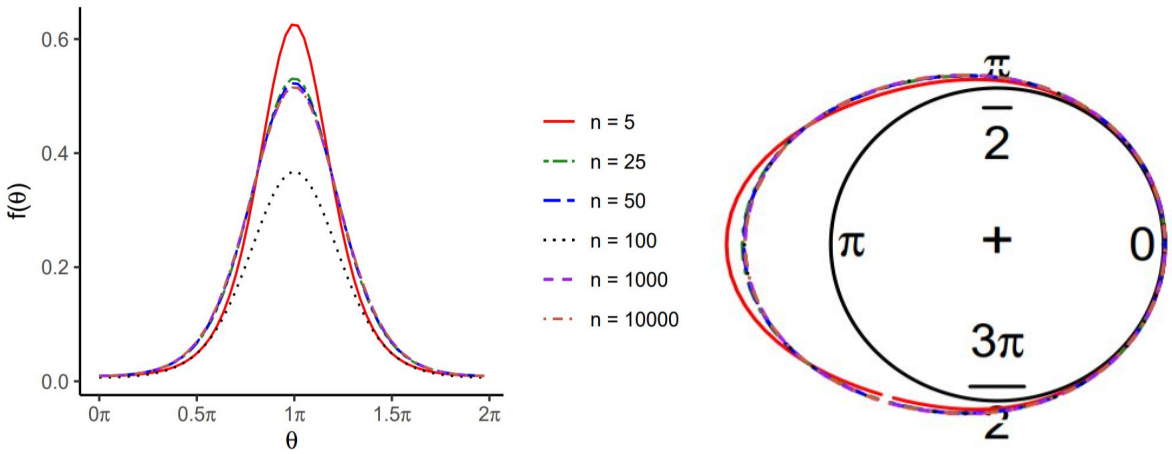


Figure 2.4: Density plots of the cardioid-t distribution for $\kappa = 2$ and $\mu = 0$, varying over n.

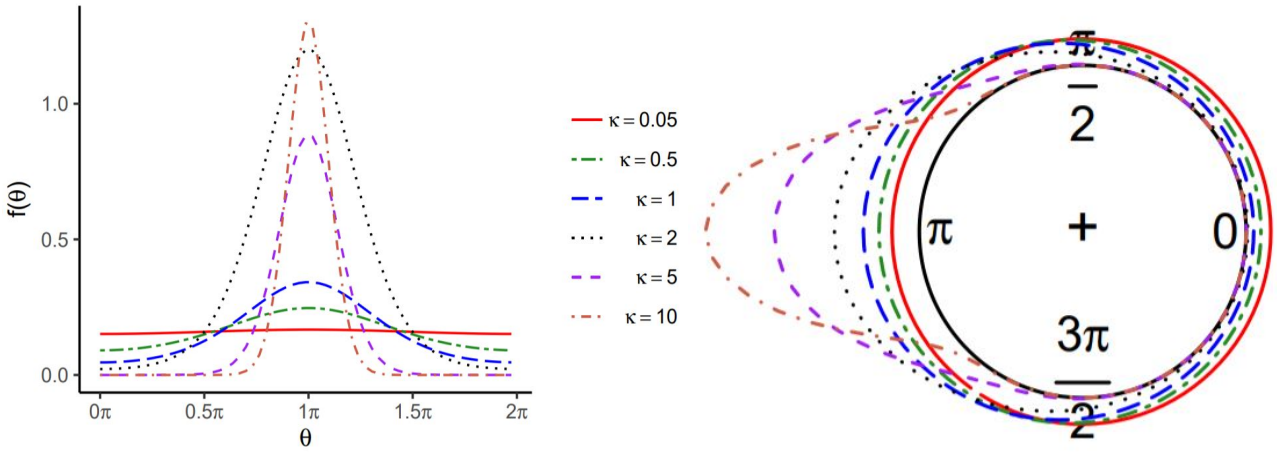


Figure 2.5: Density plots of the cardioid-t distribution for $n = 100$ and $\mu = 0$, varying over κ.

Chapter 3

Maximum likelihood estimation and simulation study

3.1 Maximum likelihood estimation

3.1.1 Maximum likelihood estimation for the cardioid-von Mises distribution

We now consider the maximum likelihood estimation for a vector parameter $\nu = (\alpha, \mu)$ in density (2.16). For $\theta_1, \theta_2, ..., \theta_n$, a random sample of size n from the cardioid-von Mises distribution. the log-likelihood function can then be expressed as

$$\ln L(\theta|\nu) = n \ln(c_{cvm}) + \sum_{i=1}^{n} \ln(1 + 2\alpha \cos(\theta_i - \mu)) + \alpha \sum_{i=1}^{n} \cos(\theta_i - \mu). \tag{3.1}$$

Taking partial derivatives with respect to α and μ and equating them to zero, will lead to finding the likelihood equations:

$$\frac{\partial \ln L(\theta|\nu)}{\partial \alpha} = \sum_{i=1}^{n} \frac{2\cos(\theta_i - \mu)}{1 + 2\alpha \cos(\theta_i - \mu)} + \sum_{i=1}^{n} \cos(\theta_i - \mu) = 0 \tag{3.2}$$

$$\frac{\partial \ln L(\theta|\nu)}{\partial \mu} = \sum_{i=1}^{n} \frac{2\alpha \sin(\theta_i - \mu)}{1 + 2\alpha \cos(\theta_i - \mu)} + \alpha \sum_{i=1}^{n} \sin(\theta_i - \mu) = 0 \tag{3.3}$$

From (3.2) and (3.3) it is possible to obtain the information matrix as the negative expected value of the Hessian matrix with the following elements:

$$l_{\alpha\alpha} = \frac{\partial^2 \ln L(\theta|\nu)}{\partial \alpha \partial \alpha} = \sum_{i=1}^{n} \frac{-4\cos^2(\theta_i - \mu)}{(1 + 2\alpha \cos(\theta_i - \mu))^2} \tag{3.4}$$

$$l_{\alpha\mu} = l_{\mu\alpha} = \frac{\partial^2 \ln L(\theta|\nu)}{\partial\alpha\partial\mu} = \sum_{i=1}^{n} \frac{2\sin(\theta_i - \mu)}{(1 + 2\alpha\cos(\theta_i - \mu))^2} + \sum_{i=1}^{n} \sin(\theta_i - \mu) \tag{3.5}$$

$$l_{\mu\mu} = \frac{\partial^2 \ln L(\theta|\nu)}{\partial\mu\partial\mu} = \sum_{i=1}^{n} \frac{-2\alpha\left(\cos(\theta_i - \mu) + 2\alpha\right)}{(1 + 2\alpha\cos(\theta_i - \mu))^2} - \alpha\sum_{i=1}^{n} \cos(\theta_i - \mu). \tag{3.6}$$

Generally, no closed form expressions exist for the maximum likelihood estimates and it is necessary to make use of nonlinear optimization methods for solving the problems numerically.

3.1.2 Maximum likelihood estimation for the cardioid-t distribution

We now consider the maximum likelihood estimation for a vector parameter $\nu = (n, \kappa, \mu)$ in density (2.32). For $\theta_1, \theta_2, ..., \theta_m$, a random sample of size n from the cardioid-t distribution, the log-likelihood function can be expressed as

$$\ln L(\theta|\nu) = m\ln(c_{ct}) - mn\sum_{i=1}^{m} \ln\left(1 + \frac{k}{n}\cos(\theta_i - \mu)\right). \tag{3.7}$$

Taking partial derivatives with respect to κ and μ and equating them to zero, will lead to finding the likelihood equations:

$$\frac{\partial \ln L(\theta|\nu)}{\partial\kappa} = -m\sum_{i=1}^{m} \frac{\cos(\theta_i - \mu)}{\left(1 + \frac{\kappa}{n}\cos(\theta_i - \mu)\right)} = 0 \tag{3.8}$$

$$\frac{\partial \ln L(\theta|\nu)}{\partial\mu} = -mk\sum_{i=1}^{m} \frac{\sin(\theta_i - \mu)}{\left(1 + \frac{\kappa}{n}\cos(\theta_i - \mu)\right)} = 0 \tag{3.9}$$

From (3.8) and (3.9) it is possible to obtain the information matrix as the negative expected value of the Hessian matrix with the following elements:

$$l_{\kappa\kappa} = \frac{\partial^2 \ln L(\theta|\nu)}{\partial\kappa\partial\kappa} = \frac{m}{n}\sum_{i=1}^{m} \frac{\cos^2(\theta_i - \mu)}{\left(1 + \frac{\kappa}{n}\cos(\theta_i - \mu)\right)^2} \tag{3.10}$$

$$l_{\kappa\mu} = l_{\mu\kappa} = \frac{\partial^2 \ln L(\theta|\nu)}{\partial\kappa\partial\mu} = -m\sum_{i=1}^{m} \frac{\sin(\theta_i - \mu)}{\left(1 + \frac{\kappa}{n}\cos(\theta_i - \mu)\right)^2} \tag{3.11}$$

$$l_{\mu\mu} = \frac{\partial^2 \ln L(\theta|\nu)}{\partial\mu\partial\mu} = mk\sum_{i=1}^{m} \frac{\cos(\theta_i - \mu) + \frac{\kappa}{n}}{\left(1 + \frac{\kappa}{n}\cos(\theta_i - \mu)\right)^2} \tag{3.12}$$

Generally, no closed form expressions exist for the maximum likelihood estimates and it is necessary to make use of nonlinear optimization methods for solving the problems numerically.

3.2 Simulation

In this section two different methods are used for simulating values from the cardioid-von Mises and the cardioid-t distribution [4]. To facilitate the simulations, the R packages $gibbs.met$ and $LearnBayes$ were used and the results were compared. From the $gibbs.met$ package the $gibbs\_met$ function uses Gibbs sampling with a Metropolis update from a Gaussian proposal distribution centerd at the previous state [21]. From the $LearnBayes$ package the $rwmetrop$ function simulates iterates of a random walk Metropolis chain for any real-valued previously defined posterior density [3]. In both the cases of the cardioid-von Mises as well as the cardioid-t distribution, a sample size of $m = 1000$ is generated. The results were captured in Tables 3.1 and 3.2 where the maximum likelihood estimates of the parameters, bias and mean squared error (MSE) of the estimates were found based on the Monte Carlo method where 200 replications for the specified $m = 1000$ were generated. Furthermore, the trace plots and compare-partial plots, where the last 10% of the values in the chain is compared to the whole chain [11], were also generated. In all cases the run time for simulating a single sample of $m = 1000$ for both simulation approaches were also calculated to compare the computational efficiency of the two methods.

3.2.1 Simulation for cardioid-von Mises distribution

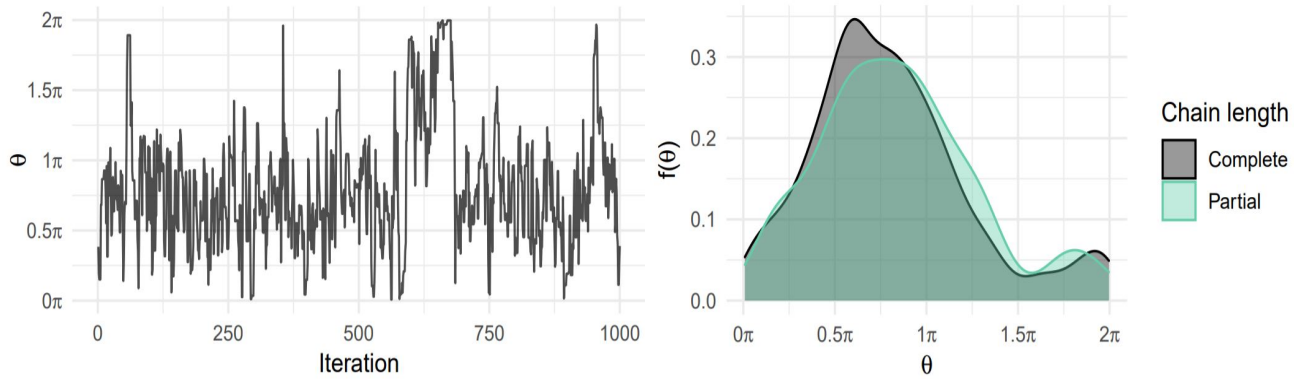
The $rwmetrop$ and $gibbs\_met$ functions in R were used to simulate a sample of $m = 1000$ from the cardioid-von Mises distribution with density as described in (2.16). The parameters were given as $\alpha = 0.4$ and $\mu = 2$. From Table 3.1 it is clear that the $gibbs\_met$ function delivers better results, even though the execution time is somewhat longer. It can be seen from Figure 3.2 and 3.1 that the compare-partial plots of the two approaches seem to fit the generated chains quite well but that the trace plot of the $gibbs\_met$ approach seems more random, as it should be, than that of the $rwmetrop$ approach.

Table 3.1: Simulation results for the cardioid-von Mises distribution.

Method	$\hat{\alpha}$	$\hat{\mu}$	Bias($\hat{\alpha}$)	Bias($\hat{\mu}$)	MSE($\hat{\alpha}$)	MSE($\hat{\mu}$)	Execution time (seconds)
$rwmetrop$	0.4013	2.0049	0.0013	0.0049	0.0017	0.0083	1.36
$gibbs\_met$	0.4008	2.0003	0.0008	0.0003	0.0009	0.0025	2.17

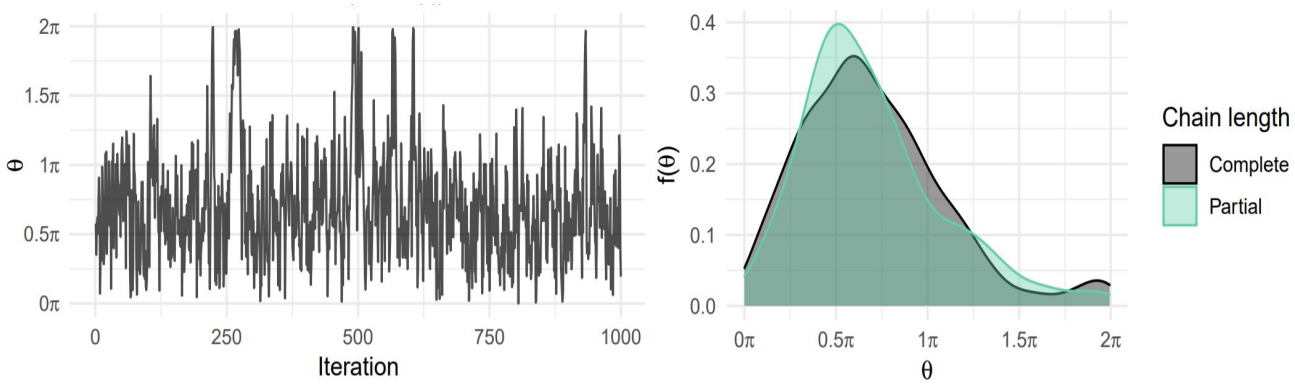
3.2.2 Simulation for cardioid-t distribution

Once again the $rwmetrop$ and $gibbs\_met$ functions in R were used to simulate a sample of $m = 1000$ from the cardioid-t distribution with density as described in (2.32). The parameters were given as $\kappa = 0.5$ and $\mu = 3$, together a pre-specified parameter $n = 1$. From Table 3.2 it can be seen that the $rwmetrop$ function delivers better results overall with a somewhat shorter execution time than that of the $gibbs\_met$ approach. The compare-partial plots of the two approaches do not fit the generated chains as well as one would have



(a) rwmetrop: m=1000 (b) rwmetrop: m=1000

Figure 3.1: Cardioid-von Mises simulation results for method $rwmetrop$.



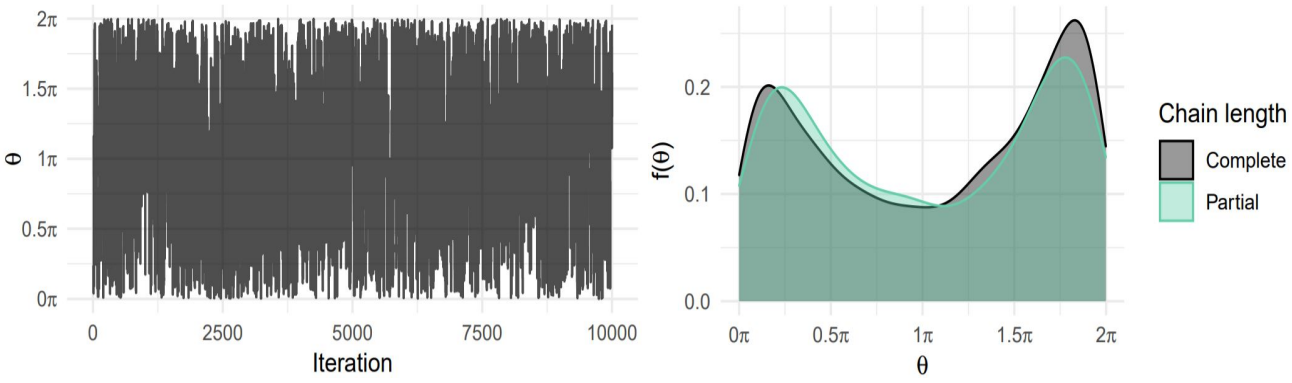
(a) gibbs_met: m=1000 (b) gibbs_met: m=1000

Figure 3.2: Cardioid-von Mises simulation results for method $gibbs\_met$.

hoped when the sample size is small. The results are, however, much better when the sample size is increased, as shown in Figure 3.4 and 3.3 for $m = 10000$. The trace plots of the two approaches seem similar and quite random.

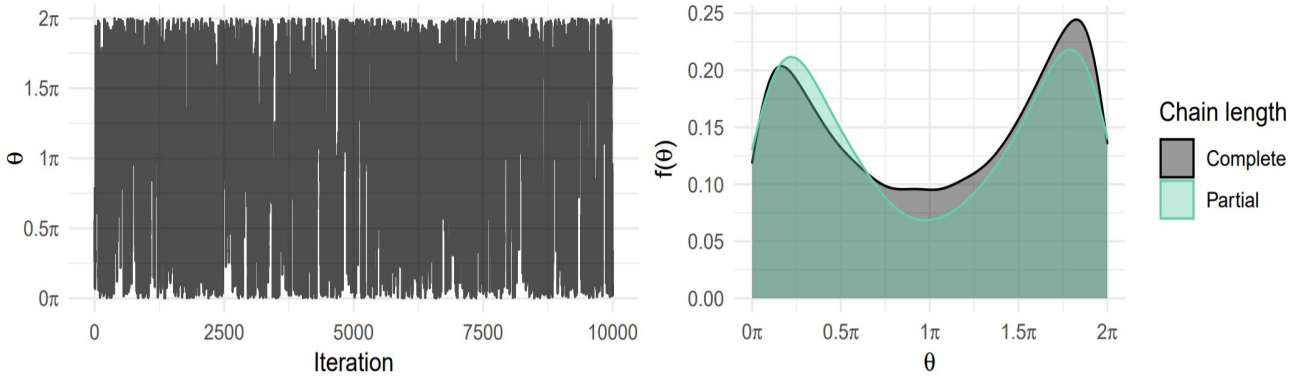
Table 3.2: Simulation results for the cardioid-t distribution.

Method	$\hat{\kappa}$	$\hat{\mu}$	Bias($\hat{\kappa}$)	Bias($\hat{\mu}$)	MSE($\hat{\kappa}$)	MSE($\hat{\mu}$)	Execution time (seconds)
$rwmetrop$	0.5054	2.9944	0.0054	-0.0056	0.0051	0.0478	1.03
$gibbs\_met$	0.5151	3.0172	0.0151	0.0172	0.0033	0.0547	1.76



(a) rwmetrop: m=10 000 (b) rwmetrop: m=10 000

Figure 3.3: Cardioid-t simulation results for method *rwmetrop*.



(a) gibbs_met: m=10 000 (b) gibbs_met: m=10 000

Figure 3.4: Cardioid-t simulation results for method *gibbs_met*.

Chapter 4

Application and performance evaluation

To determine the efficiency of the unifying cardioid generator class, it is important to fit some distributions from this class to real data as an illustrative example of the performance of the models under consideration. In this section we have considered three different data sets: the occasions on which thunder was heard at Kew, the vanishing angles of British mallards and rainfall of 1 inch or more per hour in the United States of America. Five different distributions were fit to the mentioned data sets and the maximum likelihood estimates were captured together with the Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) to evaluate the performance of the different distributions. Density plots were also created for a visual representation.

4.1 Thunder heard at Kew

Any time period can be converted into angles. In this particular case, the length of the time period under consideration is a day. The data set comprising of the number of occurrences on which thunder was heard during each two hourly interval of a day at Kew, England, was captured between the years 1910 and 1935 and each 15° corresponds to an hour [22, 8]. The time-to-angles converted data is shown in Table 6.1 in the Appendix. In Figures 4.1 and 4.2, a histogram with a kernel density of the data is displayed, together with a circular kernel density. From the histogram it seems that the data is unimodal, with a distinct concentration of points between π and $\frac{3\pi}{2}$. This is also prominent in the circular density, with a concentration of data points between the above-mentioned angles. Six different models from the generalized cardioid class were fitted to this data set: the cardioid-t (2.32), cardioid-von Mises (2.16), cardioid (1.3), Von Mises (1.9), wrapped Cauchy (1.12) and generalized t (1.13). The maximum likelihood estimates as well as the Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) was captured in Table 4.1 to evaluate the different model fits. From Table 4.1 it can be seen that the generalized t-distribution with parameters $\hat{\psi} = 0.66$, $\hat{\kappa_1} = 1.39$, $\hat{\kappa_2} = 0.42$, $\hat{\mu} = 3.90$ and $\hat{\nu} = 0.49$ outperforms the other distributions by generating slightly

smaller AIC and BIC values than the other distributions. It can also be seen from Figures 4.1 and 4.2 that the fitted generalized t-distribution is closely related to the kernel density and is hence considered to be the optimal model for the data.

Table 4.1: Maximum likelihood estimation of parameters for different distributions for *The number of occasions on which thunder was heard at Kew in the summers of 1910-1935* data set, Table 6.1.

Distribution	$\hat{\alpha}$	$\hat{\kappa}$	$\hat{\mu}$	$\hat{n}$		AIC	BIC
Cardioid-t (2.32)		1.15	0.68	5		2256.10	2265.28
Cardioid-von Mises (2.16)	0.35		3.90			2281.76	2290.93
Cardioid (1.3)	0.83		3.96			2316.61	2325.78
Von Mises (1.9)		1.20	3.85			2259.39	2268.56
Wrapped Cauchy (1.12)	0.47		3.76			2269.08	2278.25
Distribution	$\hat{\psi}$	$\hat{\kappa_1}$	$\hat{\kappa_2}$	$\hat{\mu}$	$\hat{\nu}$	AIC	BIC
Generalized t (1.13)	**0.66**	**1.39**	**0.42**	**3.90**	**0.49**	**2234.32**	**2243.49**

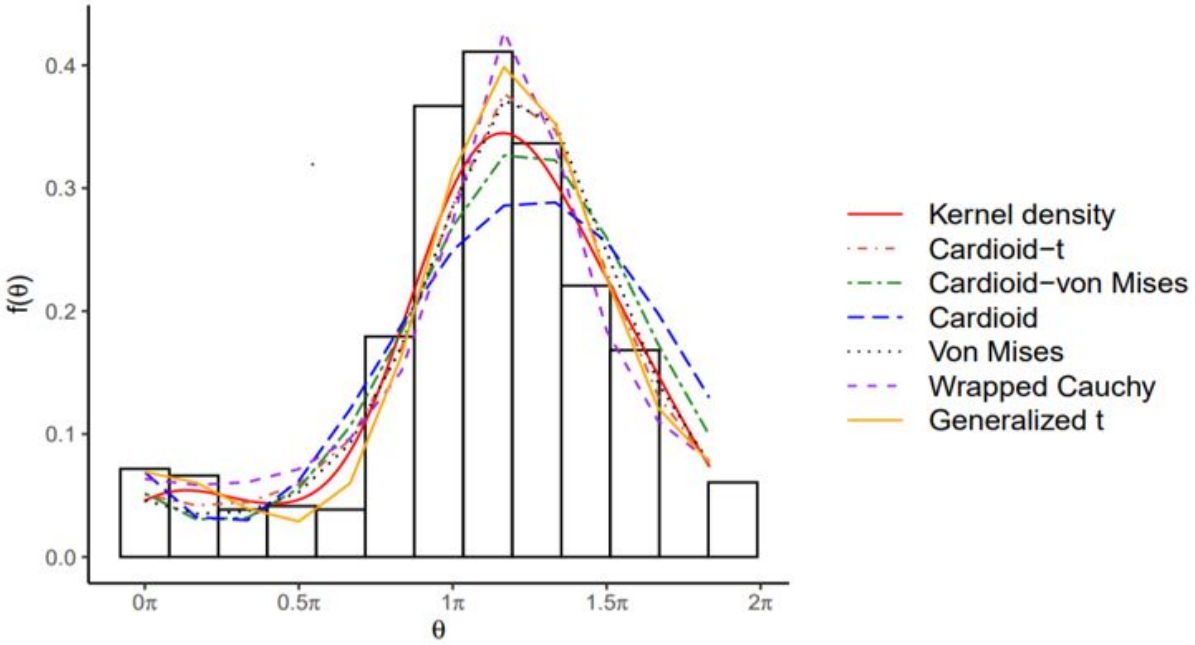


Figure 4.1: Histogram of data and density plot of different models fitted to the data set *The number of occasions on which thunder was heard at Kew in the summers of 1910-1935*, Table 6.1 in the Appendix.

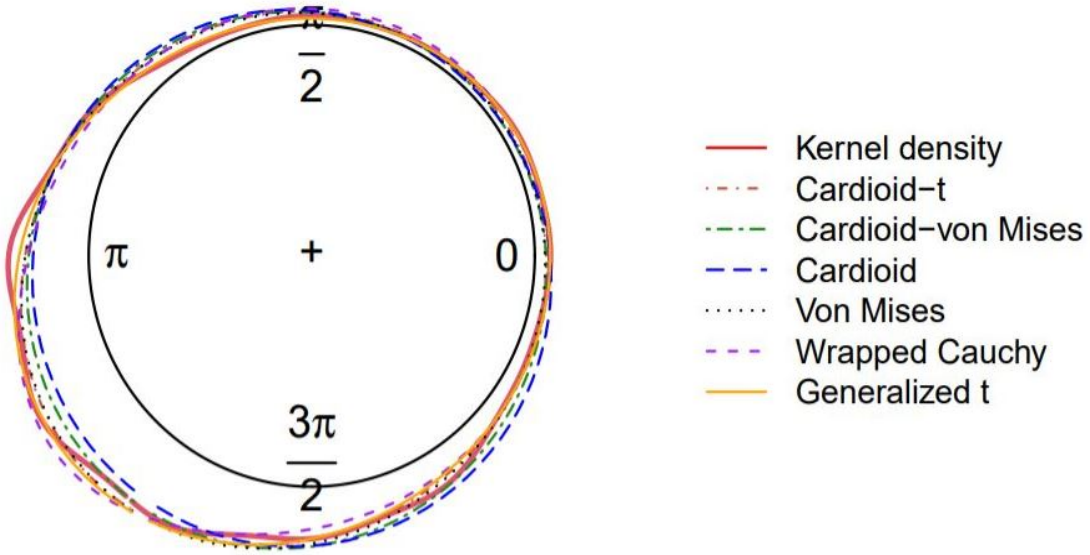


Figure 4.2: Kernel density plot of data and circular density plot of different models fitted to the data set *The number of occasions on which thunder was heard at Kew in the summers of 1910-1935*, Table 6.1 in the Appendix.

4.2 Vanishing angles of British mallards

A grouped data set of size $n = 714$, as captured in Table 6.2 of the Appendix, is considered. This data set shows the frequencies of vanishing angles of non-migratory British mallards, where 0° is considered as the northerly direction [23, 24]. The mallards were crowded out in different directions under clear skies by distances of between 30 km and 250 km from Slimbridge, Gloucestershire over a period of one year [24]. In Figure 4.3, a histogram with a kernel density of the data is displayed, together with a circular kernel density. From the histogram it seems that the data is slightly bimodal, with a distinct peak around $\frac{3\pi}{2}$ and a smaller peak at 0. Six different models from the generalized cardioid class were fitted to this data set: the newly proposed cardioid-t (2.32), cardioid-von Mises (2.16), cardioid (1.3), Von Mises (1.9), wrapped Cauchy (1.12) and generalized t (1.13). The maximum likelihood estimates as well as the AIC and BIC was captured in Table 4.2 to evaluate the different model fits. From Table 4.2 it can be seen that the generalized t-distribution with parameters $\hat{\psi} = 0.01$, $\hat{\kappa_1} = 1.84$, $\hat{\kappa_2} = 0.32$, $\hat{\mu} = 5.38$ and $\hat{\nu} = 1.95$ outperforms the other distributions by generating slightly smaller AIC and BIC values than the other distributions. It can also be seen from Figures 4.3 and 4.4 that the fitted generalized t-distribution is very closely related to the kernel density and is hence considered to be the optimal model for the data.

Table 4.2: Maximum likelihood estimation of parameters for different distributions for the data set *Vanishing angles of 714 British mallards*, Table 6.2.

Distribution	$\hat{\alpha}$	$\hat{\kappa}$	$\hat{\mu}$	$\hat{n}$	AIC	BIC
Cardioid-t (2.32)		1.86	2.16	5	1745.62	1754.76
Cardioid-von Mises (2.16)	0.49		5.33		1851.10	1860.24
Cardioid (1.3)	0.99		5.37		1967.11	1976.25
Von Mises (1.9)		2.12	5.31		1758.92	1768.06
Wrapped Cauchy (1.12)	0.64		5.29		1782.84	1790.98

Distribution	$\hat{\psi}$	$\hat{\kappa_1}$	$\hat{\kappa_2}$	$\hat{\mu}$	$\hat{\nu}$	AIC	BIC
Generalized t (1.13)	**0.01**	**1.84**	**0.32**	**5.38**	**1.95**	**1744.14**	**1753.28**

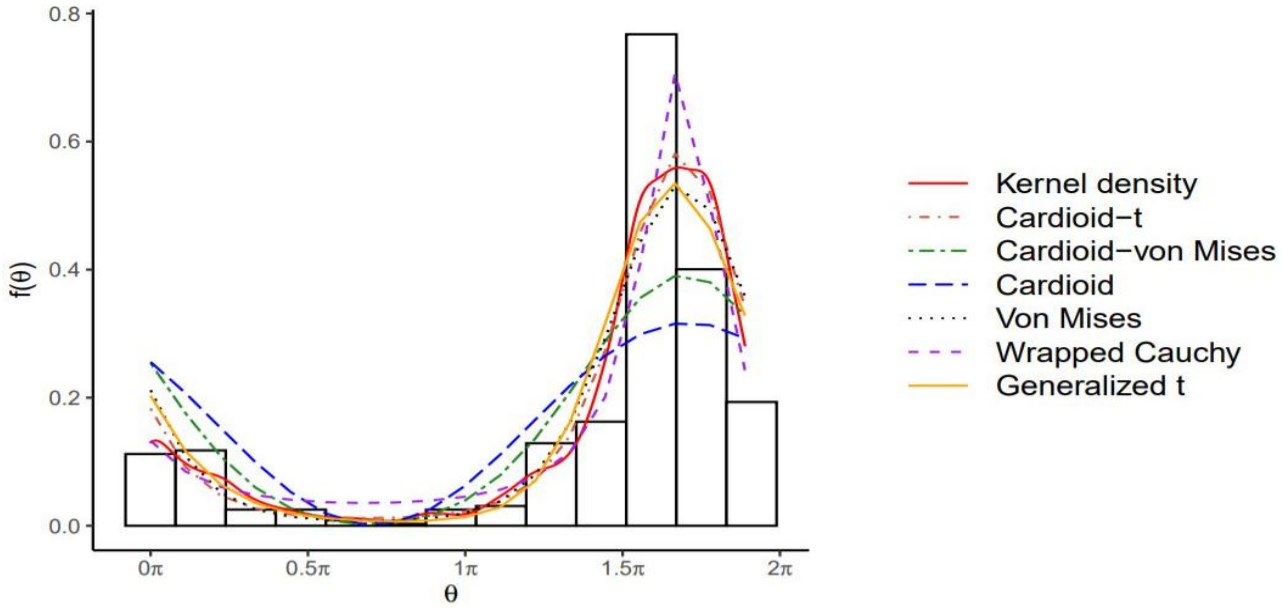


Figure 4.3: Histogram of the data and density plot of different models fitted to the data set *Vanishing angles of 714 British mallards*, Table 6.2 in the Appendix.

4.3 Rainfall in the United States of America

The data set comprising of the number of occurrences on which 1 inch or more rain per hour was recorded in an expansive collection of States in the United States of America between the years 1908 and 1937 [22, 9] is considered [22, 9]. In the case of rainfall measures, the length of the time period under consideration is a year in this particular case. The time-to-angles converted data is shown in Table 6.3 in the Appendix. In Figure 4.5, a histogram with a kernel density of the data is displayed, together with a circular kernel density. From the histogram it seems that the data could be unimodal, but the density plots indicate multimodality,

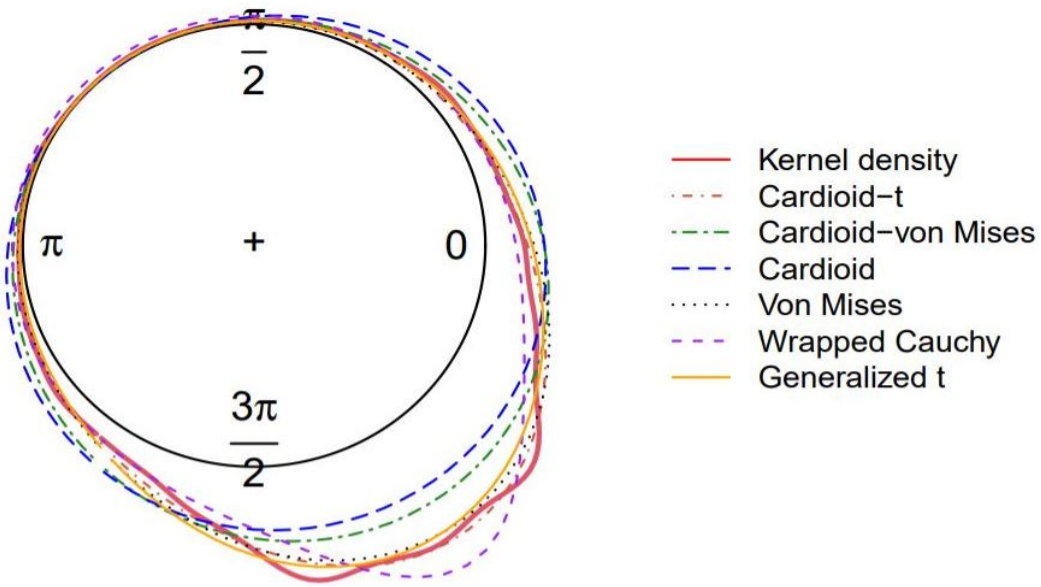


Figure 4.4: Kernel density plot of data and circular density plot of different models fitted to the data set *Vanishing angles of 714 British mallards*, Table 6.2 in the Appendix.

with the most distinct concentration of points around π. Six different models from the generalized cardioid class were fitted to this data set: the newly proposed cardioid-t (2.32), cardioid-von Mises (2.16), cardioid (1.3), Von Mises (1.9), wrapped Cauchy (1.12) and generalized t (1.13). The maximum likelihood estimates as well as the AIC and BIC was captured in Table 4.3 to evaluate the different model fits. From Table 4.3 it can be seen that the generalized t-distribution with parameters $\hat{\psi} = 0.01$, $\hat{\kappa_1} = 1.36$, $\hat{\kappa_2} = 0.07$, $\hat{\mu} = 3.17$ and $\hat{\nu} = 0.89$ outperforms the other distributions by generating slightly smaller AIC and BIC values than the other distributions. It can be seen from Figures 4.5 and 4.6 that the generalized t-distribution is more closely related to the kernel density than the rest and is hence considered to be the best fit between the five models for the data. The underlying multimodality is, however, not addressed by any of the proposed models and hence it is possible that a better model does exist.

Table 4.3: Maximum likelihood estimation of parameters for different distributions for the data set *The number of occurrences of rainfall of 1 inch or more per hour in the U.S.A.*, Table 6.3.

Distribution	$\hat{\alpha}$	$\hat{\kappa}$	$\hat{\mu}$	$\hat{n}$		AIC	BIC
Cardioid-t (2.32)		1.25	0.07	5		21734.20	21747.98
Cardioid-von Mises (2.16)	0.41		3.20			21825.70	21839.48
Cardioid (1.3)	0.92		3.20			22193.04	22206.81
Von Mises (1.9)		1.35	3.20			21711.37	21725.15
Wrapped Cauchy (1.12)	0.4906		3.222			22066.61	22080.38
Distribution	$\hat{\psi}$	$\hat{\kappa_1}$	$\hat{\kappa_2}$	$\hat{\mu}$	$\hat{\nu}$	AIC	BIC
Generalized t (1.13)	**0.01**	**1.36**	**0.07**	**3.17**	**0.89**	**21699.11**	**21712.89**

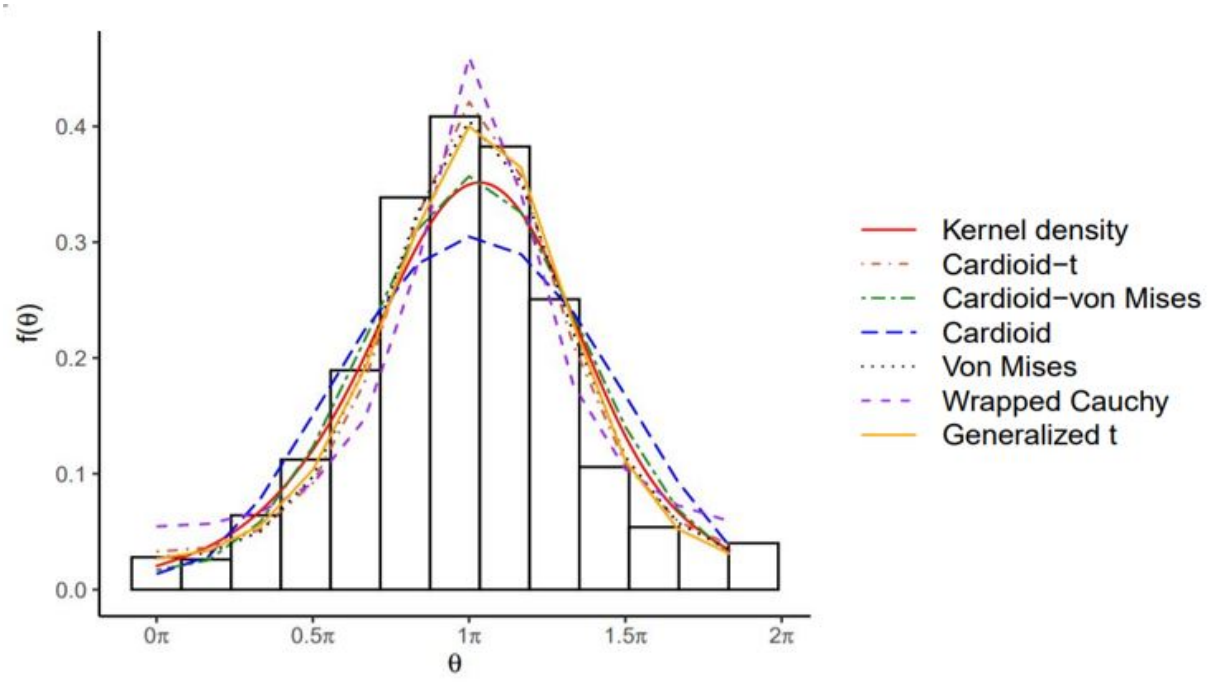


Figure 4.5: Histogram of the data and density plot of different models fitted to the data set *The number of occurrences of rainfall of 1 inch or more per hour in the U.S.A.*, Table 6.3 in the Appendix.

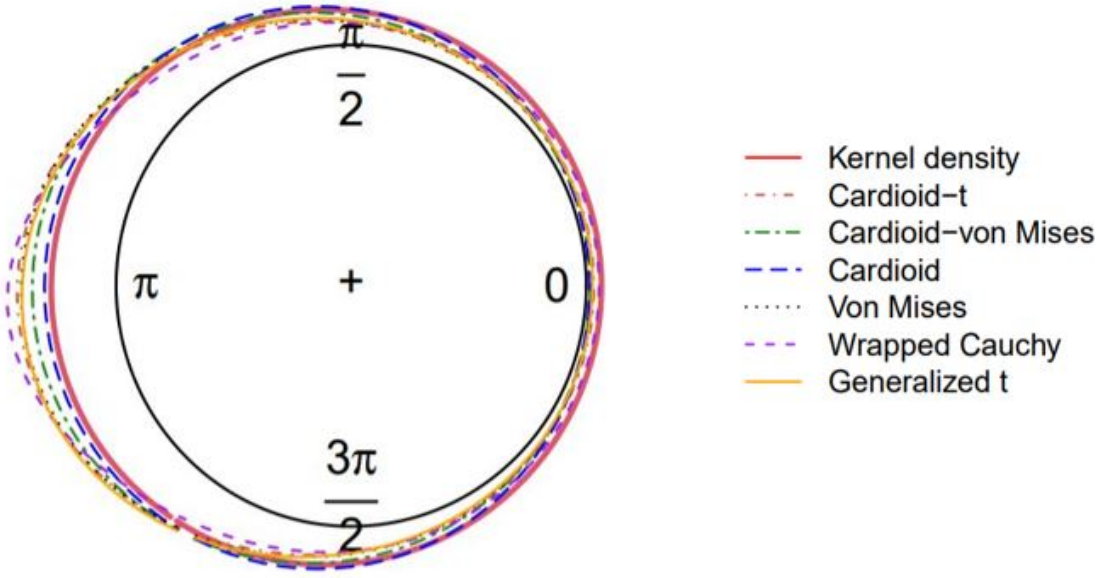


Figure 4.6: Kernel density plot of data and circular density plot of different models fitted to the data set *The number of occurrences of rainfall of 1 inch or more per hour in the U.S.A.*, Table 6.3 in the Appendix.

Chapter 5

Conclusion

Through a thorough understanding of the cardioid and other well-known symmetric circular distributions, as well as existing generalizations, a unifying generator for cardioid modelling in circular statistics have been proposed. Special cases of the proposed generator class, such as the newly introduced cardioid-t as well as the cardioid-von Mises distributions, were structurally investigated through graphic representations of their densities. Maximum likelihood estimation was investigated and simulation studies were conducted to determine the behaviour of some special cases of the proposed generator class. Two MCMC methods were used in the simulation study, but other methods (such as the Markov Chain Monte Carlo Metropolise, Metropolise, Metropolise gaussian and Metropolise-Hastings methods) could possibly lead to different results and insights, which might be worthwhile to consider in future work. In order to evaluate the performance of some of the distributions that form part of the unifying generator class, an application to real data was conducted and the results were compared based on the AIC and BIC values for the different model fits.

It was found that the generator class encompasses many of the existing symmetric circular distributions as described by Mardia [22], Abe et al [1, 2] and Paula et al [25]. The unifying model also leaves room for new distributions, such as the cardioid-t distribution, to be introduced based on the same concepts of the cardioid generator function used in the other instances. The models that were investigated proved te work best when the data is unimodal. The newly introduced cardioid-t distribution proved to be a worthy competitor and outperformed many of the other distributions in the generator class in the three real data environments. The cardioid-t distribution, as a special case of the generalized-t distribution introduced by Siew, Kato and Shimizu [28], was developed with an additional parameter so that the model proved to be more robust, as could be expected. It was, however, found that the generalized t-distribution proved to be the best model in all three sets of real data environments, with slightly smaller AIC and BIC values in all cases. Since the cardioid-t distribution is based on the approach of the generalized t-distribution, but with fewer parameters, this was to be expected. If the generalized-t distribution were to be left out of the study, the cardioid-t would have

been the best performing model in two of the three applications shown here. The cardioid-t distribution is a valuable addition to the class of cardioid generators and proved to be worth investigating on a more extensive level in future works.

One should be cautious in introducing linear models to a circular environment. The initial approach of developing a cardioid-t distribution, with a g-function as described in (2.22), was based on the t-distribution as it is known in the linear form. This proved to be problematic in the development of the normalizing constant, which tended to zero in all the attempted combinations of parameters. Although the newly introduced cardioid-t distribution proved to be a worthy model, it does not address the issue of multi-modality. This aspect could be worth investigating in future studies, perhaps by delving into mixtures of cardioid-t distributions, to further increase the robustness of the generator class and its special cases.

Further theoretical descriptions and additions to the newly proposed models could also form part of future work related to the models that originated in this report.

Bibliography

[1] T. Abe, A. Pewsey, and K. Shimizu. On Papakonstantinou's estension of the cardioid distribution. *Statistics and Probability Letters*, (79):2138-2147, 2009.

[2] T. Abe, A. Pewsey, and K. Shimizu. Extending circular distributions through transformation of argument. *The Institute of Statistical Mathematics*, (65):833-858, 2013.

[3] J. Albert. Functions for learning Bayesian inference, Package 'LearnBayes'. R package version 2.15.1, 2018.

[4] M. Arashi, N.N. Rad, A. Bekker and W.D. Shubert. Protein structure parameterization via Möbius distributions on the torus. *arXiv preprint arXiv:2011.13027*, 2020.

[5] E. Batschelet. *Circular Statistics in Biology*. Academic Press, London, 1981.

[6] M.E. Beck. On the probability of selecting a sample with inclination less than the true mean from a circular distribution: paleomagnetic and tectonic implications. *Tectnophysics*, 269:317-321, 1997.

[7] A. Bekker, P. Nagar, M. Arashi, and H. Rautenbach. From symmetry to asymmetry on the disk manifold: modeling of Marion Island data. *Symmetry*, 8(11):1030, 2019.

[8] B.V. Bishop. The frequency of thunderstorms at Kew observatory. *Meteorological Magazine*, 76:341-353, 1947.

[9] H.D. Dyck and W.A. Mattice. A study of excessive rainfall. *Monthly weather review*, 69(10):293-301, 1941.

[10] J.J. Fernàndez-Duràn. Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics*, 63(2):579-585, 2007.

[11] X. Fernàndez-i-Marin. ggmcmc: Analysis of MCMC samples and Bayesian inference. *Journal of Statistical Software*, 70(9):1-20, 2016.

[12] R. Gatto and S.R. Jammalamadaka. The generalized von Mises distribution. *Statistical Methodology*, 4(3):1572-3127, 2007.

[13] I.S. Gradshteyn and I.M. Rhyzik. *Table of Integrals, Series and Products*. Academic Press, London, Seventh edition, 2007.

[14] M. Hatami and M. H. Alamatsaz. Transformation of circular random variables based on circular distribution functions. *Filomat*, 32(17):5931-5947, 2018.

[15] S.R. Jammalamadaka. Nonparametric methods in directional statistics. *Handbook of Statistics*, 4:755-770, 1984.

[16] S.R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific Publishing Co. Pte. Ltd., 2001.

[17] H. Jeffreys. *Theory of Probability*. Oxford University Press, Third edition, 2003.

[18] R.H. Jones, A. Daniels, and W. Bach. Fitting a circular distribution to a histogram. *Journal of Applied Meteorology*, 15(1):94-98, 1976.

[19] S. Kato and M.C. Jones. An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli*, 1(19):154-171, 2013.

[20] S. Kim, A. SenGupta, and B.C. Arnold. A multivariate circular distribution with applications to the protein structure prediction problem. *Journal of Multivariate Analysis*, 143:374-382, 2016.

[21] L. Longhai. Naive Gibbs sampling with metropolis steps, Package 'gibbs.met'. R package version 1.1-3 .2012.

[22] K.V. Mardia and P.E. Jupp. *Directional Statistics*. John Wiley & Sons Ltd, 2000.

[23] K.V. Mardia. *Statistics of Directional Data*. Academic Press, Inc., 1972.

[24] G.V.T. Matthews. "Nonsense" orientation in mallard anas platyrhychos and its relation to experiments on bird aviation. *IBIS: International journal of avian science*, 103(a):211-230, 1961.

[25] F.V. Paula, A.D.C. Nascimento, and G.J.A Amaral. A new extended cardioid model: an application to wind data. *asXiv: Methodology*, 2017.

[26] F.V. Paula, A.D.C. Nascimento, G.J.A Amaral and G.M. Cordeiro. Generalized cardioid distributions for circular data analysis. *Stats 2021*,4(3):634-649, 2021.

[27] A. Pewsey, M. Neuhäuser, and G.D. Ruxton. *Circular Statisics in R*. Oxford Univerity Press, 2013.

[28] H. Siew, S. Kato, and K. Shimizu. The generalized t-distribution on the circle. *Japanese Applied Statistics*, 37(1):1-16, 2008.

[29] R.L. Tremblay and J.V. Castro. Circular distributions of an epiphytic herb on trees in a subtropical rain forest. *Tropical Ecology*, 50(2):211-217, 2009.

[30] M. Wang and K. Shimizu. On applying Möbius transformations to cardioid random variables. *Statistical Methodology*, 9(6):604-614, 2012.

Chapter 6

Appendix

6.1 Definitions

Definition 6.1.1 (Mean direction)**.** Suppose that the following unit vectors $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n$ has corresponding angles θ_i, $i = 1, ..., n$. The mean direction $\bar{\theta}$ of $\theta_1, ..., \theta_n$ is the direction of the resultant $\mathbf{x}_1 + \mathbf{x}_2 + ... + \mathbf{x}_n$ of $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n$ [22].

Definition 6.1.2 (Mean resultant length)**.** The Cartesian coordinates of unit vectors $\mathbf{x}_j$ corresponding angles θ_j, $j = 1, ..., n$, are given as $(\cos(\theta_j), \sin(\theta_j))$. The centre of mass is given as $\bar{\mathbf{x}}$ for unit vectors $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n$ so that the Cartesian coordinates of die centre of mass is given as

$$\bar{C} = \frac{1}{n}\sum_{j=1}^{n}\cos(\theta_j) \qquad \bar{S} = \frac{1}{n}\sum_{j=1}^{n}\sin(\theta_j)$$

so that $\bar{\theta}$ is the solution of the equations

$$\bar{C} = \bar{R}\cos(\bar{\theta}) \qquad \bar{S} = \bar{R}\sin(\bar{\theta}),$$

where $\bar{R} > 0$ is the mean resultant length [22] given as

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{\frac{1}{2}}. \tag{6.1}$$

Definition 6.1.3 (Circular variance)**.** Circular variance is defined in terms of mean resultant length (6.1) and is given as

$$V = 1 - \bar{R}.$$

Definition 6.1.4 (Bessel function of the first kind and n^{th} order)**.** [13]

$$I_n(z) = \frac{1}{2\pi}\int_0^{\pi} \cos(n\theta - z\sin(\theta))\mathrm{d}\theta.$$

Definition 6.1.5 (Modified Bessel function of the first kind and of order zero)**.** [13]

$$I_0(\rho) = \int_0^{2\pi} \exp\rho\cos(\psi - \mu)d\psi$$

Definition 6.1.6 (Taylor series)**.** The Taylor series of a function is given as

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(\theta - a)^k, \tag{6.2}$$

where $f^{(k)}(a)$ denotes the kth derivative of the function f evaluated at the point a.

Definition 6.1.7 (Euler's formula)**.** [13]

$$\cos(\theta - \mu) = \frac{\exp(i(\theta - \mu)) + \exp(-i(\theta - \mu))}{2}. \tag{6.3}$$

6.2 Theorems

Theorem 1. The trigonometric sum of order n, given by

$$T(\theta) = a_0 + \sum_{k=1}^{n} a_k\cos(k\theta) + b_k\sin(k\theta),$$

is non-negative for every real θ if and only if there exists complex numbers c_k, where $k = 0, 1, ..., n$, such that

$$a_0 = \sum_{k=0}^{n} |c_k|^2 \qquad \text{and} \qquad a_k - ib_k = 2\sum_{v=0}^{n-k} c_{v+k}\bar{c}_v$$

.

6.3 Datasets

Table 6.1: Number of occasions on which thunder was heard at Kew in the summers of 1910-1935 , [23].

G.M.T.	Angle	Frequency
00	0°	26
02	30°	24
04	60°	14
06	90°	15
08	120°	14
10	150°	65
12	180°	133
14	210°	149
16	240°	122
18	270°	80
20	300°	61
22	330°	22

Table 6.2: Vanishing angles of 714 British mallards , [23].

Direction	Number of birds
0°	40
20°	22
40°	20
60°	9
80°	6
100°	3
120°	3
140°	1
160°	6
180°	3
200°	11
220°	22
240°	24
260°	58
280°	136
300°	138
320°	143
340°	69

Table 6.3: Number of occurrences of rainfall of 1 inch or more per hour in the U.S.A., 1908-37, [23].

Month	Angle	Frequency
JAN.	0°	101
FEB.	30°	94
MARCH	60°	232
APRIL	90°	406
MAY	120°	685
JUNE	150°	1225
JULY	180°	1478
AUG.	210°	1384
SEPT.	240°	907
OCT.	270°	383
NOV.	300°	195
DEC.	330°	143

6.4 Code

See the following link to a GitHub repository for relevant sections of code with regard to Chapters 3 and 4 in a ready-to-run format: *https://github.com/u16099754/DMvanWyk_MSc_Code*