QUARTER VON MISES DISTRIBUTION

A Thesis

by

MASON MYERS

BS, Texas A&M University-Corpus Christi, 2021

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

MATHEMATICS

Texas A&M University-Corpus Christi
Corpus Christi, Texas

August 2023

QUARTER VON MISES DISTRIBUTION

A Thesis

by

MASON MYERS

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Jose Guardiola, PhD
Chair

Lei Jin, PhD
Committee Member

Kelum Gajamannage, PhD
Committee Member

Alexey Sadovksi, PhD Graduate
Committee Member

August 2023

ABSTRACT

The world today is increasingly relying on data science and statistics to analyze various types of directional data, such as text data, health studies, image processing, wireless sensor networks, environmental monitoring, robotics, and materials science. In many cases, these data exhibit positive orientation and require probability distributions that are confined to positive regions, such as the positive quarter of the unit circle. These facts highlight the main objective of this thesis, which is to propose a new transformation of the von Mises distribution specifically tailored for the positive quarter of the unit circle. Currently, no such distribution exists. The newly introduced distribution, referred to as the Quarter von Mises Distribution, has been thoroughly investigated in this work. The research includes characterizing the distribution through moments and developing its main properties. Additionally, methods for estimating the distribution parameters using maximum likelihood estimation are presented, along with a hypothesis testing approach using the likelihood ratio test. Furthermore, practical data applications are demonstrated to showcase the effectiveness of these methods. Overall, this thesis contributes to the field of data science and statistics by providing a novel distribution that can accurately model directional data restricted to the positive quarter of the unit circle.

DEDICATION

To the one who gave me the strength and peace to get through it all. My Lord, my God, I praise you and thank you for your many blessings in my life.

ACKNOWLEDGEMENTS

I am extremely grateful to my advisor and chair Dr. Jose Guardiola for his continued support and hours of dedication insuring my success in this endeavour. I am also greatly appreciative of my committee members Dr. Lei Jin and Dr. Kelum Gajamannage for their encouragement, interest and support during this process. Special thanks to my professors Dr. Jose Baca, Dr. Bheemaiah Rao, and Dr. Aubrey Rhoden for the inspiration to pursue this path. Last but certainly not least, thank you to my Mom for being my biggest cheerleader and encouraging me never to give up and to trust in the Lord, to my Dad for coming back to life to enjoy and appreciate these next big moments of life with, to my Maw Maw for always offering to help me in school even if she did not know how, to my aunt Dawn for her dedication to my education and caring for me as a second son, to the man of many titles to me and to whom I look up to greatly, my teacher, coach, pastor, friend, and mentor Chris, to my family for their faith in me to accomplish my dreams, to my Posie for the 18 years of happiness and always waiting for me to come home, and to my Starlight for the love, comfort and motivation she shared with me during the journey. Thank you all for the lasting impact you have made in my life leading up to this moment and certainly beyond.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

The world today is looking towards data science and statistics for identifying trends within topics involving directional data such as text data, health studies, image processing, wireless sensor networks, environmental monitoring, robotics, materials science, and more. This field of statistics is referred to as circular statistics, which is similar to the commonly known regular statistics. Circular statistics parallel the regular theory of statistics including distributions having summary statistics, characterizing functions, moments, inferential methods such as maximum likelihood, and also to perform statistical tests. Examples of circular statistics problems can be found at Fisher (1993). There are some new concepts like dispersion which is equivalent to variance but is a measure of concentration. Probably the most recognized distribution within regular statistics is the normal distribution, and in circular statistics, there is an equivalent distribution called the von Mises distribution. This distribution is essentially the normal distribution wrapped around the unit circle and it is the maximum entropy distribution. Since it also serves as a foundational distribution for circular statistics it has properties that are useful for data analysis. One major feature is that it is able to handle real-world angular data which cannot be analyzed through regular methods of analysis. For instance, some of these types of data are positively oriented and require distributions that are restricted on the positive portions of spaces like the positive quarter of the unit circle, or in the positive orthant of the hypersphere such as the on the one proposed by Guardiola (2020). This brings us to the focus and importance of this thesis. The focus is to develop a new version of the von Mises distribution restricted to the positive quarter of the unit circle called here after the Quarter von Mises Distribution (QVMD). Currently, there does not exists any distribution for the quarter-circle. The new distribution properties will be developed to find the main characteristics of this distribution such as the moments. Inference of the distribution using methods for estimating the concentration and mean direction parameters using maximum likelihood estimation will be included, as well as developing tests to perform hypothesis testing using the likelihood ratio test. Data collected from baseball statistics, which has substantial amounts of zero components when considering a ball's trajectory on the field after a batter hits it, is used alongside some simulated data to assess the modeling capabilities of this proposed distribution. Finally, there will be some discussion concerning future research for this new distribution and a summary of the distribution findings and its potential.

CHAPTER 2: THE QUARTER VON MISES DISTRIBUTION

2.1 Section 1: Definition

Let us begin by transforming the original von Mises Distribution by halving the angle to obtain an axial von Mises Distribution. Below in 2.1, the probability density function for the von Mises Distribution is defined.

$$f(x|\mu,\kappa) = \frac{e^{\kappa cos(x-\mu)}}{2\pi I_0(\kappa)} \text{ for } \mu \in \mathbb{R} \text{ ; } \kappa > 0\text{; } x \text{ is any interval of length } 2\pi \tag{2.1}$$

where

$$I_0(z) = \frac{1}{\pi}\int_0^{\pi} e^{\pm z cos(\theta)}\, d\theta, \tag{2.2}$$

$$I_n(z) = \frac{1}{\pi}\int_0^{\pi} e^{z cos(\theta)}\, cos(n\theta)\, d\theta, \tag{2.3}$$

and x is the random variable, κ is a concentration parameter, μ is the mean direction, and I_o is the modified Bessel function of the first order as seen in 2.2.

The transformation technique for one variable of a continuous distribution described in Miller (2004) is used to transform the direction parameter.

$$\begin{aligned} &\text{If } y = u(x)\text{ , } x = w(y)\text{ , and } u'(x) \neq 0 \\ &\text{then } g(y) = f(w(y))|w'(y)| = f(w(y))\left|\frac{dx}{dy}\right| \end{aligned} \tag{2.4}$$

Starting with (2.1) and using (2.4) to transform the angle x to one-half of its interval length the transformation can be calculated as follows.

$$\text{Let } \gamma = \frac{x}{2} \text{ such that } \gamma \in [0\text{ , }\pi) \text{ when } x \in [0\text{ , }2\pi),$$

taking the derivative for the transformation results in

$$x = 2\gamma\text{, and } dx = 2d\gamma\text{, or } \frac{dx}{d\gamma} = 2,$$

now, inserting this back into the original distribution we have

$$f(\gamma) = \frac{e^{\kappa cos(2(\gamma-\omega))}}{2\pi e^{\kappa} I_0(\kappa)}|2| \text{ for } \kappa > 0\text{ , } \gamma \in [0\text{ , }\pi),$$

which results in

$$f(\gamma) = \frac{e^{\kappa cos(2(\gamma-\omega))}}{\pi e^{\kappa} I_0(\kappa)} \text{ for } \kappa > 0\ ,\ \gamma \in [0\ ,\ \pi). \tag{2.5}$$

Or, using some trigonometric identities to simplify the previous expression,

$$f(\gamma) = \frac{e^{2\kappa cos^2(\gamma-\omega)}}{\pi e^{\kappa} I_0(\kappa)} \text{ for } \kappa > 0\ ,\ \gamma \in [0\ ,\ \pi), \tag{2.6}$$

which is similar to the axial distribution developed by Guardiola et al. (2006). From the axial distribution (2.6), a new distribution is derived by halving the angle parameter interval again

$$\text{Suppose } \delta = \frac{\gamma}{2} \text{ such that } \delta \in [0\ ,\ \pi/2) \text{ when } \gamma \in [0\ ,\ \pi),$$

taking the derivative for the transformation results in

$$d\gamma = 2d\delta, \text{ or } \frac{d\gamma}{d\delta} = 2,$$

now, inserting the derivative into the axial distribution (2.6) we have

$$f(\delta) = \frac{e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} I_0(\kappa)} |2| \text{ for } \kappa > 0\ ,\ \delta \in [0\ ,\ \pi/2),$$

which results in

$$f(\delta) = \frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} I_0(\kappa)} \text{ for } \kappa > 0\ ,\ \delta \in [0\ ,\ \pi/2), \tag{2.7}$$

(2.7) is a new probability density function (PDF) with location parameter ω and dispersion parameter κ that from now on we will refer to it as the Quarter von Mises Distribution or QVMD.

In Figures 2.1 and 2.2, we show several plots of the PDF for the QVMD using different values for the dispersion parameter κ fixed at 2 and the location parameter ω varying with the values shown. Graphs are developed using the library in Pewsey et al. (2013). It is clear that the ω parameter is the mode as it adjusts the location of the maximum and centers the plot at $\pi/4$, which is why this parameter will be referred as the mode or location parameter, which is not necessarily the mean of the distribution except when the distribution is symmetric. In Figures 2.3 and 2.4, the dispersion parameter κ which adjusts the height of the curve is varied while the location parameter ω keeps fixed at $\pi/4$. It is apparent that as κ goes to infinity, the maximum values of the QVMD

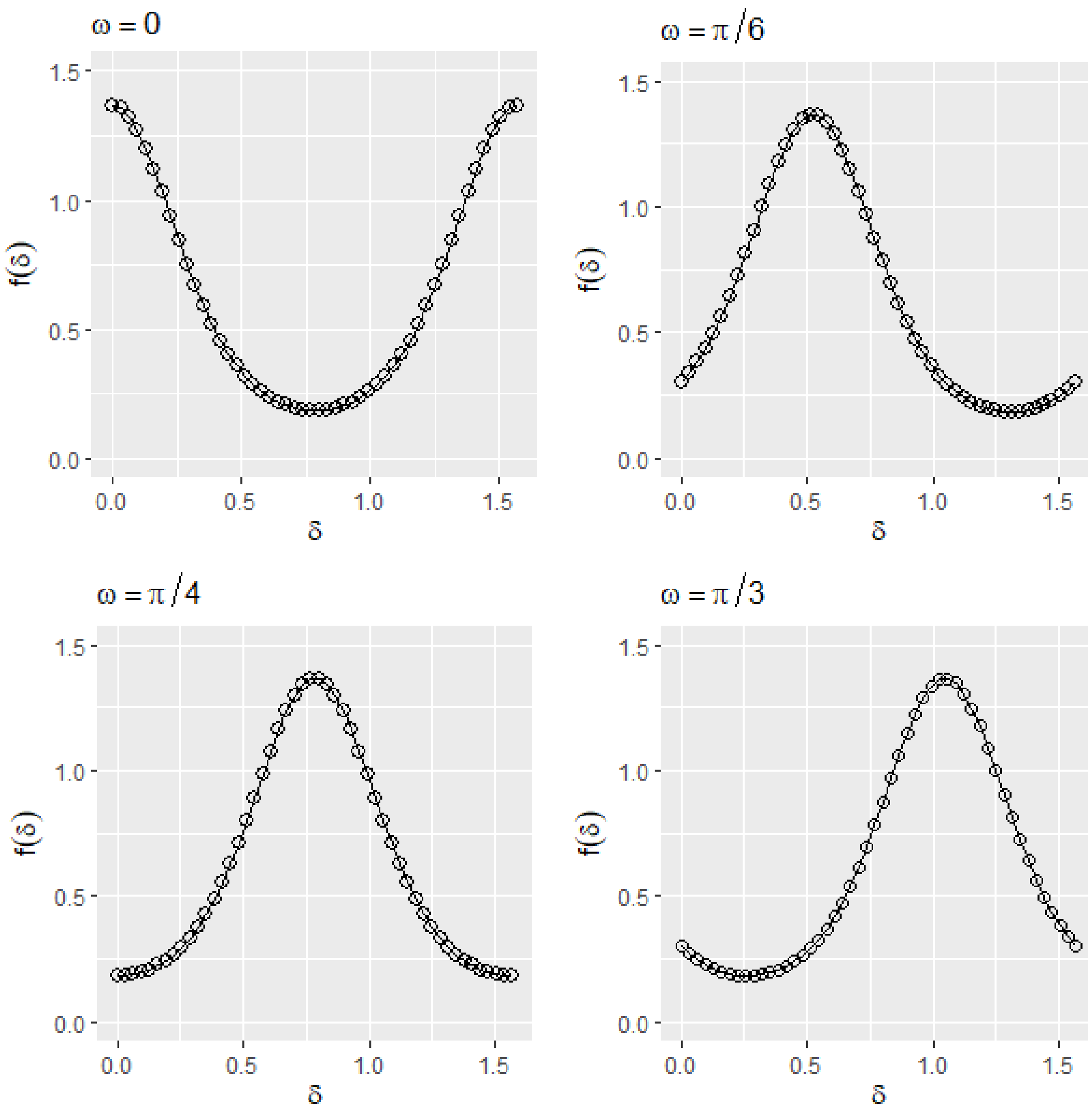

Figure 2.1
QVMD PDF with varying location parameter ω=(0, $\pi/6$, $\pi/4$, $\pi/3$), fixed dispersion parameter κ=1, and 50 equally-spaced directions δ from 0 to $\pi/4$.

rapidly go to infinity as well and the distribution narrows. Another important feature is that when κ equals zero the distribution becomes the uniform distribution as can be seen in the first plot in Figure 2.3.

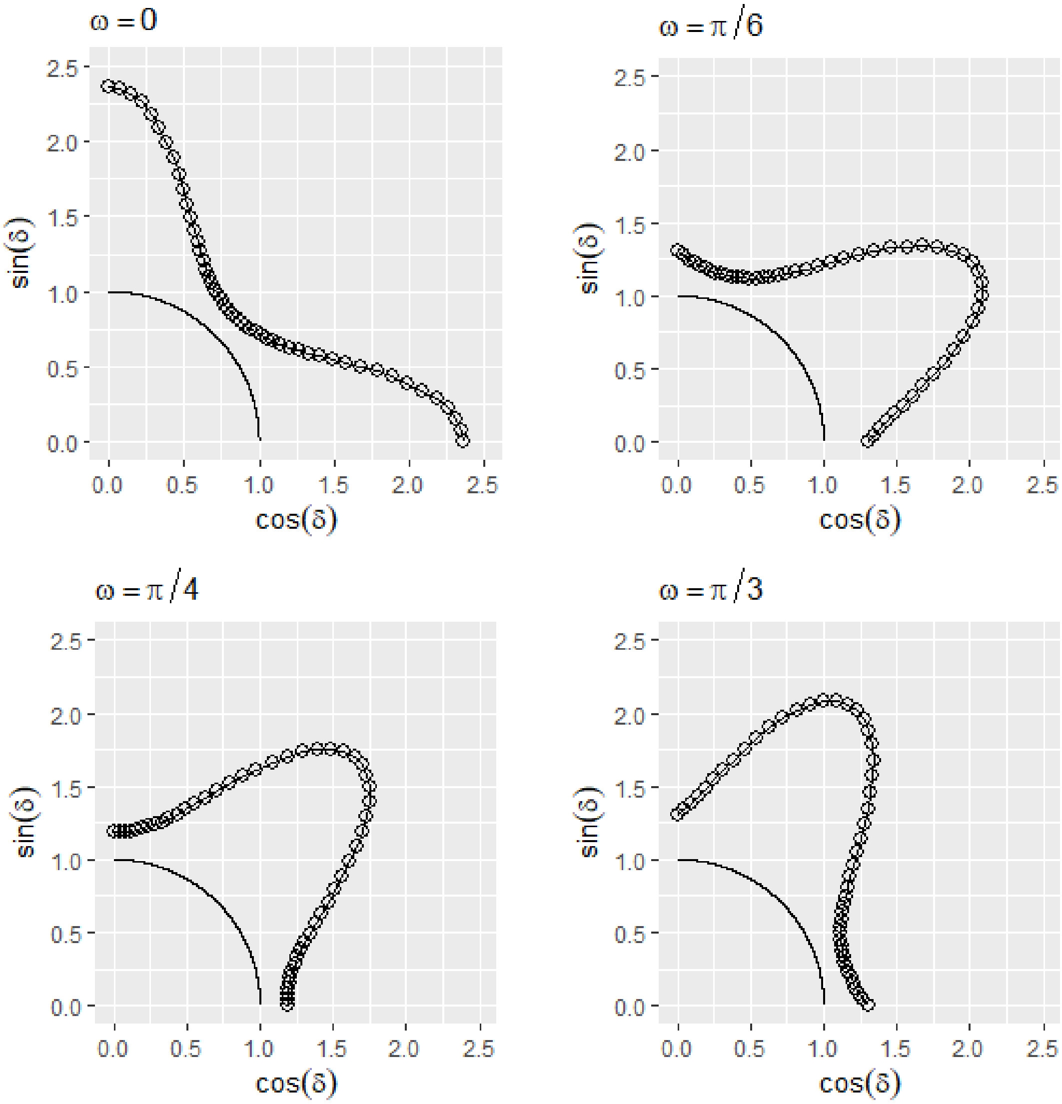

Figure 2.2
QVMD PDF with varying location parameter ω=(0, $\pi/6$, $\pi/4$, $\pi/3$), fixed dispersion parameter κ=1, and 50 equally-spaced directions δ from 0 to $\pi/4$.

2.2 Section 2: Cumulative Distribution Function

There is no explicit cumulative distribution function for this distribution, but it can be numerically computed for any angle as follows.

$$F(\delta) = \int_0^\delta \frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^\kappa I_0(\kappa)} d\delta \text{ for } \kappa > 0 \text{ , } \delta \in [0 \text{ , } \pi/2) \tag{2.8}$$

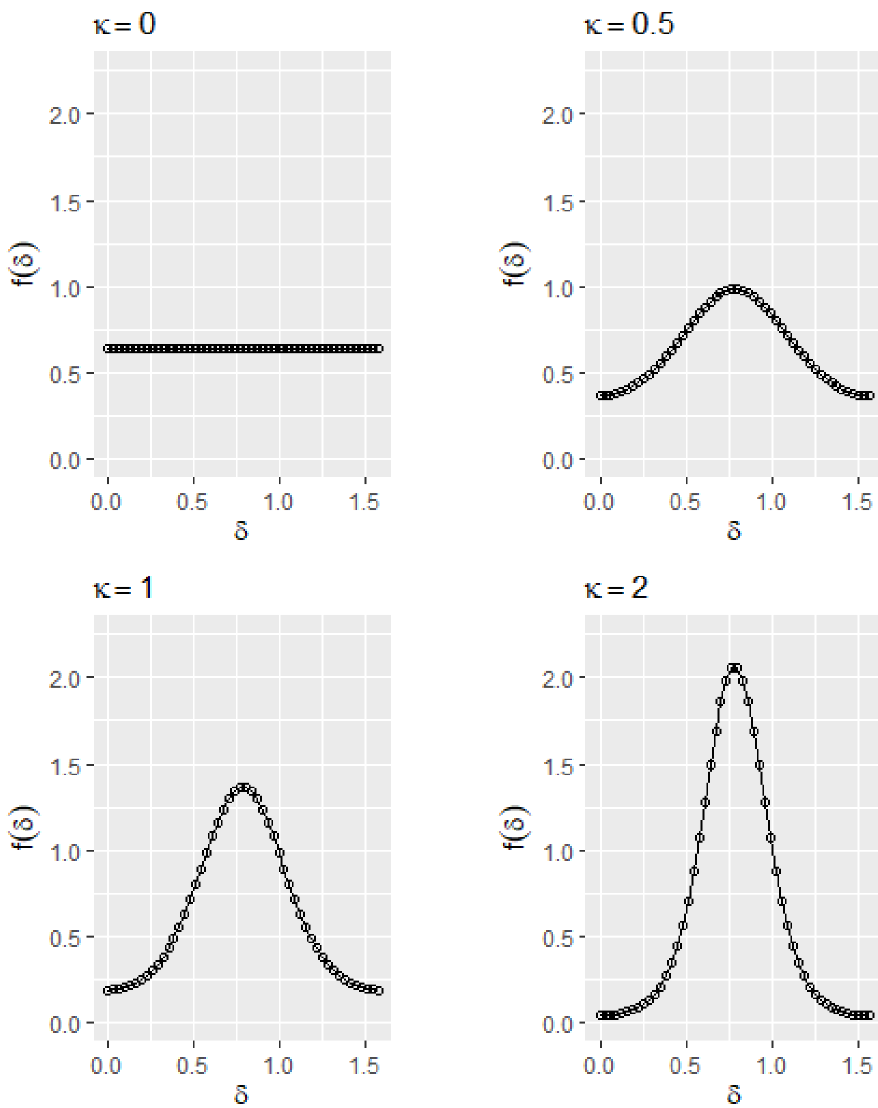

Figure 2.3
QVMD PDF with varying dispersion parameter κ=(0, 0.5, 1, 2), location parameter ω=$\pi/4$, and 50 equally-spaced directions δ from 0 to $\pi/4$.

2.3 Section 3: Trigonometric Moments

Trigonometric moments characterize the distribution as detailed in Mardia & Jupp (2000). In contrast to distributions from regular statistics, any distribution on the circle is determined by its moments. The p^{th} trigonometric moment about the mean direction is

$$\bar{\alpha}_p + i\bar{\beta}_p \tag{2.9}$$

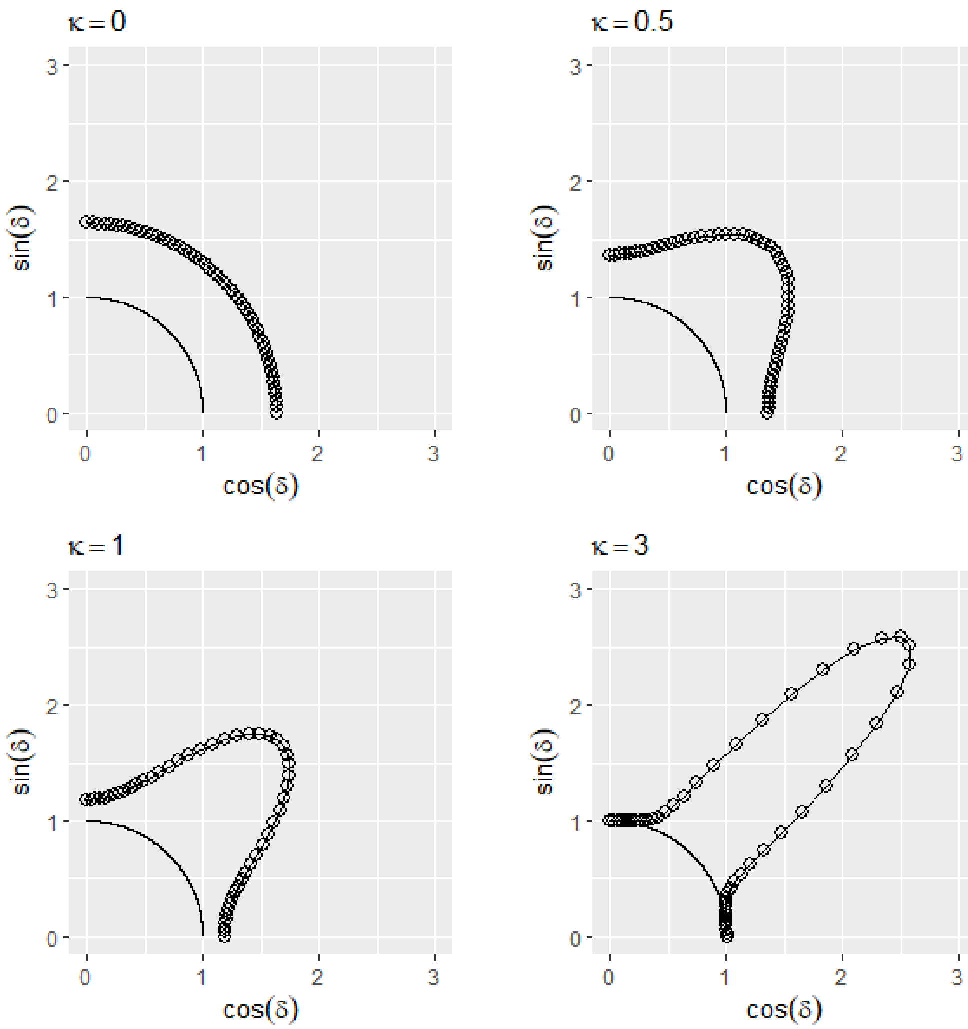

Figure 2.4
QVMD PDF with varying dispersion parameter κ=(0, 0.5, 1, 2), location parameter ω=$\pi/4$, and 50 equally-spaced directions δ from 0 to $\pi/4$.

where each component can be calculated as

$$\bar{\alpha}_p = E[cos(p(\theta-\omega))] \ , \ \bar{\beta}_p = E[sin(p(\theta-\omega))] \tag{2.10}$$

$$\alpha_p = \int_0^{\pi/2} cos(p(\delta-\omega)) \frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} \cdot I_0(\kappa)} d\delta \text{ for } \kappa > 0 \ ; \ \delta \in [0 \ , \ \pi/2) \tag{2.11}$$

$$\beta_p = \int_0^{\pi/2} sin(p(\delta-\omega)) \frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} I_0(\kappa)} d\delta \text{ for } \kappa > 0 \ ; \ \delta \in [0 \ , \ \pi/2) \tag{2.12}$$

2.3.1 Subsection 1: Mean and Variance

Using the trigonometric moment equation, for p=1 the mean resultant length and circular variance can be derived as follows:

$$\alpha_1 = E[cos(\delta)]\ ,\ \beta_1 = E[sin(\delta)] \tag{2.13}$$

$$\alpha_1 = \int_0^{\pi/2} cos(\delta)\frac{2e^{2\kappa cos^2(2\delta)}}{\pi e^{\kappa}\cdot I_0(\kappa)}d\delta \text{ for } \kappa > 0\ ;\ \delta \epsilon\ [0\ ,\ \pi/2) \tag{2.14}$$

$$\beta_1 = \int_0^{\pi/2} sin(\delta)\frac{2e^{2\kappa cos^2(2\delta)}}{\pi e^{\kappa} I_0(\kappa)}d\delta \text{ for } \kappa > 0\ ;\ \delta \epsilon\ [0\ ,\ \pi/2) \tag{2.15}$$

No explicit solution can be found for (2.14) and (2.15), but they can be computed numerically. The mean resultant length ρ is calculated using these two moment components and determines how tightly clustered (ρ closer to 1) or dispersed (ρ closer to 0) the directions are within a data set, which is why ρ is also considered a concentration parameter.

$$\rho = \sqrt{\alpha_1^2 + \beta_1^2} \tag{2.16}$$

$$0 \le \rho \le 1 \tag{2.17}$$

The circular variance can be calculated using ρ, but regardless of the similar name to the homonymous term in regular statistics, it is a different parameter with unique characteristics.
The circular variance defined as

$$v = 1 - \rho, \tag{2.18}$$

measures the dispersion from the mean direction of the data set.

2.3.2 Subsection 2: Skewness and Kurtosis

Asymmetry of the distribution can be measured by the skewness s as follows

$$s = \frac{\bar{\beta_2}}{(1-\rho)^{3/2}}, \tag{2.19}$$

where

$$\bar{\beta_2} = \int_0^{\pi/2} sin(2(\delta-\omega))\frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} I_0(\kappa)}d\delta. \tag{2.20}$$

Peakedness of the distribution can be measured by kurtosis K as follows

$$K = \frac{\bar{\alpha}_2 - \rho^4}{(1-\rho)^2}, \tag{2.21}$$

where

$$\bar{\alpha_2} = \int_0^{\pi/2} cos(2(\delta-\omega)) \frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} \cdot I_0(\kappa)} d\delta. \tag{2.22}$$

There is no explicit solution but both previous expressions (2.19) and (2.21) can be computed numerically.

2.4 Section 4: Examples

Now that the characterization of the distribution is complete, it is time to see how it can be used in practice. For this purpose, we will discuss one example from a simulated data set and another example using a data set from a real application.

2.4.1 Subsection 1: Simulated Data

The simulated data was created from randomly generated angles from the QVMD using κ=3 and ω=π/3. A plot of the data on the unit circle can be seen in Figure 2.5. Utilizing the expressions above, we compute the first moment components for the simulated data using (2.14) for α_1 and (2.15) for β_1, we get α_1=0.4967 and β_1=0.8515. Using (2.16) we obtain ρ=0.9857, and using (2.18) the circular variance is v=0.0143. Also, using (2.19) a skewness s=-3.5031 which is in agreement with a left-skewed distribution, and finally using (2.21) we obtain the kurtosis K=5.4447. The results are shown in Table 2.1

Table 2.1
Simulated data summary statistics

α_1	β_1	ρ	v	s	K
0.4967	0.8515	0.9857	0.0143	-3.5031	5.4447

2.4.2 Subsection 2: Real Data

Data from the site Savant (2023), which compiles and analyzes data from Major League Baseball, is utilized for the real data section. This data is from 2023 and contains the direction at which the ball travel vertically after being hit by the batter as can be seen in Figure 2.6 and how

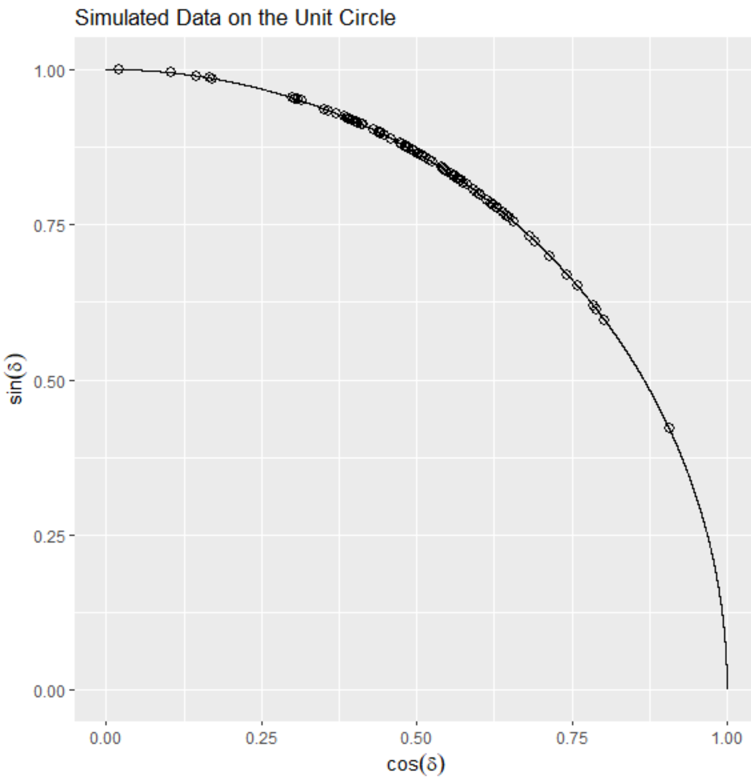

Figure 2.5
Simulated data generated from the QVMD function made in R using a location parameter $\omega = \pi/3$ and dispersion parameter $\kappa = 3$ has its direction values δ plotted on the unit circle. Most of the directions appear to be centered around the location parameter $\omega = \pi/3$.

many total hits are at each angle. The data ranges from negative ninety degrees to positive ninety degrees. For the purpose of this thesis, the data has been reduced down from zero to ninety degrees as the number of negative values is relatively small. In Figure 2.7 the directional data is displayed on the unit circle.

The MLE method shown in the next chapter was used to estimate κ. The location parameter represented by ω of the distribution will be estimated. Utilizing the expressions above (2.14) and (2.15), the simulated data has first moment components α_1=0.9042 and β_1=0.3328, using (2.16) ρ=0.9567, using (2.18) a circular variance of 0.0365 is obtained, using (2.19) a skewness of 8.5076 which is in line with how the distribution has a right-sided tail, and finally using (2.21) a kurtosis value of 15.2136 is obtained. Results are shown in Table 2.2.

Table 2.2
Real data summary statistics

α_1	β_1	ρ	v	s	K
0.9042	0.3328	0.9635	0.0365	8.5076	15.2136

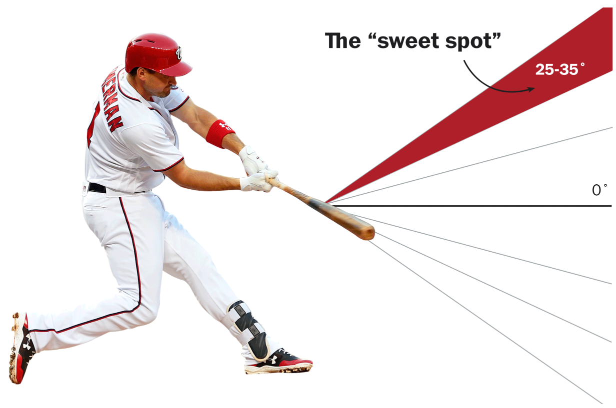

Figure 2.6
An example of how the launch angle of a baseball upon impact with a bat can be modeled in two dimensions with positive directional data. The "sweet spot" is likely where the directions of the ball are concentrated indicating the location parameter value would be around $\pi/6$.

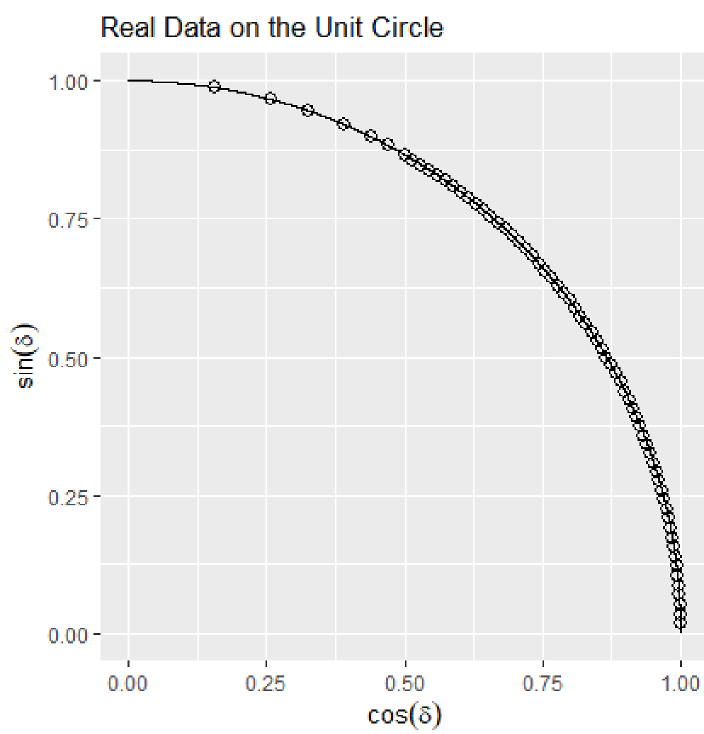

Figure 2.7
Vertical trajectories δ of a baseball after being hit are plotted on the unit circle. Most of the directions appear to be spread between 0 and $\pi/3$, which includes the "sweet spot" of around $\pi/6$.

CHAPTER 3: INFERENCE

3.1 Section 1: Method of Moments

Method of Moments could be considered the simplest yet sometimes least accurate method of estimation for parameters depending on the distribution it is working with. The idea is to consider the moments of the distribution as a sort of building blocks to find estimates for parameters. For instance, the first moment can give the estimate for the mean of a distribution. The second moment would give an estimate of the variance utilizing the previous estimate for the mean. This can continue to build up to find each parameter you need to estimate. The following is how the methods could be utilized for estimating the parameters ω and κ.

$$\frac{1}{n}\sum_{i}^{n} cos(\delta_j) = E\left[cos(\delta)\right] = \int_0^{\pi/2} cos(\delta)\frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} I_0(\kappa)}d\delta \tag{3.1}$$

$$\frac{1}{n}\sum_{i}^{n} sin(\delta_j) = E\left[sin(\delta)\right] = \int_0^{\pi/2} sin(\delta)\frac{2e^{2\kappa cos^2(2(\delta-\omega))}}{\pi e^{\kappa} I_0(\kappa)}d\delta \tag{3.2}$$

The previous expressions need to be simultaneously solved to obtain the MOM estimators of ω and κ. It is not a practical method in this case and will not be attempted as the next section provides better estimators.

3.2 Section 2: Maximum Likelihood Estimation

Maximum likelihood estimation finds the estimated values of the parameters that maximize the likelihood of observing the current data set. In this case, there are two parameters to estimate, the location ω and the dispersion parameter κ. Suppose that we have data that after visual inspection is determined that it can be fitted using the QVMD. The maximum likelihood estimation method provides an efficient method for fitting data sets to the QVMD. First, we need to write the likelihood function as described in Casella & Berger (2002).

Likelihood function:

$$\text{Let } L(\theta|x) = f(x|\theta), \text{ where } L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta), \tag{3.3}$$

the likelihood function for the sample assuming the data is distributed as a QVMD can be expressed as follows

$$L(\omega,\kappa|\delta)=\prod_{i=1}^{n} f(\delta_i|\omega,\kappa)=\prod_{i=1}^{n}\frac{2e^{2\kappa cos^2(2(\delta_i-\omega))}}{\pi e^{\kappa}I_0(\kappa)}, \tag{3.4}$$

as it is customary, it is easier to work with the Log-Likelihood function

$$Ln(L(\delta))=Ln\left[\frac{2^n}{\pi^n e^{n\kappa}(I_0(\kappa))^n}\prod_{i=1}^{n} e^{2\kappa cos^2(2(\delta_i-\omega))}\right] \tag{3.5}$$

$$=nLn(2)-nLn(\pi)-n\kappa-nLn(I_0(\kappa))+\sum_{i}^{n}2\kappa cos^2(2(\delta_i-\omega)).$$

Eliminating unnecessary constants that do not influence the optimization process we get

$$LnL(\delta)=-n\kappa-nLn(I_0(\kappa))+\sum_{i}^{n}2\kappa cos^2(2(\delta_i-\omega)).$$

Maximizing the previous expression for finding the MLE for ω and κ is equivalent to minimizing the negative of the log-likelihood function as follows

$$-LnL(\delta)=n\kappa+nLn(I_0(\kappa))-\sum_{i}^{n}2\kappa cos^2(2(\delta_i-\omega)). \tag{3.6}$$

In the following sections we will be using (3.6) to perform the inference for both the simulated data and real data.

3.2.1 Subsection 1: Simulated Data

We start with the simulated data to perform the estimation procedure. Let us set up the values ω and κ as $\pi/3$ and 3 respectively. The expectation is that the MLE procedure will find values that are close enough to the true values. Plotting the negative log-likelihood function for the simulated data that can be seen in Figure 3.1, it appears visually that the minimum value is located around 1 for ω (represented here as lmv) and around 3 for κ (represented here as kv in the plot). Initial values of ω=2 and κ=2 are used to start with the optimization process in MATLAB. The resulting MLE values that are found for ω and κ are 1.0185 and 2.7094 respectively, which both are close enough to their true parameter values of $\pi/3$ and 3. It can be seen that the MLE method works reasonably well for estimating the parameters of the QVMD. Plots of the QVMD with the estimated parameters can be seen in Figures 3.2 and 3.3.

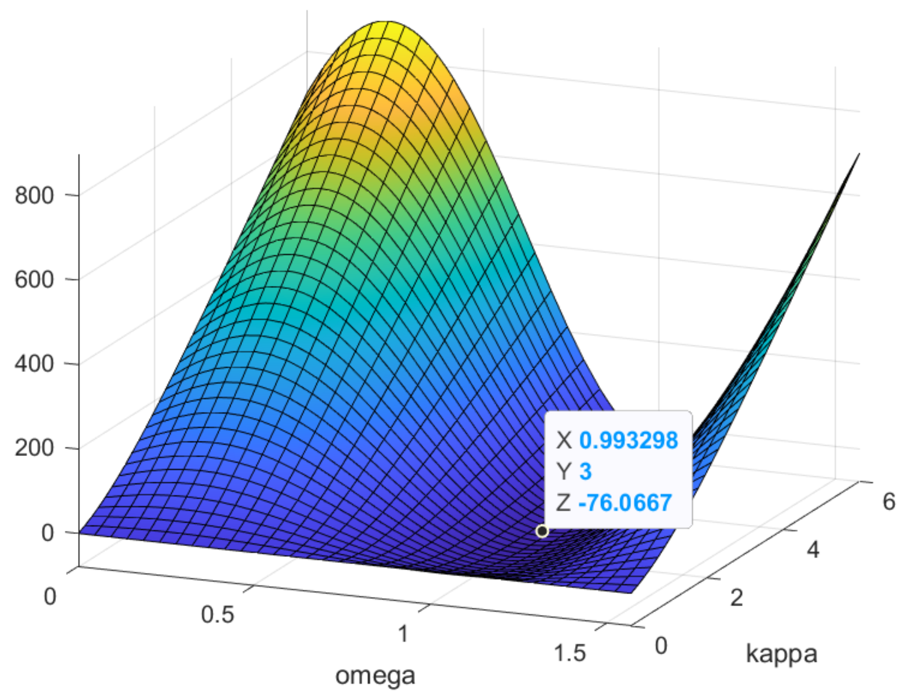

Figure 3.1
The simulated data is inserted into the negative log-likelihood function from before and plotted. The minimum of the function appears to be at approximately $\omega = 1$ and $\kappa = 3$.

Table 3.1
True vs estimated parameter values for the simulated data

True ω	Est. ω	% Error	True κ	Est. κ	% Error
1.0472	1.0186	2.73%	3	2.7094	9.69%

3.2.2 Subsection 2: Real Data

Now, the parameters for the real data will be estimated using the MLE procedure described before. Plotting the negative log-likelihood function we may visually locate approximately where the minimum point occurs in Figure 3.2 and how it matches our estimation. The minimum can be seen approximately around 0.3 for the ω value, and about 3 for the κ value. Using the same optimization method in MATLAB and starting with an initial guess of ω=0.5 and κ=2, the maximum likelihood estimators for the parameters are 0.295 for ω and 2.77 for κ. Plots of the QVMD using these results is shown in Figures 3.5 and 3.6.

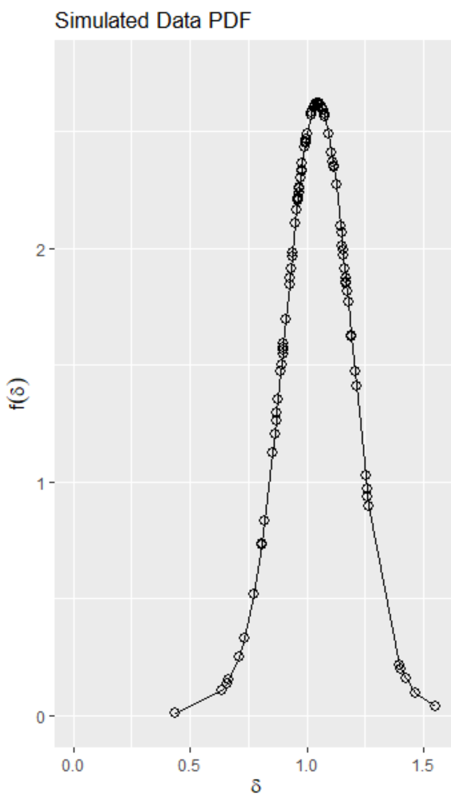

Figure 3.2
QVMD PDF plot for the simulated data using the estimated parameters $\omega = 1.0185$ and $\kappa = 2.7094$.

Table 3.2
Estimated parameter values for the real data

Est. ω	Est. κ
0.2950	2.77

3.3 Section 3: Likelihood Ratio Test

From Hogg R. (2019), the Likelihood Ratio Test (LRT) is defined as

$$\lambda(x) = \frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)} \tag{3.7}$$

where $\hat{\theta}$ is from the unrestricted maximization of the likelihood function and $\hat{\theta}_0$ is from the restricted maximization of the likelihood function on a subset of the parameter space. This ratio is usually difficult to assess as the exact distribution of the LRT is unknown, and in some cases it can be approximated using the asymptotic behavior suggested in Casella & Berger (2002). This

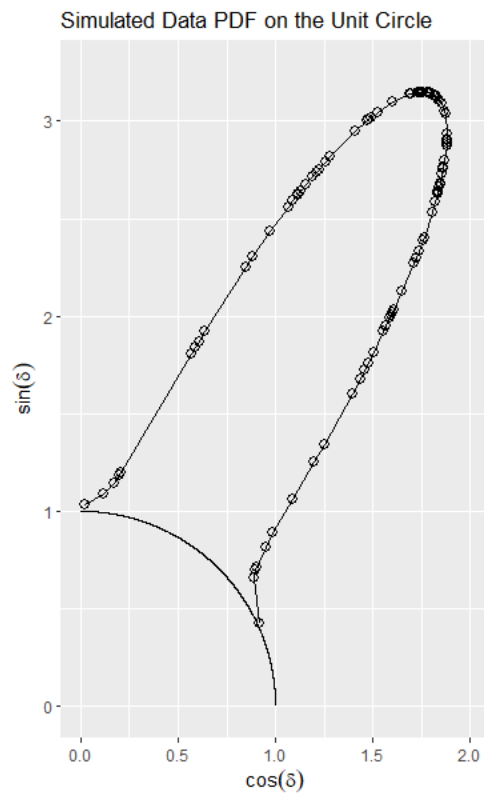

Figure 3.3
QVMD PDF plot on the unit circle for the simulated data using the estimated parameters $\omega = 1.0185$ and $\kappa = 2.7094$.

approximation seen in (3.8), requires the LRT to be under certain regularity conditions so that it can be approximately distributed as a Chi-Square distribution with ν degrees of freedom (the number of parameters to be estimated). In the three upcoming tests there is only one parameter that is being estimated, so the LRT approximation is to a chi-square distribution with one degree of freedom.

$$-2Ln(\lambda(x)) = -2Ln\left[\frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)}\right] \sim \chi^2_\nu \tag{3.8}$$

3.3.1 Subsection 1: Hypothesis test for the location parameter

We want to test the location parameter using the LRT asymptotic approximation to the chi-square which can be seen in (3.8)

$$\begin{aligned} &H_0 : \omega = \omega_0 \text{ (null hypothesis)} \\ &H_1 : \omega \neq \omega_0 \text{ or } \omega = \hat{\omega} \text{ (alternative hypothesis)} \end{aligned} \tag{3.9}$$

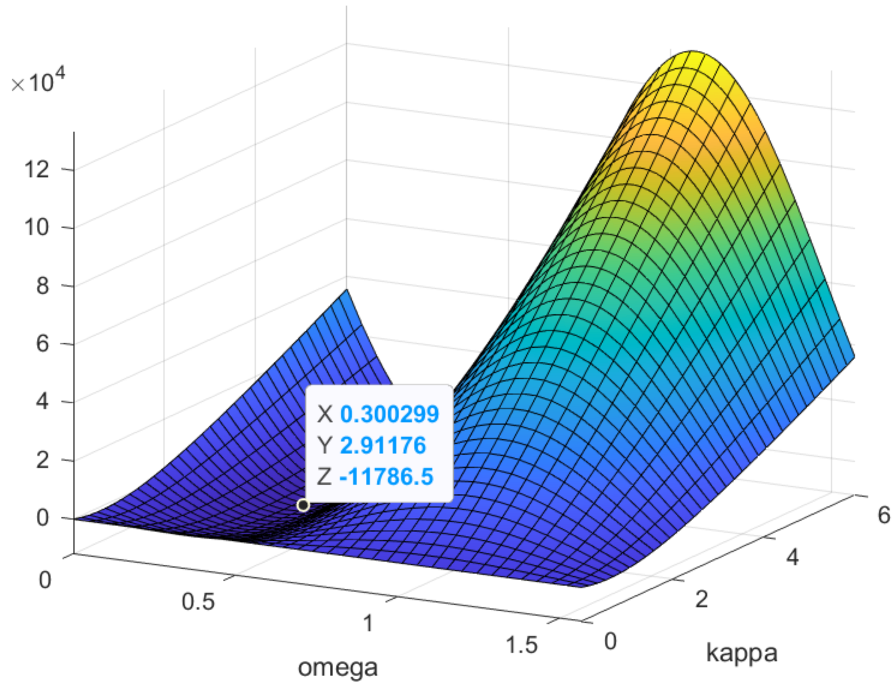

Figure 3.4
The real data is inserted into the negative log-likelihood function from before and plotted. The minimum of the function appears to be at approximately $\omega = 0.3$ and $\kappa = 3$.

$$\lambda(\delta) = \frac{L(\omega_0, \kappa|\delta_i)}{L(\hat{\omega}, \kappa|\delta_i)} \tag{3.10}$$

$$= \frac{\prod_{i=1}^{n} \frac{2e^{2\kappa cos^2(2(\delta_i - \omega_0))}}{\pi e^{\kappa} I_0(\kappa)}}{\prod_{i=1}^{n} \frac{2e^{2\kappa cos^2(2(\delta_i - \hat{\omega}))}}{\pi e^{\kappa} I_0(\kappa)}}$$

$$= \frac{\prod_{i=1}^{n} e^{2\kappa cos^2(2(\delta_i - \omega_0))}}{\prod_{i=1}^{n} e^{2\kappa cos^2(2(\delta_i - \hat{\omega}))}}$$

$$-2Ln(\lambda(\delta)) = -2Ln\left[\frac{\prod_{i=1}^{n} e^{2\kappa cos^2(2(\delta_i - \omega_0))}}{\prod_{i=1}^{n} e^{2\kappa cos^2(2(\delta_i - \hat{\omega}))}}\right]$$

$$-2Ln(\lambda(\delta)) = 4\kappa\left(\sum_{i=1}^{n} cos^2(2(\delta_i - \hat{\omega})) - \sum_{i=1}^{n} cos^2(2(\delta_i - \omega_0))\right) \tag{3.11}$$

After some trials, we determine that the regularity conditions are not necessarily met and the chi-square approximation does not hold for extreme values, but it can still perform well under certain conditions.

Figure 3.5
QVMD PDF plot for the real data using the estimated parameters $\omega = 0.2950$ and $\kappa = 2.77$.

3.3.2 Subsection 2: Hypothesis test for the dispersion parameter

We want to test the dispersion κ parameter using the theory for the LRT and the corresponding chi-square approximation (3.8)

$$\begin{aligned} H_0 &: \kappa = \kappa_0 \text{ (null hypothesis)} \\ H_1 &: \kappa \neq \kappa_0 \text{ or } (\kappa = \hat{\kappa}) \text{ (alternative hypothesis)} \end{aligned} \tag{3.12}$$

$$\lambda(\delta) = \frac{L(\omega, \kappa_0 | \delta_i)}{L(\omega, \hat{\kappa} | \delta_i)} \tag{3.13}$$

$$= \frac{\prod_{i=1}^{n} \frac{2e^{2\kappa_0 cos^2(2(\delta_i - \omega))}}{\pi e^{\kappa_0} I_0(\kappa_0)}}{\prod_{i=1}^{n} \frac{2e^{2\hat{\kappa} cos^2(2(\delta_i - \omega))}}{\pi e^{\hat{\kappa}} I_0(\hat{\kappa})}}$$

$$= \frac{\frac{e^{2\kappa_0 \sum_{i=1}^{n} cos^2(2(\delta_i - \omega))}}{e^{\kappa_0} I_0(\kappa_0)}}{\frac{e^{2\hat{\kappa} \sum_{i=1}^{n} cos^2(2(\delta_i - \omega))}}{e^{\hat{\kappa}} I_0(\hat{\kappa})}}$$

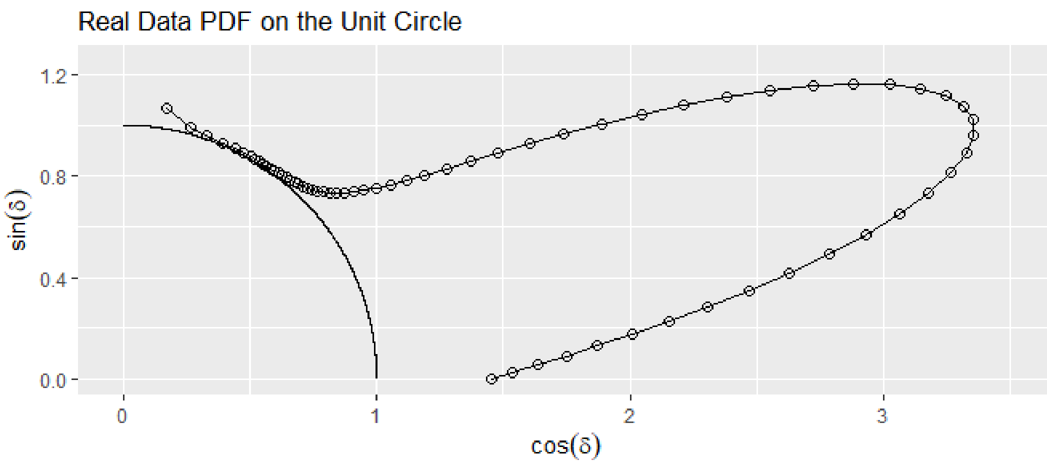

Figure 3.6
QVMD PDF plot on the unit circle for the real data using the estimated parameters $\omega = 1.0185$ and $\kappa = 2.7094$.

$$-2Ln(\lambda) = 2(\kappa_0 + \hat{\kappa})\left(n - 2\sum_{i=1}^{n} cos^2(2(\delta_i - \omega))\right) + 2nLn(\frac{(I_0(\kappa_0))}{(I_0(\hat{\kappa})}) \sim \chi_1^2 \tag{3.14}$$

that can be computed numerically.

3.3.3 Subsection 3: Uniformity test

This is a particular case of the hypothesis testing for the dispersion parameter κ when κ=0, in this case the QVMD is uniform. We can develop a hypothesis test for uniformity.

$$\begin{aligned} &H_0 : \kappa = 0 \text{ (null hypothesis, the distribution is uniform)} \\ &H_1 : \kappa \neq 0 \text{ or } (\kappa = \hat{\kappa}) \text{ (alternative hypothesis, the distribution is not uniform)} \end{aligned} \tag{3.15}$$

$$\begin{aligned} \lambda(\kappa|\delta) &= \frac{\prod_{i=1}^{n} \frac{2e^{2(0)cos^2(2\delta_i)}}{\pi e^0 I_0(0)}}{\prod_{i=1}^{n} \frac{2e^{2\hat{\kappa}cos^2(2\delta_i)}}{\pi e^{\hat{\kappa}} I_0(\hat{\kappa})}} \\ &= \frac{1}{e^{n\hat{\kappa}}(I_0(\hat{\kappa}))^n} \prod_{i=1}^{n} e^{2\hat{\kappa}cos^2(2\delta_i)} \end{aligned}$$

now using (3.8)

$$-2Ln(\lambda(x)) = -2n\hat{\kappa} - 2nLn(I_0(\hat{\kappa})) + 4\sum_{1}^{n} \hat{\kappa}cos(\delta_i)^2 \sim \chi_1^2, \tag{3.16}$$

again, the previous expression can be tested numerically.

3.3.4 Subsection 4: Examples

Each of the previous hypothesis tests (3.11), (3.14), and (3.16) will now be applied to the data. The statistical conclusions for these tests will be drawn based on the following decision rules. Sample size n affects the strength of our results, if a sample size is large enough the resulting test statistic will be large and thus the conclusions drawn from the hypothesis test will be strong. Positive results greater than the one degree-of-freedom chi-square critical value of 3.841 indicate that there is sufficient evidence to reject the alternative hypothesis. Results less than the critical value including negative values indicate that there is not sufficient evidence to reject the null hypothesis.

3.3.4.1 Simulated Data

For the first test, the true location parameter value of ω=π/3 is tested against a second value of ω=π/4 to determine if the null hypothesis can be rejected suggesting that the true location parameter value is not π/3.

$$\begin{aligned} &H_0 : \omega = \pi/3 \text{ (null hypothesis)} \\ &H_1 : \omega \neq \pi/3 \ (\omega = \pi/4) \text{ (alternative hypothesis)} \end{aligned} \tag{3.17}$$

$$-2Ln(\lambda(\delta)) = -187.42 < 3.841 \tag{3.18}$$

The LRT approximation was less than the critical value, so we fail to reject the null hypothesis. There is not enough evidence to support the claim that the location parameter ω is different from π/3 (true value) which is the correct conclusion.

For the second test, the true dispersion parameter value of κ=3 is tested against a second value of κ=2 to determine if the null hypothesis can be rejected suggesting that the true dispersion parameter value is not 3.

$$\begin{aligned} &H_0 : \kappa = 3 \text{ (null hypothesis)} \\ &H_1 : \kappa \neq 3 \ (\kappa = 2) \text{ (alternative hypothesis)} \end{aligned} \tag{3.19}$$

$$-2Ln(\lambda(\delta)) = -3.90 < 3.841 \tag{3.20}$$

The LRT approximation was less than the critical value, so we fail to reject the null hypothesis. There is not enough evidence to support the claim that the dispersion parameter κ is different from 3 (true value) which is the correct conclusion.

For the third test, the dispersion parameter value of κ=0 is tested against a second value of κ=3 to determine if the null hypothesis can be rejected suggesting that the simulated data is not distributed uniformly.

$$\begin{aligned} &H_0 : \kappa = 0 \text{ (null hypothesis)} \\ &H_1 : \kappa \neq 0 \ (\kappa = 3) \text{ (alternative hypothesis)} \end{aligned} \tag{3.21}$$

$$-2Ln(\lambda(\delta)) = 151.41 > 3.841 \tag{3.22}$$

The LRT approximation was greater than the critical value, so we reject the null hypothesis. There is enough evidence to support the claim that the dispersion parameter κ is different from 0 which suggests the simulated data is not distributed uniformly which is the correct conclusion.

3.3.4.2 Real Data

For the first test, the estimated location parameter value of ω=0.3 is tested against a second value of ω=1 to determine if the null hypothesis can be rejected suggesting that the estimated location parameter value is not 0.3.

$$\begin{aligned} &H_0 : \omega = 0.3 \text{ (null hypothesis)} \\ &H_1 : \omega \neq 0.3 \ (\omega = 1) \text{ (alternative hypothesis)} \end{aligned} \tag{3.23}$$

$$-2Ln(\lambda(\delta)) = -127,450 < 3.841 \tag{3.24}$$

The LRT approximation was less than the critical value, so we fail to reject the null hypothesis. There is not enough evidence to support the claim that the location parameter ω is different from 0.3 (estimated value) which is the correct conclusion.

For the second test, the estimated dispersion parameter value of κ=2.77 is tested against a second value of κ=2 to determine if the null hypothesis can be rejected suggesting that the estimated dispersion parameter value is not 2.77.

$$\begin{aligned} &H_0 : \kappa = 2.77 \text{ (null hypothesis)} \\ &H_1 : \kappa \neq 0 \ (\kappa = 2) \text{ (alternative hypothesis)} \end{aligned} \tag{3.25}$$

$$-2Ln(\lambda(\delta)) = -969.04 < 3.841 \tag{3.26}$$

The LRT approximation was less than the critical value, so we fail to reject the null hypothesis. There is not enough evidence to support the claim that the dispersion parameter κ is different from 2.77 (estimated value) which is the correct conclusion.

For the third test, the dispersion parameter value of κ=0 is tested against a second value of κ=3 to determine if the null hypothesis can be rejected suggesting that the real data is not distributed uniformly.

$$\begin{aligned} &H_0 : \kappa = 0 \text{ (null hypothesis)} \\ &H_1 : \kappa \neq 0 \ (\kappa = 2.77) \text{ (alternative hypothesis)} \end{aligned} \tag{3.27}$$

$$-2Ln(\lambda(\delta)) = 23,613 > 3.841 \tag{3.28}$$

The LRT approximation was greater than the critical value, so we reject the null hypothesis. There is enough evidence to support the claim that the dispersion parameter κ is different from 0 suggesting the real data is not distributed uniformly which is the correct conclusion.

3.4 Section 4: Discussion of Results

From the simulated data inference, it was shown that the proposed method of estimation produces acceptable results for the values of the parameters with the percentage of errors of about 2.73% for the location parameter ω and about 9.69% for the dispersion parameter κ. Even though no closed-form solution exists for this estimation procedure, it can be programmed in a mathematical or statistical package to obtain a fast and reliable solution. We have demonstrated that the estimation process yields accurate estimation of the true values of the parameter through the simulated data example, and that the QVMD is a reasonable model to fit circular data located on the positive quadrant of the unit circle. Moreover, the statistical tests developed from the LRT approximation of the Chi-square distribution produced reasonable results. The LRT approximation for the location parameter ω for the examples shown above reached the correct conclusion. The LRT approximation for the dispersion parameter κ and the tests for uniformity behaved well and yielded the correct conclusions.

CHAPTER 4: CONCLUSION AND FUTURE RESEARCH

Overall, the work done in this thesis contributes to develop a new and useful distribution in the field of circular statistics which can model data pertaining to the positive quarter of the unit circle as can be seen in previous examples. In chapter two, we characterized the distribution yielding summary statistics such as the mean, variance, skewness, and kurtosis which described numerically the main characteristics of the data set. The simulated data showed a distribution that was skewed to the left and the real data set showed a distribution that was skewed to the right. There were also interesting features inherited from the full circle von Mises distribution such as displaying a uniform distribution when κ equals zero. We can also see when the QVMD parameter values change it exhibits a rich variety of shapes. Moreover, even though the parent full circle von Mises distribution is strictly symmetric, the QVMD allows us to fit skewed data. Using the methods for maximum likelihood estimation described in chapter three, a procedure was developed for estimating parameters for real data. Further uses for the QVMD will include the eventual transformation of the univariate QVMD into a multivariate version suitable to be used in the hypersphere, allowing the modeling of complex data sets in the form of unit vectors in the hyperspace such as data found in gene expressions and text mining, which provides an outlook on future research prospects.

REFERENCES

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. The Wadsworth Group.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge, UK.

Guardiola, J. H. (2020). The Spherical-Dirichlet Distribution [Journal Article]. *Journal of Statistical Distributions and Applications*, *7*(6), 14. (Retrieved from https://doi.org/10.1186/s40488-020-00106-9)

Guardiola, J. H., Stamey, J., Seaman, J., & Elsalloukh, H. (2006). The Semicircular Normal Distribution. *Far East Journal of Theoretical Statistics*, *20*(2), 207–216. (Retrieved from pphmj.com/journals/fjts.htm)

Hogg R., C. A., McKean J. (2019). *Introduction to Mathematical Statistics, 8th Ed.* Pearson, Boston MA.

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. John Wiley and Sons, Ltd.

Miller, I. (2004). *John E. Freund's Mathematical Statistics: with Applications*. Pearson.

Pewsey, A., Neuhauser, M., & Ruxtion, G. D. (2013). *Circular Statistics in R*. Oxford University Press.

Savant. (2023). *Hit Probabilities Broken Down By Exit Velocity and Launch Angle.* https://baseballsavant.mlb.com/statcast_hit_probability?year=2023&type=la. (Accessed: 2023-07-07)

APPENDIX A:

MATLAB CODE

```
% Quarter von Mises distribution moments
% and properties numeric computations

% Constants of directional data
clear vars
% Simulated data parameter values:
% kappa = 3
% omega = pi/3
% n = 100
% Real data parameter values:
kappa = 2.77
omega = 0.295
n = 14910

% Reciprocal of integration constant
Cr = pi*besselj(0,kappa*1i)*exp(kappa)/2
% Probability density function of the QVMD
fn=@(delta,kappa,omega) exp(2*kappa*...
    ((cos(2.*(delta-omega))).^2))/Cr;

% Simple plot of distribution
plot(y, fn(y,kappa,omega))

% Integration over 0 to pi/2 should equal 1
CDF = integral(@(delta) fn(delta,kappa,omega),0,pi/2)
```

```
% 1st moment alpha component function
alpha1 = @(delta,kappa,omega) cos(delta).*...
    exp(2*kappa*((cos(2.*(delta-omega))).^2))/Cr;
% Numeric integration of the 1st moment alpha component
a1 = integral(@(y) alpha1(y,kappa,omega),0,pi/2)

% 1st moment beta component function
beta1=@(delta,kappa,omega) sin(delta).*...
    exp(2*kappa*((cos(2.*(delta-omega))).^2))/Cr;
% Numeric integration of the the 1st moment beta component
b1 =integral(@(y) beta1(y,kappa,omega),0,pi/2)

% Mean Resultant Length calculation
rho = sqrt(a1^2+b1^2)

% Circular Variance calculation
cv = 1 - rho

% Skewness
% 2nd moment beta component function
beta2 = @(delta,kappa,omega) sin(2*(delta-omega)).*...
    exp(2*kappa*((cos(2.*(delta-omega))).^2))/Cr;
% Numeric integration of the 2nd moment beta component
b2 = integral(@(y) beta2(y,kappa,omega),0,pi/2)
% Skewness calculation
s = b2/((1-rho)^(3/2))
```

```
% Kurtosis
% 2nd moment alpha component function
alpha2=@(delta,kappa,omega) cos(2*(delta-omega)).*...
    exp(2*kappa*((cos(2.*(delta-omega))).^2))/Cr;
% Numeric integration of the 2nd moment alpha component
a2 =integral(@(y) alpha2(y,kappa,omega),0,pi/2)
% Kurtosis calculation
K = (a2-rho^4)/((1-rho)^2)
% Quarter von Mises distribution MLE and Likelihood Ratio Tests

clear vars

% Simulated angles from QVMD distribution in R
% with omega=pi/3 and kappa=3
x=csvread("simdata.csv");

% Baseball data
y=csvread("baseballdata.csv")*pi/180;

%%
% MAXIMUM LIKELIHOOD ESTIMATION

% Plot the negative log likelihood function
% for the simulated data and replace n with the
% sample size of the real data and x with y for real data
n=14910 % n=14910
```

```
z = @(omega,kappa) n.*kappa +n.*log(besselj(0,kappa.*1i)) -...
                   2.*kappa.*sum((cos(2.*(y-omega)).^2),'all')
fsurf(z,[0 pi/2 0 6])
xlabel omega
ylabel kappa

% Define the function z
z = @(omega, kappa) n.*kappa + n.*log(besselj(0, kappa.*1i)) -...
    2.*kappa.*sum((cos(2.*(y-omega)).^2), 'all');

% Define initial values for lmv and kv
initial_omega = 1;
initial_kappa = 2;

% Define the objective function to be minimized
objective = @(params) z(params(1), params(2));

% Minimize the objective function using fminsearch
A=[];
b=[];
Aeq=[];
beq=[];
lb=[0,0];
ub=[pi/2,inf];
optimal_params =...
    fmincon(objective,[initial_omega, initial_kappa],...
            A,b,Aeq,beq,lb,ub)
```

```
% Retrieve the optimal values for omega and kappa
optimal_omega = optimal_params(1);
optimal_kappa = optimal_params(2);

% Display the optimal values
fprintf('Optimal omega: %f\n', optimal_omega);
fprintf('Optimal kappa: %f\n', optimal_kappa);

%%
% HYPOTHESIS TESTING



% TEST FOR THE LOCATION PARAMETER


% Test H0 omega = pi/3 vs H1 omega not = pi/3 (omega = pi/4) and
%      H0 omega = 0.295 vs H1 omega not = 0.295 (omega = 1)


% LRT approximation function -2Ln(lambda)
LRTAL = @(omega0,omegahat,kappa,delta) -4.*kappa.*...
    sum((cos(2.*(delta-omega0))).^2 -...
    ((cos(2.*(delta-omegahat))).^2),'all')

% Simulated data result:
LRTAL(pi/3,pi/4,3,x)
% Real data result:
LRTAL(0.295,1,2.77,y)
```

```
% Reject in both cases, but it is noted that
% the approximation is sensitive
% to large differences between the values


% TEST FOR DISPERSION PARAMETER KAPPA

% Test H0 kappa = 3 vs H1 kappa not = 3 (kappa = 2) and
%       H0 kappa = 2.77 vs H1 kappa not = 2.77 (kappa = 2)

% LRT approximation function -2Ln(lambda)
LRTAD = @(omega, kappa0, kappahat, delta, n)...
 2.*n.*log(besselj(0, kappa0.*1i)/besselj(0,kappahat.*1i))+...
 2.*(kappa0-kappahat).*...
 (n-2.*sum(((cos(2.*(delta-omega))).^2), 'all'))

% Simulated data result:
LRTAD(pi/3,3,2,x,100)
% Real data result:
LRTAD(0.295,2.77,2,y, 14910)
% We should fail to reject in both cases


% TEST FOR UNIFORMLY DISTRIBUTED DATA

% Test H0 kappa = 0 vs H1 kappa not = 0 (kappa = 3) and
% H0 kappa = 0 vs H1 kappa not = 0 (kappa = 2.77)
```

```
% LRT approximation function -2Ln(lambda)
LRTAU = @(omega, kappa,delta,n) -2.*n.*kappa -...
    2.*n.*log(besselj(0, kappa.*1i)) +...
    4.*kappa.*sum(((cos(2.*(delta-omega))).^2), 'all')

% Simulated data result:
LRTAU(pi/3,3,x,100)
% Real data result:
LRTAU(0.295,2.77,y,14910)
% We should reject in both cases
```

APPENDIX B:

R CODE

```
library(circular)
library(dplyr)
library(tidyr)
library(ggplot2)
library(gridExtra)

# Probability density function for the QVMD
dqvmd <- function(delta,omega,kappa){
  qvmpdf <- 2*(exp(2*kappa*((cos(2*(delta - omega)))^2)))/
    (pi * exp(kappa) *
       besselI(x = kappa, nu = 0, expon.scaled =F))
  return(data.frame(delta,qvmpdf))
}

# Unit circle function for ggplots
circleFun <- function(center = c(0,0),
                      diameter = 1, npoints = 10000){
  r = diameter / 2
  tt <- seq(0,pi/2,length.out = npoints)
  xx <- center[1] + r * cos(tt)
  yy <- center[2] + r * sin(tt)
  return(data.frame(x = xx, y = yy))
}
dat <- circleFun(diameter=2)
```

```
# QVMD PDF with varying omega
n=50
delta=seq(from=0,to=pi/2,length.out=n)
omegac=c('0','pi/6','pi/4','pi/3')
omega=c(0,pi/6,pi/4,pi/3)
kappa=1
n=length(omega)
qvmpdo=data.frame(delta,0,0,0,0)

for (i in 1:n){
  qvmpdo[,i+1]=dqvmd(delta,omega[i],kappa)[,2]
  colnames(qvmpdo)[i+1] <- paste0("pd", i)
  i=i+1
}

# QVMD PDF with varying omega plotted on rectangular coords
ol1=ggplot(qvmpdo, aes(x = delta,y = pd1)) +
  geom_point(col="black",fill=NA,shape=21,size=2) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  ggtitle(expression(omega==0)) +
  geom_path(data=qvmpdo,aes(x = delta,y = pd1))+
  coord_fixed(xlim=c(0,pi/2),ylim=c(0,1.5))
ol2=ggplot(qvmpdo, aes(x = delta,y = pd2)) +
  geom_point(col="black",fill=NA,shape=21,size=2) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
```

```
  ggtitle(expression(omega==pi/6)) +
  geom_path(data=qvmpdo, aes(x = delta, y = pd2))+
  coord_fixed(xlim=c(0, pi/2), ylim=c(0, 1.5))
ol3=ggplot(qvmpdo, aes(x = delta, y = pd3)) +
  geom_point(col="black", fill=NA, shape=21, size=2) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  xlim(0, pi/2)+ylim(0, 1.5)+
  ggtitle(expression(omega==pi/4)) +
  geom_path(data=qvmpdo, aes(x = delta, y = pd3))+
  coord_fixed(xlim=c(0, pi/2), ylim=c(0, 1.5))
ol4=ggplot(qvmpdo, aes(x = delta, y = pd4)) +
  geom_point(col="black", fill=NA, shape=21, size=2) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  ggtitle(expression(omega==pi/3)) +
  geom_path(data=qvmpdo, aes(x = delta, y = pd4))+
  coord_fixed(xlim=c(0, pi/2), ylim=c(0, 1.5))
grid.arrange(ol1, ol2, ol3, ol4, ncol=2, nrow=2)

# QVMD PDF with varying omega plotted on polar coords
o1=ggplot(qvmpdo, aes(x=(1+pd1)*cos(delta),
                      y=(1+pd1)*sin(delta)))+
    geom_point(col="black", fill=NA, shape=21, size=2.5)+
    xlab(expression(cos(delta)))+
    ylab(expression(sin(delta)))+
    ggtitle(expression(omega==0))+
```

```
    geom_path(data=dat, aes(x,y), col="black")+
    geom_path(data=qvmpdo, aes(x=(1+pd1)*cos(delta),
                               y=(1+pd1)*sin(delta)))+
    coord_fixed(xlim=c(0,2.5), ylim=c(0,2.5))
o2=ggplot(qvmpdo, aes(x=(1+pd2)*cos(delta),
                      y=(1+pd2)*sin(delta)))+
    geom_point(col="black", fill=NA, shape=21, size=2.5)+
    xlab(expression(cos(delta)))+
    ylab(expression(sin(delta)))+
    ggtitle(expression(omega==pi/6)) +
    geom_path(data=dat, aes(x,y), col="black")+
    geom_path(data=qvmpdo, aes(x=(1+pd2)*cos(delta),
                               y=(1+pd2)*sin(delta)))+
    coord_fixed(xlim=c(0,2.5), ylim=c(0,2.5))
o3=ggplot(qvmpdo, aes(x=(1+pd3)*cos(delta),
                      y=(1+pd3)*sin(delta)))+
    geom_point(col="black", fill=NA, shape=21, size=2.5)+
    xlab(expression(cos(delta)))+
    ylab(expression(sin(delta)))+
    ggtitle(expression(omega==pi/4))+
    geom_path(data=dat, aes(x,y), col="black")+
    geom_path(data=qvmpdo, aes(x=(1+pd3)*cos(delta),
                               y=(1+pd3)*sin(delta)))+
   coord_fixed(xlim=c(0,2.5), ylim=c(0,2.5))
o4=ggplot(qvmpdo, aes(x=(1+pd4)*cos(delta),
                      y=(1+pd4)*sin(delta)))+
    geom_point(col="black", fill=NA, shape=21, size=2.5)+
```

```
    xlab(expression(cos(delta)))+
    ylab(expression(sin(delta)))+
    ggtitle(expression(omega==pi/3))+
    geom_path(data=dat, aes(x,y), col="black")+
    geom_path(data=qvmpdo, aes(x=(1+pd4)*cos(delta),
                               y=(1+pd4)*sin(delta)))+
    coord_fixed(xlim=c(0,2.5), ylim=c(0,2.5))
grid.arrange(o1,o2,o3,o4,ncol=2,nrow=2)


# QVMD PDF with varying kappa
omega=pi/4
kappa=c(0,0.5,1,2)
qvmpdk=data.frame(delta,0,0,0,0)
n=length(kappa)
for (i in 1:n){
  qvmpdk[,i+1]=dqvmd(delta,omega,kappa[i])[,2]
  colnames(qvmpdk)[i+1] <- paste0("pd", i)
  i=i+1
}

# QVMD PDF with varying kappa plotted on rectangular coords
k11=ggplot(qvmpdk, aes(x = delta,y = pd1)) +
  geom_point(col="black", fill=NA, shape=21, size=1.5) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  xlim(0, pi/2)+ylim(0,3)+
```

```
  ggtitle(expression(kappa==0)) +
  geom_path(data=qvmpdk,aes(x = delta,y = pd1))+
  coord_fixed(ratio=1,xlim=c(0,pi/2),ylim=c(0,2.25))
kl2=ggplot(qvmpdk, aes(x = delta,y = pd2)) +
  geom_point(col="black",fill=NA,shape=21,size=1.5) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  xlim(0,pi/2)+ylim(0,3)+
  ggtitle(expression(kappa==0.5)) +
  geom_path(data=qvmpdk,aes(x = delta,y = pd2))+
  coord_fixed(ratio=1,xlim=c(0,pi/2),ylim=c(0,2.25))
kl3=ggplot(qvmpdk, aes(x = delta,y = pd3)) +
  geom_point(col="black",fill=NA,shape=21,size=1.5) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  xlim(0,pi/2)+ylim(0,3)+
  ggtitle(expression(kappa==1)) +
  geom_path(data=qvmpdk,aes(x = delta,y = pd3))+
  coord_fixed(ratio=1,xlim=c(0,pi/2),ylim=c(0,2.25))
kl4=ggplot(qvmpdk, aes(x = delta,y = pd4)) +
  geom_point(col="black",fill=NA,shape=21,size=1.5) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  ggtitle(expression(kappa==2)) +
  geom_path(data=qvmpdk,aes(x = delta,y = pd4))+
  coord_fixed(ratio=1,xlim=c(0,pi/2),ylim=c(0,2.25))
grid.arrange(kl1,kl2,kl3,kl4,ncol=2,nrow=2)
```

```
# QVMD PDF with varying kappa plotted on polar coords
k1=ggplot(qvmpdk, aes(x=(1+pd1)*cos(delta),
                        y=(1+pd1)*sin(delta)))+
   geom_point(col="black", fill=NA, shape=21, size=2.5)+
   xlab(expression(cos(delta)))+
   ylab(expression(sin(delta)))+
   ggtitle(expression(kappa==0))+
   geom_path(data=dat, aes(x,y), col="black")+
   geom_path(data=qvmpdk, aes(x=(1+pd1)*cos(delta),
                               y=(1+pd1)*sin(delta)))+
   coord_fixed(ratio=1, xlim=c(0,3), ylim=c(0,3))
k2=ggplot(qvmpdk, aes(x=(1+pd2)*cos(delta),
                        y=(1+pd2)*sin(delta)))+
   geom_point(col="black", fill=NA, shape=21, size=2.5)+
   xlab(expression(cos(delta)))+
   ylab(expression(sin(delta)))+
   ggtitle(expression(kappa==0.5))+
   geom_path(data=dat, aes(x,y), col="black")+
   geom_path(data=qvmpdk, aes(x=(1+pd2)*cos(delta),
                               y=(1+pd2)*sin(delta)))+
   coord_fixed(ratio=1, xlim=c(0,3), ylim=c(0,3))
k3=ggplot(qvmpdk, aes(x=(1+pd3)*cos(delta),
                        y=(1+pd3)*sin(delta)))+
   geom_point(col="black", fill=NA, shape=21, size=2.5)+
   xlab(expression(cos(delta)))+
   ylab(expression(sin(delta)))+
```

```
    ggtitle(expression(kappa==1))+
    geom_path(data=dat, aes(x,y), col="black")+
    geom_path(data=qvmpdk, aes(x=(1+pd3)*cos(delta),
                               y=(1+pd3)*sin(delta)))+
    coord_fixed(ratio=1,xlim=c(0,3),ylim=c(0,3))
k4=ggplot(qvmpdk, aes(x=(1+pd4)*cos(delta),
                      y=(1+pd4)*sin(delta)))+
    geom_point(col="black", fill=NA, shape=21, size=2.5)+
    xlab(expression(cos(delta)))+
    ylab(expression(sin(delta)))+
    ggtitle(expression(kappa==3))+
    geom_path(data=dat, aes(x,y), col="black")+
    geom_path(data=qvmpdk, aes(x = (1+pd4)*cos(delta),
                               y = (1+pd4)*sin(delta)))+
    coord_fixed(ratio=1,xlim=c(0,3),ylim=c(0,3))
grid.arrange(k1, k2, k3, k4,ncol = 2, nrow = 2)

# Simulated data
simdat=dqvmd(read.csv("simdat2.csv")[,2],pi/3,3)

# Simulated data plotted on polar coordinates
ggplot(simdat, aes(x = cos(delta), y = sin(delta))) +
  geom_point(col="black", fill=NA, shape=21, size=2.5) +
  xlab(expression(cos(delta)))+
  ylab(expression(sin(delta)))+
  ggtitle("Simulated Data on the Unit Circle") +
  geom_path(data=dat, aes(x,y), col="black") +
```

```
  coord_fixed()

# Simulated data plotted on rectangular coordinates
ggplot(simdat, aes(x=delta, y=qvmpdf)) +
  geom_point(col="black", fill=NA, shape=21, size=2.5) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  ggtitle("Simulated Data PDF") +
  geom_path(data=simdat, aes(x=delta, y=qvmpdf), col="black") +
  coord_fixed(ratio=1, xlim=c(0, pi/2), ylim=c(0, 2.75))

# Simulated Data PDF plotted on polar coordinates
ggplot(simdat, aes(x=(1+qvmpdf)*cos(delta),
                   y=(1+qvmpdf)*sin(delta))) +
  geom_point(col="black", fill=NA, shape=21, size=2.5) +
  xlab(expression(cos(delta)))+
  ylab(expression(sin(delta)))+
  ggtitle("Simulated Data PDF on the Unit Circle") +
  geom_path(data=dat, aes(x, y), col="black") +
  geom_path(data=simdat,
            aes(x = (1+qvmpdf)*cos(delta),
                y = (1+qvmpdf)*sin(delta)),
                col="black") +
  coord_fixed(ratio=1, xlim=c(0, 2), ylim=c(0, 3.25))


# Real data
```

```r
rqvmd=read.csv("baseballdata.csv")
rqvmd=rqvmd[2:91,1:2]
colnames(rqvmd)=c("delta","hits")
rqvmd=uncount(data=rqvmd, weights=hits)*pi/180
rqvmpdf=dqvmd(rqvmd,0.295,2.77)
rqvmpdf=arrange(rqvmpdf, rqvmpdf$delta)
colnames(rqvmpdf)=c("delta","qvmpdf")

# Real data plotted on polar coordinates
ggplot(rqvmpdf, aes(x = cos(delta), y = sin(delta))) +
  geom_point(size=2.5,shape=21,fill=NA) +
  xlab(expression(cos(delta)))+
  ylab(expression(sin(delta)))+
  ggtitle("Real Data on the Unit Circle") +
  geom_path(data=dat, aes(x,y)) +
  coord_fixed()

# Real data PDF plotted on rectangular coordinates
ggplot(rqvmpdf, aes(x = delta, y = qvmpdf)) +
  geom_point(size=2.5,shape=21,fill=NA) +
  xlab(expression(delta))+
  ylab(expression(f(delta)))+
  ggtitle("Real Data PDF") +
  geom_path(data=rqvmpdf, aes(x = delta, y = qvmpdf)) +
  coord_fixed()

# Real data PDF plotted on polar coordinates
```

```
ggplot(rqvmpdf,  aes(x=(1+qvmpdf)*cos(delta),
                    y=(1+qvmpdf)*sin(delta)))+
  geom_point(size=2,shape=21,fill=NA)+
  xlab(expression(cos(delta)))+
  ylab(expression(sin(delta)))+
  ggtitle("Real Data PDF on the Unit Circle")+
  geom_path(data=dat,aes(x,y))+
  geom_path(data=rqvmpdf,aes(x=(1+qvmpdf)*cos(delta),
                              y=(1+qvmpdf)*sin(delta)))+
  coord_fixed(ratio=1,xlim=c(0,3.5),ylim=c(0,1.25))
```