Best Approach Direction For Spherical Random Variables

Technical Report No. ASU/2017/13
Dated 16 August, 2017

Jayant Jha

Applied Statistics Unit
Indian Statistical Institute
Kolkata 700 108

Best Approach Direction For Spherical Random Variables

Jayant Jha
Applied Statistics Unit, Indian Statistical Institute
203 B. T. Road, Kolkata - 700 108, India
E-mail: jayantjha@gmail.com
Phone: +91-9038016983

Abstract

In this paper, we attempt the estimation of the directional quantiles, where the random variable takes values on the surface of a unit sphere. We first form a notion of directional quantiles in this regard. On the basis of the directional quantiles, we define the concept of the best approach direction. After that, we discuss some properties of the population best approach direction and discuss the inferential procedures related to its sample variant. Then, we study the asymptotic distribution of the best directional quantile estimator using the known asymptotic distribution of a directional estimator. Simulation and real data analyses are done to illustrate the results.

Some keywords: Quantiles; Spherical Random Variables; von Mises distribution; Wrapped Cauchy distribution.

1 Introduction

In R^p, a unit sphere is given by $S^{p-1} = \{x : x \in R^p; ||x|| = 1\}$. A random variable X is called a spherical random variable if the support of X is a subset of S^{p-1}. Spherical random variables have wide application in geology (e.g. Chang, 1986,1989), astronomy (e.g. Hall *et al.*, 1987), shape analysis (e.g. Bryner *et al.*, 2012) and crystallography (e.g. Chapman *et al.*, 1995), among others. In two dimensions, the support of the random variable is on the circumference of a circle and the random variable is called a circular or directional random variable. Random variables taking values in angles such as wind direction and direction of migration of birds are the most common examples of circular random variable. Other not so obvious examples of circular random variables are the time of arrival of patients in a hospital and date of the occurrence of an event in a year.

Quantiles of a random variable are very informative statistics because the distribution of data over the sample space can be estimated using the quantiles. When the random variable takes value on R and is continuous, the definition of the quantile is widely known. In this case, a quantile τ_α is the smallest number which satisfies $F(\tau_\alpha) >= \alpha$ where $\alpha \in (0,1)$ and $F(.)$ is the distribution function. Using this definition, the spherical quantiles for the spherical random variable was defined in Ley et al.(2014). In the paper, given the direction of the median of the population, different quantiles were defined on the diameter which contains the median. Some asymoptotic Bahadur (1966)- type representations were derived for the quantile estimators. A drawback of the quantile definition was that it gave equal weightage to the observations on both sides of the diameter and hence, one could not determine for a particular quantile on which side of the diameter the frequency of the observations is more. In this paper, an attempt to address the issue is undertaken where for every direction, directional quantiles are defined and then, one can understand the distribution of the spherical random variable more vividly by the sample directional quantile estimates.
Suppose, we want to optimise the use of traffic signal in such a way that the traffic light is on when most of the accidents occur and we want the traffic lights to be switched on for the minimum span of time. Another similar problem is determining the working hours of a doctor in a hospital. We want to optimise the working hours of a doctor by time span such that the doctor is present in the hospital for the least span of time in which a predefined proportion of patients arrive. Take for another example, one wants to launch a product which is marketed for the least number of days in a year but is able to target a predfined proportion of the product users. All of these examples can be converted to problems regarding circular random variables where the random variable (time of accident, time of arrival

of patients in a hospital and the date of use of the product) are circular in nature. Then, we can define the shortest time-interval in terms of the time span (S) and a mean time(T) such that $Pr(Y \in [T - S/2, T + S/2] \geq \alpha)$. The problem of choosing the best T such that S is minimized is basically a problem of choosing the point on the circumference (surface) of a circle (sphere) or the unit vector so that if one moves along the line starting from the point towards the origin and project all the observations on this line, one would cover α proportion of observations the fastest. We have used the term Best Approach Direction for the optimum mean direction (T) and we attempt to solve such problems given historical datasets. In Section 2, we introduce the definitions on which we shall be working throughout the paper. Section 3 covers the inferential aspects of the problem and some properties of the best approach direction. Section 4 deals with some asymptotic results. We have shown some simulation results on 2- and 3-dimensional sphere in Section 5. Section 6 deals with the illustration of the theory by real datasets. Section 7 concludes.

2 Definitions

2.1 Population Directional Quantiles

Directions on a unit sphere can be determined by unit vectors. Let X be a spherical random variable. Denoting a unit vector by β and the quantile by τ and using a definition similar to the one given in Ley et al.(2014) for the quantile along the median direction, the directional quantile $C_{\tau,\beta} \in [-1, 1]$ along the direction β can be defined by $C_{\tau,\beta}\beta$ such that the region, bounded by the hyperplane orthogonal to the unit vector β and passing through $C_{\tau,\beta}\beta$ and the unit sphere in the direction of β, denoted by $A*$ is the smallest region among all the regions formed by hyperplanes orthogonal to β and between the unit sphere in the direction of β denoted by A such that $P(X \in A) \geq \tau$. Now, this hyperplane C_A divides the unit sphere in two parts $C^+_{\tau,\beta} = \{x : x \in S^{p-1}; x'\beta \geq k\}$ and $C^-_{\tau,\beta} == \{x : x \in S^{p-1}; x'\beta \leq k\}$ for some k and if X follows a continuous distribution on the sphere, then at least one of $P(X \in C^+_{\tau,\beta})$ or $P(X \in C^-_{\tau,\beta})$ exactly equals τ.

2.2 Empirical Directional Quantiles

Let, $x_1, x_2 \dots x_n$ are the spherical observations. Then, for a fixed direction β, there is a straightforward representation of the empirical directional quantile. The empirical directional quantile $\hat{C}_{\tau,\beta} \in [-1, 1]$ can be defined in the following way:

$$\hat{C}_{\tau,\beta} = \inf\{C : card(x_i : x_i \in R_{p,\beta}) \geq n\tau\}$$

where $R_{p,\beta}$ is the region bounded by the hyperplane orthogonal to β, passing through $\hat{C}_{\tau,\beta}\beta$ and the unit sphere in the direction of β and $card(A)$ is the cardinality of set A.
When the direction β is fixed, the directional quantiles behave like the quantiles of a linear random variable and one can directly get the Bahadur-type representation as well as under the assumption of regularity conditions, asymptotic distributions of the quantiles.
It should be noted here that, whenever more than one observations are on the hyperplane orthogonal to β and passing through C_β, some of the empirical quantiles coincide.

2.3 Population BAD

The best approaching direction for a population, given a quantile τ, is the unit vector $U_\tau \in S^{p-1}$ such that $C_{\tau,U_\tau} \geq C_{\tau,\alpha}\ \forall \alpha \in S^{p-1}$ i.e. for a given τ, U_τ represents the unit vector from which if one moves inside the sphere along the line joining U_τ and origin, one may cover the region having probability measure τ the fastest in terms of the length of line segment between U_τ and $C_{\tau,U_\tau}U_\tau$. Thus,

$$PBAD = \{U_\tau \in S^{p-1} : C_{\tau,U_\tau} \geq C_{\tau,V_\tau} \forall V_\tau \in S^{p-1}\}$$

2.4 Sample BAD

The sample version of BAD is given by the direction $\hat{u}_\tau$ such that if one approaches the inside of the sphere from the direction $\hat{u}_\tau$, one covers τ proportion of the sample the fastest. Thus,

$$SBAD = \{\hat{u}_\tau \in S^{p-1} : \hat{C}_{\tau,\hat{u}_\tau} \geq \hat{C}_{\tau,V_\tau} \forall V_\tau \in S^{p-1}\}$$

3 Finding SBAD

Theorem 1: If the points $x_1, x_2, \dots x_n \in S^{p-1}$ are in general position, then given a region bounded between a hyperplane and the unit sphere which contains m out of n observations, we always get a smaller region between a hyperplane passing through p out of the m points which also contains the same m points $\{x_1, \dots x_m\}$.

Proof: Let P be a hyperplane such that $\{x_1, \dots x_m\} \in C_P$, where C_P represents the area between P and S^{p-1} which contains $\{x_1, \dots x_m\}$. Also, if we define $S_P = C_P \cap S^{p-1}$ to be the part of the surface of the sphere which contains $\{x_1, \dots x_m\}$ and is bounded by P. Then, due to the convexity of C_P,

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots \alpha_m x_m \in C_P \, \forall \alpha_i \geq 0; \sum_{i=1}^{m} \alpha_i = 1$$

Let, $x_1, x_2, \dots x_r \in P$ and $x_{r+1}, \dots x_m \notin P$ where, $r < p$. If r=0, then we can always get a smaller region containing at least one point out of the m points which has a smaller area by moving the plane parallelly such that it passes through one of $\{x_1, \dots x_m\}$ and it will have a smaller area. Thus, we take r¿0. Also assume that Q is another hyperplane such that $\{x_1, \dots x_m\} \in C_Q$ and without loss of generality, $x_1, x_2, \dots x_p \in Q$. Then, $\exists \epsilon > 0 : B_\epsilon(x_p) \in C_P$ where $B_\epsilon(x_p) = \{x : x \in S_{p-1}; ||x - x_p|| < \epsilon\}$. But, $x_p \in Q \Rightarrow B_\epsilon(x_p) \notin C_Q$. Thus, $C_P - C_Q \neq \phi$.

Case 1: Let $O \notin C_P, C_Q$ where O is the origin. Then,

$$\frac{\alpha_1 x_1 + \alpha_2 x_2 + \dots \alpha_m x_m}{||\alpha_1 x_1 + \alpha_2 x_2 + \dots \alpha_m x_m||} \in S^{p-1} \, \forall \alpha_i \geq 0; \sum_{i=1}^{m} \alpha_i = 1 \tag{3.1}$$

As $x_1, x_2, \dots x_p \in Q$, hence, if $z \in C_Q \cap S^{p-1}$, then $z = \frac{\sum_{i=1}^{p} \alpha_i x_i}{||\sum_{i=1}^{p} \alpha_i x_i||}$ for some α_i's such that $\sum_{i=1}^{p} \alpha_i = 1$ by equation (3.1). Also, because x_p is an interior point of C_P, $z \in C_P \cap S^{p-1}$ Thus, $C_Q \cap S^{p-1} \subset C_P \cap S^{p-1}$ which implies $C_Q \in C_P$.

Case 2: let $O \in C_P, C_Q$. Then, we may look at the region C_P^C, C_Q^C, the complements of C_P, C_Q and show that $C_P^C \subset C_Q^C$. Let $z \in C_Q^C \cap S_{p-1}$. Then, $z = \frac{\sum_{i=1}^{p} \alpha_i x_i}{||\sum_{i=1}^{p} \alpha_i x_i||}$ for some α_i's such that $\sum_{i=1}^{p} \alpha_i = 1$. Now, let $x_1, \dots x_r, x_{r+1}*, \dots x_p* \in P$ where $x_j* \in S^{p-1}$ and $x_{r+1}*, \dots x_p* \notin C_Q$. Now, because $x_j* \in C_Q^C \cap S_{p-1}$, thus, $\forall j = \{r+1, \dots p\}$, $x_j* \in C_Q^C \cap S_{p-1}$. Now, $z \in C_P^C \cap S^{p-1}$ implies, $z = \frac{\sum_{i=1}^{r} \beta_i x_i + \sum_{i=r+1}^{p} \beta_i x_i*}{||\sum_{i=1}^{r} \beta_i x_i + \sum_{i=r+1}^{p} \beta_i x_i*||}$. But, $x_j* \in C_Q^C \cap S_{p-1}$ implies $x_j* = \frac{\sum_{i=1}^{p} \alpha_i x_i}{||\sum_{i=1}^{p} \alpha_i x_i||}$ for some α_i's such that $\sum_{i=1}^{p} \alpha_i = 1$ for all $j = \{r+1, \dots p\}$. Thus, $C_P^C \cap S^{p-1} \subset C_Q^C \cap S^{p-1}$ which implies $C_Q \in C_P$. Now, $O \in P$ but $O \notin Q$ implies that the area of C_Q is less than the area of C_P trivially and thus, Case 1 and Case 2 are sufficient for the proof. Q.E.D. □

Using the theorem, given $x_1, \ldots x_n \in S^{p-1}$, one can make the following algorithm to find $SBAD$ for τ where, $\tau = \frac{m}{n}$.

Algorithm 1:

Let $dis_{max} = -\infty$, P be $X_1 = 0$.

Step 1: Take any set of p points out of n points.

Step 2: Check if the plane passing through the p points cuts the sphere in such a way that one side of the plane contains more than m points.

Step 3: If there are more than m points on one side of the plane, find the distance between the plane and the origin and denote it by dis and the plane by P_{temp}.

Step 4: If the region on the side of the plane which contains more than m points contains origin too, $dis_{temp} = dis + 1$, otherwise, $dis_{temp} = 1 - dis$.

Step 5: If $dis_{temp} < dis_{max}$, $dis_{max} = dis_{temp}$ and $P = P_{temp}$

Thus, the unit vector orthogonal to P in the direction of the region containing the m points gives SBAD.

3.1 O(n) algorithm for finding SBAD in S^1

The complexity of Algorithm 1 is $\binom{n}{p}$. When $x_1, x_2, \ldots x_n \in S^1$, the complexity of the algorithm can be reduced further because of a sense of ordering in the points $x_1, x_2, \ldots x_n$ because in this case, the points can be represented in terms of angles. Let, $x_1, \ldots x_n \in [0, 2\pi)$ and if $x_{(1)}, \ldots x_{(n)}$ be the ordered angles such that $0 \leq x_{(1)} < x_{(2)} \ldots x_{(n)} < 2\pi$. Then, to find SBAD, one can start from $x_{(1)}$ and then measure the distance between $x_{(1)}$ and $x_{(m)}$. In the similar way one may move forward and take the distance between $x_{(r)}$ and $x_{(r+m-1)modn}$ till $r = n$. SBAD is then determined by the unit vector perpendicular to the plane which has the least distance among these n distances and thus, the problem can be solved in O(n) when the observations are on a circle.

3.2 Properties of BAD

Property 1: Equivalence of spherical mean and SBAD on a family of distributions

Theorem 2: For the class of spherical probability density function $f(x;\theta) = g(x'\theta)$ where $\theta \in S^{p-1}$ and $g(y)$ is a monotonically increasing function of y and for any $\tau \in (0,1)$, the PBAD is θ.

Proof: Let, $\delta \neq \theta \in S^{p-1}$. and $c \in [-1,1]$. Let, $A_2 = \{x : x'\theta \geq c\}$ and $A_1 = \{x : x'\delta \geq c\}$

$$\int_{A_1} f(x;\theta)d\Omega = \int_{A_1 \cap A_2} f(x;\theta)d\Omega + \int_{A_1 \cap A_2^C} f(x;\theta)d\Omega$$

Similarly,

$$\int_{A_2} f(x;\theta)d\Omega = \int_{A_1 \cap A_2} f(x;\theta)d\Omega + \int_{A_1^C \cap A_2} f(x;\theta)d\Omega$$

Now, $\forall x \in A_1 \cap A_2^C, x'\theta < c$ and $\forall x \in A_1^C \cap A_2, x'\theta \geq c$. Thus, $\sup_{A_1 \cap A_2^C} f(x;\theta) \leq \inf_{A_1^C \cap A_2} f(x;\theta)$. Also, the surface area covered by A_1 is the same as the surface area covered by A_2. Thus, $\int_{A_1} f(x;\theta)d\Omega < \int_{A_2} f(x;\theta)d\Omega$. Thus, PBAD is θ. Q.E.D. □

The family of pdfs on a sphere satisfying the condition of Theorem 2 includes von Mises Fisher distribution, exit distribution etc. Theorem 2 establishes the equivalence between the spherical mean, spherical median and PBAD in case of symmetrically distributed monotonic distributions.

Property 2: PBAD and SBAD are rotationally equivariant for rotationally equiavariant family of distributions on the sphere i.e. $PBAD(AX) = A \times PBAD(X)$ where $A \in O(p)$ and $PBAD(X)$ is the PBAD based on X.

This is because for any unit vector v and $A \in O(p)$, $v^\top x = (Av)^\top Ax$. Now, for the vector v and $\tau \in [0,1]$, let the set $\{x : x^\top v > c\}$ has probability τ. Then, for the random variable $Y = AX$, the set $\{y : y^\top Av > c\}$ has the same probability. Thus, $PBAD(AX) = A \times PBAD(X)$ and $C_{\tau,PBAD(X)} = C_{\tau,PBAD(Y)}$. Similarly, the results hold for SBAD.

4 Some Asymptotic Results

4.1 a.e. Convergence of SBAD

Let, $U_\tau \in S^{p-1}$ be the PBAD for a fixed $\tau \in [0,1]$. Then, under the following assumptions, almost sure convergence of SBAD can be established.

Assumptions:

A1. $f_{X_n}(A) \xrightarrow{a.e.} f_X(A)\ \forall A \in S^{p-1}$

A2. $\exists \tau \in [0,1] : \exists \epsilon(\tau) > 0 : \sup_{B_\epsilon(u)} C_\tau(u) > \inf_{B_\epsilon(u)} C_\tau(u) \geq \sup_{B_\epsilon^C} C_\tau(u)$

Theorem 3: Under A1 and for any $\tau \in [0,1]$ for which A2 is satisfied, $\arg\max_{S^{p-1}} C_n(\tau) \in B_\epsilon(u)$ almost surely.

Proof: For any $v \in S^{p-1}$, $Z(v) : \Omega \to [-1,1]$ is a random variable which is a function of the random variable X such that $Z(v) = X'v$. Then, $C_\tau(v)$ is the population quantile of $-Z(v)$. Now, sample quantile converges almost surely to population quantile under A1. Hence, $C_{n,\tau}(v) \xrightarrow{a.e.} C_\tau(v)$.

Now, if we take τ such that it satisfies A1 and A2, then we can write $\sup_{B^C_{\epsilon(u)}} C_\tau(u) = A$; $\sup_{B_{\epsilon(u)}} C_\tau(u) = B$ and $B > A$.Then, $\forall \delta > 0$, $\exists N : \forall n \geq N$, $A - \delta \leq \sup_{B^C_{\epsilon(u)}} C_n(u) \leq A + \delta$ and $B - \delta \leq \sup_{B_{\epsilon(u)}} C_n(u) \leq B + \delta$. But, δ is arbitrary and $B > A$. Thus, $\exists N : \forall n \geq N, \sup_{B^C_{\epsilon(u)}} C_n < \sup_{B_{\epsilon(u)}} C_n$. Thus, for any τ satisfying A2, $\forall n \geq N$, $\arg\max C_n \in B_\epsilon(u)$ almost surely. Q.E.D. □

4.2 Asymptotic Distribution of SBAD

Given an asymptotic distribution for mean or median and using the result established in Theorem 2, we can find the asymptotic distribution of SBAD for circular random variables taking values on S^1. Denoting the population mean or median by $\bar{\theta}$ and a consistent estimator of mean or median by $\hat{\bar{\theta}}_n$ and under the following assumptions and the assumptions A1 and A2, the asymptotic results may be established for SBAD. Let, there is an estimator $\hat{\bar{\theta}}_n$ of θ where $\bar{\theta}$ can be mean or median such that

A3. $\hat{\bar{\theta}}_n \xrightarrow{p} \bar{\theta}$

A4. $\sqrt{n}(\hat{\bar{\theta}}_n - \bar{\theta}) \xrightarrow{d} O_p(1)$

A unique unit complex number z can be defined for every angle θ by the transformation $z : \theta \to \mathbb{C}$ such that $z(\theta) = e^{i\theta}$. Now, A3 and Theorem 2 implies $\hat{\bar{\theta}}_n \xrightarrow{p} \bar{\theta}$ and $SBAD \xrightarrow{p} PBAD$. We denote $z_\theta = e^{i\bar{\theta}}$, $z_{\theta *} = e^{iPBAD}$, $\hat{z}_\theta = e^{i\hat{\bar{\theta}}_n}$ and $\hat{z}_{\theta *} = e^{iSBAD}$. Also, let, $Bz_\theta = z_{\theta *}$ and $\hat{B}\hat{z}_{\theta *} = \hat{z}_{\theta *}$ where, $B, \hat{B} \in \mathbb{C}; |B|, |\hat{B}| = 1$. Thus, $\hat{B} \xrightarrow{p} B$ by Slutsky's Theorem.

Now, writing the 2-dimensional representation for SBAD as $\hat{\theta *}_2 = (\cos\hat{\theta *}, \sin\hat{\theta *})$ and the same for PBAD as $\theta *_2 = (\cos\theta *, \sin\theta *)$. Also, let $b, \hat{b} \in O(2)$, where O(2) is the special orthogonal group of matrices of dimension 2×2 uniquely determined by $B, \hat{B}$ respectively. Now, representing $\theta m_2 = (\cos\bar{\theta}, \sin\bar{\theta})$ and $\hat{\theta m}_2 = (\cos\hat{\bar{\theta}}, \sin\hat{\bar{\theta}})$. Then,

$$\sqrt{n}(\hat{b}\hat{\theta m}_2 - b\theta m_2) = \sqrt{n}(\hat{b}\hat{\theta m}_2 - b\hat{\theta m}_2 + b\hat{\theta m}_2 - b\theta m_2)$$

$$= \sqrt{n}(\hat{b} - b)(\hat{\theta} m_2) + \sqrt{n}b(\hat{\theta} m_2 - \theta m_2)$$

$$= o_p(1) + F'$$

where, F' is the distribution of $b\sqrt{n}(\hat{\theta} m_2 - \theta m_2)$.
Now, if $\sqrt{n}(\hat{\theta} m_2 - \theta m_2) \sim N(\vec{0}, \Sigma)$, then,

$$\sqrt{n}(\hat{b}\hat{\theta m}_2 - b\theta m_2) \sim N(\vec{0}, b\Sigma b')$$

. Thus, if we know a consistent estimator of population mean or median (or any directional statistic) whose asymptotic distribution is known, we can compute the asymptotic distribution of SBAD.

4.3 Asymptotic Distribution of $\hat{C}_{\tau,SBAD}$

We denote SBAD by $\hat{\beta}$ and PBAD by β.

Assumptions:

A5.$||\frac{d^2 C_{\tau,z}}{dz^2}|| < M$ for some finite M at $\hat{\beta}$.

Under A5, we may derive the asymptotic distribution for $\hat{C}_{\tau,\hat{\beta}}$ by the Taylor Series expansion of $C_{\tau,\hat{\beta}}$ at $\hat{\beta}$ in the following way:

$$\sqrt{n}(\hat{C}_{\tau,\hat{\beta}} - C_{\tau,\beta}) = \sqrt{n}(\hat{C}_{\tau,\hat{\beta}} - \hat{C}_{\tau,\beta} + \hat{C}_{\tau,\beta} - C_{\tau,\beta})$$

$$= \sqrt{n}(\hat{\beta} - \beta)' C_{\tau,\hat{\beta}}"(\hat{\beta} - \beta) + \sqrt{n}(\hat{C}_{\tau,\beta} - C_{\tau,\beta})$$

Now, due to the asymptotic distribution of SBAD, $\sqrt{n}(\hat{\beta} - \beta)' C_{\tau,\hat{\beta}}"(\hat{\beta} - \beta) \to o_p(1)$. Also, we may find the distribution of $\hat{C}_{\tau,\beta}$ by the distribution of $X'\beta$ which can be approximated by the distribution of $X'\hat{\beta}$.

5 Simulation

We carried out simulations at two sample sizes for the cases when $X \in S^1, S^2$. The random variable $X \sim vM(0, \kappa)$ (von Mises distribution) when X is a circular random variable and $X \sim vMF(\mu, \kappa)$ (von Mises Fisher distribution) with $\mu = (1, 0, 0)$ when $X \in S^2$ at two different values of κ. From Theorem 3, we know that the PBAD in these cases will be equal to the mean direction for all values of τ. We calculated SBAD at three levels of τ in 10000 replications for each simulation experiment. In the tables, we reported the spherical (circular) mean and spherical (circular) dispersion of SBAD and the mean and standard error of $\hat{C}_\tau$ over the 10000 simulations. Denoting $z = \frac{x_1 + x_2 + \ldots + x_n}{n}$, the spherical mean direction of $x_1, x_2, \ldots x_n$ is defined as $\frac{z}{||z||}$ and the spherical dispersion is defined as $1 - ||z||$. For the circular case, we have tabulated the circular mean defined as $arg(z)$ when z is expressed in the form of complex number. SBAD is found in the circular case based on the algorithm mentioned in section 3.1 and SBAD is calculated in the spherical case using Algorithm 1.
The results of the simulations for circular cases are presented in Table 1 and Table 2, while the same for the spherical cases are presented in Table 3 and Table 4. We find that the mean direction of SBAD in all these cases are very close to the mean direction of the true parameters and the dispersion is also very low.

Table 1: Simulation results: SBAD and $\hat{C}_\tau$ (standard errors in parentheses) for circular random variable at $n = 50, 200$, $\mu = 0, \kappa = 3.5$ at different τ's

$n = 50$			
τ	0.3	0.6	0.9
SBAD	-0.0007 (0.0259)	-0.0007 (0.0131)	-0.0001 (0.0126)
$\hat{C}_\tau$	0.0144(5.6490×10^{-5})	0.0938 (0.0003)	0.3925 (0.0009)
$n = 200$			
SBAD	0.0021 (0.0118)	0.0016 (0.0054)	0.0003 (0.0051)
$\hat{C}_\tau$	0.0188(3.7346×10^{-5})	0.1023 (0.0001)	0.4061 (0.0005)

Table 2: Simulation results: SBAD and $\hat{C}_\tau$ (standard errors in parentheses) for circular random variable at $n = 50, 200$, $\mu = 0, \kappa = 1.5$ at different τ's

$n = 50$			
τ	0.3	0.6	0.9
SBAD	0.0024 (0.0627)	9.5023×10^{-6} (0.0338)	-0.0015 (0.0589)
$\hat{C}_\tau$	0.0412(0.0002)	0.2621 (0.0008)	1.0734 (0.0023)
$n = 200$			
SBAD	-0.0019 (0.0274)	-0.0002 (0.0135)	-0.0004 (0.0212)
$\hat{C}_\tau$	0.0523(0.0001)	0.2831 (0.0004)	1.1222 (0.0013)

Table 3: Simulation results: SBAD and $\hat{C}_\tau$ (standard errors in parentheses) for spherical random variable at $n = 50, 100$, $\mu = (1, 0, 0), \kappa = 3.5$ at different τ's

$n = 50$			
τ	0.3	0.6	0.9
SBAD	(1,0.0031,-0.0018) (0.0476)	(1,0.0025,0.0007) (0.0265)	(1,-0.0009,-0.0036) (0.0312)
$\hat{C}_\tau$	0.0656(0.0002)	0.2088 (0.0377)	0.5342 (0.0009)
$n = 100$			
SBAD	(1,-0.0020,0) (0.0281)	(1,-0.0003,0.0010) (0.0173)	(1,-0.0017,0.0009) (0.0199)
$\hat{C}_\tau$	0.0788(0.0001)	0.2268 (0.0003)	0.5745 (0.0007)

Table 4: Simulation results: SBAD and $\hat{C}_\tau$ (standard errors in parentheses) for spherical random variable at $n = 50, 100$, $\mu = (1, 0, 0), \kappa = 1.5$ at different τ's

$n = 50$			
τ	0.3	0.6	0.9
SBAD	(1,-0.0027,0.0021) (0.1178)	(1,0.0023,-0.0025) (0.0768)	(1,0,-0.0026) (0.1154)
$\hat{C}_\tau$	0.1433(0.0003)	0.4484 (0.0008)	1.0641 (0.0014)
$n = 100$			
SBAD	(1,0.0012,-0.0039) (0.0734)	(1,-0.0009,0.0004) (0.0480)	(1,0.0025,-0.0042) (0.0760)
$\hat{C}_\tau$	0.1700(0.0003)	0.4873 (0.0006)	1.1385 (0.0010)

6 Examples

6.1 Leukamia Data Example

In table 1.4 of Mardia and Jupp (2002), the month of onset of cases for lymphatic leukamia data in UK during 1946-1960 is shown. The data is grouped and we have the number of observations corresponding to each month of the year. To make the data useful for the illustration of the theory presented in the previous sections, we first converted the data into circular data by taking the span for a month as $[(i-1)\frac{\pi}{6}, i\frac{\pi}{6})$ and taking $i = 1$ for January and $i = 12$ for December moving in anti-clockwise fashion. Then, we uniformly took f_i observations between $[(i-1)\frac{\pi}{6}, i\frac{\pi}{6})$ for each i, where f_i represents the number of observations for the ith month. Thus, the sample space is the circumference of the circle. Then, we calculated SBAD based on quantiles $\tau_1 = 150/506$, $\tau_2 = 300/506$ and $\tau_3 = 450/506$ close to 0.3,0.6 and 0.9 respectively. The results of the data analysis are presented in Table 5. The analysis on such datasets help the health officials to target the disease appropriately using minimum monitoring (in terms of time-span) so that a pre-defined % of patients can be taken care of. SBAD and $\hat{C}_\tau$ for the cases are mentioned in Table 5 . We have reported the standard error of the SBAD assuming the joint asymptotic distribution for $\cos(X_i), \sin(X_i)$ as multivariate normal as shown in equation (6.5) and using delta method.

$$\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} \cos x_i, \quad \sum_{i=1}^{n} \sin x_i\right] \sim N(\phi, \Sigma) \tag{6.2}$$

Table 5: Data Analysis for leukamia data

τ	$\frac{150}{506}$	$\frac{300}{506}$	$\frac{450}{506}$
SBAD	3.3696 (0.0315)	3.5854 (0.0307)	4.3700 (0.0282)
$\hat{C}_\tau$	0.2427	1.0506	1.8781

Table 6: Data Analysis for Ant data

τ	0.3	0.6	0.9
SBAD	3.2446 (0.0602)	3.3406 (0.0601)	1.9836 (0.0252)
$\hat{C}_\tau$	0.0167	0.1279	1.4810

where, $\phi = E(X)$ As we do not know the true distribution for C_τ, hence the standard error for C_τ is not reported in the table.

6.2 Ant Data Example

In Bentley (2006), dataset on the direction of movement of ant is considered. The dataset is a modification of the data used in Fisher (1993) (p.243) which is sampled from a large dataset from Jander (1957. It has been argued in Bentley (2006) that the data comes from a mixture of vonMises distribution and a uniform distribution. Theorem 3 holds for this distribution also as uniform distribution adds a fixed constant to the equal area on sphere surface. We have used this dataset and calculated SBAD and $\hat{C}_\tau$ for $\tau = 0.3, 0.6, 0.9$. We have estimated the standard error of SBAD similar to the previous example. The results are then tabulated in Table . It can be seen from the results that for $\tau = 0.3$ and $\tau = 0.6$, the SBAD is near the mean direction of the von Mises fit as the effect of mixing does not take place but when $\tau = 0.9$, the mean direction shifts a lot and this may be be due to the non-uniformity of the assumed uniform distribution as the mixture.

7 Conclusion

In this paper, we have defined directional quantiles for spherical random variables and on the basis of it, we formed a notion of best approach direction. If we define in the same sense, the worst approach direction, then also most of the properties can be derived similarly.

Section 4.2 was restricted to the case of circular random variable. This was due to the non-uniqueness of B in dimensions greater than 2. There is a scope to study the asymptotics in higher dimensions. Property 1 can also be used to develop tests of unimodality. The method could also be generalised for general shapes by embedding the shapes inside a sphere.

References

Bahadur, R. R. (1966). A note on quantiles in large samples. *Annals of Mathematical Statistics*, 37, 577–580.

Bentley,J (2006). Modelling circular data using a mixture of von Mises and uniform distributions. *M.Sc. Project Simon Fraser University*

Bryner, D., Klassen, E. and Srivastava, A. (2012). Affine invariant elastic shape analysis of planar contours. *IEEE conference on computer vision and pattern recognition* 2012, 390-397.

Chang, T. (1986). Spherical regression. *Annals of Statistics* 14, 907-924.

Chang, T. (1989). Spherical regression with errors in variables. *Annals of Statistics* 17, 293-306.

Chapman, G.R., Chen, G. and Kim, P.T. (1995). Assessing geometrical integrity through spherical regression techniques. *Statistica Sinica* 5, 173-220.

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

Jander, R. (1957). Die opstische Richtangsorientierung der roten Waldameise. (Formicarufa. L.) Z. vergl. Physiologie, 40:162–238

Hall, P., Watson, G.S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* 74, 751-762.

Ley, Christophe, Sabbah, Camille and Verdebout, Thomas (2014). A new concept of quantiles for directional data and the angular Mahalanobis depth. *Electronic Journal of Statistics* 8,795-816.

Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics.* John Wiley and Sons, Chichester.