Statistical methods for random rotations

by

Bryan Adams Stanfill

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:

Ulrike Genschel, Major Professor

Heike Hofmann

Dan Nordman

Stephen Vardeman

Jonathan Smith

Iowa State University

Ames, Iowa

2014

DEDICATION

To Mimi and Pops.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Many people have helped to make this dissertation possible and I would like to take this opportunity to acknowledge them. I would like to thank my advisers Ulrike Genschel and Heike Hofmann for being so thoughtful and patient with me. Both of your lives have changed significantly over the past few years and yet you have always made time for me. I truly appreciate that. I also would like to thank my committee, Dan Nordman, Stephen Vardeman and Jonathan Smith, for your questions, comments and suggestions.

I would like to thank my family for their constant support and encouragement, especially my wife, grandparents, mom, dad, step-mom and brothers. I would also like to thank my aunt Helen, uncle Ed and cousins Emily and Erica for understanding when I can't make it home for Thanksgiving. Many thanks to Dave too for taking me on walks and letting me dump work on you so I could finish this dissertation.

ABSTRACT

The analysis of orientation data is a growing field in statistics. Though the rotationally symmetric location model for orientation data is simple, statistical methods for estimation and inference for the location parameter, $\boldsymbol{S}$ are limited. In this dissertation we develop point estimation and confidence region methods for the central orientation.

Both extrinsic and intrinsic approaches to estimating the central orientation $\boldsymbol{S}$ have been proposed in the literature, but no rigorous comparison of the approaches is available. In Chapter 2 we consider both intrinsic and extrinsic estimators of the central orientation and compare their statistical properties in a simulation study. In particular we consider the projected mean, geometric mean and geometric median. In addition we introduce the projected median as a novel robust estimator of the location parameter. The results of a simulation study suggest the projected median is the preferred estimator because of its low bias and mean square error.

Non-parametric confidence regions for the central orientation have been proposed in the literature, but they have undesirable coverage rates for small samples. In Chapter 3 we propose a nonparametric pivotal bootstrap to calibrate confidence regions for the central orientation. We demonstrate the benefits of using calibrated confidence regions in a simulation study and prove the proposed bootstrap method is consistent.

Robust statistical methods for estimating the central orientation has received very little attention. In Chapter 4 we explore the finite sample and asymptotic properties of the projected median. In particular we derive the asymptotic distribution of the projected median and show it is SB-robust for the Cayley and matrix Fisher distributions. Confidence regions for the central orientation $\boldsymbol{S}$ are proposed, which can be shown to have preferable finite sample coverage rates compared to those based on the projected mean.

Finally the `rotations` package is developed in Chapter 5, which contains functions for the statistical analysis of rotation data in $SO(3)$.

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

Estimation and inference for location parameters is a fundamental problem in statistics. For parameter spaces such as the real line or two-dimensional plane, the statistical tools used to learn about the population location are many and varied. Additionally, guidelines exist to suggest which statistical tool is most appropriate in a given situation. For example, it is typically suggested the sample median be used to estimate the population location rather than the sample mean if the data are skewed. For parameter spaces like the rotation group, however, not only are the statistical tools limited in number, it is not well known how those tools behave or which tool is best for a given situation. In this dissertation we provide empirical and theoretical evidence that both expands the statistical methodology available for the analysis of rotation data and suggest which tool is most appropriate.

Let $SO(3)$ denote the collection of all 3×3 rotation matrices called the rotation group. We consider the random sample $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n \in SO(3)$ from the location model

$$\boldsymbol{R}_i = \boldsymbol{S}\boldsymbol{E}_i \quad i = 1, \ldots, n \tag{1.1}$$

where $\boldsymbol{S} \in SO(3)$ is the fixed parameter of interest indicating an orientation of central tendency, and $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n \in SO(3)$ denote i.i.d. random rotations which symmetrically perturb $\boldsymbol{S}$. This dissertation examines the empirical and theoretical behavior of point and region estimators for the central orientation parameter $\boldsymbol{S}$.

There are two approaches to analyzing $SO(3)$ data; each approach results in a different class of estimators for the central orientation $\boldsymbol{S}$. The intrinsic approach uses the innate topology of the rotation group $SO(3)$ to define distance and ultimately estimators for the parameters. Alternatively, the extrinsic approach embeds the $SO(3)$ space into the space of all 3×3 ma-

trices $\mathbb{R}^{3\times3}$ then standard Euclidean metrics to define distance and therefore estimators are implemented. As of this writing there has been no clear reason presented in the literature as to which choice results in the best analysis. In fact, a majority of the literature only examines one method of analysis and ignores the other entirely.

In what follows we consider both approaches when deciding which estimator to use to estimate the central orientation $\boldsymbol{S}$. An in-depth examination of the extrinsic estimators leads to confidence region estimates and robustness properties. Some computing considerations involved in $SO(3)$ data analysis are also detailed.

1.2 Thesis Organization

This dissertation is comprised or four manuscripts related to the statistical analysis of three-dimensional rotations. Chapter 2 includes a literature review of the approaches used to compute a point estimate for the central orientation and their small sample behavior is investigated in a simulation study. The asymptotic behavior of the extrinsic mean is defined in Chapter 3 and a pivotal bootstrap confidence region procedure is proposed. In Chapter 4 the extrinsic mean is investigated in detail, including its asymptotic distribution and robustness properties. The computer code used throughout this dissertation has been collected into the R package `rotations`, which is detailed in Chapter 5. Finally, general conclusions based on the entire work and future directions are discussed in Chapter 6.

CHAPTER 2. POINT ESTIMATION OF THE CENTRAL ORIENTATION OF RANDOM ROTATIONS

A paper published in *Technometrics*

Bryan Stanfill, Ulrike Genschel, Heike Hofmann

Abstract

Data as three-dimensional rotations have application in computer science, kinematics and materials sciences, among other areas. Estimating the central orientation from a sample of such data is an important problem, which is complicated by the fact that several different approaches exist for this, motivated by various geometrical and decision-theoretic considerations. However, little is known about how such estimators compare, especially on common distributions for location models with random rotations. We examine four location estimators, three of which are commonly found in different literatures and the fourth estimator (a projected median) is newly introduced. Our study unifies existing literature and provides a detailed numerical investigation of location estimators for three commonly used rotation distributions in statistics and materials science. While the data-generating model influences the best choice of an estimator, the proposed projected median emerges as an overall good performer, which can be suggested without particular distributional assumptions. We illustrate the estimators and our findings with data from a materials science study by approximating the central orientation of cubic crystals on the micro-surface of a metal. Accompanying supplementary materials are available online.

Keywords: Cayley distribution, Electronic Backscatter Diffraction, Geodesic distance, Matrix Fisher distribution, Projected median, Rotation Group

2.1 Introduction

Data in the form of 3×3 rotation matrices find application in several scientific areas, such as biomedical engineering, computer visioning, and geological and materials sciences, where such data represent the positions of objects within some three-dimensional reference frame. For example, Rancourt et al. (2000) examine rotation matrix data in studying body positions whilst operating machinery. Fletcher et al. (2009) consider this type of orientation data in magnetic resonance imaging and in shape analysis; similar examples can be found in Schwartz and Rozumalski (2005), Pierrynowski and Ball (2009), Dai et al. (2010), or Hadani and Singer (2011). The data in our illustrative example to follow arise from a study in materials science, where 3×3 rotations represent the orientations of cubic crystals on the micro-surface of a metal specimen as measured through electron backscatter diffraction (EBSD) and "grains" within metals are composed of crystals which roughly share a common orientation; see Randle (2003) for details on EBSD data.

From a sample of orientations, an important interest is often the estimation of a main or central orientation $\boldsymbol{S}$. That is, letting the rotation group $SO(3)$ denote the collection of all 3×3 rotation matrices, observations $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n \in SO(3)$ can be conceptualized as a random sample from a *location model*

$$\mathbf{R}_i = \boldsymbol{S}\boldsymbol{E}_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\boldsymbol{S} \in SO(3)$ is the *fixed* parameter of interest indicating an orientation of central tendency, and $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n \in SO(3)$ denote i.i.d. *random* rotations which symmetrically perturb $\boldsymbol{S}$. The data-generating model in (2.1) is a rotation-matrix analog of a location model for scalar data $Y_i = \mu + e_i$, where $\mu \in \mathbb{R}$ denotes a mean and $e_i \in \mathbb{R}$ denotes an additive error symmetrically distributed around zero. This representation (2.1) for orientations is quite common and, in fact, a variety of parametric models exist for describing symmetrically distributed rotations $\boldsymbol{E}_i$, such as the symmetric matrix Fisher distribution (Downs, 1972), the symmetric Cayley distribution (León et al., 2006) and the circular-von Mises-based rotation distribution (Bingham et al., 2009) in the statistics literature, as well as Bunge's Gaussian distribution (Bunge, 1982), the isotropic Gaussian distribution (Matthies et al., 1988; Savyolova and Nikolayev, 1995) and the

de la Vallée Poussin distribution (Schaeben, 1997) in the materials science literature. Our goal in this paper is to summarize and compare the most frequently proposed approaches for the point estimation of $\boldsymbol{S}$ based on a sample of orientation data generated by (2.1). Depending on the scientific literature, the approaches can be quite different.

The topic of location estimation has received considerable attention for directional data on circles or spheres (see Fisher 1953; Karcher 1977; Khatri and Mardia 1977; Fisher 1985; Ducharme and Milasevic 1987; Bajaj 1988; Liu et al. 1992; Chan and He 1993; Mardia and Jupp 2000), but less is known about estimator properties with rotation data. As a compounding factor, several current approaches to estimating $\boldsymbol{S}$ have arisen out of literatures having differing statistical and geometrical emphases. In the applied sciences literature, estimators of $\boldsymbol{S}$ are typically based on *non-Euclidean* (i.e., Riemannian) geometry, such as the *geometric mean* (Moakher, 2002; Manton, 2004) or, more recently, the *geometric median* (Fletcher et al., 2008, 2009). Preferences may depend on potential outliers in the data, but such suggestions for estimating $\boldsymbol{S}$ often do *not* consider the potential impact of the underlying data-generating mechanism. On the other hand, approaches in the statistics literature tend to motivate an estimator for $\boldsymbol{S}$ through likelihood or moment-estimation principles applied to a specifically assumed distributional model (e.g., matrix Fisher or Cayley distribution) for the symmetric rotation errors $\boldsymbol{E}_i$ (Downs, 1972; Jupp and Mardia, 1979; León et al., 2006; Bingham et al., 2010b). Almost always, this estimator turns out to be a *projected arithmetic mean* based on *Euclidean* geometry. Hence, in addition to possible distributional assumptions, more fundamental divisions in estimation approaches may be attributable to different geometrical perspectives with rotation data.

Considering the potential effects of an underlying data generation model as well as the choice of geometry (i.e., Euclidean vs. Riemannian), the above discussion indicates a need to investigate and identify good point estimators for rotation data. In particular, because estimators in the applied sciences literatures are often selected without decision-theoretical considerations based on underlying distributions, it is of interest to understand how different location estimators behave across common distributions for rotations. In this paper, we evaluate four estimators for $\boldsymbol{S}$ in the context of the location model (2.1). These are either mean- or median-

type estimators and based either on Euclidean or Riemannian geometry; the Euclidean-based median estimator is introduced for the first time for $SO(3)$ data. Its inclusion is natural and its performance can be generally quite good so that this estimator may be broadly recommendable (as will be demonstrated). Through simulation, we compare how these estimators perform with respect to three common probability models for symmetric rotation errors as defined in (2.1), namely the circular-von Mises-based distribution, the symmetric matrix Fisher distribution and the symmetric Cayley distribution. The matrix Fisher is arguably the most common distribution in the statistics literature (see Chikuse 2003). While not noted previously, the symmetric Cayley and the de la Vallée Poussin distribution are in fact the same; the de la Vallée Poussin distribution has been advocated in the materials science literature (Schaeben, 1997). The circular-von Mises-based distribution is included because it is often applied to EBSD data (Bingham et al., 2009). We describe how properties of error distributions for rotation data, in particular their variability and tail behavior, translate into performance differences among point estimators of $\boldsymbol{S}$.

The remainder of the manuscript is organized as follows. Section 2.2 provides a brief background on the geometry of rotations and different distance metrics that can be used to assess overall estimation bias. Section 2.3 describes the location estimators for rotation data, introduces the projected median as a novel measure of location and compares their geometric underpinnings, which serves to unify some of the existing estimation literature. Section 4 explains the design of the simulation study followed by a summary of our main findings in Section 5. Section 6 provides an illustration of the estimation methods for EBSD data in a materials science application. We provide concluding remarks and future research possibilities in Section 7. Accompanying supplementary materials are available online and the R package `rotations` is currently under development to be distributed via CRAN.

2.2 Background

2.2.1 Geometry of Three-dimensional Orientations

Three-dimensional orientation data consist of observations belonging to the group $SO(3)$, where an element $\boldsymbol{R}$ in $SO(3)$ is an orthogonal 3×3 matrix (i.e., $\boldsymbol{R}^\top\boldsymbol{R} = \boldsymbol{I}_{3\times3}$) with determinant one. As $SO(3)$ is a Lie group, its elements live on a differentiable manifold. This fact is helpful in understanding the two different geometric approaches for estimating the central location $\boldsymbol{S} \in SO(3)$ from a sample of orientation data, referred to here as the *intrinsic* and the *embedding* estimation approaches (see also Jupp and Mardia 1989 and Mardia and Jupp 2000 for analogs with directional data).

The rotation group $SO(3)$ is not closed under routine addition or scalar multiplication (i.e., operations natural to statisticians). Hence, statistical estimation approaches often *embed* the rotation group into the higher-dimensional linear space consisting of all 3×3 real matrices, denoted as $\mathcal{M}(3)$. Doing so enables the use of the familiar Euclidean geometry (and "averaging" notions) to define standard distance measures and loss criteria for obtaining location estimators (see Section 2.2.2 and the estimators given in Sections 2.3.1 and 2.3.2). This embedding technique has been largely applied by statisticians, typically resulting in the projected arithmetic mean of Section 2.3.1. See, for example, Downs (1972); Khatri and Mardia (1977) and Jupp and Mardia (1979, 1989). The Bayesian estimator used in Bingham et al. (2010b) is another concrete example of this approach as is the median-type estimator we propose in Section 2.3.2.

Alternatively, *intrinsic* estimation approaches use Riemannian geometry to define distances that account for the innate topology or curvature of the space $SO(3)$. In the intrinsic approach, each rotation from $SO(3)$ is associated with a skew-symmetric matrix $\boldsymbol{\Phi}(\boldsymbol{W})$, defined as

$$\boldsymbol{\Phi}(\boldsymbol{W}) = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix}$$

for $\boldsymbol{W} = (w_1, w_2, w_3)^\top \in \mathbb{R}^3$. That is, through a so-called exponential operator, we map $\boldsymbol{\Phi}(\boldsymbol{W})$

to a rotation matrix as

$$\exp[\boldsymbol{\Phi}(\boldsymbol{W})] = \sum_{k=0}^{\infty} \frac{[\boldsymbol{\Phi}(\boldsymbol{W})]^k}{k!} = \cos(r)\boldsymbol{I}_{3\times 3} + \sin(r)\boldsymbol{\Phi}(\boldsymbol{U}) + (1-\cos r)\boldsymbol{U}\boldsymbol{U}^\top$$

where $r = \|\boldsymbol{W}\|$ and $\boldsymbol{U} = \boldsymbol{W}/\|\boldsymbol{W}\|$. The space $\mathfrak{so}(3)$ of all skew-symmetric matrices forms the tangent space (Lie-algebra) of $SO(3)$, which is closed under familiar summation and scalar multiplication operations in the usual (i.e., element-wise) manner. The fact that $SO(3)$ is a differentiable manifold allows a distance measure (the geodesic distance in Section 2.2.2) to be defined between points in $SO(3)$ according to Riemannian geometry. The resulting geodesic distance provides the basis for the "geometric" location estimators commonly found in computer science (Fletcher et al., 2008, 2009; Hartley et al., 2011) and engineering applications (Manton, 2004); see Sections 2.3.3 and 2.3.4.

Before leaving this section, it is helpful to note that each rotation matrix $\boldsymbol{R}$ can be associated with an angle-axis pair $(r, \boldsymbol{U})$, where $r \in (-\pi, \pi]$ and $\boldsymbol{U} \in \mathbb{R}^3$, $\|\boldsymbol{U}\| = 1$, through

$$\boldsymbol{R} = \boldsymbol{R}(r, \boldsymbol{U}) = \exp[\boldsymbol{\Phi}(\boldsymbol{U}r)] \in SO(3). \tag{2.2}$$

This is Euler's axis-angle representation of $\boldsymbol{R}$, where $\boldsymbol{R}$ is represented by rotating the coordinate axis $\boldsymbol{I}_{3\times 3}$ about the axis $\boldsymbol{U} \in \mathbb{R}^3$ by the angle r. In the materials science literature, $\boldsymbol{U}$ and r are commonly referred to as the misorientation axis and misorientation angle of $\boldsymbol{R}$ with respect to $\boldsymbol{I}_{3\times 3}$; see Randle (2003).

2.2.2 Choice of Distance Metrics

The choice of geometry, i.e. Riemannian or Euclidean, results in two different metrics to measure the distance between two rotation matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2 \in SO(3)$. Under the embedding approach, the natural distance metric between two random matrices is the Euclidean distance, d_E, which is induced by the Frobenius norm

$$d_E(\boldsymbol{R}_1, \boldsymbol{R}_2) = \|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_F, \tag{2.3}$$

where $\|\boldsymbol{A}\|_F = \sqrt{\mathbf{tr}(\boldsymbol{A}^\top \boldsymbol{A})}$ denotes the Frobenius norm of a matrix $\boldsymbol{A}$ and $\mathbf{tr}(\cdot)$ denotes the trace of a matrix. The Euclidean distance between two rotation matrices corresponds to the

shortest chord in $\mathcal{M}(3)$ that connects both matrices. (For an illustrative example of $d_E(\boldsymbol{R}_1, \boldsymbol{R}_2)$, we refer to Figure 2.1 where $d_E(\boldsymbol{R}_1, \boldsymbol{R}_2)$ corresponds to the gray line.) If $r \in (-\pi, \pi]$ denotes the misorientation angle in the angle-axis representation (2.2) of $\boldsymbol{R}_1^\top \boldsymbol{R}_2 \equiv \boldsymbol{R}_1^\top \boldsymbol{R}_2(r, \boldsymbol{U})$ (so that $\mathbf{tr}(\boldsymbol{R}_1^\top \boldsymbol{R}_2) = 1 + 2\cos r$), then $d_E(\boldsymbol{R}_1, \boldsymbol{R}_2) = 2\sqrt{2}\sin(|r|/2)$ holds (see the supplementary material online for a short proof of this).

By staying in the Riemannian space $SO(3)$ under the intrinsic approach, the natural distance metric becomes the Riemannian (or geodesic) distance, d_R, between two rotations $\boldsymbol{R}_1, \boldsymbol{R}_2 \in SO(3)$ defined as

$$d_R(\boldsymbol{R}_1, \boldsymbol{R}_2) = \frac{1}{\sqrt{2}} ||\mathrm{Log}(\boldsymbol{R}_1^\top \boldsymbol{R}_2)||_F = |r|, \tag{2.4}$$

where $\mathrm{Log}(\boldsymbol{R})$ denotes the principle logarithm of $\boldsymbol{R}$ (i.e., $\mathrm{Log}(\boldsymbol{R}) = \mathrm{Log}(\boldsymbol{R}(\boldsymbol{U}, r)) = \boldsymbol{\Phi}(r\boldsymbol{U})$ in (2.2)) and $r \in (-\pi, \pi]$ is the misorientation angle of $\boldsymbol{R}_1^\top \boldsymbol{R}_2$. The Riemannian distance corresponds to the length of the shortest path that connects $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ *within* the space $SO(3)$; see Figure 2.1 for an illustration. For this reason, the Riemannian distance is often considered the more natural metric on $SO(3)$; see Moakher (2002) for this discussion along with more details on exponential and logarithmic operators related to $SO(3)$.

2.3 Location Estimators

This section describes four estimators for the location parameter $\boldsymbol{S} \in SO(3)$ corresponding to orientation data generated by the model in (2.1). The estimators are based on two different choices. First, the choice whether to use the embedding approach, i.e. to base the estimator on the Euclidean distance metric defined in (2.3) or, alternatively, to use the intrinsic approach by employing the Riemannian distance metric as defined in (2.4). The second choice concerns the decision-theoretic loss functions, i.e., either using squared deviations (an L_2-norm) or absolute deviations (an L_1-norm). The extent to which the choice of geometry or loss function matters in the estimation of $\boldsymbol{S}$ will be an important aspect explored in Section 2.4. We provide an overview of three commonly used estimators and the newly defined projected median along with their properties in Table 2.1.

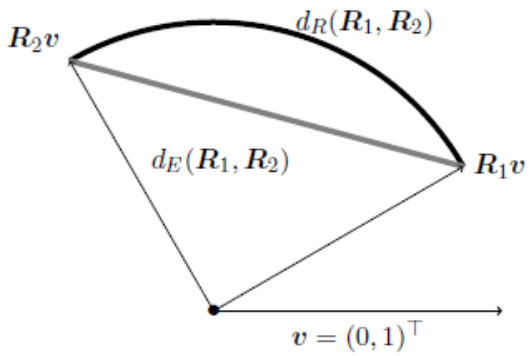


Figure 2.1: An illustration of the Euclidean and Riemannian distance metric, where to simplify the visualization, we use $SO(2)$ (rotations of points on the $\mathbb{R}^2$ unit circle) in place of $SO(3)$. Here $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ are 2×2 rotation matrices in $SO(2)$, where $\boldsymbol{R}_1\boldsymbol{v}$ and $\boldsymbol{R}_2\boldsymbol{v}$ are points on the $\mathbb{R}^2$ unit circle after rotating $\boldsymbol{v} = (0, 1)^\top$ by $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, respectively. $d_R(\boldsymbol{R}_1, \boldsymbol{R}_2)$ is displayed by the curved line (black), $d_E(\boldsymbol{R}_1, \boldsymbol{R}_2)$ by the straight line (gray).

2.3.1 The Projected Arithmetic Mean

We begin with the definition of the arithmetic mean for orientation data, as its analog is most frequently encountered in the statistical literature for directional data (e.g., see Mardia and Jupp 2000). For a sample of n random rotations $\boldsymbol{R}_i \in SO(3)$, $i = 1, 2, \ldots, n$, this mean-type estimator is defined as

$$\widehat{\boldsymbol{S}}_E = \underset{\boldsymbol{S} \in SO(3)}{\arg\min} \sum_{i=1}^{n} d_E^2(\boldsymbol{R}_i, \boldsymbol{S}) = \underset{\boldsymbol{S} \in SO(3)}{\arg\max}\, \mathbf{tr}(\boldsymbol{S}^\top \bar{\boldsymbol{R}}) \tag{2.5}$$

where $\bar{\boldsymbol{R}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{R}_i$. The estimator is obtained by minimizing the sum of the squared deviations in the Euclidean sense in the ambient space $\mathcal{M}(3)$, which then is projected back into $SO(3)$. Moakher (2002), who studied the mathematical characteristics of this estimator in detail, therefore refers to it as the *projected arithmetic mean*. This estimator's appeal lies in its simplicity and statistically intuitive nature, though it has been noted that the estimator is not invariant under rigid transformations (see Moakher 2002). However, the estimator does correspond to the maximum likelihood estimator of $\boldsymbol{S}$ when the symmetrically distributed rotation errors in (2.1) follow a matrix Fisher distribution (Jupp and Mardia, 1979). León

Table 2.1: An overview of the estimators and their underlying geometry and loss function.

Estimator name	**Denoted**	**Distance metric**	**Minimizer of loss function**
Projected Arithmetic Mean	$\widehat{\boldsymbol{S}}_E$	Euclidean	$\sum_{i=1}^n d_E^2$
Projected Median	$\widetilde{\boldsymbol{S}}_E$	Euclidean	$\sum_{i=1}^n d_E$
Geometric Mean	$\widehat{\boldsymbol{S}}_R$	Riemannian	$\sum_{i=1}^n d_R^2$
Geometric Median	$\widetilde{\boldsymbol{S}}_R$	Riemannian	$\sum_{i=1}^n d_R$

et al. (2006) also derived this estimator as the method of moment estimator under a Cayley distribution, and Bingham et al. (2009) showed that the projected arithmetic mean corresponds to the maximum quasi-likelihood estimator for orientation data with rotation errors arising from the circular-von Mises-based distribution. For a numerical implementation of $\widehat{\boldsymbol{S}}_E$ we refer to algorithms proposed by Arun et al. (1987) and Horn et al. (1988) as well as to Umeyama (1991) for refinements of their solutions including special cases such as $\det(\bar{\boldsymbol{R}}) = 0$.

2.3.2 The Projected Median

Previously proposed estimators for $X_1, \ldots, X_n \in \mathbb{R}^p$ include, for example, the Euclidean median (also known as the Weber point (Bajaj, 1988)), the mediancentre (Gower, 1974) or the projection median (Durocher and Kirkpatrick, 2009). For directional and spherical data, exemplary estimators include the circular median (Mardia, 1972), the normalized spatial median (Ducharme and Milasevic, 1987) and the Fisher median (Fisher, 1985) also known as the spherical median. Chan and He (1993) compare the performance of the normalized spatial median, an L_1 estimator on the sphere by He and Simpson (1992) and the Fisher median for the central direction for spherical data following the von Mises-Fisher distribution. They conclude that the normalized spatial median estimator is preferable for spherical data under the von Mises-Fisher model.

A modification of the estimator from Section 2.3.1 in line with these proposals is obtained by replacing the squared distances in (2.5) with absolute distances, leading to a median-type

estimator defined as

$$\widetilde{\boldsymbol{S}}_E = \underset{\boldsymbol{S}\in SO(3)}{\arg\min} \sum_{i=1}^{n} d_E(\boldsymbol{R}_i, \boldsymbol{S}). \tag{2.6}$$

We will refer to this estimator of $\boldsymbol{S}$ as the *projected median*. Although median-type estimators exist for high dimensional and directional high dimensional data, such estimators have not been defined for rotational data.

We next propose an algorithm to compute the projected median (2.6). We base our method on the Weiszfeld algorithm originally given by Weiszfeld (1937). The algorithm requires an initial value that does not equal any sample point. Note that the solution is generally not sensitive to the choice of starting points unless the data exhibit extreme spread.

1. Set $\widetilde{\boldsymbol{S}} = \widehat{\boldsymbol{S}}_E$ and choose an arbitrarily small stopping rule ε.

2. For $i = 1, \ldots, n$ compute $\boldsymbol{s}_i = \boldsymbol{R}_i - \widetilde{\boldsymbol{S}}$.

3. Calculate

$$\bar{\boldsymbol{R}}_W = \frac{\sum_{i=1}^{n} \boldsymbol{R}_i / ||\boldsymbol{s}_i||_F}{\sum_{i=1}^{n} 1/||\boldsymbol{s}_i||_F}$$

   which we call the weighted mean with respect to $\widetilde{\boldsymbol{S}}$. Note that in theory the probability that $||\boldsymbol{s}_i||_F = 0$ equals zero, but in practice we impose a lower limit on $||\boldsymbol{s}_i||_F = \delta > 0$ (for some small $\delta \in \mathbb{R}$) to avoid an undefined result.

4. Define $\widetilde{\boldsymbol{S}}_{\text{new}} = \arg\max_{\boldsymbol{S}\in SO(3)} \mathbf{tr}(\boldsymbol{S}^\top \bar{\boldsymbol{R}}_W)$ as the $\mathcal{M}(3)$ projection of $\bar{\boldsymbol{R}}_W$; see (2.5).

5. If $\varepsilon > ||\widetilde{\boldsymbol{S}} - \widetilde{\boldsymbol{S}}_{\text{new}}||_F$ return $\widetilde{\boldsymbol{S}}_E = \widetilde{\boldsymbol{S}}_{\text{new}}$; otherwise set $\widetilde{\boldsymbol{S}} = \widetilde{\boldsymbol{S}}_{\text{new}}$ and return to step 2.

For $\mathbb{R}^d$ data steps 2 and 3 of the algorithm agree with the Weiszfeld algorithm, which has been shown to converge to the d-dimensional median for convex cost functions such as the Frobenius norm (c.f. Weiszfeld (1937) for details). Because $\bar{\boldsymbol{R}}_W$ does not lie in $SO(3)$ but its ambient space $\mathcal{M}(3)$, $\bar{\boldsymbol{R}}_W$ must be projected into $SO(3)$ (step 4). Step 5 evaluates the convergence of the algorithm in $SO(3)$. A short argument that this algorithm indeed converges to $\widetilde{\boldsymbol{S}}_E$ as defined in (2.6) is provided in the online supplementary material.

2.3.3 The Geometric Mean

As sketched in Section 2.2.1, the Lie group property of $SO(3)$ provides us with a convenient transform from $SO(3)$ into the tangent space $\mathfrak{so}(3)$ that is closed under addition and scalar multiplication. Obtaining the median or mean in this transformed space and projecting the result back to $SO(3)$ corresponds to the rotation that minimizes the first and second order Riemannian distances, respectively (Karcher, 1977; Moakher, 2002; Fletcher et al., 2008, 2009). Karcher (1977) made use of Riemannian manifolds to compute what is often called the Riemannian center of mass. Moakher (2002) applied Karcher's ideas to rotation matrices and defined

$$\widehat{\boldsymbol{S}}_R = \underset{\boldsymbol{S}\in SO(3)}{\arg\min} \sum_{i=1}^{n} d_R^2(\boldsymbol{R}_i, \boldsymbol{S}), \tag{2.7}$$

which was termed as the *geometric mean*. Note that the solution to (2.7) may not be unique. Uniqueness is tied to the property of geodesic convexity of the objective function in (2.7). For more information, we refer to Moakher (2002). Additionally, (2.7) generally does not have a closed-form solution making this estimator much more computationally intensive than its Euclidean counterpart (the projected arithmetic mean of Section 2.3.1). We used the algorithm proposed by Manton (2004) for implementation in our simulation study.

2.3.4 The Geometric Median

The median-type counterpart to the geometric mean was defined first in the context of spherical data by Fisher (1985) as the point on the sphere that minimizes the sum of the arc lengths to all observations in the sample. For this type of data, the resulting estimator is known as the spherical median, which is a special case of the generalized median in $\mathbb{R}^d$ proposed by Gower (1974). For spherical data, an alternative formulation to the spherical median has been given by Liu et al. (1992) in the framework of data depth leading, however, to the same solution. We give an adaptation of the spherical median to rotation matrices. Recall that the shortest geodesic path between two rotations $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ is given by the Riemannian distance $d_R(\boldsymbol{R}_1, \boldsymbol{R}_2)$. Thus the rotation matrix analog of the Fisher (1985) spherical median can be defined as

$$\widetilde{\boldsymbol{S}}_R = \underset{\boldsymbol{S}\in SO(3)}{\arg\min} \sum_{i=1}^{n} d_R(\boldsymbol{R}_i, \boldsymbol{S});$$

see also Fletcher et al. (2008, 2009). We refer to this estimator of $\boldsymbol{S}$ as the *geometric median.* Hartley et al. (2011) offers an algorithm to find the geometric median in $SO(3)$.

2.4 Simulation Study

Section 2.4.1 gives an outline of the simulation design. Section 2.4.2 briefly describes the parametric distributional models for describing symmetric rotation errors (2.1) of differing variability used in the study.

2.4.1 Design of Simulation Study

To compare the performance of the proposed location estimators for determining the central direction $\boldsymbol{S}$ given a sample of size n, we generated random rotation error samples $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n$ in model (2.1) with sizes $n = 10, 50, 100$ and 300. Without loss of generality, we set the location parameter $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$ (the identity matrix). To compare the performance of the estimators for different probability models for random rotations exhibiting the same spread, we consider varying circular variances $\nu = 0.25$, 0.50 and 0.75 (described in Section 2.4.2). For each combination of sample size, ν and choice of distribution, we generated 1,000 samples and for each sample estimated the central direction $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$ using each of the four proposed estimators. We continue with an introduction to the considered distributions for rotations in the next section.

2.4.2 Generating Random Rotations in the Location Model

We wish to compare estimators of the (fixed) location parameter $\boldsymbol{S} \in SO(3)$ under three common distributional models for describing symmetric rotation errors $\boldsymbol{E} \in SO(3)$ in a data model $\boldsymbol{R} = \boldsymbol{SE}$ (cf. eqn. 2.1): the symmetric matrix Fisher (Langevin, 1905; Downs, 1972; Khatri and Mardia, 1977; Jupp and Mardia, 1979), the symmetric Cayley (Schaeben, 1997; León et al., 2006) and the circular-von Mises-based distribution (Bingham et al., 2009). A general construction approach exists for random rotations that are symmetrically distributed around the identity matrix $\boldsymbol{I}_{3\times 3}$; see Watson (1983); Bingham et al. (2009) and Hielscher et al. (2010). To this end, let $\boldsymbol{U} \in \mathbb{R}^3$ represent a point chosen uniformly on the unit sphere and,

independently, generate a random angle r according to some circular density $C(r|\kappa)$ on $(-\pi, \pi]$, which is symmetric around 0 and where κ denotes a concentration parameter governing the spread of the circular distribution. Then, define a random rotation as $\boldsymbol{E} = \boldsymbol{E}(\boldsymbol{U}, r)$ using the constructive definition (2.2) (i.e., $\boldsymbol{E}$ represents the position of $\boldsymbol{I}_{3\times 3}$ upon rotating the standard coordinate frame in $\mathbb{R}^3$ about the random axis $\boldsymbol{U}$ by the random angle r). The resulting rotation $\boldsymbol{E}$ will be symmetrically distributed and its distributional type (i.e., matrix Fisher, Cayley or circular-von Mises-based) is determined by the form of the circular density $C(r|\kappa)$ for the (misorientation) angle r. The circular densities for the three distributions on rotations above are given in Table 2.2, where $\mathrm{I}_p(\cdot)$ denotes the Bessel function of order p defined as $\mathrm{I}_p(\kappa) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(pr)e^{\kappa\cos r}dr$. The variability in the density $C(r|\kappa)$ of the angle r controls the variability in the resulting constructed rotations and, for consistency, we use the circular variance defined as $\nu = 1-\rho$ as a measure of spread for the circular densities for r in Table 2.2. (Note that $\rho = \mathrm{E}[\cos(r)]$ is commonly referred to as the mean resultant length, confined to the range $[0, 1]$ and directly related to κ.) The values of κ corresponding to the chosen circular variances ν are given in Table 2.3.

Table 2.2: Circular densities with respect to the Lebesgue measure and circular variance ν.

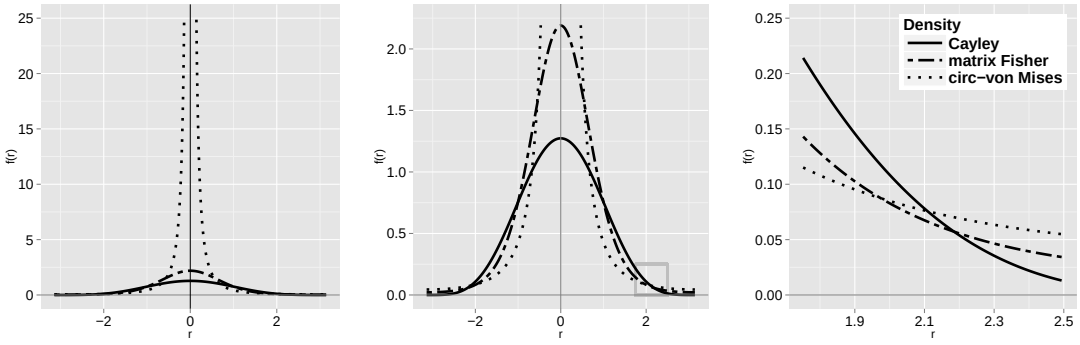
Name	**Density** $C(r\|\kappa)$	**Circular variance** ν
Cayley	$\frac{1}{\sqrt{\pi}}\frac{\Gamma(\kappa+2)}{\Gamma(\kappa+1/2)}2^{-(\kappa+1)}(1+\cos r)^{\kappa}(1-\cos r)$	$\frac{3}{\kappa+2}$
matrix Fisher	$\frac{1}{2\pi[\mathrm{I}_0(2\kappa)-\mathrm{I}_1(2\kappa)]}e^{2\kappa\cos(r)}[1-\cos(r)]$	$\frac{3\mathrm{I}_0(2\kappa)-4\mathrm{I}_1(2\kappa)+\mathrm{I}_2(2\kappa)}{2[\mathrm{I}_0(2\kappa)-\mathrm{I}_1(2\kappa)]}$
circular-von Mises	$\frac{1}{2\pi\mathrm{I}_0(\kappa)}e^{\kappa\cos(r)}$	$\frac{\mathrm{I}_0(\kappa)-\mathrm{I}_1(\kappa)}{\mathrm{I}_0(\kappa)}$

Table 2.3: Values of κ for each rotational distribution corresponding to the circular variances.

Distribution	**Circular variance** ν		
	0.25	0.50	0.75
Cayley	10.00	4.00	2.00
matrix Fisher	3.17	1.71	1.15
circular-von Mises	2.40	1.16	0.52

The density, with respect to the Haar measure, for each distribution of a random rotation

given a circular variance of 0.75 is plotted in Figure 2.3. The Haar measure (or uniform distribution on $SO(3)$) acts as the dominating measure for rotations and the symmetric nature of the random rotation $\boldsymbol{E}_i = \boldsymbol{E}_i(\boldsymbol{U}_i, r_i)$ means that its density $f(\boldsymbol{E}_i|\nu) = f(r_i|\nu)$ can be plotted in terms of the misorientation angle r_i of $\boldsymbol{E}_i$ in (2.2), which is common in materials science (Matthies et al., 1988; Savyolova and Nikolayev, 1995). Density plots with the other circular variances considered in our simulations are similar.

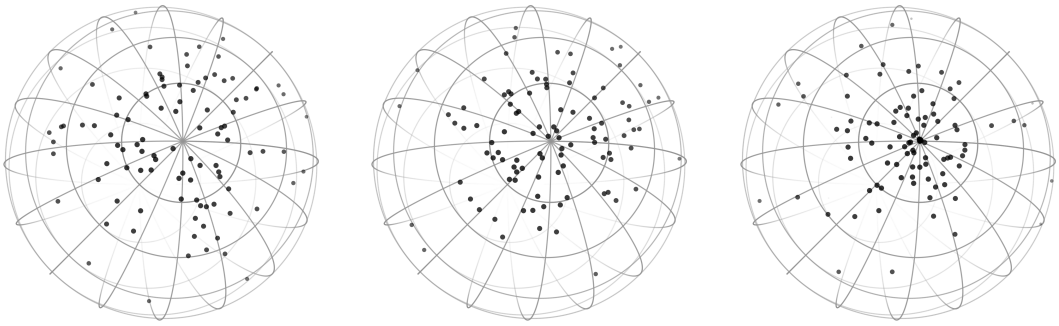


(a) Overview (b) Mode Behavior (c) Tail Behavior

Figure 2.3: Density comparison for rotation distributions with $\nu = 0.75$. The circular-von Mises based-distribution has the highest concentration, but also the heaviest tail.

In Figures 2.3a and 2.3b, we see that the circular-von Mises-based distribution has the most mass around the mode of zero whereas the Cayley distribution has the least mass around its mode. Note that the circular variances for all three distributions are indeed the same, which would visually be more clear if the densities were evaluated and plotted with respect to Lebesgue measure instead of the Haar measure. Figure 2.3c offers a better view of the tail behavior of the distributions. The circular-von Mises-based distribution has the heaviest tail, while the Cayley distribution has the least mass in the tails. This will become important in the discussion of the results. A visualization of a random sample of 100 rotations, one sample for each of the three angular distributions, is given in the sphere plots in Figure 2.5. The samples are adjusted to have a circular variance of $\nu = 0.25$. Note that Figure 2.5 shows only the first of the three columns of each rotation matrix. We refer to the supplementary material online for a complete visualization of the three samples.

In the simulation study to follow, for generating random rotation errors based on the con-



(a) Cayley (b) matrix Fisher (c) circular-von Mises

Figure 2.5: Sphere plots of the first column (x-axis) for randomly generated rotations with different distributions having $\nu = 0.25$

struction above, we randomly generate angles $r \in (-\pi, \pi]$ from a given circular density, recalling that the form of $C(r|\kappa)$ depends on the intended symmetric distribution for the rotation errors $\boldsymbol{E}$, see also the supplementary material online.

2.5 Simulation Results

In this section we summarize and present the main findings of the simulation study for estimating the central direction $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$ with the four proposed estimators of Section 2.3. We quantify the estimation error between the true location $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$ and an estimate $\widehat{\boldsymbol{S}}$ using the geodesic distance, i.e.

$$d_R(\boldsymbol{S}, \widehat{\boldsymbol{S}}) = \frac{1}{\sqrt{2}}||\text{Log}(\boldsymbol{S}^\top \widehat{\boldsymbol{S}})||_F, \quad \text{where} \quad \widehat{\boldsymbol{S}} = \widehat{\boldsymbol{S}}_E,\ \widehat{\boldsymbol{S}}_R,\ \widetilde{\boldsymbol{S}}_E \text{ or } \widetilde{\boldsymbol{S}}_R. \tag{2.8}$$

Results using d_E would prove equivalent, albeit on a smaller scale as noted in Section 2.2.2. Figure 2.6 displays side-by-side boxplots showing the estimation errors of all four estimators for a given choice of rotation distribution and circular spread ν when $n = 100$. For a tabular summary of this figure including the root mean square error (RMSE) as well as the *mean estimation error* and estimated standard errors we refer to Table 5 in the online supplement.

First and foremost the results suggest that different location estimators emerge as preferable depending on the type of distribution assumed for the rotation errors in (2.1). For the circular-von Mises-based distribution both median-type estimators ($\widetilde{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_R$) are superior with

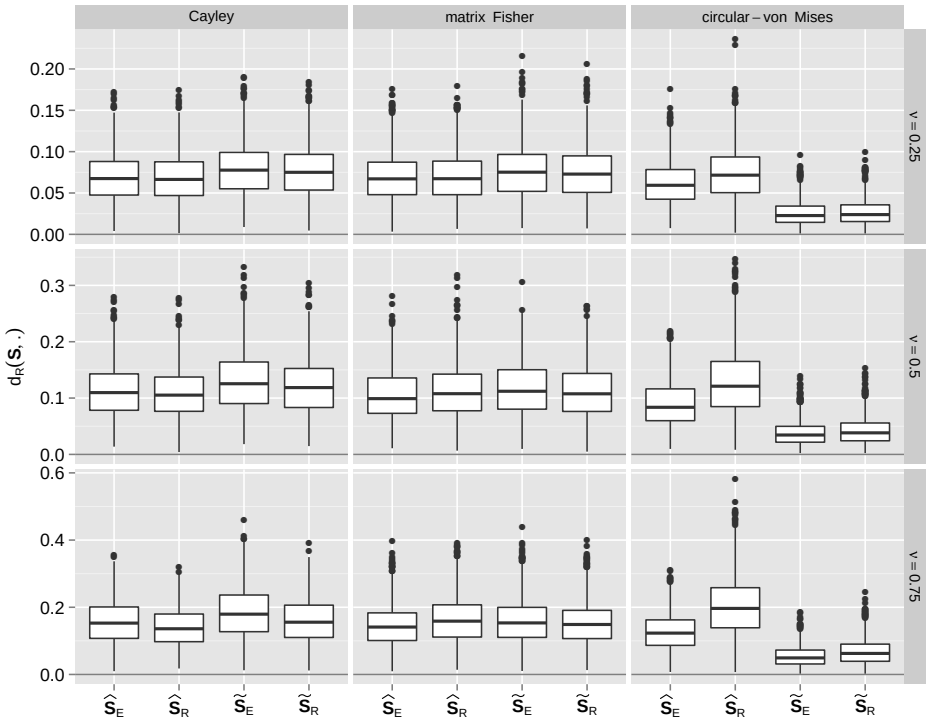


Figure 2.6: Boxplots of the estimation errors for each rotation distribution and level of ν, $n = 100$.

respect to the estimation error. For the Cayley and matrix Fisher models the mean-type estimators ($\widehat{\boldsymbol{S}}_E$ and $\widehat{\boldsymbol{S}}_R$) perform slightly better, though on a less pronounced scale. Figure 2.6 further shows that the estimation error is a function of the circular spread ν; as ν decreases the range of the observed estimation errors decreases within each rotation model and for each of the four estimators. The same holds for the mean estimation error and RMSE. Similarly, the estimation error decreases as the sample size n increases. This result is shown in Figure 10 in the online supplement.

While preferences within the median- and mean-type estimators are visible, these generally disappear as the variability in the data, i.e. ν decreases. For the Cayley and the matrix Fisher distribution the overall pattern of estimation is similar. $\widehat{\boldsymbol{S}}_E$ and $\widehat{\boldsymbol{S}}_R$ typically exhibit less spread and a lower average estimation error than $\widetilde{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_R$ with differences between the estimators lessening as ν becomes smaller.

Figure 2.7 illustrates the extent to which the mean estimation error and RMSE as a function

of sample size differ with respect to estimator choice for the circular-von Mises-based distribution when $\nu = 0.75$. We can see more clearly the advantage of the median estimators over the mean estimators across sample sizes.

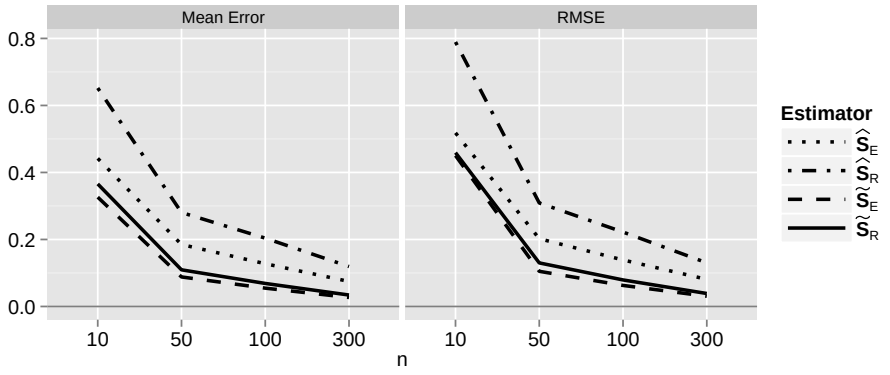


Figure 2.7: Plot of the estimation error for all levels of n for the circular-von Mises-based distribution, $\nu = 0.75$.

The previous findings raise the question why, unlike the Cayley and the Fisher matrix distribution, the circular-von Mises-based distribution so clearly distinguishes between the mean- and median-type estimators. A first insight can be obtained from Figures 2.3b and 2.3c which reveal that out of the three distributions the circular-von Mises-based distribution exhibits the heaviest tail. Thus, we can expect a larger proportion of *more extreme* observations to be sampled under the circular-von Mises-based distribution suggesting that a median-type estimator is more favorable. We use Figure 2.8 to examine the extent to which the tail-behavior indeed accounts for the observed differences in the mean- and median-type estimators. Figure 2.8 displays for each simulated sample of size $n = 300$ the proportion of observations in the sample considered to come from the tail of the distribution plotted against the difference in errors for the mean- and median-type estimators. The results shown in Figure 2.8 are with respect to the Euclidean geometry-based estimators $\widehat{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_E$. Similar results are obtained for the Riemannian geometry-based estimators $\widehat{\boldsymbol{S}}_R$ and $\widetilde{\boldsymbol{S}}_R$ and therefore are omitted. Note that we define the tail to begin at the location obtained by averaging the three pairwise crossing points in Figure 2.2c at which the densities cross for the second time. From Figure 2.8 we can see

that with an increase in the proportion of tail observations the error of the mean estimator indeed increases at a higher rate than does the error of the median estimator, i.e. the relative difference in the errors plotted on the y-axis increases. As a result the projected median is preferable to the projected mean more often as the tail becomes heavier.

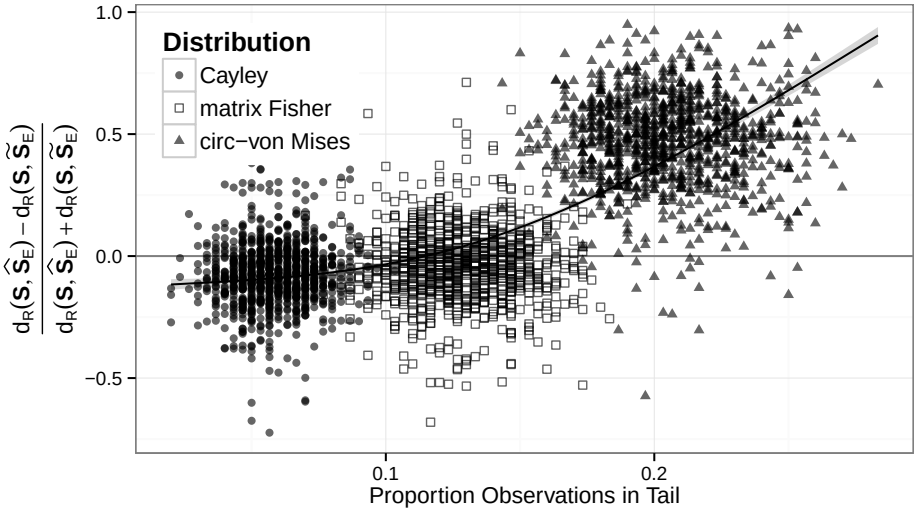
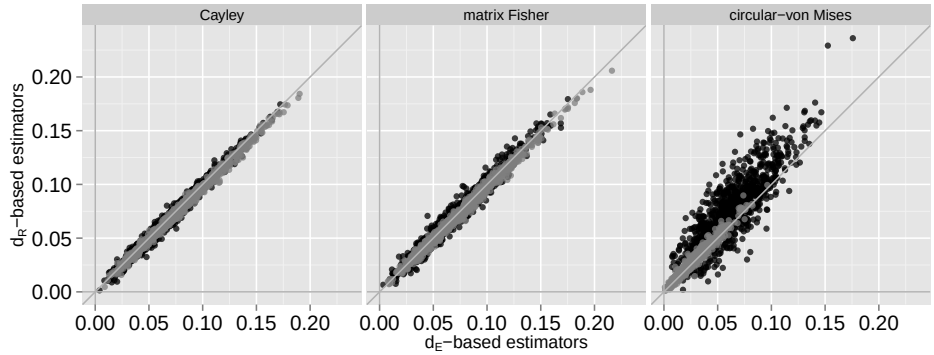


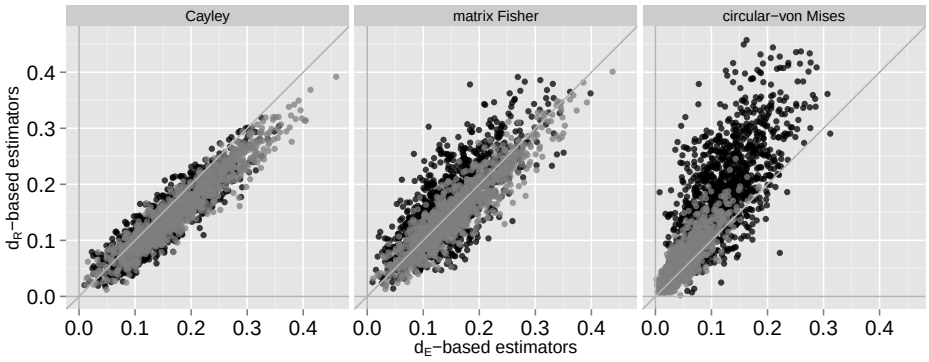
Figure 2.8: The proportion of observations in the tail against the difference in projected mean and median errors for simulated data with $n = 300$. Different symbols indicate different error distributions.

We next explore how the choice of geometric distance (Euclidean d_E or Riemannian d_R) affects the estimation error for both types of loss functions (i.e., L_2-norm or L_1-norm yielding a mean- or median type estimator, respectively). To provide more insight into the observed differences we plot in Figure 2.10 on the x-axis, for each type of loss function, the estimation errors resulting from using d_E versus the corresponding estimation errors resulting from using d_R (y-axis) for $n = 100$ and $\nu = 0.25$, $\nu = 0.75$, respectively. That is, Figure 2.10 plots estimation errors in pairs (error with $\widehat{\boldsymbol{S}}_E$, error with $\widehat{\boldsymbol{S}}_R$) and (error with $\widetilde{\boldsymbol{S}}_E$, error with $\widetilde{\boldsymbol{S}}_R$). Paired estimators based on the L_2-norm $(\widehat{\boldsymbol{S}}_E, \widehat{\boldsymbol{S}}_R)$ are represented by black dots whereas the L_1-norm $(\widetilde{\boldsymbol{S}}_E, \widetilde{\boldsymbol{S}}_R)$ based estimators correspond to light gray dots. For example, Figure 2.10 suggests that $\widetilde{\boldsymbol{S}}_R$ tends to yield less estimation error than $\widetilde{\boldsymbol{S}}_E$ for the Cayley distribution as most of the points fall below the identity line while the Riemannian distance-based estimators $\widetilde{\boldsymbol{S}}_R$ and $\widehat{\boldsymbol{S}}_R$ result in greater errors for $\boldsymbol{S}$ than their Euclidean distance based counterparts for

the circular-von Mises-based distribution. These findings support similar results for $\widetilde{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_R$ seen earlier in Figure 2.6.



(a) $\nu = 0.25$



(b) $\nu = 0.75$

Figure 2.10: Comparison of the estimation errors resulting from d_E (x-axis) and d_R (y-axis) approaches based on simulated random samples of rotations with $n = 100$. Dots in black are estimation error pairs from L_2-norm based estimators, light gray dots denote estimation error pairs for L_1-norm based estimators.

Tables 7 and 8 of the online supplement support Figure 2.10 with an exact count (expressed as a percentage) of how often d_R resulted in a smaller estimation error than d_E. Additionally, we show the average amount of error by which the Riemannian d_R- and Euclidean d_E-based estimates deviate from one another. Earlier results suggested the use of median-type estimators for the circular-von Mises-based distribution. Taking the findings with respect to the choice of geometry into account, we consider $\widetilde{\boldsymbol{S}}_E$ preferable for the circular-von Mises-based distribution as $d_R(\widetilde{\boldsymbol{S}}_E, \boldsymbol{S}) < d_R(\widetilde{\boldsymbol{S}}_R, \boldsymbol{S})$ most of the time. For the Cayley distribution our preference regard-

ing the geometry is reversed; although differences are subtle d_R is the preferred metric for the Cayley distribution especially when ν is large and overall $\widehat{\boldsymbol{S}}_R$ typically exhibits the least spread for this distribution. For the matrix Fisher distribution the preference is less clear, especially for less variable data, but as ν increases the Euclidean-based mean yields generally a smaller estimation error more often. In summary,

- the choice of location estimator can depend on the rotation error distribution in the location model (2.1). For the matrix Fisher and the Cayley distribution the projected arithmetic mean $\widehat{\boldsymbol{S}}_E$ and the geometric mean $\widehat{\boldsymbol{S}}_R$ are, respectively, preferable though $\widetilde{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_R$ are not far behind especially when the circular spread is smaller. For the circular-von Mises-based distribution the projected median $\widetilde{\boldsymbol{S}}_E$ should be used.

- a significant finding of the simulation results is that the (Euclidean-based) projected median $\widetilde{\boldsymbol{S}}_E$ is typically a good location estimator across rotation error models. For the circular-von Mises-based estimation, this generally has the best performance, while for the Cayley or matrix Fisher distributions, this estimator is often quite comparable to the best estimator. In other words, $\widetilde{\boldsymbol{S}}_E$, an estimator not previously considered for rotation matrices in the literature, appears to be a good overall choice, particularly in small samples and without knowledge of the underlying rotation error distribution.

2.6 Data Application

We consider electron backscatter diffraction (EBSD) data obtained by scanning a fixed 12.5 μm $\times$ 10 μm nickel surface at individual locations spaced 0.2 μm apart. This scan was repeated 14 times for each location yielding a total of 3,449 observations (Bingham et al., 2009, 2010a). Every observation corresponds to the orientation, expressed as a rotation matrix, of a cubic crystal on the metal surface at a particular location. One goal of processing ESBD data is to identify the main orientation of cubic crystals in the metal, where regions of cubic crystals with similar orientations constitute "grains" on the metal surface. In the material sciences literature mean-type estimators are commonly used for this problem (Humbert et al., 1996; Cho et al., 2005; Bachmann et al., 2010), thus making the estimation of the main direction $\boldsymbol{S}$ for a sample

of rotations relevant here.

At every location, using the 14 repeat scans we computed the misorientation angle $|r| = d_R(\widehat{\boldsymbol{S}}, \boldsymbol{I}_{3\times3})$ of the four estimators ($\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{S}}_E$, $\widetilde{\boldsymbol{S}}_E$, $\widehat{\boldsymbol{S}}_R$, and $\widetilde{\boldsymbol{S}}_R$) to compare resulting differences in the corresponding estimates. In the following we will focus particularly on the differences between $\widehat{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_E$ as the Riemannian estimators $\widetilde{\boldsymbol{S}}_R$ and $\widehat{\boldsymbol{S}}_R$ largely agree with their Euclidean counterparts.

Figure 2.11a illustrates the implementation of $\widetilde{\boldsymbol{S}}_E$: each location on the plot is colored according to the mis-orientation angle between the estimate $\widetilde{\boldsymbol{S}}_E$ and the identity $\boldsymbol{I}_{3\times3}$ (as an arbitrarily chosen reference point). The plot shows a distinct spatial structure resembling a grain map. In Figure 2.11b we illustrate the difference between $\widehat{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_E$ at each location. The difference in estimates is again defined with respect to the mis-orientation angle between both estimates and locations are colored accordingly.

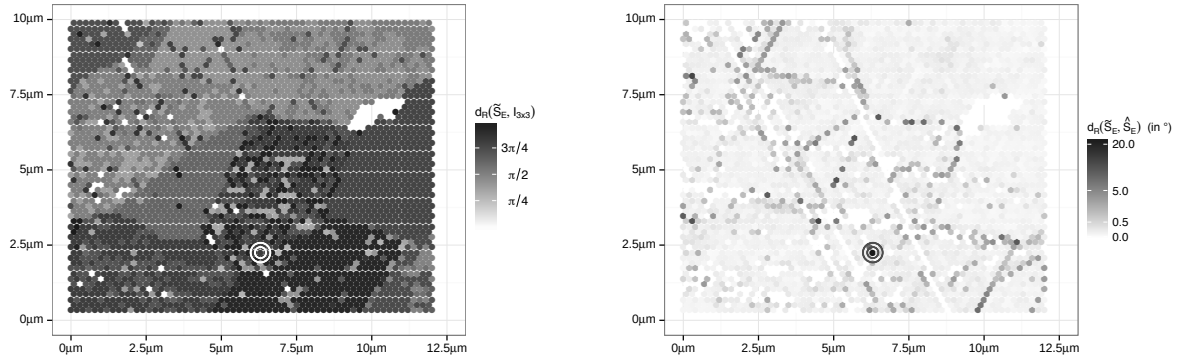


Figure 2.11: Display of all locations of the investigated nickel surface. Each dot corresponds to one location and shading reflects angles. The location of the the largest difference between estimators $\widetilde{\boldsymbol{S}}_E$ and $\widehat{\boldsymbol{S}}_E$ is circled. Panel (a) on the left shows angles of $\widetilde{\boldsymbol{S}}_E$ with respect to identify. Panel (b) on the right shows locations shaded by the difference (in degrees) between $\widetilde{\boldsymbol{S}}_E$ and $\widehat{\boldsymbol{S}}_E$. Distances of 0.5° or more are generally considered to suggest different main orientations. Note that the mapping of distance to color shading is on a square-root scale.

Note that the literature, e.g. Bingham et al. (2010a), suggests that distances of 0.5° are indicative of different grains. In our example, about 10% of the locations result in a difference between $\widehat{\boldsymbol{S}}_E$ and $\widetilde{\boldsymbol{S}}_E$ estimates of at least that size. Differences tend to be largest along boundaries between the spatial structures in Figure 2.11b resulting in a different allocation of a location to grain depending on the choice of estimation method.

As an example, Table 2.4 contains the observed orientations (the collection of all nine coefficients x_{ij}, $1 \leq i, j \leq 3$, of the 3×3 rotation matrix) for each of the 14 repeated scans at the location with the largest observed difference between $\widetilde{\boldsymbol{S}}_E$ and $\widehat{\boldsymbol{S}}_E$, namely 22.3° (circled in Figure 2.11).

The scans have been ordered to better illustrate the clustering of the rotations observed at this particular location. This clustering is also visible in the parallel coordinate plot in Figure 2.12a. The clusters suggest that, for this location on a "grain boundary," a subset of the scans likely picked up the orientation of a neighboring cubic crystal belonging to a different grain making this a situation in which a median-type estimator is more suitable than a mean-type estimator.

Table 2.4: List of all rotations in the location with the largest difference between mean and median estimators. We observe one main cluster and one smaller cluster with three additional rotations in the proximity.

scan	x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}
1	-0.646	-0.552	-0.527	0.464	-0.833	0.303	-0.606	-0.049	0.794
2	-0.641	-0.550	-0.535	0.459	-0.834	0.307	-0.615	-0.048	0.787
3	-0.640	-0.549	-0.537	0.457	-0.834	0.309	-0.618	-0.048	0.785
4	-0.639	-0.546	-0.542	0.462	-0.836	0.297	-0.615	-0.061	0.787
5	-0.639	-0.547	-0.540	0.456	-0.835	0.307	-0.619	-0.050	0.783
6	-0.638	-0.550	-0.540	0.459	-0.834	0.307	-0.619	-0.052	0.784
7	-0.637	-0.551	-0.540	0.459	-0.833	0.309	-0.620	-0.051	0.783
8	-0.633	-0.554	-0.540	0.464	-0.830	0.309	-0.619	-0.055	0.783
9	-0.068	-0.422	-0.904	0.994	-0.105	0.025	-0.084	-0.900	0.427
10	-0.017	-0.633	-0.774	0.961	-0.224	0.162	-0.276	-0.741	0.612
11	-0.005	-0.551	-0.834	0.982	-0.158	0.099	-0.186	-0.819	0.542
12	-0.002	-0.587	-0.809	0.978	-0.167	0.124	-0.208	-0.792	0.574
13	-0.002	-0.595	-0.804	0.974	-0.182	0.132	-0.225	-0.783	0.580
14	-0.727	-0.475	-0.496	0.138	-0.809	0.572	-0.672	-0.348	0.653

We visualize Table 2.4 and the estimates resulting from applying $\widetilde{\boldsymbol{S}}_E$ and $\widehat{\boldsymbol{S}}_E$ at this location in Figure 2.12a using a parallel coordinate plot: for every scan each of the nine coefficients x_{ij} ($1 \leq i, j \leq 3$) is plotted separately. Coefficients that correspond to the same scan are connected by a line. Note that the coefficient values are jittered using a small perturbation in form of a rotation matrix to avoid over-plotting and to illustrate cluster sizes. The light- and dark-green colored lines represent the $\widetilde{\boldsymbol{S}}_E$ and $\widehat{\boldsymbol{S}}_E$ estimates of the main direction based on the 14 scans. This example illustrates that the median estimator, $\widetilde{\boldsymbol{S}}_E$, is the more robust estimate of

the main direction in the presence of potential anomalies on the spatial boundaries of grains, as this estimate is located within the largest cluster of rotations, while the projected mean is pulled into a location between the clusters.

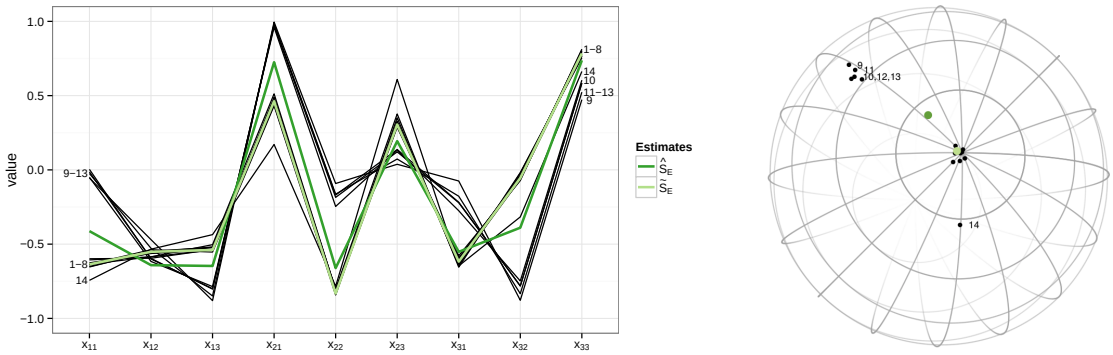


Figure 2.12: Parallel coordinated plot (a) on the left, and a sphere plot of the y axis (b) on the right show all fourteen rotation matrices of table 2.4, the location with the largest angle difference between projected mean and median estimator.

2.7 Recommendations and Conclusions

The scientific literature suggests a variety of approaches to estimate the central orientation $\boldsymbol{S}$ given a random sample of three-dimensional orientations from (2.1). These approaches differ largely with respect to the geometry (Riemannian vs. Euclidean) of estimation, assumptions about the underlying data-generating mechanism, and the choice of loss function when defining suitable estimators. The main goal of this paper was to explore the extent to which these choices affect the estimation of $\boldsymbol{S}$. Our simulation study showed that the underlying data-generating mechanism guides our choice of loss function. For the circular-von Mises-based model median-type estimators perform better while for the Cayley and matrix Fisher model the mean-type estimators show less estimation error and variability. As noted in Section 2.1 the applied sciences generally pursue estimation of $\boldsymbol{S}$ without considering the distributional underpinnings. Restricting ourselves to the three rotation distributions under consideration, if indeed nothing is known about the underlying data-generating mechanism we suggest to use either median-type estimator, where the proposed median, the Euclidean based estimator $\widetilde{\boldsymbol{S}}_E$,

emerges as a good overall choice. Its overall estimation error, even under mis-specification, will be much less than the potential estimation error resulting from either mean-type estimator. The extent to which all four estimators disagree depends on the variability in the rotation data; the estimators differ more when the circular variance ν is large and tend to agree more as the data become more concentrated.

Further studies could be extended to include location estimation in non-symmetric distributional models for rotations as we considered common models for symmetric perturbations around $\boldsymbol{S}$ in (2.1). Another important consideration is the extension of the studied point estimators to building confidence regions for the location parameter $\boldsymbol{S}$. One possibility would be resampling-based confidence regions, but this requires theoretical development for the estimator's sampling distributions and improvements in computing time before this can be practically implemented.

Supplementary Material

Supplementary document: This document first proves that the relationship between d_E and d_R is as stated in Section 2.2.2 and that the algorithm in Section 2.3.2 converges to $\widetilde{\boldsymbol{S}}_E$. Then we extend our discussion of the simulation results from Section 2.5. (Supplement.pdf)

Code and data files: The R code and data files necessary to make the figures and tables in this manuscript are packaged here. (Supplement.zip)

Acknowledgements: The authors wish to thank Dr. Melissa Bingham and the Ames Laboratory for collecting and providing the EBSD data. We also gratefully acknowledge the suggestions made by the reviewers, the Associate Editor and Editor, all of which significantly improved the manuscript.

2.8 Supplementary Material

Proofs of Claims

Relationship Between Riemannian and Euclidean Distance in $SO(3)$

Claim 1. For rotation matrices $\boldsymbol{R}_i \in SO(3)$, $i = 1, 2, \ldots, n$ and distances d_E and d_R as defined in the main manuscript in equations (3) and (4), respectively, it holds that

$$d_E(\boldsymbol{R}_1, \boldsymbol{R}_2) = 2^{3/2} \sin\left(\frac{d_R(\boldsymbol{R}_1, \boldsymbol{R}_2)}{2}\right). \tag{2.9}$$

Proof:

For two rotations $\boldsymbol{R}_1, \boldsymbol{R}_2 \in SO(3)$ recall that if $\mathbf{tr}(\boldsymbol{R}_1^\top \boldsymbol{R}_2) = 1 + 2\cos(r)$ then $|r| = d_R(\boldsymbol{R}_1, \boldsymbol{R}_2)$. When $|r| = 0$, the statement in (2.9) follows directly. Consider the case $|r| > 0$. By definition we know

$$\begin{aligned}
d_E(\boldsymbol{R}_1, \boldsymbol{R}_2)^2 &= ||\boldsymbol{R}_1 - \boldsymbol{R}_2||_F^2 \\
&= \mathbf{tr}\left[(\boldsymbol{R}_1 - \boldsymbol{R}_2)^\top (\boldsymbol{R}_1 - \boldsymbol{R}_2)\right] \\
&= \mathbf{tr}\left[\boldsymbol{R}_1^\top \boldsymbol{R}_1 + \boldsymbol{R}_2^\top \boldsymbol{R}_2 - \boldsymbol{R}_2^\top \boldsymbol{R}_1 - \boldsymbol{R}_1^\top \boldsymbol{R}_2\right] \\
&= \mathbf{tr}\left[2\boldsymbol{I} - \boldsymbol{R}_2^\top \boldsymbol{R}_1 - \boldsymbol{R}_1^\top \boldsymbol{R}_2\right] \\
&= 2\mathbf{tr}(\boldsymbol{I}) - \mathbf{tr}(\boldsymbol{R}_2^\top \boldsymbol{R}_1) - \mathbf{tr}(\boldsymbol{R}_1^\top \boldsymbol{R}_2) \\
&= 6 - 2\mathbf{tr}(\boldsymbol{R}_1^\top \boldsymbol{R}_2) \\
&= 4 - 4\cos(|r|) \\
&= 8\left(\frac{1 - \cos(|r|)}{2}\right) \\
&= 8\sin^2\left(\frac{|r|}{2}\right) \\
&= \left[2^{3/2}\sin\left(\frac{|r|}{2}\right)\right]^2 \\
&= \left[2^{3/2}\sin\left(\frac{d_R(\boldsymbol{R}_1, \boldsymbol{R}_2)}{2}\right)\right]^2.
\end{aligned}$$

Taking square root on both sides gives (2.9).

Convergence of $\widetilde{\boldsymbol{S}}_E$ Algorithm

Claim 2. For a sample n of rotation matrices, $\boldsymbol{R}_i \in SO(3)$ $i = 1, 2, \ldots, n$, define $\widetilde{\boldsymbol{M}}$ to be the median in $\mathcal{M}(3)$, i.e. $\widetilde{\boldsymbol{M}} = \arg\min_{M \in \mathcal{M}(3)} \sum_{i=1}^{n} ||\boldsymbol{R}_i - \boldsymbol{M}||_F$. Then

$$\widetilde{\boldsymbol{S}}_E := \underset{\boldsymbol{S} \in SO(3)}{\arg\min} \sum_{i=1}^{n} ||\boldsymbol{R}_i - \boldsymbol{S}||_F \overset{!}{=} \underset{\boldsymbol{S} \in SO(3)}{\arg\min} ||\widetilde{\boldsymbol{M}} - \boldsymbol{S}||_F =: \widetilde{\boldsymbol{S}}_E^*.$$

Proof:

From the delta inequality of a distance measure we know that the following holds:

$$||\boldsymbol{R}_i - \widetilde{\boldsymbol{S}}_E||_F = ||\boldsymbol{R}_i - \widetilde{\boldsymbol{M}} + \widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E||_F \leq ||\boldsymbol{R}_i - \widetilde{\boldsymbol{M}}||_F + ||\widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E||_F.$$

Also, by definition of $\widetilde{\boldsymbol{S}}_E$, $\sum_{i=1}^{n} ||\boldsymbol{R}_i - \widetilde{\boldsymbol{S}}_E||_F \leq \sum_{i=1}^{n} ||\boldsymbol{R}_i - \widetilde{\boldsymbol{S}}_E^*||_F$. Therefore,

$$\begin{aligned}
0 \leq & \sum_{i=1}^{n} ||\boldsymbol{R}_i - \widetilde{\boldsymbol{S}}_E^*||_F - \sum_{i=1}^{n} ||\boldsymbol{R}_i - \widetilde{\boldsymbol{S}}_E||_F \\
\leq & \sum_{i=1}^{n} \left(||\boldsymbol{R}_i - \widetilde{\boldsymbol{M}}||_F + ||\widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E^*||_F - ||\boldsymbol{R}_i - \widetilde{\boldsymbol{M}}||_F - ||\widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E||_F \right) \\
= & n||\widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E^*||_F - n||\widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E||_F \leq 0 \qquad \text{(by definition of } \widetilde{\boldsymbol{S}}_E^*) \\
\Longrightarrow & ||\widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E^*||_F = ||\widetilde{\boldsymbol{M}} - \widetilde{\boldsymbol{S}}_E||_F \qquad (2.10)
\end{aligned}$$

The result follows if the projection of $\widetilde{\boldsymbol{M}}$ onto $SO(3)$ is unique, which holds as long as $\widetilde{\boldsymbol{M}} \neq \boldsymbol{0}_{3\times 3}$.

Sampling Processes

In the following subsection we will briefly illustrate how a sample of random rotations from each of the three rotational distributions (circular-von Mises-based, Cayley and matrix-Fisher) is obtained for the purpose of the simulation study.

Circular-von Mises-based distribution

To simulate a random sample of rotation angles from the circular-von Mises-based distribution we follow the algorithm proposed by Best and Fisher (1979). The algorithm is available in the IMSL Library (1991) and is implemented as follows. Let $\mu = 0$ denote the mean of the target angular distribution and κ its concentration parameter. We define constants a, b and

d as $a \equiv 1 + \sqrt{1 + 4\kappa^2}$, $\quad b \equiv (a - \sqrt{2a})$, $\quad d \equiv (1 + b^2)/2b$. In steps one, two and four we generate three new observations u_1, u_2 and u_3, each from a uniform distribution defined over the interval $(0, 1)$.

1. Set $z = \cos(\pi u_1)$, $f = (1 + dz)/(z + d)$ and $c = \kappa(d - f)$.
2. If $c(2 - c) - u_2 > 0$ go to step 4.
3. If $\log(c/u_2) + 1 - c < 0$ return to step 1.
4. Set $r = \text{sign}(u_3 - 0.5) \cos^{-1}(f)$.

It follows that r is distributed according to the circular-von Mises(κ) distribution.

Cayley distribution

To simulate rotation angles from a Cayley distribution we make use of a result given in León et al., (2006). If the angle r follows a Cayley distribution it holds that $\frac{1+\cos r}{2} \sim \text{Beta}(\kappa + 1/2, 3/2)$. Hence, angles following the Cayley distribution can be simulated through composition:

1. Generate $Z \sim$Bernoulli(0.5) and set $Y = 1 - 2Z$.
2. Independently generate $X \sim \text{Beta}(\kappa + 1/2, 3/2)$.
3. Set $r = \frac{Y}{2} \cos^{-1}(2X - 1)$.

Angles r simulated in this fashion follow a Cayley(κ) distribution.

matrix Fisher distribution

Simulation from the matrix Fisher distribution is achieved through a rejection algorithm. Let $\text{C}_\text{F}(r|\kappa)$ denote the matrix Fisher density.

1. Define $M = \frac{1}{2\kappa} e^{2\kappa - 1} \frac{1}{\mathbf{I}_0(2\kappa) - \mathbf{I}_1(2\kappa)}$.
2. Generate $U \sim \text{Uniform}(0, 1)$ and $Y \sim \text{Uniform}(-\pi, \pi]$, where U and Y are independent.
3. If $U < \frac{1}{M} \text{C}_\text{F}(Y|\kappa)$, accept Y; otherwise return to step (2).

Additional Simulation Study Results

We expand on some of the results given in Section 5 of the main manuscript by providing additional numerical results to support graphical displays as well as to further clarify the relationship between the different estimators.

Figure 4 of the main manuscript showed boxplots of the estimation errors for a each of a 1,000 samples of size $n = 100$ for all three distributions and choices of the circular variance ν. We accompany this figure with Table 2.5, which provides numerical summaries of the errors displayed in each boxplot showing the mean estimation error $\overline{d_R}(\boldsymbol{S}, \widehat{\boldsymbol{S}})$ in the $1,000$ simulation runs, the estimated standard error $SE(\overline{d_R})$ and the estimated $RMSE$ for each estimator. Although the boxplots of the estimation errors look very similar in Figure 4, the estimated standard errors in Table 2.5 suggest that on average some of the estimators differ significantly.

Table 2.5: Mean estimation error, respective standard error and RMSE for $n = 100$ based on 1,000 simulation runs. Despite skewness in some of the plotted error distributions the *median estimation error* was quantitatively similar to the mean estimation error and therefore is not reported.

		Cayley		**matrix Fisher**		**circular-von Mises**	
ν	Estimator	$\overline{d_R}(\boldsymbol{S}, \widehat{\boldsymbol{S}})$ $SE(\overline{d_R})$	RMSE	$\overline{d_R}(\boldsymbol{S}, \widehat{\boldsymbol{S}})$ $SE(\overline{d_R})$	RMSE	$\overline{d_R}(\boldsymbol{S}, \widehat{\boldsymbol{S}})$ $SE(\overline{d_R})$	RMSE
0.25	$\widehat{\boldsymbol{S}}_R$	0.0690 (0.0009)	0.0752	0.0699 (0.0010)	0.0761	0.0744 (0.0010)	0.0811
	$\widehat{\boldsymbol{S}}_E$	0.0698 (0.0009)	0.0759	0.0695 (0.0009)	0.0756	0.0617 (0.0008)	0.0671
	$\widetilde{\boldsymbol{S}}_R$	0.0769 (0.0010)	0.0834	0.0747 (0.0010)	0.0813	0.0269 (0.0005)	0.0310
	$\widetilde{\boldsymbol{S}}_E$	0.0791 (0.0011)	0.0858	0.0766 (0.0010)	0.0832	0.0256 (0.0005)	0.0296
0.50	$\widehat{\boldsymbol{S}}_R$	0.1086 (0.0014)	0.1174	0.1121 (0.0015)	0.1219	0.1279 (0.0018)	0.1406
	$\widehat{\boldsymbol{S}}_E$	0.1129 (0.0015)	0.1222	0.1054 (0.0014)	0.1143	0.0894 (0.0012)	0.0976
	$\widetilde{\boldsymbol{S}}_R$	0.1210 (0.0016)	0.1313	0.1113 (0.0015)	0.1211	0.0426 (0.0008)	0.0491
	$\widetilde{\boldsymbol{S}}_E$	0.1295 (0.0017)	0.1407	0.1160 (0.0016)	0.1262	0.0379 (0.0007)	0.0438
0.75	$\widehat{\boldsymbol{S}}_R$	0.1398 (0.0018)	0.1514	0.1703 (0.0045)	0.2225	0.2039 (0.0028)	0.2221
	$\widehat{\boldsymbol{S}}_E$	0.1567 (0.0020)	0.1695	0.1462 (0.0020)	0.1588	0.1276 (0.0017)	0.1388
	$\widetilde{\boldsymbol{S}}_R$	0.1597 (0.0021)	0.1729	0.1527 (0.0021)	0.1660	0.0687 (0.0012)	0.0792
	$\widetilde{\boldsymbol{S}}_E$	0.1847 (0.0024)	0.2000	0.1597 (0.0022)	0.1736	0.0547 (0.0010)	0.0628

To establish significant differences more formally we conducted *matched pair t-tests* (two-sided) for all six pairwise comparisons of the four estimators within a specific simulation setting. The results are given in Table 2.6. We tested the null hypothesis that the difference in the resulting estimates, on average, is zero against the alternative hypothesis that the difference, on average, is non-zero. Because differences in the estimation error seem less obvious for the

Cayley and matrix-Fisher distribution we conducted tests for both distributions for samples of size 10 and 100 and circular variances of 0.25 and 0.75. We suspect that a *matched pair t-test* based on all of the $B = 1,000$ simulations will likely yield statistically significant differences that are an artifact of the size of B as opposed to meaningful and practical differences we also provide the results for (arbitrary) choices of $B = 50$ and $B = 100$. For $B = 50$ and $B = 100$ we base the test on the estimation results for B randomly selected simulation runs and repeated this process 100 times. The reported *p-value* then corresponds to the average *p-value* of these 100 runs. For $B = 1,000$ the reported *p-value* is based on all original 1,000 runs. We further adjusted the level of significance within each row of Table 2.6 using a Bonferroni correction for multiple comparisons and therefore, within a set of all six pairwise comparisons, consider *p-values* less than $0.05 \div 6 = 0.0083$. We chose the Bonferroni adjustment because of its simplicity.

From Table 2.6 we can conclude that for the Cayley distribution the choice of geometry is important when using median type estimators (column 1), but not as important for mean type estimators (column 2) unless the circular variance and sample size is large. For a given geometry, the difference between mean and median type estimator depends on the level of variability in the data as is the case in the one dimension case (columns 3 and 4). As for the matrix-Fisher distribution, the choice of estimator does not appear to be overtly significant for any sample size or circular variance. This is likely due to the fact that this distribution closely resembles the normal distribution on the range $[-\pi, \pi)$ in which case the mean and median are equivalent.

Figure 2.13 illustrates the behavior of the estimators as a function of the sample size. Results are displayed for $\nu = 0.75$ at each sample size. As to be expected, the estimation error decreases as the sample size increases for all three distributions. For small samples, e.g., $n = 10$, the estimator exhibiting the largest amount of variability is the geodesic mean $\widehat{\boldsymbol{S}}_R$. This behavior is consistent for all three distributions. While the estimator's variability lessens considerably for the Cayley and matrix Fisher distribution as n increases, the estimator remains the most variable estimator for the circular-von Mises-based distribution. A possible explanation for this behavior is that the algorithm to estimate $\widehat{\boldsymbol{S}}_R$ uses a random sample point in its initiating

Table 2.6: P-values for matched pair t-tests on the average differences between estimators. Unless B equals 1,000 the reported p-values correspond to the average of 100 p-values based on random samples of size B from 1,000 simulation runs.

Distribution	ν	n	B	$\widetilde{\boldsymbol{S}}_E - \widetilde{\boldsymbol{S}}_R$	$\widehat{\boldsymbol{S}}_E - \widehat{\boldsymbol{S}}_R$	$\widehat{\boldsymbol{S}}_E - \widetilde{\boldsymbol{S}}_E$	$\widehat{\boldsymbol{S}}_R - \widetilde{\boldsymbol{S}}_R$	$\widehat{\boldsymbol{S}}_E - \widetilde{\boldsymbol{S}}_R$	$\widetilde{\boldsymbol{S}}_E - \widehat{\boldsymbol{S}}_R$
Cayley	0.25	10	50	**0.0012**	0.4106	0.0181	0.0591	0.0385	0.0298
			100	**<0.0001**	0.3438	**0.0004**	**0.0040**	**0.0022**	**0.0008**
			1000	**<0.0001**	**0.0022**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
		100	50	**<0.0001**	0.2937	0.0110	0.0360	0.0248	0.0179
			100	**<0.0001**	0.1356	**0.0002**	**0.0012**	**0.0007**	**0.0003**
			1000	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
	0.75	10	50	**0.0001**	0.0343	**0.0008**	0.0376	0.3306	**0.0005**
			100	**<0.0001**	**0.0015**	**<0.0001**	**0.0018**	0.2848	**<0.0001**
			1000	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
		100	50	**<0.0001**	**0.0017**	**0.0001**	**0.0024**	0.3292	**<0.0001**
			100	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0.3046	**<0.0001**
			1000	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
matrix-Fisher	0.25	10	50	0.0116	0.5032	0.0791	0.2135	0.1338	0.1394
			100	**0.0012**	0.3815	0.0213	0.1171	0.0540	0.0586
			1000	**<0.0001**	**0.0008**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
		100	50	**0.0072**	0.4314	0.0773	0.2313	0.1200	0.1620
			100	**0.0001**	0.4250	**0.0047**	0.0603	0.0143	0.0244
			1000	**<0.0001**	0.0303	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
	0.75	10	50	0.1027	0.1864	0.0842	0.3483	0.3509	0.4652
			100	0.0205	0.0730	**0.0075**	0.2476	0.1666	0.5001
			1000	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0.1050
		100	50	0.1475	0.0586	0.0559	0.2005	0.2296	0.4738
			100	0.0419	**0.0081**	**0.0056**	0.0747	0.0768	0.4437
			1000	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0.0198

step. For small samples or samples with several extreme observations it is likely to start the algorithm far from the true center, which in turn may cause the algorithm to get stuck in a local minimum and to fail to converge globally. In practice we suggest the algorithm be started at some other location estimate of $\boldsymbol{S}$ such as the $\widehat{\boldsymbol{S}}_E$. In simulations where $\widehat{\boldsymbol{S}}_E$ was used as a starting point for the algorithm we observed similar results with less variability in the estimates of $\widehat{\boldsymbol{S}}_R$, but for the purpose of a fair comparison in this study we started the algorithm at a random sample point.

In Figure 7 of the main manuscript we examined how the choice of geometry (Riemannian

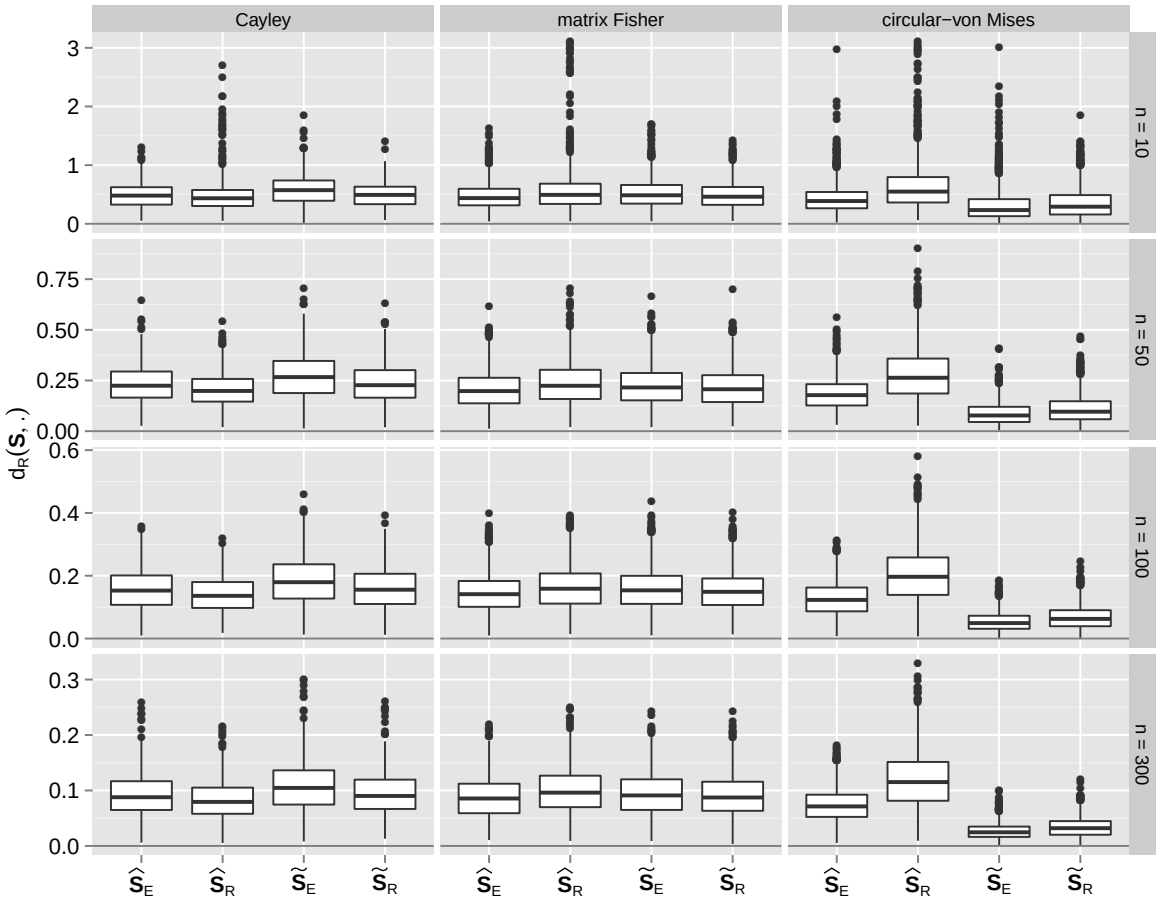


Figure 2.13: Boxplots of the estimation error for each rotation distribution and level of n, $\nu = 0.75$.

versus Euclidean) affected the estimation error under both types of loss functions L_1 and L_2. Tables 2.7 and 2.8 support Figure 7 with an exact count (expressed as a percentage) of how often d_R resulted in a smaller estimation error than d_E. Additionally, we give the average amount by which the d_R- and d_E-based estimates deviate from one another (along with a standard error estimate). We denote the latter quantity by $\bar{\delta}$ where $\delta = d_R(\widetilde{\boldsymbol{S}}_E, \boldsymbol{S}) - d_R(\widetilde{\boldsymbol{S}}_R, \boldsymbol{S})$.

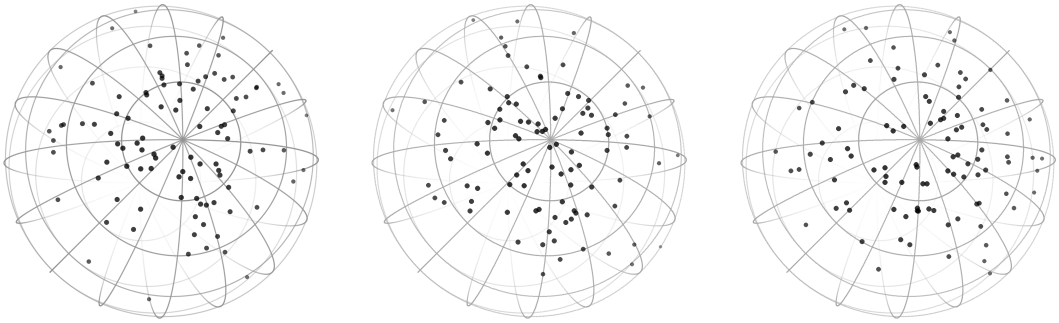
Sphere plots

Figures 2.15 to 2.19 show sphere plots for 100 samples from each of the Cayley, matrix Fisher and circular-von Mises distribution. The concentration parameter κ in each distribution

Table 2.7: Average reduction in estimation error by using $\widetilde{\boldsymbol{S}}_R$ instead of $\widetilde{\boldsymbol{S}}_E$, $\delta = d_R(\widetilde{\boldsymbol{S}}_E, \boldsymbol{S}) - d_R(\widetilde{\boldsymbol{S}}_R, \boldsymbol{S})$ with standard error and percentage of samples for which $d_R(\widetilde{\boldsymbol{S}}_R, \boldsymbol{S}) < d_R(\widetilde{\boldsymbol{S}}_E, \boldsymbol{S})$.

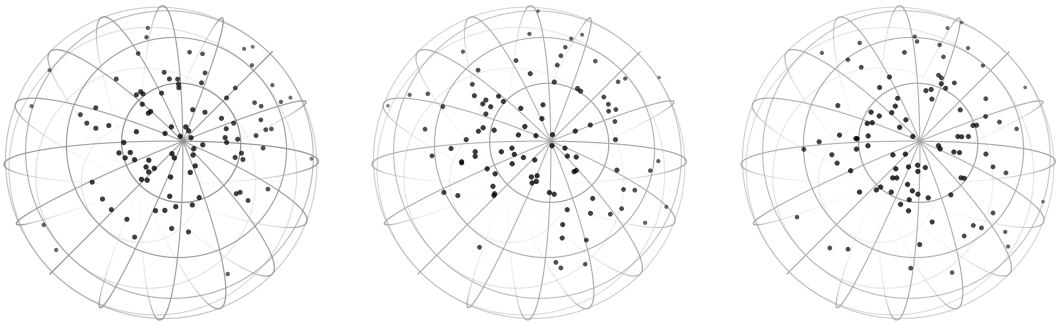
		Cayley		**matrix Fisher**		**circular-von Mises**	
ν	n	$\bar{\delta}$ (SE)	%	$\bar{\delta}$ (SE)	%	$\bar{\delta}$ (SE)	%
0.25	10	0.0078 (0.0004)	0.7430	0.0060 (0.0004)	0.7250	-0.0053 (0.0006)	0.3280
	50	0.0031 (0.0001)	0.7830	0.0024 (0.0001)	0.6970	-0.0018 (0.0001)	0.3270
	100	0.0022 (0.0001)	0.7890	0.0018 (0.0001)	0.7120	-0.0013 (0.0001)	0.3080
	300	0.0012 (0.0001)	0.7810	0.0010 (0.0001)	0.7110	-0.0008 (0.0001)	0.2840
0.50	10	0.0315 (0.0016)	0.7720	0.0171 (0.0017)	0.6620	-0.0192 (0.0020)	0.3350
	50	0.0126 (0.0005)	0.8110	0.0055 (0.0006)	0.6200	-0.0081 (0.0005)	0.2820
	100	0.0085 (0.0003)	0.8090	0.0047 (0.0004)	0.6600	-0.0047 (0.0003)	0.3020
	300	0.0049 (0.0002)	0.8040	0.0024 (0.0002)	0.6580	-0.0027 (0.0001)	0.2550
0.75	10	0.0895 (0.0042)	0.8210	0.0340 (0.0040)	0.6330	-0.0396 (0.0062)	0.3220
	50	0.0366 (0.0011)	0.8660	0.0093 (0.0013)	0.5970	-0.0213 (0.0011)	0.2380
	100	0.0250 (0.0008)	0.8580	0.0069 (0.0009)	0.6030	-0.0140 (0.0007)	0.2400
	300	0.0140 (0.0004)	0.8500	0.0033 (0.0005)	0.5890	-0.0072 (0.0003)	0.2180

is chosen such that the samples have a circular variance of $\nu = 0.25$. The concentration of rotation matrices under circular-von Mises sampling is much higher to the origin (center of the circles) than for the two other distributions.



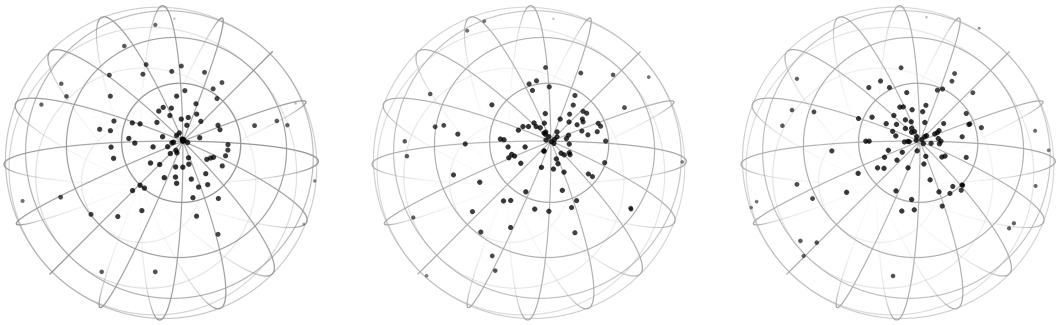
(a) x axis (b) y axis (c) z axis

Figure 2.15: Sphere plots for a sample of 100 rotations from a Cayley distribution with circular variance $\nu = 0.25$



(a) x axis (b) y axis (c) z axis

Figure 2.17: Sphere plots for a sample of 100 rotations from a matrix Fisher distribution with circular variance $\nu = 0.25$



(a) x axis (b) y axis (c) z axis

Figure 2.19: Sphere plots for a sample of 100 rotations from a circular-von Mises distribution with circular variance $\nu = 0.25$

Table 2.8: Average reduction in estimation error by using $\widehat{\boldsymbol{S}}_R$ instead of $\widehat{\boldsymbol{S}}_E$, $\delta = d_R(\widehat{\boldsymbol{S}}_E, \boldsymbol{S}) - d_R(\widehat{\boldsymbol{S}}_R, \boldsymbol{S})$ with standard error and percentage of samples for which $d_R(\widehat{\boldsymbol{S}}_R, \boldsymbol{S}) < d_R(\widehat{\boldsymbol{S}}_E, \boldsymbol{S})$.

		Cayley		**matrix Fisher**		**circular-von Mises**	
ν	n	$\bar{\delta}$ (SE)	%	$\bar{\delta}$ (SE)	%	$\bar{\delta}$ (SE)	%
0.25	10	0.0011 (0.0004)	0.5210	-0.0016 (0.0005)	0.4500	-0.0344 (0.0030)	0.1280
	50	0.0007 (0.0002)	0.5310	-0.0011 (0.0003)	0.4350	-0.0156 (0.0007)	0.2090
	100	0.0007 (0.0001)	0.5650	-0.0004 (0.0002)	0.4690	-0.0126 (0.0005)	0.2010
	300	0.0005 (0.0001)	0.5880	-0.0003 (0.0001)	0.4860	-0.0070 (0.0003)	0.2390
0.50	10	0.0099 (0.0013)	0.5920	-0.0178 (0.0027)	0.4340	-0.1011 (0.0051)	0.1570
	50	0.0069 (0.0006)	0.6450	-0.0107 (0.0011)	0.3920	-0.0545 (0.0018)	0.1450
	100	0.0043 (0.0004)	0.6420	-0.0067 (0.0008)	0.3930	-0.0385 (0.0013)	0.1620
	300	0.0025 (0.0002)	0.6420	-0.0040 (0.0004)	0.3930	-0.0234 (0.0007)	0.1570
0.75	10	0.0163 (0.0062)	0.6680	-0.0958 (0.0105)	0.3570	-0.2101 (0.0113)	0.1710
	50	0.0257 (0.0012)	0.7410	-0.0356 (0.0045)	0.3380	-0.0955 (0.0032)	0.1500
	100	0.0169 (0.0009)	0.7350	-0.0240 (0.0043)	0.3460	-0.0763 (0.0021)	0.1110
	300	0.0091 (0.0005)	0.7280	-0.0124 (0.0009)	0.3440	-0.0446 (0.0012)	0.1270

CHAPTER 3. NONPARAMETRIC CONFIDENCE REGIONS FOR THE CENTRAL ORIENTATION OF RANDOM ROTATIONS

A paper submitted to *Statistica Sinica*

Bryan Stanfill, Ulrike Genschel, Heike Hofmann, Dan Nordman

Abstract

Three-dimensional orientation data, with observations as 3×3 rotation matrices, have applications in areas such as computer science, kinematics and materials sciences, where it is often of interest to estimate a central orientation parameter $\boldsymbol{S}$ represented by a 3×3 rotation matrix. A well-known estimator of this parameter is the projected arithmetic mean and, based on this statistic, two nonparametric methods for setting confidence regions for $\boldsymbol{S}$ exist. Both of these methods involve large-sample normal theory, with one approach based on a data-transformation of rotations to directions (four-dimensional unit vectors) prior to analysis. However, both of these nonparametric methods may result in poor coverage accuracy in small samples. As a remedy, we consider two bootstrap methods for approximating the sampling distribution of the projected mean statistic and calibrating nonparametric confidence regions for the central orientation parameter $\boldsymbol{S}$. As with normal approximations, one bootstrap method is based on the rotation data directly while the other bootstrap approach involves a data-transformation of rotations into directions. Both bootstraps are shown to be valid for approximating sampling distributions and calibrating confidence regions based on the projected mean statistic. A simulation study compares the performance of the normal theory and proposed bootstrap confidence regions for $\boldsymbol{S}$, based on common data-generating models for symmetric orientations. The bootstrap methods are shown to exhibit good coverage accuracies, thus providing an improvement over normal theory approximations especially for small sample sizes. The bootstrap

methods are also illustrated with a real data example from materials science.

*K*eywords: Influence Function, Orientation Data, Pivotal Bootstrap, Projected Arithmetic Mean, Quaternions

3.1 Introduction

Orientation data frequently arise in areas such as computer science, structural geology, materials sciences and biomedical imaging, where technological advancements in recent years have increased the availability of such data (cf. Humbert et al. 1996; Preisig and Kragic 2006; Fletcher et al. 2009; Bingham et al. 2010a). Orientation data, given as 3×3 orthogonal rotation matrices with determinant one, are often used to represent the positions of objects with respect to a three-dimensional reference frame. We denote the collection of all such 3×3 matrices as $SO(3)$. Given a random sample of orientations, a quantity of interest is often the main or central orientation parameter $\boldsymbol{S} \in SO(3)$. That is, observations $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n \in SO(3)$ are commonly conceptualized as a random sample from a *location model*

$$\boldsymbol{R}_i = \boldsymbol{S}\boldsymbol{E}_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where $\boldsymbol{S} \in SO(3)$ is a *fixed* parameter of interest called the central orientation, and $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n \in SO(3)$ denote independent and identically (symmetrically) distributed *random* rotations. The model in (3.1) provides a rotation analog of a location model for scalar data $Y = \mu + \varepsilon$ where $\mu \in \mathbb{R}$ denotes a mean and $\varepsilon \in \mathbb{R}$ denotes an additive error symmetrically distributed around zero. The symmetry assumption in (3.1) is common in location estimation (cf. Chang and Rivest 2001), where the data are then interpreted as random perturbations of an underlying central orientation $\boldsymbol{S}$. In the following, we consider the projected arithmetic mean $\widehat{\boldsymbol{S}}_n$, which is perhaps the most common and popular estimator of the location $\boldsymbol{S}$ parameter for rotation data (cf. Moakher 2002; Fletcher et al. 2003; Bhattacharya and Patrangenaru 2003; Bachmann et al. 2010); see Stanfill et al. (2013) for a discussion of point estimators for the central orientation. This estimator has a least-square motivation, corresponding to the rotation (a statistic) that minimizes a sum of squared Euclidean distances from the data. Section 3.2 provides more details.

Based on the projected mean $\widehat{\boldsymbol{S}}_n$, nonparametric large sample confidence regions for the central orientation $\boldsymbol{S}$ were first considered by Prentice (1986). These confidence regions are based on a transformation of $SO(3)$ rotation data to directional data (i.e., unit vectors) in $\mathbb{R}^4$ combined with limiting distributional results for directional statistics; a confidence region for $\boldsymbol{S}$ results by back transforming $\mathbb{R}^4$-directional observations into $SO(3)$ rotations. Chang and Rivest (2001) also explored nonparametric confidence regions for $\boldsymbol{S}$ based on the rotation data directly, using M-estimation theory (i.e., related to least squares) and the limiting distribution of $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$. While these nonparametric regions require few distributional assumptions for their validity, both methods are based on large-sample normal theory and can exhibit poor finite-sample coverage accuracy, especially with small samples.

As a remedy, we propose two bootstrap approaches for calibrating nonparametric confidence regions for $\boldsymbol{S}$ based on the projected mean $\widehat{\boldsymbol{S}}_n$, in place of these two normal theory methods. The first bootstrap method uses rotation data directly to approximate the sampling distribution of $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$ and this bootstrap is shown to provide consistent distributional estimation under weak assumptions about the generation of rotation data. The second bootstrap method is based on the transformation approach of Prentice (1986), where the orientation data are first mapped into directional data in $\mathbb{R}^4$. This bootstrap applied to rotation data is an extension of a resampling method proposed by Fisher et al. (1996) for general directional data in $\mathbb{R}^d$, but with special tailoring for application to rotation matrices in $SO(3)$. We then compare the finite sample behavior of the existing large sample confidence regions and the proposed bootstrap approaches through simulation. This simulation study appears to provide the first numerical comparison of the normal theory confidence regions for the location parameter $\boldsymbol{S}$ based on direct use of rotations (e.g., Chang and Rivest 2001) or directional-based transformations (Prentice, 1986); the study also provides the first evaluation of the bootstrap method of Fisher et al. (1996) for calibrating confidence regions based on rotation data (after transforming these to directional data in $\mathbb{R}^4$). The numerical evidence suggests that the bootstrap confidence regions exhibit much better coverage accuracy than the normal theory ones, especially for small sample sizes and in the presence of greater variability. In particular, the proposed bootstrap which uses rotation data directly to approximate the sampling distribution of $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$ is shown to have

the best performance in a variety of settings.

The remainder of this manuscript is organized as follows. Section 3.2 briefly provides distributional background for random rotations (3.1) as well as the definition of the projected mean $\widehat{\boldsymbol{S}}_n$. Section 3.3 reviews confidence region methodology based on large sample normal theory. In Section 3.4, we present the two distinct bootstrap methods for calibrating confidence regions for the location parameter $\boldsymbol{S}$ based on $\widehat{\boldsymbol{S}}_n$. Section 3.5 describes a simulation study, which provides empirical justification for the asymptotic results and compares the finite sample behavior of the proposed confidence regions to those based on normal approximations. We demonstrate these results with several common distributions for $SO(3)$ data, involving the Cayley, circular-von Mises and matrix Fisher distributions (cf. Stanfill et al. 2013). In Section 3.6, we apply our bootstrap to a real data set from materials science involving rotation data from electron backscatter diffraction. Concluding remarks and directions for future research are given in Section 3.7. A supplemental appendix contains proofs of the main results.

3.2 Rotation Data and the projected arithmetic mean

We briefly establish some notation for describing rotation data and the distributional results to follow. Let $\mathfrak{so}(3)$ represent the space of all 3×3 skew-symmetric matrices, i.e., $\mathfrak{so}(3) = \{\boldsymbol{X} \in \mathbb{R}^{3\times 3} : \boldsymbol{X}^\top = -\boldsymbol{X}\}$. Then, for each rotation $\boldsymbol{R} \in SO(3)$ we can define a skew-symmetric matrix $\boldsymbol{\Phi}(\boldsymbol{W}) \in \mathfrak{so}(3)$

$$\boldsymbol{\Phi}(\boldsymbol{W}) = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix}, \tag{3.2}$$

$\boldsymbol{W} = (w_1, w_2, w_3)^\top \in \mathbb{R}^3$ and, through the exponential operator, map $\boldsymbol{\Phi}(\boldsymbol{W})$ into $SO(3)$ as

$$\exp[\boldsymbol{\Phi}(\boldsymbol{W})] = \sum_{k=0}^{\infty} \frac{1}{k!}[\boldsymbol{\Phi}(\boldsymbol{W})]^k \quad \in \quad SO(3) \tag{3.3}$$

(Moakher, 2002). By properties of skew-symmetric matrices, (3.3) can be rewritten as

$$\exp[\boldsymbol{\Phi}(\boldsymbol{W})] = \cos(r)\boldsymbol{I}_{3\times 3} + \sin(r)\boldsymbol{\Phi}(\boldsymbol{U}) + (1-\cos r)\boldsymbol{U}\boldsymbol{U}^\top$$

where the angle $r \in (-\pi, \pi]$ is defined as $r = \|\boldsymbol{W}\| (\mathrm{mod} 2\pi)$ and $\boldsymbol{U} = \boldsymbol{W}/\|\boldsymbol{W}\|$ represents an axis in $\mathbb{R}^3$ with $\|\boldsymbol{U}\| = 1$. Each non-identity rotation matrix $\boldsymbol{R}$ can then be associated with an angle-axis pair $(r, \boldsymbol{U})$ as above (uniquely up to the sign $(r, \boldsymbol{U}) = (-r, -\boldsymbol{U})$) such that

$$\boldsymbol{R} = \boldsymbol{R}(r, \boldsymbol{U}) = \exp[\boldsymbol{\Phi}(r\boldsymbol{U})] \quad \in \quad SO(3). \tag{3.4}$$

In this angle-axis representation, the columns of $\boldsymbol{R}$ are interpreted as the positions of standard coordinate axes $\boldsymbol{I}_{3\times 3}$ after a rotation of the reference frame $\boldsymbol{I}_{3\times 3}$ about an axis $\boldsymbol{U} \in \mathbb{R}^3$ through an angle r. In the materials science literature, $\boldsymbol{U}$ and r are commonly referred to as the misorientation axis and misorientation angle of $\boldsymbol{R}$ with respect to $\boldsymbol{I}_{3\times 3}$, respectively; see Randle (2003).

In the data model in (3.1), each *random* rotation matrix $\boldsymbol{R} = \boldsymbol{S}\boldsymbol{E}$ has a simple construction based on the angle-axis representation (3.4). Namely, a symmetrically distributed random rotation matrix

$$\boldsymbol{E} = \boldsymbol{E}(r, \boldsymbol{U}) \tag{3.5}$$

defined by a *random* angle-axis pair $(r, \boldsymbol{U})$ in (3.4), where $\boldsymbol{U}$ is uniformly distributed on the unit sphere in $\mathbb{R}^3$ and, independently from $\boldsymbol{U}$, the angle $r \in (-\pi, \pi]$ is symmetrically distributed about zero. Distributional models for r are commonly parameterized through a concentration parameter $\kappa > 0$ controlling the variability of r and, hence, the variability of the random rotation (3.5); see Bingham et al. (2009) and Section 3.3.1 for more details. Because $\boldsymbol{E}$ provides a rotationally symmetric random rotation of the coordinate axes $\boldsymbol{I}_{3\times 3}$, any observation $\boldsymbol{R} = \boldsymbol{S}\boldsymbol{E}$ can be interpreted as a random perturbation of the location parameter $\boldsymbol{S} \in SO(3)$.

Based on a random sample $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$, we next describe the projected mean $\widehat{\boldsymbol{S}}_n$ as a common estimator of the central orientation $\boldsymbol{S}$ in the location model in (3.1). The projected mean is standardly defined as the minimizer of sum of the squared Euclidean distances

$$\widehat{\boldsymbol{S}}_n = \underset{\boldsymbol{S} \in SO(3)}{\arg\min} \sum_{i=1}^{n} d_E^2(\boldsymbol{R}_i, \boldsymbol{S}) \tag{3.6}$$

where, in terms of the matrix Frobenius norm $\|\cdot\|_F$, the squared distance between two rotations $\boldsymbol{R}_1, \boldsymbol{R}_2$ is given as

$$d_E^2(\boldsymbol{R}_1, \boldsymbol{R}_2) \equiv \|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_F = \left[6 - 2\mathbf{tr}\left(\boldsymbol{R}_1^\top \boldsymbol{R}_2\right)\right]; \tag{3.7}$$

and $\mathbf{tr}(\cdot)$ above denoting the matrix trace. Hence, $\widehat{\boldsymbol{S}}_n$ corresponds to a type of M-estimator also known as an extrinsic mean, (cf. Bhattacharya and Patrangenaru 2003, 2005). From equation (3.7), note that the mean as defined in (3.6) is equivalent to the rotation that maximizes $\mathbf{tr}(\boldsymbol{S}^\top \overline{\boldsymbol{R}})$ over $\boldsymbol{S} \in SO(3)$ where $\overline{\boldsymbol{R}} = \sum_{i=1}^n \boldsymbol{R}_i/n$ denotes the sample mean of the rotation data. Computational algorithms for $\widehat{\boldsymbol{S}}_n$ have been studied at length (see Arun et al. 1987; Horn et al. 1988; Umeyama 1991; Moakher 2002) as well as this point estimator's performance under different distributional assumptions (see Jupp and Mardia 1979; León et al. 2006; Bingham et al. 2009; Stanfill et al. 2013).

3.3 Confidence Regions based on Large Sample Normal Approximations

This section describes existing nonparametric confidence regions for the central orientation $\boldsymbol{S}$ based on the mean estimator $\widehat{\boldsymbol{S}}_n$ and large sample normal theory. Section 3.3.1 explains a direct calibration approach (i.e., involving rotation data directly) based on large sample distributional theory for $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$. Section 3.3.2 describes a confidence region approach based on first transforming the rotation data into $\mathbb{R}^4$ directions for inference.

3.3.1 Direct Large Sample Approach

A first large-sample normal theory approach to setting confidence regions for $\boldsymbol{S}$ based on the project mean $\widehat{\boldsymbol{S}}_n$ (again determined from rotation data directly as in (3.6)) is based on determining the limiting distribution of $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$. In order to do so, it is useful to write $\exp[\boldsymbol{\Phi}(\widehat{\boldsymbol{h}}_n)] = \boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$ where we associate a vector $\widehat{\boldsymbol{h}}_n \in \mathbb{R}^3$ with the rotation $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$ through the exponential map operator in (3.3). Confidence regions for $\boldsymbol{S}$ based on the limiting distribution of $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$ can then be framed in terms of the large-sample distribution of $\sqrt{n}\widehat{\boldsymbol{h}}_n$. Note that Chang and Rivest (2001) developed a general theory for M-estimation for location parameters in statistical group models, which include $SO(3)$ data as a special case. Their general formulation may be applied to determine a normal limit law for $\sqrt{n}\widehat{\boldsymbol{h}}_n$ (and the induced law of $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$), but the translation of their results (cf. Proposition 2 there) to the location model for rotation data is not immediate and an explicit distributional result for the projected mean

$\widehat{\boldsymbol{S}}_n$, as required here, does not appear in that work. The following Proposition 3.1 provides the necessary distributional result for normal theory confidence regions for $\boldsymbol{S}$ from $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$, showing in particular that a scaled version of $\|\sqrt{n}\widehat{\boldsymbol{h}}_n\|^2$ has a chi-square limit with 3 degrees of freedom. The proof is based on a novel argument involving influence functions; see the supplementary material for details.

Proposition 3.1. *Let $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ denote a random sample of independently and identically distributed observations from (3.1). Further, let $\widehat{\boldsymbol{S}}_n$ denote the projected mean (cf. (3.6)) and let $\widehat{\boldsymbol{h}}_n \in \mathbb{R}^3$ such that $\exp[\boldsymbol{\Phi}(\widehat{\boldsymbol{h}}_n)] = \boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$. Then, as $n \longrightarrow \infty$,*

$$\frac{2na_2^2}{a_1} \left\| \widehat{\boldsymbol{h}}_n \right\|^2 \xrightarrow{d} \chi^2_3 \tag{3.8}$$

for constants

$$a_1 = \frac{2}{3}\mathrm{E}\left[1 - \cos^2(r)\right] \qquad \textit{and} \qquad a_2 = \frac{1}{3}\mathrm{E}\left[1 + 2\cos(r)\right], \tag{3.9}$$

based on the distribution of the random angle $r \in (-\pi, \pi]$ defining random rotations in (3.5).

By Proposition 3.1, a $100(1-\alpha)\%$ large sample confidence region for the central orientation $\boldsymbol{S} \in SO(3)$ can be calibrated as the set of all 3×3 rotation matrices defined by

$$\left\{ \boldsymbol{S} = \exp[\boldsymbol{\Phi}(\boldsymbol{h})]^\top \widehat{\boldsymbol{S}}_n : \frac{2na_2^2}{a_1} \|\boldsymbol{h}\|^2 < \chi^2_{3,1-\alpha} \right\}.$$

However, this region is not distribution-free in that the population quantities a_1 and a_2 depend on the underling distribution of the rotation data. For illustration, exact values of a_1 and a_2 in (3.9) are given in Table 3.1 for several different models for random rotations (e.g., Cayley, circular-von Mises and matrix Fisher distributions) which involve a concentration parameter $\kappa > 0$.

To obtain a nonparametric confidence region for $\boldsymbol{S}$, the unknown constants a_1 and a_2 require estimation. For a random sample $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ consistent estimators (by the law of large numbers) of a_1 and a_2, respectively, are given by

$$\hat{a}_{1_n} = \frac{1}{6n}\sum_{i=1}^{n}\left\{3 - \mathbf{tr}\left[\left(\widehat{\boldsymbol{S}}_n^\top \boldsymbol{R}_i\right)^2\right]\right\} \text{ and } \hat{a}_{2_n} = \frac{1}{3n}\sum_{i=1}^{n}\mathbf{tr}\left(\widehat{\boldsymbol{S}}_n^\top \boldsymbol{R}_i\right). \tag{3.10}$$

Then, a nonparametric approximate $100(1-\alpha)\%$ confidence region for location parameter $\boldsymbol{S} \in SO(3)$ is given by the set of all 3×3 rotation matrices satisfying

$$\left\{ \boldsymbol{S} = \exp[\boldsymbol{\Phi}(\boldsymbol{h})]^{\top} \widehat{\boldsymbol{S}}_n : \frac{2n\hat{a}_{2_n}^2}{\hat{a}_{1_n}} \|\boldsymbol{h}\|^2 < \chi_{3,1-\alpha}^2 \right\}. \tag{3.11}$$

The region (3.11) provides a rotationally symmetric confidence region about the projected mean $\widehat{\boldsymbol{S}}_n$, corresponding to the set of all rotations within a geodesic distance of $[\chi_{3,1-\alpha}^2 \hat{a}_{1_n}/(2n\hat{a}_{2_n}^2)]^{1/2} (\mathrm{mod}\pi)$ of $\widehat{\boldsymbol{S}}_n$.

In the remainder, we will refer to this normal theory method of constructing confidence regions for $\boldsymbol{S}$ as a *direct* approach as it is based on using the rotation data $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ directly to formulate the region in (3.11).

Table 3.1: Constants a_1 and a_2 defined in (3.9) for the Cayley, circular-von Mises and matrix Fisher distribution on $SO(3)$. The symbol $I_i(\cdot)$ is the modified Bessel function of the first kind of order i.

Distribution	a_1	a_2
Cayley	$\frac{4\kappa+2}{\kappa^2+5\kappa+6}$	$\frac{\kappa}{\kappa+2}$
circular-von Mises	$\frac{I_0(\kappa)-I_2(\kappa)}{3I_0(\kappa)}$	$\frac{I_0(\kappa)+2I_1(\kappa)}{3I_0(\kappa)}$
matrix-Fisher	$\frac{(\kappa+1)I_1(2\kappa)-\kappa I_0(2\kappa)}{3\kappa^2[I_0(2\kappa)-I_1(2\kappa)]}$	$\frac{(\kappa+1)I_1(2\kappa)-\kappa I_0(2\kappa)}{3\kappa[I_0(2\kappa)-I_1(2\kappa)]}$

3.3.2 Transformation-based Large Sample Approach

A second approach to formulate large sample normal theory confidence regions for $\boldsymbol{S}$ based on $\widehat{\boldsymbol{S}}_n$ is by Prentice (1986). This method involves a transformation of rotations in $SO(3)$ into unsigned directional data (i.e., unit vectors) in $\mathbb{R}^4$. In particular, for every unit vector $\boldsymbol{x} = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, $\|\boldsymbol{x}\| = 1$, there exists a map

$$\boldsymbol{R} = \mu(\boldsymbol{x}) \equiv \begin{bmatrix} x_1^2+x_4^2-x_2^2-x_3^2 & 2(x_1x_2-x_3x_4) & 2(x_2x_4+x_1x_3) \\ 2(x_3x_4+x_1x_2) & x_2^2+x_4^2-x_1^2-x_3^2 & 2(x_2x_3-x_1x_4) \\ 2(x_1x_3-x_2x_4) & 2(x_2x_3+x_1x_4) & x_3^2+x_4^2-x_1^2-x_2^2 \end{bmatrix} \tag{3.12}$$

into a rotation $\boldsymbol{R} \in SO(3)$. Likewise, for any rotation $\boldsymbol{R} \in SO(3)$, there is an inverse map $\mu^{-1}(\boldsymbol{R}) = \pm\boldsymbol{x}$ back to unit vectors in $\mathbb{R}^4$, which is unique up to the sign of $\boldsymbol{x} \in \mathbb{R}^4, \|\boldsymbol{x}\| = 1$.

With this transformation, the location parameter of interest $\boldsymbol{S} \in SO(3)$ can be associated with a unit vector $\boldsymbol{m} \in \mathbb{R}^4$ through $\boldsymbol{S} = \mu(\boldsymbol{m})$ and $\boldsymbol{m}$ may be estimated after transforming the rotation data to directions in $\mathbb{R}^4$. Namely, for a random rotation $\boldsymbol{R} = \boldsymbol{SE}$ under the location model in (3.1), it holds that $\boldsymbol{m}$ (a mean polar axis) corresponds to the eigenvector associated with the largest eigenvalue of $\mathrm{E}(\boldsymbol{x}\boldsymbol{x}^\top)$ where $\boldsymbol{x} = \mu^{-1}(\boldsymbol{R})$ (Prentice, 1986). Hence, if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ denote the corresponding directions in $\mathbb{R}^4$ from the transformation $\boldsymbol{x}_i = \mu^{-1}(\boldsymbol{R}_i)$ of rotation data $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$, then the project mean is given by $\widehat{\boldsymbol{S}}_n = \mu(\widehat{\boldsymbol{m}}_n)$, where $\widehat{\boldsymbol{m}}_n$ is the eigenvector associated with the largest eigenvalue of the sample moment of inertia matrix $\widehat{\boldsymbol{V}}_n \equiv (1/n)\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\top$.

The limiting normal distribution of $\sqrt{n}(\widehat{\boldsymbol{m}}_n - \boldsymbol{m})$ can be justified using results established by Davis (1977) and Tyler (1981) on the large sample behavior of principal components in non-normal populations. For calibrating confidence regions, Prentice (1984) derived that $n\boldsymbol{m}^\top \widehat{\boldsymbol{F}}_n^- \boldsymbol{m}$ has a limiting χ^2_3 distribution, where $\widehat{\boldsymbol{F}}_n$ denotes an estimator of the limiting variance matrix $\boldsymbol{F}$ of $\sqrt{n}(\widehat{\boldsymbol{m}}_n - \boldsymbol{m})$ and $\widehat{\boldsymbol{F}}_n^-$ denotes the generalized inverse. If the sample moment matrix $\widehat{\boldsymbol{V}}_n \equiv (1/n)\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\top = \widehat{\boldsymbol{A}}_n\widehat{\boldsymbol{\Lambda}}_n\widehat{\boldsymbol{A}}_n^\top$ is spectrally decomposed into a 4×4 rotation matrix $\widehat{\boldsymbol{A}}_n$ and a diagonal (eigenvalue) matrix $\widehat{\boldsymbol{\Lambda}}_n = \mathrm{diag}(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3, \widehat{\lambda}_4)$, where $\widehat{\lambda}_1 < \widehat{\lambda}_2 < \widehat{\lambda}_3 < \widehat{\lambda}_4$, then Prentice (1984) suggests a consistent (method of moments) estimator

$$\widehat{\boldsymbol{F}}_n \equiv \widehat{\boldsymbol{A}}_{4n}\mathrm{diag}(\widehat{b}_{1n}, \widehat{b}_{2n}, \widehat{b}_{3n})\widehat{\boldsymbol{A}}_{4n}^\top, \tag{3.13}$$

where $\widehat{\boldsymbol{A}}_{4n}$ denotes the resulting 4×3 submatrix after removing the 4th column of $\widehat{\boldsymbol{A}}_n$ and $\widehat{b}_{jn} = \widehat{c}_j/(\widehat{\lambda}_4 - \widehat{\lambda}_j)$, $j = 1, 2, 3$ for $\widehat{c}_j$ denoting the jth diagonal entry of $\widehat{\boldsymbol{C}}_n = \sum_i \boldsymbol{x}_i^{(2)}\boldsymbol{x}_i^{(2)\top}/n$ with $\boldsymbol{x}_i^{(2)} \in \mathbb{R}^3$ as the vector of the diagonal entries of $(\widehat{\boldsymbol{A}}_n\boldsymbol{x}_i)(\widehat{\boldsymbol{A}}_n\boldsymbol{x}_i)^\top$.

Hence, by Prentice (1984) and Prentice (1986), an asymptotic $100(1-\alpha)\%$ confidence region for $\boldsymbol{S}$ is the collection of all 3×3 rotation matrices defined by the mapped set

$$\left\{\boldsymbol{S} = \mu(\boldsymbol{m}) : \boldsymbol{m} \in \mathbb{R}^4,\ \|\boldsymbol{m}\| = 1,\ n\boldsymbol{m}^\top \widehat{\boldsymbol{F}}_n^- \boldsymbol{m} < \chi^2_{3,1-\alpha}\right\} \tag{3.14}$$

using $\mu(\cdot)$ from (3.12). Note that, in the above decomposition $\widehat{\boldsymbol{V}}_n = \widehat{\boldsymbol{A}}_n\widehat{\boldsymbol{\Lambda}}_n\widehat{\boldsymbol{A}}_n^\top$, the projected mean corresponds to $\widehat{\boldsymbol{S}}_n = \mu(\widehat{\boldsymbol{m}}_n)$, where $\widehat{\boldsymbol{m}}_n$ is the 4th column of $\widehat{\boldsymbol{A}}_n$. Because of this and by

the definition of $\widehat{\boldsymbol{F}}_n^- = \widehat{\boldsymbol{A}}_{4n}\text{diag}(1/\widehat{b}_{1n}, 1/\widehat{b}_{2n}, 1/\widehat{b}_{3n})\widehat{\boldsymbol{A}}_{4n}^\top$, it holds that $n\widehat{\boldsymbol{m}}_n^\top \widehat{\boldsymbol{F}}_n^- \widehat{\boldsymbol{m}}_n = 0$. Hence, the confidence region (3.14) contains the projected mean $\widehat{\boldsymbol{S}}_n$ at its center.

As the large sample normal theory confidence region for $\boldsymbol{S}$ in Prentice (1986) is based on transformation of rotations into $\mathbb{R}^d$-directional data in, we refer to this confidence region method as the *transformation-based* approach. We also note that Prentice (1986) proposed a second method for setting confidence regions for the central orientation $\boldsymbol{S}$, which provides a simplification of (3.14) under symmetry assumptions on the perturbations $\boldsymbol{E}_i$. This second region approximates the confidence region defined in (3.11) of Section 3.3.1 and is therefore not considered further.

3.4 Bootstrap Confidence Regions

As alternatives to the asymptotic normal theory approaches of Section 3.3, we propose two bootstrap versions for constructing confidence regions for $\boldsymbol{S} \in SO(3)$ based on the projected mean $\widehat{\boldsymbol{S}}_n$. These provide analogs of the regions in (3.11) and (3.14) where bootstrap approximations replace chi-square limit calibrations. The first bootstrap method, described in Section 3.4.1, resamples rotation data directly, while the second method, described in Section 3.4.2, is based on transforming rotations to directional data in $\mathbb{R}^4$ prior to resampling.

3.4.1 Direct Bootstrap Approach

Recall that the normal theory regions (3.11) of Section 3.3.1 are based on the limiting chi-square distribution of $(2na_2^2/a_1)\|\widehat{\boldsymbol{h}}_n\|^2$ under Proposition 3.1, or its studentized version

$$D_n \equiv \frac{2n\hat{a}_{2_n}^2}{\hat{a}_{1_n}}\|\widehat{\boldsymbol{h}}_n\|^2 \tag{3.15}$$

with estimators $\hat{a}_{1_n}, \hat{a}_{2_n}$ from (3.10), where the vector $\widehat{\boldsymbol{h}}_n \in \mathbb{R}^3$ is related to the projected mean through $\exp[\boldsymbol{\Phi}(\widehat{\boldsymbol{h}}_n)] = \boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$ via the exponential operator (3.3). We seek to approximate the sampling distribution of the asymptotic pivot D_n in (3.15) with a bootstrap counterpart. To this end, given rotation observations $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$, we generate a bootstrap sample $\boldsymbol{R}_1^*, \ldots, \boldsymbol{R}_n^*$ by randomly selecting n rotation matrices with replacement from $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$. The bootstrap sample has a corresponding projected mean $\widehat{\boldsymbol{S}}_n^*$ (as the bootstrap version of (3.6)) and we

define the bootstrap version $\widehat{\boldsymbol{h}}_n^*$ of $\widehat{\boldsymbol{h}}_n$ through the bootstrap analog $\exp[\boldsymbol{\Phi}(\widehat{\boldsymbol{h}}_n^*)] = \widehat{\boldsymbol{S}}_n^\top \widehat{\boldsymbol{S}}_n^*$ of $\exp[\boldsymbol{\Phi}(\widehat{\boldsymbol{h}}_n)] = \boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$; here $\widehat{\boldsymbol{S}}_n$ plays the role of the location parameter $\boldsymbol{S} \in SO(3)$ in the bootstrap world. We then define the bootstrap rendition of $D_n \equiv (2n\hat{a}_{2_n}^2/\hat{a}_{1_n})\|\widehat{\boldsymbol{h}}_n\|^2$ as

$$D_n^* \equiv \frac{2n(\widehat{a}_{2n}^*)^2}{\widehat{a}_{1n}^*}\|\widehat{\boldsymbol{h}}_n^*\|^2, \tag{3.16}$$

where $\widehat{a}_{1n}^*, \widehat{a}_{2n}^*$ are computed as in (3.10) based on the bootstrap sample $\boldsymbol{R}_1^*, \ldots, \boldsymbol{R}_n^*$.

The following strong consistency result establishes that this bootstrap validly approximates the sampling distribution of D_n, as the cornerstone of subsequent bootstrap confidence regions for $\boldsymbol{S}$; see the supplemental material for its proof. Let P_* denote bootstrap probability conditional on the data.

Theorem 3.2. *Let $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ be a random sample of i.i.d. observations from the location model in (3.1). For the bootstrap version D_n^* in (3.16) of D_n in (3.15), as $n \to \infty$,*

$$\sup_{x \in \mathbb{R}} |P(D_n \le x) - P_*(D_n^* \le x)| \longrightarrow 0 \quad \textit{with probability one.}$$

As a consequence of Theorem 3.2, an approximate $100(1-\alpha)\%$ bootstrap confidence region for the central orientation $\boldsymbol{S}$ is given by

$$\left\{ \boldsymbol{S} = \exp[\boldsymbol{\Phi}(\boldsymbol{h})]^\top \widehat{\boldsymbol{S}}_n : \frac{2n\hat{a}_{2_n}^2}{\hat{a}_{1_n}} \|\boldsymbol{h}\|^2 < d_{1-\alpha} \right\} \tag{3.17}$$

where $d_{1-\alpha}$ denotes the (lower) $(1-\alpha)$ percentile of the bootstrap distribution of D_n^*. That is, the region in (3.17) is guaranteed to have asymptotically correct coverage by Theorem 3.2, again providing a bootstrap version of the normal theory region in (3.11). Because this bootstrap method is based on direct resampling of the rotation data, we refer to this confidence region approach for $\boldsymbol{S}$ as the *direct bootstrap* method.

We end this section by mentioning other possible bootstrap versions, not pursued here. If $\rho(\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n)$ denotes a distance assessment between $\boldsymbol{S}$ and $\widehat{\boldsymbol{S}}_n$ based on some smooth function $\rho(\cdot)$ of $\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n$, such as the Frobenius norm $\|\boldsymbol{S} - \widehat{\boldsymbol{S}}_n\|_F$ in (3.7) or geodesic distance $2^{-1/2}\|\log \boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n\|_F \equiv \sin\|\widehat{\boldsymbol{h}}_n\|$, it can be shown that the same resampling scheme for rotation data produces a bootstrap version $n\rho(\widehat{\boldsymbol{S}}_n^\top \widehat{\boldsymbol{S}}_n^*)$ that validly approximates the distribution of $n\rho(\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n)$. This allows confidence regions to be formulated as $\{\boldsymbol{S} : n\rho(\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n) < q_{1-\alpha}\}$, where

$q_{1-\alpha}$ denotes a percentile from the bootstrap distribution of $n\rho(\widehat{\boldsymbol{S}}_n^\top \widehat{\boldsymbol{S}}_n^*)$. However, unlike the bootstrap regions in (3.17), such confidence regions are not based on asymptotically pivotal quantities as $n\rho(\boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n)$ does not generally have a distribution-free limit law (e.g., in the cases of distances $\rho(\cdot)$ mentioned above). Bootstrap-based methodology involving pivotal quantities (e.g., D_n in (3.15)), especially for mean-like statistics $\widehat{\boldsymbol{S}}_n$, is often expected to provide better distributional approximations than bootstrap versions applied to non-pivotal quantities Hall (1992). Numerical studies (not reported here) indicate that this indeed holds for bootstrap regions for the location parameter $\boldsymbol{S} \in SO(3)$, so that we focused our development on the bootstrap regions in (3.17).

3.4.2 Transformation-based Bootstrap Approach

As in the transformation-based normal theory approach of Section 3.3.2, we next consider a bootstrap confidence region method based on transforming rotations to $\mathbb{R}^4$-directional data prior to resampling. That is, as in Section 3.3.2, we map the rotation data $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ to unit vectors $\boldsymbol{x}_i = \mu^{-1}(\boldsymbol{R}_i) \in \mathbb{R}^4$, $\|\boldsymbol{x}_i\| = 1$, $i = 1, \ldots, n$, through the inverse of the mapping $\mu(\cdot)$ in (3.12). Recall that the transformation-based normal theory regions in (3.14) are based on the limiting chi-square distribution of

$$T_n \equiv n\boldsymbol{m}^\top \widehat{\boldsymbol{F}}_n^- \boldsymbol{m}, \tag{3.18}$$

where $\widehat{\boldsymbol{F}}_n$ denotes a variance estimator from (3.13) (derived from $\widehat{\boldsymbol{V}} = \sum_{i=1}^n \boldsymbol{x}_i^\top \boldsymbol{x}_i/n$) and $\boldsymbol{m} \in \mathbb{R}^4$, $\|\boldsymbol{m}\| = 1$ denotes the direction defining the location parameter $\boldsymbol{S} = \mu(\boldsymbol{m})$ (Prentice 1984, 1986).

To produce a bootstrap version T_n^* of T_n, we resample with replacement from the observed directions $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ to produce a bootstrap data set $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*$. This leads to a version of the statistic $\widehat{\boldsymbol{V}}_n^* = \sum_{i=1}^n \boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}/n$ used to define a bootstrap analog $\widehat{\boldsymbol{F}}_n^*$ of the estimator $\widehat{\boldsymbol{F}}_n$ in (3.13). In place of the unknown (mean polar axis) parameter $\boldsymbol{m} \in \mathbb{R}^4$, $\boldsymbol{S} = \mu(\boldsymbol{m})$, we use its estimator $\widehat{\boldsymbol{m}}_n$ related to the project mean via $\mu(\widehat{\boldsymbol{m}}_n) = \widehat{\boldsymbol{S}}_n$ (cf. Section 3.3.2). We then define a bootstrap version of T_n in (3.18) as

$$T_n^* \equiv n\widehat{\boldsymbol{m}}_n^\top \widehat{\boldsymbol{F}}_n^{*-} \widehat{\boldsymbol{m}}_n. \tag{3.19}$$

Theorem 3.3 shows that this bootstrap based on transformed rotations results in consistent estimation for the sampling distribution of T_n; see the supplemental material for its proof. This result is primarily an extension of a pivotal bootstrap method, proposed by Fisher et al. (1996), for resampling directional data in $\mathbb{R}^d$ and calibrating conference regions for such data. Our emphasis differs, however, in that our goal is distributional approximations and regions for rotation data.

Theorem 3.3. *Let $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ be a random sample of i.i.d. observations from the location model in (3.1). For the bootstrap version T_n^* in (3.19) of T_n in (3.18), as $n \to \infty$,*

$$\sup_{x \in \mathbb{R}} |P\left(T_n \leq x\right) - P_*\left(T_n^* \leq x\right)| \longrightarrow 0 \quad \textit{with probability one.}$$

Hence, as an alternative to the normal theory region (3.14), Theorem 3.3 justifies an approximate $100(1-\alpha)\%$ bootstrap confidence region for the central orientation $\boldsymbol{S}$ as

$$\left\{ \boldsymbol{S} = \mu(\boldsymbol{m}) : \boldsymbol{m} \in \mathbb{R}^4,\ \|\boldsymbol{m}\| = 1,\ n\boldsymbol{m}^\top \widehat{\boldsymbol{F}}_n^- \boldsymbol{m} < t_{1-\alpha} \right\} \tag{3.20}$$

where $t_{1-\alpha}$ denotes the (lower) $(1-\alpha)$ percentile of the bootstrap distribution of T_n^*. The next section considers the performance of bootstrap confidence region methods through numerical studies.

3.5 Simulation Study

Here we summarize a comprehensive simulation study of finite-sample coverage accuracy of confidence regions for the location parameter $\boldsymbol{S} \in SO(3)$ of rotation data, based on both normal theory and bootstrap methods from Sections 3.3 and 3.4, respectively.

To provide some initial motivation for the bootstrap methods, recall that the normal theory region (3.10) from the direct approach, for example, depends on a limiting χ^2_3 approximation for the sampling distribution of the studentized statistic $D_n \equiv (2n\hat{a}_{2_n}^2/\hat{a}_{1_n})\|\widehat{\boldsymbol{h}}_n\|^2$ (cf. Proposition 3.1). To assess the quality of the chi-square approximation, we generated rotation data under the location model (3.1) using a Cayley distribution (other common distributions such as the matrix Fisher and circular-von Mises distributions produced similar results), considering samples of size $n = 10, 20, 50$ and 100 along with concentration parameters $\kappa = 1$ and 8

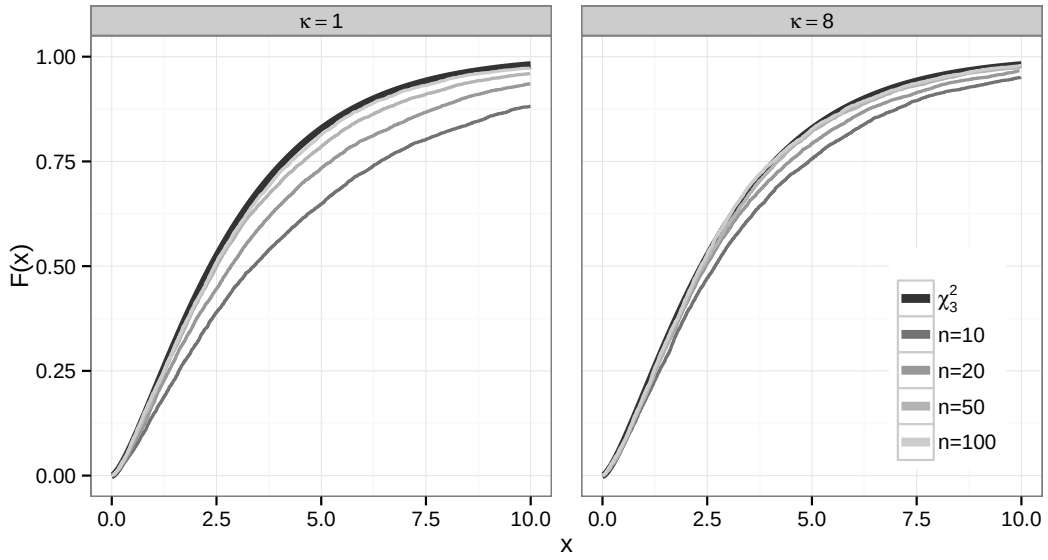


Figure 3.1: Empirical (cumulative) distribution function of $x = 2n\hat{a}_{2_n}^2 \|\widehat{\boldsymbol{h}}_n\|^2/\hat{a}_{1_n}$ for data generated from the Cayley distribution with central orientation $\boldsymbol{I}_{3\times 3}$ and concentration parameter $\kappa = 1$ (left) and 8 (right) plotted over the limiting χ^2_3 (cumulative) distribution function.

(impacting the dispersion of each distributional model). For each (n, κ) combination, the pivotal quantity D_n was computed over 5,000 generated samples. The resulting (approximated) distribution function of D_n is plotted in Figure 3.1 along with the theoretical limiting χ^2_3 distribution function. For highly concentrated data (right panel of Figure 3.1), the distribution of D_n follows the chi-square limit well for samples as small as $n = 50$, but the distribution approximation is less adequate for smaller concentration values κ (left panel of Figure 3.1). Figure 3.1 additionally indicates that normal theory confidence regions (3.10) can often be expected to be too small to achieve a nominal coverage rate. For example, from (3.11) a 90% confidence region for $\boldsymbol{S}$ based on a sample of size $n = 10$ from the Cayley distribution involves a calibration cutoff of $\chi^2_{3,0.9} = 6.25$. However, this cutoff only corresponds to the 74^{th} and 84^{th} percentiles of the sampling distribution of D_n for $\kappa = 1$ and 8, respectively, whereas the true 90^{th} percentile values are approximately 10.86 for $\kappa = 1$ and 7.60 for $\kappa = 8$. This motivates the bootstrap for improving distribution approximations with D_n and calibrating confidence regions as in (3.17), particularly for small samples with low concentration.

To evaluate the relative performance of the bootstrap regions compared to normal theory

approaches, we generated 10,000 samples of rotations $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ according to model (3.1) by generating random perturbations $\boldsymbol{E}_i$ from each of three distributional models (Cayley, circular-von Mises and matrix Fisher distributions) and setting $\boldsymbol{R}_i = \boldsymbol{S}\boldsymbol{E}_i$, $i = 1, \ldots, n$ for $n = 10, 20, 50$ and 100. (Without loss of generality, we set $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$). Because the concentration parameter κ does not translate exactly across distributional models, we considered generating data and reporting results in terms of the circular variance $\nu = 1 - \mathrm{E}[\cos(r)]$, rather than κ. In particular, we considered circular variances $\nu = 0.25$, 0.50 and 0.75. Table 3.2 illustrates how the concentration parameter κ and the circular variance are related for these distributions.

Table 3.2: Values of κ for each rotational distribution corresponding to the circular variances.

Distribution	**Circular variance ν**		
	0.25	0.50	0.75
Cayley	10.00	4.00	2.00
circular-von Mises	2.40	1.16	0.52
matrix Fisher	3.17	1.71	1.15

In Figure 3.2 and Table 3.3 (see appendix) we present coverage rates of the two normal theory and two bootstrap confidence regions of Sections 3.3 and 3.4, with nominal confidence level 90%. Namely, for each combination of sample size, distribution and circular variance combination, 10,000 samples were generated and, for each sample, it was determined if the true central orientation was contained in each confidence region. Following the recommendations in Hall (1986), we determined critical values for the bootstrap confidence regions based on 300 resampled data sets. Results for $\nu = 0.25$ and 0.75 are summarized graphically in Figure 3.2; results for $\nu = 0.5$ overall followed a similar pattern and are therefore omitted. In this figure, the distributions are displayed according to the proportion of observations in the tail: the Cayley distribution has the least heavy tail while the circular-von Mises distribution has the heaviest. Exact numerical results for all values of ν and each of the distributions are given in Table 3.3. For every combination of the circular variance ν and the distributional models, all four methods converge to the nominal coverage rate as n increases.

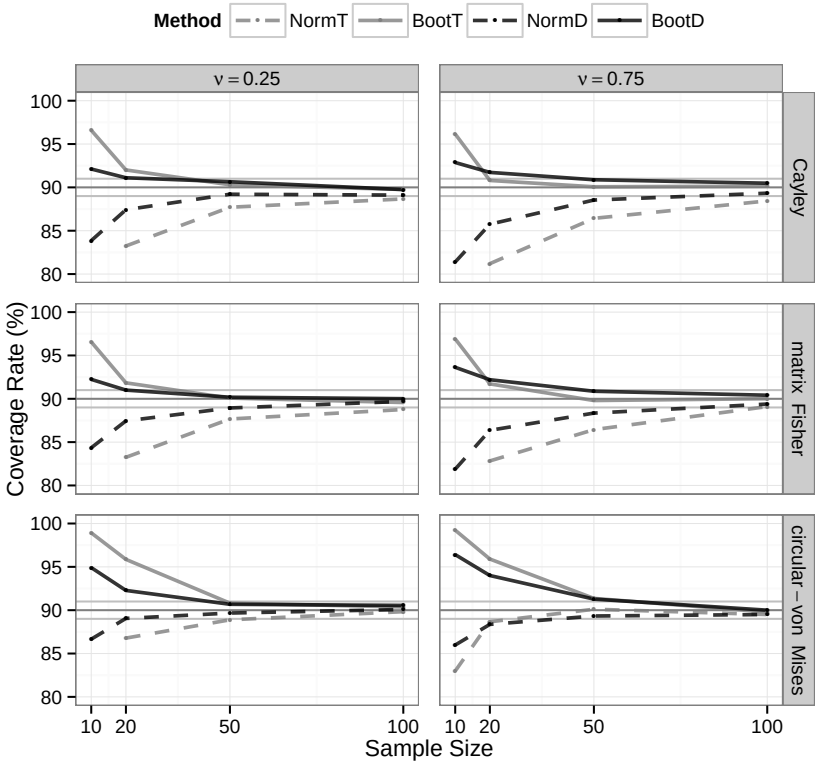


Figure 3.2: Confidence region coverage rates as a function of sample size for different circular variances (ν) and distributions. The horizontal line indicates the nominal coverage rate 90%. See Table 3.3 for a tabular account of these results.

In comparing the normal theory and bootstrap calibrated confidence regions, in agreement with Figure 3.1, the normal theory regions (dashed lines) can exhibit coverage rates below the nominal level. In contrast, bootstrap calibrated confidence regions (solid lines) tend to be conservative with coverage rates often converging quickly to the nominal level as the sample size increases. Within the normal theory methods, the coverage rates of the direct approach converge more quickly to the nominal level as a function of n compared to the transformation-based approach. We conclude that, in general, the bootstrap calibrated confidence regions are preferred to the large sample normal theory-based ones, particularly for small samples.

We next compare the direct and transformation-based bootstrap approaches. Confidence regions calibrated by the direct bootstrap method (in black) achieve coverage rates closer to

the nominal 90% faster than those based on the transformation-based method (in gray) for small samples. For highly variable samples (e.g. $\nu = 0.75$) however, regions based on the transformation-based bootstrap appear to converge slightly faster to the nominal coverage rate compared to the direct boostrap counterparts, as a function of sample size. These numerical results suggest, however, that the direct bootstrap approach is the generally preferred approach to setting a confidence region for the rotation location parameter.

3.6 Data Example

In Sections 3.3.1 and 3.4.1, we explained that confidence regions using the direct approach can be interpreted as the set of all rotations within a fixed geodesic distance of the projected mean $\widehat{\boldsymbol{S}}_n$. This interpretation also leads to a natural definition of confidence region size. In particular, the size of a $100(1-\alpha)\%$ normal theory-based confidence region (3.10) can be measured by $[\chi^2_{3,1-\alpha}\hat{a}_{1_n}/(2n\hat{a}^2_{2_n})]^{1/2}$, which is a function of sample size, confidence level and within–sample variability. The following data example uses the notion of confidence region size to estimate the precision of an orientation measurement tool common in material sciences.

			LSA		Bootstrap	
Distribution	ν	n	Directional	Direct	Directional	Direct
Cayley	0.25	10	73.71	83.86	96.59	92.12
		20	83.24	87.42	92.01	91.11
		50	87.71	89.21	90.30	90.65
		100	88.67	89.11	89.76	89.75
	0.50	10	74.12	83.55	96.54	92.54
		20	82.53	86.73	91.23	91.18
		50	87.03	88.55	89.70	90.30
		100	88.37	89.31	89.77	89.96
	0.75	10	71.73	81.39	96.16	92.87
		20	81.20	85.75	90.80	91.73
		50	86.43	88.54	90.07	90.88
		100	88.44	89.33	90.14	90.48
matrix Fisher	0.25	10	74.73	84.35	96.56	92.24
		20	83.25	87.45	91.84	91.00
		50	87.67	88.94	90.08	90.18
		100	88.76	89.70	89.57	90.01
	0.50	10	74.11	83.14	96.71	92.68
		20	83.47	87.65	92.11	91.72
		50	87.29	88.73	89.95	89.96
		100	88.93	89.38	89.73	89.92
	0.75	10	73.51	81.88	96.91	93.63
		20	82.83	86.36	91.72	92.20
		50	86.44	88.35	89.80	90.87
		100	89.10	89.39	90.01	90.41
circular- von Mises	0.25	10	79.10	86.66	98.92	94.91
		20	86.78	89.05	95.86	92.28
		50	88.89	89.67	90.81	90.68
		100	89.83	90.08	90.37	90.55
	0.50	10	82.36	86.66	99.02	95.54
		20	87.71	88.47	95.39	92.24
		50	89.29	89.29	90.78	90.38
		100	89.81	89.93	89.96	90.10
	0.75	10	82.98	85.98	99.23	96.41
		20	88.66	88.36	95.90	94.00
		50	90.11	89.33	91.36	91.28
		100	89.50	89.50	89.79	90.00

Table 3.3: Coverage rates for different confidence region methods for $\boldsymbol{S}$ compared to the nominal coverage rate of 90%. See Figure 3.2 for a graphical representation of this table.

We consider electron backscatter diffraction (EBSD) data obtained by scanning a fixed 12.5 μm $\times$ 10 μm nickel surface at individual locations spaced 0.2 μm apart (Bingham et al., 2009, 2010a). Every observation corresponds to the orientation of a cubic crystal on the metal surface at a particular location expressed as a rotation matrix. One goal of processing EBSD data is to estimate the within-grain precision of the EBSD instrument where a grain is defined as a region of cubic crystals with similar orientations on the metal surface, i.e. regions that share a common central orientation $\boldsymbol{S}$. Studies currently in the literature have reported a within-grain variability of 1° though the methods used to arrive at this estimate of precision are not always statistically formal or obvious (Demirel et al., 2000; Wilson and Spanos, 2001; Bingham et al., 2009). In the following, we investigate the claim of 1° precision by constructing confidence regions based on direct normal-theory and bootstrap approaches.

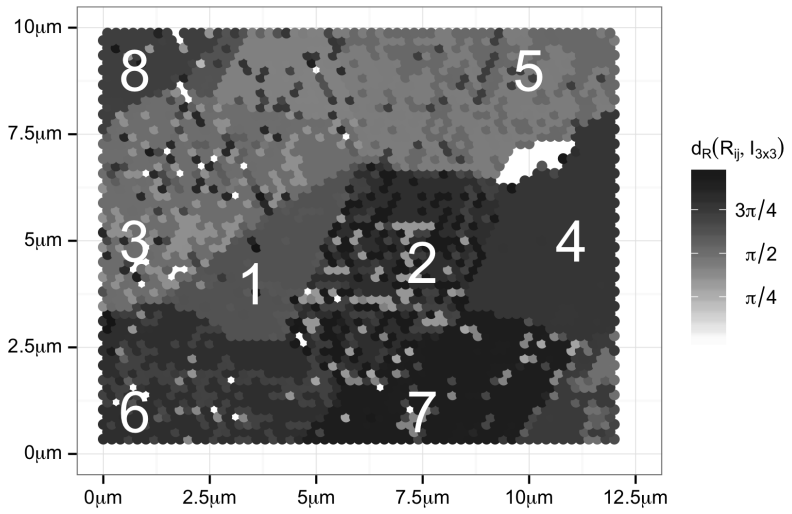


Figure 3.3: Display of all locations of the investigated nickel surface. Shading reflects the misorientation angle r_{g_i} of the observed cubic crystal orientation $\boldsymbol{R}_{g_i}$ with respect to the identity rotation $\boldsymbol{I}_{3\times3}$. Eight distinct grains are investigated.

Figure 3.3 displays the results of the aforementioned scan. Let $\boldsymbol{R}_{g_i}$ represent the random variable associated with the observed rotation matrix at location i within grain g. At each measured location, the shading represents the size of the misorientation angle r_{g_i} (in degrees)

defined in (3.4) for every rotation $\boldsymbol{R}_{g_i}$ with respect to the identity rotation $\boldsymbol{I}_{3\times 3}$, i.e. $r_{g_i} = 90\arccos[\mathbf{tr}(\boldsymbol{I}_{3\times 3}^{\top}\boldsymbol{R}_{g_i}) - 1]/\pi$. A distinct spatial structure resembling a grain map is present; eight distinct spatial areas with similar orientations were identified and are marked. Each grain is assumed to have a distinct central orientation $\boldsymbol{S}_g$ for $g = 1, \ldots, 8$ and each observation is assumed to stem from the data model (3.1), $\boldsymbol{R}_{g_i} = \boldsymbol{S}_g\boldsymbol{E}_{g_i}$, where $i = 1, \ldots, n_g$ indicates the observation number within grain g. The value n_g represents the total number of observations within grain g and the perturbation matrices $\boldsymbol{E}_{g_i}$ follow a rotationally symmetric distribution on $SO(3)$. The exact form of the distribution for the random perturbations $\boldsymbol{E}_{g_i}$ need not be specified since the confidence regions presented here are non-parametric. For each of the eight identified grains, normal theory (3.10) and bootstrap (3.17) regions based on the direct approach are used to form a 99% confidence region for $\boldsymbol{S}$.

Table 3.4: Size (in degrees) of the 99% confidence regions for $\boldsymbol{S}_g$ for each grain and adjusted for sample size based on the direct large sample theory (NormD) and bootstrap (BootD) approach. The grains are identified in Figure 3.3.

Grain (g)	n_g	NormD	BootD
1	27	0.770	1.017
2	46	0.770	0.869
3	255	0.953	0.989
4	354	0.738	0.723
5	145	1.042	1.078
6	246	1.370	1.571
7	272	0.903	0.915
8	99	1.091	1.138

The size (in degrees) of the large sample theory and bootstrap calibrated confidence regions for each grain g are adjusted for grain size and reported in Table 3.4. To illustrate, consider the normal theory confidence region for grain $g = 1$. Let $\boldsymbol{R}_{1_1}, \ldots, \boldsymbol{R}_{1_{27}}$ represent the $n_1 = 27$ observed rotations in grain one and $\widehat{\boldsymbol{S}}_{1_{27}}$ represent the estimate of the true central orientation associated with grain one, $\boldsymbol{S}_1$, based on the 27 observations. The size of the 99% large sample normal theory confidence region for $\boldsymbol{S}_1$ based on $\widehat{\boldsymbol{S}}_{1_{27}}$ is given by $\{\chi^2_{3,1-\alpha}\widehat{a}_{1,g}/[2(n_g)\widehat{a}^2_{2,g}]\}^{1/2} = \{\chi^2_{3,.01}\widehat{a}_{1,1}/[2(27)\widehat{a}^2_{2,1}]\}^{1/2}$ where $\widehat{a}_{1,1}$ and $\widehat{a}_{2,1}$ are the sample estimates of a_1 and a_2 for grain one computed by (3.10). Multiplying the confidence region size by $180/\pi$ translates it into

degrees, followed by adjusting for sample size through multiplication by $\sqrt{n_1} = \sqrt{27}$.

From Table 3.4, we see that the bootstrap method generally returns a more conservative (larger) confidence region (which supports the findings of the simulation study), though this is not always the case. In general, it appears however that the reported 1° within-grain variability is reasonable based on the bootstrap confidence regions.

3.7 Discussion

For the location model in $SO(3)$, current methods to compute confidence regions for the location parameter $\boldsymbol{S}$ are based on large sample normal approximations, which may suffer in approximation performance in small samples. In this manuscript, we examined two bootstrap confidence region approaches for the central orientation $\boldsymbol{S}$. Like their normal theory counterparts, one bootstrap method was based on rotation data directly and the second bootstrap involved a transformation of rotations to directional data for inference. Theoretical results established that the bootstrap provides consistent estimators of sampling distributions, needed for calibrating confidence regions. Simulation studies also showed that the bootstrap methods typically exhibited better coverage accuracies than the normal theory approaches and did not overstate the level of confidence. In particular, the newly proposed direct bootstrap method emerged as a generally preferred approach to confidence region estimation for location parameters in the rotation data model.

The confidence regions in this manuscript are all based on the projected mean estimator which, like mean estimators in other contexts, can be a non-robust estimator of the central orientation $\boldsymbol{S}$. In future work, the authors plan to develop confidence regions for the location parameter based on robust estimators, such as the projected median in $SO(3)$.

3.8 Appendix

In Section 3.8 we prove the asymptotic normality of the extrinsic mean and that the proposed bootstrap is consistent.

Proofs

Proof of Proposition 3.1.

We first develop an expansion of the trace $\mathbf{tr}(\tilde{\boldsymbol{S}}^\top \boldsymbol{R}_i), i = 1, \ldots, n$ requiring maximization over all $\tilde{\boldsymbol{S}} \in SO(3)$ (specifically over all $\boldsymbol{h} \in \mathbb{R}^3$ where $\tilde{\boldsymbol{S}} = \boldsymbol{S}\exp[\boldsymbol{\Phi}(\boldsymbol{h})]$). This is accomplished in the following, where we develop and justify an expansion of $f_i(\boldsymbol{h}) = \mathbf{tr}[\exp(\boldsymbol{\Phi}[\boldsymbol{h}])^\top \boldsymbol{S}^\top \boldsymbol{R}_i]$, $i = 1, \ldots, n$, making use of properties of quadratic forms and powers of skew-symmetric matrices. We then show that the vector $\widehat{\boldsymbol{h}}_n$ that maximizes $g_n(\boldsymbol{h}) = n^{-1}\sum_i f_i(\boldsymbol{h})$ (and therefore is a solution to $\partial g_n(\boldsymbol{h})/\partial \boldsymbol{h} = \mathbf{0}_3$) has a trivariate normal limiting distribution. To accomplish this we first rewrite $\partial g_n(\widehat{\boldsymbol{h}}_n)/\partial \boldsymbol{h}$ as a linear function of $\widehat{\boldsymbol{h}}_n$ and higher order terms. Then, based on the linearization of $\partial g_n(\widehat{\boldsymbol{h}}_n)/\partial \boldsymbol{h}$, we use the Strong Law of Large Numbers (SLLN), Central Limit Theorem (CLT) and Slutsky's Theorem to prove the main result.

We require some properties of skew-symmetric matrices. If $\boldsymbol{u} = (u_1, u_2, u_3)^\top \in \mathbb{R}^3$, then

$$\boldsymbol{\Phi}(\boldsymbol{u}) = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \tag{3.21}$$

is skew-symmetric so that $\mathbf{tr}[\boldsymbol{\Phi}(\boldsymbol{u})] = 0$ and, for any $\boldsymbol{x} = (x_1, x_2, x_3)^\top \in \mathbb{R}^3$,

$$\mathbf{tr}(\boldsymbol{x}^\top \boldsymbol{\Phi}(\boldsymbol{u})\boldsymbol{x}) = x_1(-u_3x_2 + u_2x_3) + x_2(u_3x_1 - u_1x_3) + x_3(u_1x_2 - u_2x_1) = 0. \tag{3.22}$$

Additionally, for $\boldsymbol{u} \in \mathbb{R}^3$, $\|\boldsymbol{u}\| = 1$, and integer $k \geq 1$, it holds that

$$[\boldsymbol{\Phi}(\boldsymbol{u})]^{2k} = (-1)^k(\boldsymbol{I}_{3\times 3} - \boldsymbol{u}\boldsymbol{u}^\top), \qquad [\boldsymbol{\Phi}(\boldsymbol{u})]^{2k+1} = (-1)^k\boldsymbol{\Phi}(\boldsymbol{u}). \tag{3.23}$$

Now for $\boldsymbol{h} \neq 0 \in \mathbb{R}^3$, write

$$\exp[\boldsymbol{\Phi}(\boldsymbol{h})] = \sum_{k=0}^{\infty} \frac{[\boldsymbol{\Phi}(\boldsymbol{h})]^k}{k!} = \sum_{k=0}^{\infty} \frac{[\|\boldsymbol{h}\|\boldsymbol{\Phi}(\boldsymbol{w})]^k}{k!} \tag{3.24}$$

where $\boldsymbol{w} = \boldsymbol{h}/\|\boldsymbol{h}\|$ with $\|\boldsymbol{w}\| = 1$. Replacing the even and odd powers of $[\boldsymbol{\Phi}(\boldsymbol{w})]$ in (3.24) via (3.23) gives

$$\begin{aligned}\exp\left[\boldsymbol{\Phi}(\boldsymbol{h})\right] &= \boldsymbol{I}_{3\times 3} + \sum_{k=0}^{\infty} \frac{(-1)^k \|\boldsymbol{h}\|^{2k+1}}{(2k+1)!} \boldsymbol{\Phi}(\boldsymbol{w}) + \sum_{k=1}^{\infty} \frac{(-1)^k \|\boldsymbol{h}\|^{2k}}{(2k)!} (\boldsymbol{I}_{3\times 3} - \boldsymbol{w}\boldsymbol{w}^\top) \\ &= \boldsymbol{I}_{3\times 3} + \boldsymbol{\Phi}(\boldsymbol{h}) \sum_{k=0}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^k}{(2k+1)!} \\ &\quad + (\boldsymbol{I}_{3\times 3} \boldsymbol{h}^\top \boldsymbol{h} - \boldsymbol{h}\boldsymbol{h}^\top) \sum_{k=1}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^{k-1}}{(2k)!} \end{aligned} \tag{3.25}$$

using $\|\boldsymbol{h}\|^2 = \boldsymbol{h}^\top \boldsymbol{h}$, $\|\boldsymbol{h}\| \boldsymbol{\Phi}(\boldsymbol{w}) = \boldsymbol{\Phi}(\boldsymbol{h})$. The expansion (3.25) holds for all $\boldsymbol{h} \in \mathbb{R}^3$ including $\boldsymbol{h} = \boldsymbol{0}_3$ and is continuously differentiable in $\boldsymbol{h}$.

The data are i.i.d. $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ where $\boldsymbol{R}_i = \boldsymbol{S}\boldsymbol{E}_i$ for $\boldsymbol{E}_i \equiv \boldsymbol{E}_i(r_i, \boldsymbol{U}_i)$ defined by i.i.d $r_1, \ldots, r_n \in (-\pi, \pi]$ distributed symmetrically around 0 and independent of $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ being uniformly distributed on $\mathbb{R}^3$-sphere; see (5) of the main manuscript. For $\boldsymbol{h} \in \mathbb{R}^3$, define a function $f_i : \mathbb{R}^3 \to \mathbb{R}$ for each $i = 1, \ldots, n$ as $f_i(\boldsymbol{h}) = \mathbf{tr}\left[\exp\left(\boldsymbol{\Phi}[\boldsymbol{h}]\right)^\top \boldsymbol{S}^\top \boldsymbol{R}_i\right]$. Without loss of generality, we assume $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$. By Rodrigues' formula (cf. Sec. 2), $\boldsymbol{R}_i$ can then be written as $\boldsymbol{R}_i = \boldsymbol{u}_i \boldsymbol{u}_i^\top (1 - \cos r_i) + \boldsymbol{I}_{3\times 3} \cos r_i + \boldsymbol{\Phi}(\boldsymbol{u}_i) \sin r_i$ with $\mathbf{tr}(\boldsymbol{R}_i) = 1 + 2\cos(r_i)$ by (3.21). From (3.22), it follows for $\boldsymbol{w} \in \mathbb{R}^3$, $\|\boldsymbol{w}\| = 1$, that

$$\begin{aligned} \mathbf{tr}\left[\boldsymbol{\Phi}(\boldsymbol{w}) \boldsymbol{R}_i\right] &= \mathbf{tr}\left[\sin r_i \boldsymbol{\Phi}(\boldsymbol{w}) \boldsymbol{\Phi}(\boldsymbol{u}_i)\right] \\ &= -2 \sin r_i (\boldsymbol{w}^\top \boldsymbol{u}_i) \\ \mathbf{tr}\left[(\boldsymbol{I}_{3\times 3} - \boldsymbol{w}\boldsymbol{w}^\top) \boldsymbol{R}_i\right] &= \mathbf{tr}(\boldsymbol{R}_i) - \mathbf{tr}(\boldsymbol{w}^\top \boldsymbol{R}_i \boldsymbol{w}) \\ &= 1 + 2\cos r_i - (1 - \cos r_i)(\boldsymbol{w}^\top \boldsymbol{u}_i)^2 - \cos r_i \\ &= 1 + \cos r_i - (1 - \cos r_i)(\boldsymbol{w}^\top \boldsymbol{u}_i)^2. \end{aligned}$$

From the above and (3.25), we may re-write for each $i = 1, \ldots, n$:

$$\begin{aligned} f_i(\boldsymbol{h}) &= \mathbf{tr}\left[\exp\left(\boldsymbol{\Phi}[\boldsymbol{h}]\right)^\top \boldsymbol{R}_i\right] \\ &= \mathbf{tr}\left[\exp\left(\boldsymbol{\Phi}[-\boldsymbol{h}]\right) \boldsymbol{R}_i\right] \\ &= 1 + 2\cos r_i + 2(\sin r_i)(\boldsymbol{h}^\top \boldsymbol{u}_i) \sum_{k=0}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^k}{(2k+1)!} \\ &\quad + \left[(\boldsymbol{h}^\top \boldsymbol{h})(1+\cos r_i) - (1-\cos r_i)(\boldsymbol{h}^\top \boldsymbol{u}_i)^2\right] \sum_{k=1}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^{k-1}}{(2k)!} \end{aligned} \tag{3.26}$$

for any $\boldsymbol{h} \in \mathbb{R}^3$.

Then, $\widehat{\boldsymbol{S}}_n = \exp\left[\boldsymbol{\Phi}(\widehat{\boldsymbol{h}}_n)\right]$ estimates $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$, where $\widehat{\boldsymbol{h}}_n \in \mathbb{R}^3$ maximizes the function $g_n(\boldsymbol{h}) = n^{-1} \sum_{i=1}^n f_i(\boldsymbol{h})$. The SLLN yields $\widehat{\boldsymbol{S}}_n \to \boldsymbol{I}_{3\times 3}$ with probability 1 (w.p.1), implying that $\widehat{\boldsymbol{h}}_n \to \mathbf{0}_3$ w.p.1 from the continuity of $\exp\left[\boldsymbol{\Phi}(\boldsymbol{h})\right]$ in (3.24) and the fact that $\exp\left[\boldsymbol{\Phi}(\boldsymbol{h})\right] = \boldsymbol{I}_{3\times 3}$ only if $\boldsymbol{h} = \mathbf{0}_3$. Let A_n be the event "$\|\widehat{\boldsymbol{h}}_n\| < 1/4$" and note $P(A_n) \to 1$ as $n \to \infty$ by $\widehat{\boldsymbol{h}}_n \xrightarrow{p} \mathbf{0}_3$. When event A_n holds, the function $g_n(\boldsymbol{h})$ has a maximum inside the closed ball $B(1/4) \equiv \{\boldsymbol{h} \in \mathbb{R}^3 : \|\boldsymbol{h}\| \leq 1/4\}$ (at $\widehat{\boldsymbol{h}}_n$) and, because $g_n(\boldsymbol{h})$ is differentiable, it holds that $\partial g_n(\boldsymbol{h})/\partial \boldsymbol{h}|_{\boldsymbol{h}=\widehat{\boldsymbol{h}}_n} = \mathbf{0}_3$. Based on the linearized version of $f_i(\boldsymbol{h})$ in (3.26), for $\boldsymbol{h} = (h_1, h_2, h_3)^\top$ and $\boldsymbol{u}_i = (u_{i1}, u_{i2}, u_{i3})^\top$, $i = 1, \ldots, n$, note that the partial derivatives $j = 1, 2, 3$ may be expanded as

$$\begin{aligned} \frac{\partial g_n(\boldsymbol{h})}{\partial h_j} &= \frac{1}{n}\sum_{i=1}^n (2\sin r_i)\left[u_{ij} \sum_{k=0}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^k}{(2k+1)!} + (\boldsymbol{h}^\top \boldsymbol{u}_i) \sum_{k=1}^{\infty} \frac{(-1)^k k (\boldsymbol{h}^\top \boldsymbol{h})^{k-1} 2h_j}{(2k+1)!}\right] \\ &\quad + \frac{1}{n}\sum_{i=1}^n \left[\left\{(1+\cos r_i)2h_j - (1-\cos r_i)(\boldsymbol{h}^\top \boldsymbol{u}_i)2u_{ij}\right\} \sum_{k=1}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^{k-1}}{(2k)!}\right] \\ &\quad + \frac{1}{n}\sum_{i=1}^n \left[\left\{(\boldsymbol{h}^\top \boldsymbol{h})(1+\cos r_i) - (1-\cos r_i)(\boldsymbol{h}^\top \boldsymbol{u}_i)^2\right\} \sum_{k=2}^{\infty} \frac{(-1)^k (k-1)(\boldsymbol{h}^\top \boldsymbol{h})^{k-2} 2h_j}{(2k)!}\right] \\ &= \frac{2}{n}\sum_{i=1}^n u_{ij} \sin r_i + \frac{1}{n}\sum_{i=1}^n (1-\cos r_i)(\boldsymbol{h}^\top \boldsymbol{u}_i)u_{ij} - \frac{1}{n}\sum_{i=1}^n (1+\cos r_i)h_j + \tilde{R}_{nj}(\boldsymbol{h}) \end{aligned} \tag{3.27}$$

where

$$\begin{aligned}\tilde{R}_{nj}(\boldsymbol{h}) = \frac{1}{n}\sum_{i=1}^{n}(2\sin r_i)\left[u_{ij}\sum_{k=1}^{\infty}\frac{(-1)^k(\boldsymbol{h}^\top\boldsymbol{h})^k}{(2k+1)!} + 2h_j(\boldsymbol{h}^\top\boldsymbol{u}_i)\sum_{k=1}^{\infty}\frac{(-1)^k k(\boldsymbol{h}^\top\boldsymbol{h})^{k-1}}{(2k+1)!}\right] \\ + \frac{1}{n}\sum_{i=1}^{n}\left[\left\{(1+\cos r_i)2h_j - (1-\cos r_i)(\boldsymbol{h}^\top\boldsymbol{u}_i)2u_{ij}\right\}\sum_{k=2}^{\infty}\frac{(-1)^k(\boldsymbol{h}^\top\boldsymbol{h})^{k-1}}{(2k)!}\right] \\ + \frac{1}{n}\sum_{i=1}^{n}\left[\left\{(\boldsymbol{h}^\top\boldsymbol{h})(1+\cos r_i)\right\}2h_j\sum_{k=2}^{\infty}\frac{(-1)^k(k-1)(\boldsymbol{h}^\top\boldsymbol{h})^{k-2}}{(2k)!}\right] \\ - \frac{1}{n}\sum_{i=1}^{n}\left[\left\{(1-\cos r_i)(\boldsymbol{h}^\top\boldsymbol{u}_i)^2\right\}2h_j\sum_{k=2}^{\infty}\frac{(-1)^k(k-1)(\boldsymbol{h}^\top\boldsymbol{h})^{k-2}}{(2k)!}\right].\end{aligned}$$

so that

$$\begin{aligned}|\tilde{R}_{nj}(\boldsymbol{h})| &\le 2\|\boldsymbol{h}\|^2\sum_{k=1}^{\infty}\frac{1}{(2k+1)!} + 2\|\boldsymbol{h}\|^2\sum_{k=1}^{\infty}\frac{k}{(2k+1)!} + 4\|\boldsymbol{h}\|^3\sum_{k=1}^{\infty}\frac{1}{(2k)!} \\ &\quad + 4\|\boldsymbol{h}\|^3\sum_{k=2}^{\infty}\frac{1}{(2k)!} + 4\|\boldsymbol{h}\|^2\sum_{k=2}^{\infty}\frac{k-1}{(2k)!} \\ &\le c\|\boldsymbol{h}\|^2 \qquad (3.28)\end{aligned}$$

uniformly for $\boldsymbol{h} \in B(1/4)$ and $j = 1, 2, 3$ for some $c \in (0, \infty)$. Therefore, evaluating (3.27) at $\widehat{\boldsymbol{h}}_n$ and using (3.28) gives

$$\begin{aligned}\mathbf{0}_3 &= \frac{\partial g_n(\widehat{\boldsymbol{h}}_n)}{\partial \boldsymbol{h}} \\ &= \frac{2}{n}\sum_{i=1}^{n}\sin r_i\boldsymbol{u}_i + \frac{1}{n}\sum_{i=1}^{n}(1-\cos r_i)\boldsymbol{u}_i\boldsymbol{u}_i{}^\top\widehat{\boldsymbol{h}}_n - \frac{1}{n}\sum_{i=1}^{n}(1+\cos r_i)\widehat{\boldsymbol{h}}_n + O_p(\|\widehat{\boldsymbol{h}}_n\|^2) \\ &\equiv \boldsymbol{M}_n + \boldsymbol{C}_n\widehat{\boldsymbol{h}}_n + o_p(\|\widehat{\boldsymbol{h}}_n\|) \qquad (3.29)\end{aligned}$$

for

$$\boldsymbol{M}_n = \frac{2}{n}\sum_{i=1}^{n}\sin r_i\boldsymbol{u}_i \quad \text{and} \quad \boldsymbol{C}_n = \frac{1}{n}\sum_{i=1}^{n}\left[(1-\cos r_i)\boldsymbol{u}_i\boldsymbol{u}_i{}^\top - (1+\cos r_i)\boldsymbol{I}_{3\times 3}\right].$$

In the following, define variables $(r, \boldsymbol{U}) \overset{d}{=} (r_i, \boldsymbol{U}_i)$ for simplicity. By SLLN, as r is independent of $\boldsymbol{U}$ and $\mathrm{E}(\boldsymbol{U}\boldsymbol{U}^\top) = \boldsymbol{I}_{3\times 3}/3$, it follows that $\boldsymbol{C}_n \to \boldsymbol{C}$ w.p.1 in (3.29), where

$$\boldsymbol{C} \equiv E\left[(1-\cos r)\boldsymbol{U}\boldsymbol{U}^\top - (1+\cos r)\boldsymbol{I}_{3\times 3}\right] = \frac{-2-4\mathrm{E}\cos r}{3}\boldsymbol{I}_{3\times 3}.$$

Letting $a_2 = (1+2\mathrm{E}\cos r)/3$, $\boldsymbol{C}_n^{-1} \to \boldsymbol{C}^{-1} = -\boldsymbol{I}_{3\times 3}/2a_2$ w.p.1 so that $\boldsymbol{C}_n^{-1} \xrightarrow{p} \boldsymbol{C}^{-1}$ (i.e., for large n, $\boldsymbol{C}_n$ is invertible with high probability). By the classical multivariate CLT for $\boldsymbol{M}_n$ in

(3.29),

$$\sqrt{n}\,\boldsymbol{M}_n \xrightarrow{d} N_3(\mathbf{0}_3, 4\mathrm{Var}[\sin r\boldsymbol{U}])$$

where $\mathrm{E}\sin r\boldsymbol{U} = (\mathrm{E}\sin r)(\mathrm{E}\boldsymbol{U}) = \mathbf{0}_3$, as r is symmetric around 0 and $\sin(\cdot)$ is an odd function, and

$$\mathrm{Var}(\sin r\,\boldsymbol{U}) = \mathrm{E}\left[\sin^2(r)\,\boldsymbol{U}\boldsymbol{U}^\top\right] - \mathbf{0}_3\mathbf{0}_3^\top = \mathrm{E}(\sin^2 r)E(\boldsymbol{U}\boldsymbol{U}^\top) = \frac{1}{3}\mathrm{E}\sin^2 r\boldsymbol{I}_{3\times 3}.$$

Hence, we have from (3.29) that

$$\sqrt{n}\,\widehat{\boldsymbol{h}}_n(1+o_p(1)) = \boldsymbol{C}_n^{-1}(-\sqrt{n}\,\boldsymbol{M}_n) \xrightarrow{d} N_3\left(\mathbf{0}_3, \frac{a_1}{2a_2^2}\boldsymbol{I}_{3\times 3}\right)$$

where $a_1 = 2\mathrm{E}\sin^2 r/3$. Thus $\sqrt{n}\,\widehat{\boldsymbol{h}}_n = O_p(1)$ and $\sqrt{n}\,\widehat{\boldsymbol{h}}_n o_p(1) = o_p(1)$, implying $\sqrt{n}\,\widehat{\boldsymbol{h}}_n = \boldsymbol{C}_n^{-1}(-\sqrt{n}\,\boldsymbol{M}_n) + o_p(1)$ and therefore $\sqrt{n}\,\widehat{\boldsymbol{h}}_n \xrightarrow{d} N_3(\mathbf{0}_3, a_1/2a_2^2)$ by Slutsky's theorem. □

In order to prove Theorem 3.2, we first establish a proposition.

Proposition 3.4. *Assume $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ are a sample of i.i.d. observations from a distribution F defined on $SO(3)$ by the location model (1) with central orientation parameter $\boldsymbol{S} \in SO(3)$ and concentration κ. Let $\widehat{\boldsymbol{S}}_n$ denote the rotation matrix satisfying $\arg\min_{\boldsymbol{S}\in SO(3)} \sum_{i=1}^n \|\boldsymbol{R}_i - \boldsymbol{S}\|_F^2$. Expressed terms of a (non-random) $\boldsymbol{R}$ where $\boldsymbol{S}^\top\boldsymbol{R} \equiv \boldsymbol{S}^\top\boldsymbol{R}(r,\boldsymbol{U})$ defined by an angle-axis pair $(r,\boldsymbol{U})$ as in (4), the influence function of $\widehat{\boldsymbol{S}}_n$ is given by*

$$IF_2(\boldsymbol{R}, F) = \frac{1}{a_2}\sin(r)\boldsymbol{U}$$

where $a_2 = \mathrm{E}(1+2\cos r_1)/3$ for a random angle r_1 (5) defining a random rotation $\boldsymbol{R}_1$ as above. Furthermore,

$$\mathrm{E}[IF_2(\boldsymbol{R}_1, F)] = 0 \qquad \textit{and} \qquad \mathrm{E}[IF_2(\boldsymbol{R}_1, F)IF_2(\boldsymbol{R}_1, F)^\top] = \frac{a_1}{2a_2^2}\boldsymbol{I}_{3\times 3}$$

were $a_1 = 2\mathrm{E}(\sin^2 r_1)/3$.

Proof. From Proposition 3.1, $\widehat{\boldsymbol{S}}_n$ is an M-estimator with a normal asymptotic distribution. We use standard results for M-estimators to find the influence function of $\widehat{\boldsymbol{h}}_n$ and to show it has zero first moment and finite second moment. The notation and expansions in the proof of Proposition 3.1 are used here.

From Proposition 3.1 $\widehat{\boldsymbol{S}}_n$ is an M-estimator, therefore it has influence function

$$\mathrm{IF}_2(\boldsymbol{R}, F) = [\Psi_2'(\boldsymbol{h}_0)]^{-1} \left[\psi_2(\boldsymbol{S}^\top \boldsymbol{R}, \boldsymbol{h}_0)\right]$$

where $\psi_2(\boldsymbol{S}^\top \boldsymbol{R}, \boldsymbol{h}) = \partial f(\boldsymbol{h})/\partial \boldsymbol{h}$ for a function $f(\boldsymbol{h})$ defined as in (3.26) (with $(r, \boldsymbol{U})$ defining $\boldsymbol{S}^\top \boldsymbol{R}$ here replacing $(r_i, \boldsymbol{U}_i)$ there), $\Psi_2'(\boldsymbol{h}) = \partial \mathrm{E}[\psi_2(\boldsymbol{S}^\top \boldsymbol{R}_1, \boldsymbol{h})^\top]/\partial \boldsymbol{h}$ (upon taking $\boldsymbol{R}$ distributed as $\boldsymbol{R}_1$ in the expectation), and $\boldsymbol{h}_0 = \boldsymbol{0}_3$. Without loss of generality assume the central orientation $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$.

It follows from arguments based on (3.27) that

$$\psi_2(\boldsymbol{R}, \boldsymbol{h}) = \frac{\partial}{\partial \boldsymbol{h}} f(\boldsymbol{h}) = 2\sin(r)\boldsymbol{U} + \boldsymbol{C}\boldsymbol{h} + \tilde{\boldsymbol{R}}(\boldsymbol{h}) \tag{3.30}$$

for $\boldsymbol{C} = [1 - \cos(r)]\boldsymbol{U}\boldsymbol{U}^\top - [1 + \cos(r)]\boldsymbol{I}_{3\times 3}$ so that $\psi_2(\boldsymbol{R}, \boldsymbol{h}_0) = 2\sin(r)\boldsymbol{U}$ because $\tilde{\boldsymbol{R}}(\boldsymbol{h}) = \boldsymbol{0}_3$ when $\boldsymbol{h}_0 = \boldsymbol{0}_3$. Because $\psi_2(\boldsymbol{R}, \boldsymbol{h})$ has a series power expansion in $\boldsymbol{h}$ (with $\partial \tilde{\boldsymbol{R}}(\boldsymbol{h})/\partial \boldsymbol{h} = \boldsymbol{0}_{3\times 3}$ at $\boldsymbol{h}_0 = \boldsymbol{0}_3$), the Dominated Convergence Theorem may be used to justify the interchange of expectation and differentiation may in determining $\Psi_2'(\boldsymbol{h}_0)$ based on a random rotation $\boldsymbol{R}_1$ (defined by a random angle r_1 and independent uniformly distributed random axis $\boldsymbol{U}_1 \in \mathbb{R}^3$, $\|\boldsymbol{U}_1\| = 1$) used in (3.30) as

$$\begin{aligned}
\Psi_2'(\boldsymbol{h}_0) &= \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{h}} \psi_2(\boldsymbol{R}_1, \boldsymbol{h})^\top \Big|_{\boldsymbol{h}=\boldsymbol{h}_0}\right] \\
&= \mathrm{E}\left\{\frac{\partial}{\partial \boldsymbol{h}} \left[2\sin(r_1)\boldsymbol{U}_1^\top + \boldsymbol{C}_1 \boldsymbol{h}^\top + \tilde{\boldsymbol{R}}_1^\top(\boldsymbol{h})\right]\Big|_{\boldsymbol{h}=\boldsymbol{h}_0}\right\} \\
&= \mathrm{E}\left(\boldsymbol{C}_1\right) \equiv \mathrm{E}[1 - \cos(r_1)]\boldsymbol{U}_1\boldsymbol{U}_1^\top - [1 + \cos(r_1)]\boldsymbol{I}_{3\times 3} \\
&= \frac{2}{3}\{1 + 2\mathrm{E}[\cos(r_1)]\}\boldsymbol{I}_{3\times 3},
\end{aligned}$$

recalling $\mathrm{E}(\boldsymbol{U}_1\boldsymbol{U}_1^\top) = \boldsymbol{I}_{3\times 3}/3$. Hence, we have

$$\mathrm{IF}_2(\boldsymbol{R}, F) = [\Psi_2'(\boldsymbol{h}_0)]^{-1} [\psi_2(\boldsymbol{R}, \boldsymbol{h}_0)] = \frac{3\sin(r)}{1 + 2\mathrm{E}[\cos(r_1)]}\boldsymbol{U} = \frac{1}{a_2}\sin(r)\boldsymbol{U}.$$

where $a_2 = \mathrm{E}(1 + 2\cos r_1)/3$. As r_1 is independent of $\boldsymbol{U}_1$ with $\mathrm{E}[\sin(r_1)] = 0$, it follows that

$$\mathrm{E}[\mathrm{IF}_2(\boldsymbol{R}_1, F)] = \frac{1}{a_2}\mathrm{E}[\sin(r_1)]\mathrm{E}(\boldsymbol{U}_1) = \boldsymbol{0}_3.$$

and

$$\mathrm{E}[\mathrm{IF}_2(\boldsymbol{R}_1, F)\mathrm{IF}_2(\boldsymbol{R}_1, F)^\top] = \frac{1}{a_2^2}\mathrm{E}[\sin^2(r_1)\boldsymbol{U}_1\boldsymbol{U}_1^\top] = \frac{a_1}{2a_2^2}\boldsymbol{I}_{3\times 3}$$

where $a_1 = 2\mathrm{E}[\sin^2(r_1)]/3$. □

Proof of Theorem 3.2.

From Propositions 3.1 and 3.4, $\widehat{\boldsymbol{S}}_n = \boldsymbol{S}\exp\{\boldsymbol{\Phi}(\widehat{\boldsymbol{h}}_n)\}$ is an M-estimator with continuous and bounded influence function, that can be expressed in terms of $\boldsymbol{S}^\top\boldsymbol{R} = \exp\{\boldsymbol{\Phi}(\boldsymbol{h})\}$, $\boldsymbol{h} \in \mathbb{R}^3$ across rotations $\boldsymbol{R}$. Therefore, by Theorem 5.7 of Shao (2003), $\widehat{\boldsymbol{S}}_n$ is ϱ_∞-Hadamard differentiable (see page 347). Furthermore, the influence function of $\widehat{\boldsymbol{S}}_n$ has zero first and finite second moment by Proposition 3.4. Hence, by Theorem 5.20 (page 383) of Shao (2003), as $n \to \infty$

$$\sup_{x\in\mathbb{R}}\left|P\left(2na_2^2\|\widehat{\boldsymbol{h}}_n\|^2/a_1 \le x\right) - P_*\left(2n\hat{a}_{2_n}^2\|\widehat{\boldsymbol{h}}_n^*\|^2/\hat{a}_{1_n} \le x\right)\right| \longrightarrow 0,$$

w.p.1 where P_* denotes bootstrap probability conditional on the data. □

Proof of Theorem 3.3.

From Prentice (1984),

$$n\boldsymbol{m}^\top\widehat{\boldsymbol{F}}_n^-\boldsymbol{m} \xrightarrow{d} \chi_3^2$$

holds so it is enough to show that (w.p.1) $n\widehat{\boldsymbol{m}}_n^\top\widehat{\boldsymbol{F}}_n^{*-}\widehat{\boldsymbol{m}}_n \xrightarrow{d} \chi_3^2$. By definition, $\widehat{\boldsymbol{m}}_n$ is the eigenvector associated with the largest eigenvalue of $\widehat{\boldsymbol{V}}_n = \mathrm{E}_*(\boldsymbol{x}_i^*\boldsymbol{x}_i^{*\top})$ where $\mathrm{E}_*(\cdot)$ is the bootstrap expectation conditional on the data. Similarly, $\widehat{\boldsymbol{m}}_n^*$ and $\widehat{\boldsymbol{F}}_n^*$ are functions of the spectral decomposition of $\widehat{\boldsymbol{V}}_n^*$, an estimate of $\widehat{\boldsymbol{V}}_n$ based on the bootstrap sample satisfying $\mathrm{E}_*(\widehat{\boldsymbol{F}}_n^*) = \widehat{\boldsymbol{F}}_n$. Therefore, by the SLLN and 3.2 of Prentice (1984)

$$\sqrt{n}\widehat{\boldsymbol{F}}_n^{*-1/2}(\widehat{\boldsymbol{m}}_n^* - \widehat{\boldsymbol{m}}_n) \xrightarrow{d} N_3(\boldsymbol{0}_3, \boldsymbol{I}_{3\times 3})$$

w.p.1, where $\widehat{\boldsymbol{F}}_n^{*-1/2}$ is the Cholesky decomposition of $\widehat{\boldsymbol{F}}_n^{*-}$. Because $\widehat{\boldsymbol{F}}_n^{*-}\widehat{\boldsymbol{m}}_n^* = \boldsymbol{0}_3$ by construction, the continuous mapping theorem then gives, as $n \to \infty$

$$n\widehat{\boldsymbol{m}}_n^\top\widehat{\boldsymbol{F}}_n^{*-}\widehat{\boldsymbol{m}}_n \xrightarrow{d} \chi_3^2$$

w.p.1 as and the result follows.

□

CHAPTER 4. THE EXTRINSIC MEDIAN FOR THE ROTATION GROUP

A paper in preperation

Bryan Stanfill, Ulrike Genschel, Heike Hofmann

Abstract

In this manuscript we study the properties of the extrinsic median in $SO(3)$. The median is shown to be consistent, asymptotically normal and SB-robust with respect to Fisher information for the Cayley and matrix Fisher distributions. We also show that the extrinsic mean is not SB-robust for the same distributions, but it is more efficient. Similar results have appeared for data on the circle and sphere, but they are extended here to the Stiefel manifold for the first time. In the presence of contamination we show that confidence regions for the central orientation based on the median achieve a smaller size and closer to nominal coverage rates compared to those based on the extrinsic mean. Finally we demonstrate how these results can be applied in practice with a material sciences data example.

*K*eywords: Directional symmetry, Stiefel Manifold, SB-robustness, Influence Functions, Pivotal bootstrap

4.1 Introduction

Consider the location model for the rotation group $SO(3)$, the collection of all 3×3 orthogonal matrices with determinant one, defined by

$$\mathbf{R}_i = \boldsymbol{S}\boldsymbol{E}_i, \quad i = 1, \ldots, n, \tag{4.1}$$

where $\boldsymbol{S} \in SO(3)$ is the the central orientation, and $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n \in SO(3)$ denote i.i.d. directionally symmetric random perturbations of $\boldsymbol{S}$. Model (4.1) is the $SO(3)$ analog of a location model for scalar data $Y_i = \mu + e_i$, where $\mu \in \mathbb{R}$ denotes a mean and $e_i \in \mathbb{R}$ denotes an additive error symmetrically distributed around zero.

Though $SO(3)$ is a bounded parameter space, extreme observations occur in practice and adversely affect mean estimators of the location parameter $\boldsymbol{S}$. For a random sample of rotations from the Model (4.1), Fletcher et al. (2009) and Stanfill et al. (2013) each introduced a median-type estimator for $\boldsymbol{S}$ as alternative robust estimators. In both works the small sample behavior of the respective medians was evaluated through a simulation study but no theoretical results were provided.

In this manuscript we develop the theoretical properties of the median estimator proposed by Stanfill et al. (2013). We show that the median is strongly consistent for the central orientation $\boldsymbol{S}$, asymptotically normal and SB-robust for the Cayley and matrix Fisher distributions on $SO(3)$. In addition, normal theory and pivotal bootstrap confidence regions for the central orientation based on the median are proposed. We compare the finite sample behavior confidence regions based on the median to the existing methods based on the extrinsic mean. With respect to both the regions' size and empirical coverage rate, confidence regions based on the median are preferred to those based on the mean for contaminated samples.

The remainder of this manuscript is organized as follows. Section 4.2 includes a review of robust methods for directional data, which informs the approach taken to $SO(3)$ data analysis. In Section 4.3 we give a brief description of the $SO(3)$ parameter space and median estimators. Section 4.4 includes large sample results for the median. Consistency and asymptotic normality are established first and provide the theory necessary to form robust confidence regions for $\boldsymbol{S}$. Next the efficiency and sensitivity of the extrinsic median is investigated relative to the extrinsic mean. A simulation study is detailed in Section 4.5, which demonstrates the finite sample behavior of the confidence regions based on the extrinsic median for contaminated data. Finally, a data example is described in Section 4.6. Proofs and an extended look at the simulation results are included in the Appendix.

4.2 Literature Review

Consider estimating the location parameter based on a random sample on the unit circle $r_1, \ldots, r_n$, $r_i \in [-\pi, \pi)$ for $i = 1, \ldots, n$, following a symmetric and unimodal distribution about the central direction μ with concentration parameter κ. A common estimator for μ is the circular mean defined as $\hat{\mu}_n = \tan^{-1}[\sum_i (\sin r_i) / \sum_i (\cos r_i)]$. The traditionally robust alternative to the circular mean is the circular median, which is the angle $\tilde{\mu}_n$ that minimizes $\sum_i [\pi - |\pi - |r_i - \theta||]$ over $\theta \in [-\pi, \pi)$ (Mardia and Jupp, 2000). Wehrly and Shine (1981) compared the robustness properties of the circular mean and median based on their influence functions, which were stated without proof as $\mathrm{IF}(r, \hat{\mu}_n) = \sin[(r - \phi)/\kappa]$ and $\mathrm{IF}(r, \tilde{\mu}_n) = 0.5\mathrm{sign}(r - \phi)/[f(\phi) - f(\phi + \pi)]$, respectively, where $f(\cdot)$ is the distributional model assumed for the data, $\mathrm{sign}(x)$ is 1 if $x > 0$, 0 if $x = 0$ and -1 otherwise. Otieno (2002) later proved the influence functions reported by Wehrly and Shine (1981) are correct. Wehrly and Shine (1981) states that since the influence function of the mean is bounded it is sufficiently robust and the mean is the maximum likelihood estimator for several common distributions, therefore the mean is "hard to beat" as a location estimator for circular data.

Extending the notion of robust location estimation to the d-dimensional unit sphere, consider the random sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ with $\boldsymbol{x}_i \in \mathbb{R}^d$ and $\|\boldsymbol{x}_i\| = 1$ for $i = 1, \ldots, n$ symmetrically distributed about the location parameter $\boldsymbol{\mu} \in \mathbb{R}^d$ with $\|\boldsymbol{\mu}\| = 1$ and concentration parameter κ. With respect to estimator sensitivity, Ko and Guttorp (1988) argue that because the location parameter space is bounded, the influence functions reported by Wehrly and Shine (1981) should be standardized by some measure of dispersion when considering the robustness the estimator. The measure of dispersion with which to standardize the influence function is debated still (Laha and Mahesh, 2011) but common choices are the limiting variance of the estimator (Ko and Guttorp, 1988), the Fisher information matrix (Hampel et al., 2011) or the Kullback-Leibler (KL) discrepancy (He and Simpson, 1992).

An estimator is considered robust, or standardized bias (SB) robust, if the supremum of the standardized influence function over all parameter values and data points is finite. He and Simpson (1992) showed the class of estimator based on results in Lenth (1981), including the L_1

estimator on the sphere to be the vector $\tilde{\boldsymbol{\mu}}_n$ that minimizes $\sum_i(\|\boldsymbol{x}_i - \boldsymbol{m}\|)$ over unit vectors in $\mathbb{R}^d$, is SB-robust with respect to the KL discrepancy for the von Mises family of distributions. Ko and Chang (1993) propose SB-robust M-estimators on the sphere based on the results of Hampel et al. (2011).

Another important measure of estimator robustness is efficiency. The efficiency of the spatial median, or mediancentre (Gower, 1974), and the normalized spatial median were considered by Brown (1983) and Ducharme and Milasevic (1987), respectively. The spatial median is defined as the vector $\tilde{\boldsymbol{m}}_n$ that minimizes $\sum_i(\|\boldsymbol{x}_i - \boldsymbol{m}\|)$ over d-dimensional vectors $\boldsymbol{m}$ of any length. Though Brown (1983) does not treat directional data directly, they report the asymptotic relative efficiency (ARE) of the spatial median relative to the sample mean $\bar{\boldsymbol{x}}_n = \sum_i \boldsymbol{x}_i/n$ for dimensions two through seven assuming the vectors are scaled and therefore behave like directional data. Focused solely on directional data, Ducharme and Milasevic (1987) proposed the normalized spatial median $\tilde{\boldsymbol{\mu}}_n = \tilde{\boldsymbol{m}}_n/\|\tilde{\boldsymbol{m}}_n\|$ and derived its limiting distribution, proposed confidence cones for $\boldsymbol{\mu}$ and computed the ARE of $\tilde{\boldsymbol{\mu}}_n$ relative to the maximum likelihood estimator $\hat{\boldsymbol{\mu}}_n = \bar{\boldsymbol{x}}_n/\|\bar{\boldsymbol{x}}_n\|$ for data following the von Mises-Fisher distribution. Because the von Mises-Fisher distribution for unit vectors in $d+1$-dimensions converges to the local normal distribution in d-dimensions as $\kappa \to \infty$, the ARE results of Ducharme and Milasevic (1987) match those of Brown (1983) for large κ.

With respect to both sensitivity and efficiency, Chan and He (1993) conclude the normalized spatial median is preferred to the L_1 estimator on the sphere as well as the spherical median, where the spherical median is the minimizer of $\sum_i \cos^{-1}(\boldsymbol{x}_i^\top \boldsymbol{m})$ over unit vector in $\mathbb{R}^d$ (Fisher, 1985). For more details on robustness of estimators in directional statistics see the review paper by He (1992) or Chapter 12.4 of Mardia and Jupp (2000).

4.3 Rotation Data Preliminaries

Three-dimensional rotation data consist of observations belonging to the group $SO(3)$, where every element $\boldsymbol{R} \in SO(3)$ is an orthogonal 3×3 matrix with determinant one. Let $\mathfrak{so}(3)$ represent the space of all 3×3 skew-symmetric matrices $\mathfrak{so}(3) = \{\boldsymbol{X} \in \mathbb{R}^{3\times 3} : \boldsymbol{X}^\top = -\boldsymbol{X}\}$. Then each rotation in $\boldsymbol{R} \in SO(3)$ is associated with a skew-symmetric matrix $\boldsymbol{\Phi}(\boldsymbol{W}) \in \mathfrak{so}(3)$,

defined as

$$\boldsymbol{\Phi}(\boldsymbol{W}) = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix} \tag{4.2}$$

for $\boldsymbol{W} = (w_1, w_2, w_3)^\top \in \mathbb{R}^3$. Through the exponential operator, we map $\boldsymbol{\Phi}(\boldsymbol{W})$ to a rotation matrix as

$$\exp[\boldsymbol{\Phi}(\boldsymbol{W})] = \sum_{k=0}^{\infty} \frac{1}{k!} [\boldsymbol{\Phi}(\boldsymbol{W})]^k. \tag{4.3}$$

Using properties of skew-symmetric matrices, (4.3) can be simplified to

$$\exp[\boldsymbol{\Phi}(\boldsymbol{W})] = \cos(r)\boldsymbol{I}_{3\times 3} + \sin(r)\boldsymbol{\Phi}(\boldsymbol{U}) + (1 - \cos r)\boldsymbol{U}\boldsymbol{U}^\top$$

where $r = \|\boldsymbol{W}\|$ and $\boldsymbol{U} = \boldsymbol{W}/\|\boldsymbol{W}\|$.

Therefore, each rotation matrix $\boldsymbol{R}$ is associated with an angle-axis pair $(r, \boldsymbol{U})$, where $r \in (-\pi, \pi]$ and $\boldsymbol{U} \in \mathbb{R}^3$, $\|\boldsymbol{U}\| = 1$, through

$$\boldsymbol{R} = \boldsymbol{R}(r, \boldsymbol{U}) = \exp[\boldsymbol{\Phi}(r\boldsymbol{U})] \in SO(3). \tag{4.4}$$

The rotation $\boldsymbol{R}$ can be interpreted as a rotation of the coordinate axes $\boldsymbol{I}_{3\times 3}$ about the axis $\boldsymbol{U} \in \mathbb{R}^3$ by the angle r. In the materials science literature, $\boldsymbol{U}$ and r are commonly referred to as the misorientation axis and misorientation angle of $\boldsymbol{R}$ with respect to $\boldsymbol{I}_{3\times 3}$ respectively; see Randle (2003).

For a sample of rotations $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$, an M-estimator for the central orientation $\boldsymbol{S}$ is defined as the rotation matrix that minimize the loss function $\rho(\boldsymbol{R}_i, \boldsymbol{S})$

$$\widehat{\boldsymbol{S}} = \underset{\boldsymbol{S}\in SO(3)}{\arg\min} \sum_{i=1}^{n} \rho(\boldsymbol{R}_i, \boldsymbol{S}).$$

The choice of loss function depends on distribution of the data and the goal of the researcher; see Stanfill et al. (2013) for a detailed discussion. A common choice of loss function is the p-th order Euclidean distance defined for rotations $\boldsymbol{R}_1$ and $\boldsymbol{R}_2 \in SO(3)$ as

$$d_E^p(\boldsymbol{R}_1, \boldsymbol{R}_2) = \|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_F = \left[6 - 2\mathbf{tr}\left(\boldsymbol{R}_1^\top \boldsymbol{R}_2\right)\right]^{p/2} \tag{4.5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and $\mathbf{tr}(\cdot)$ denotes the matrix trace. The Euclidean distance corresponds to the length of the shortest path in $\mathbb{R}^{3\times 3}$ that connects $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$.

Since the Euclidean distance is defined in $\mathbb{R}^{3\times 3}$ rather than $SO(3)$, the Euclidean distance is an *extrinsic* distance measure and estimators based on the Euclidean distance are referred to as extrinsic estimators. Even though extrinsic estimators are cited as being less natural for $SO(3)$ data analysis, they often have nice statistical properties (León et al., 2006; Bingham et al., 2009).

In this manuscript we will focus on the extrinsic estimators minimizing the first and second order Euclidean distances, defined next. The extrinsic median $\widetilde{\boldsymbol{S}}_n$ is defined as

$$\widetilde{\boldsymbol{S}}_n = \underset{\boldsymbol{S}\in SO(3)}{\arg\min} \sum_{i=1}^{n} d_E(\boldsymbol{R}_i, \boldsymbol{S}). \tag{4.6}$$

The extrinsic mean $\widehat{\boldsymbol{S}}_n$ is defined as

$$\widehat{\boldsymbol{S}}_n = \underset{\boldsymbol{S}\in SO(3)}{\arg\min} \sum_{i=1}^{n} d_E^2(\boldsymbol{R}_i, \boldsymbol{S}) = \underset{\boldsymbol{S}\in SO(3)}{\arg\max}\, \mathbf{tr}(\boldsymbol{S}^\top \overline{\boldsymbol{R}}) \tag{4.7}$$

where $\overline{\boldsymbol{R}} = \sum_{i=1}^{n} \boldsymbol{R}_i/n$. For a comparison of these two estimators in terms of the point estimation characteristics see Stanfill et al. (2013).

In the next section we state the limiting distribution of extrinsic median and propose methods to constructing confidence regions for the central orientation $\boldsymbol{S}$. For the remainder of this manuscript we will refer to the extrinsic median and extrinsic mean simply as the median and mean, respectively.

4.4 Large Sample Theory

In this section we study the large sample behavior of the median for i.i.d. samples of rotations from a directionally symmetric population. The median is shown to be consistent and asymptotically normal. The asymptotic variance is given in a closed form for the Cayley and matrix Fisher distributions. It is also demonstrated that the median is SB-robust for the Cayley and matrix Fisher distributions while the mean is not.

4.4.1 Consistency and Asymptotic Normality

Let $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ be a sample of i.i.d. random rotations from a rotationally symmetric distribution on $SO(3)$ with central orientation parameter $\boldsymbol{S} \in SO(3)$ and concentration $\kappa > 0$.

The following two propositions establish the consistency and asymptotic normality of the median estimator $\widetilde{\boldsymbol{S}}_n$ for the central orientation $\boldsymbol{S}$ as defined in (4.6). The proofs are left to the Appendix.

Proposition 4.1. *The median $\widetilde{\boldsymbol{S}}_n$ is a strongly consistent estimator of the true central orientation $\boldsymbol{S}$.*

Proposition 4.2. *Assume $\widetilde{\boldsymbol{S}}_n \neq \boldsymbol{R}_i$ for all i. Define $\widetilde{\boldsymbol{h}}_n \in \mathbb{R}^3$ such that $\exp\left[\boldsymbol{\Phi}\left(\widetilde{\boldsymbol{h}}_n\right)\right] = \boldsymbol{S}^\top \widetilde{\boldsymbol{S}}_n$ then*

$$\sqrt{n}\,\widetilde{\boldsymbol{h}}_n \xrightarrow{\mathcal{L}} N_3\left(\mathbf{0}_3, \frac{a_1}{2a_2^2}\boldsymbol{I}_{3\times 3}\right)$$

as $n \to \infty$ where

$$a_1 = \frac{1}{6}\mathrm{E}[1+\cos(r)] \quad \textit{and} \quad a_2 = \frac{1}{12}\mathrm{E}\left[\frac{1+3\cos(r)}{\sqrt{1-\cos(r)}}\right] \tag{4.8}$$

provided a_1 and a_2 are finite and non-zero.

Equivalent to Proposition 4.2

$$\frac{2na_2^2}{a_1}\left\|\widetilde{\boldsymbol{h}}_n\right\|^2 \xrightarrow{\mathcal{L}} \chi^2_3 \tag{4.9}$$

as $n \to \infty$. The explicit forms of the constants a_1 and a_2 in (4.8) are given for the Cayley and matrix Fisher distributions in Table 4.1. In practice a_1 and a_2 will not be known and will need to be estimated. Estimates of a_1 and a_2 based on the sample $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ are

$$\hat{a}_{1_n} = \frac{1}{12n}\sum_{i=1}^{n}\left[1+\mathbf{tr}\left(\widetilde{\boldsymbol{S}}_n^\top \boldsymbol{R}_i\right)\right] \quad \text{and} \quad \hat{a}_{2_n} = \frac{\sqrt{2}}{24n}\sum_{i=1}^{n}\frac{3\mathbf{tr}(\widetilde{\boldsymbol{S}}_n^\top \boldsymbol{R}_i)-1}{\sqrt{3-\mathbf{tr}(\widetilde{\boldsymbol{S}}_n^\top \boldsymbol{R}_i)}}. \tag{4.10}$$

By the strong law of large numbers, it follows that $\hat{a}_{1_n} \to a_1$ and $\hat{a}_{2_n} \to a_2$ both with probability one as $n \to \infty$.

4.4.2 Confidence Regions

Confidence regions for the central orientation $\boldsymbol{S}$ can be formed using the results of Proposition 4.2. These regions can be interpreted by rewriting the squared norm of the vector $\widetilde{\boldsymbol{h}}_n$ as the geodesic distance between the estimate of the central direction and the true central direction. That is, $\|\widetilde{\boldsymbol{h}}_n\|^2 = \|r_{\widetilde{\boldsymbol{h}}_n}\boldsymbol{U}_{\widetilde{\boldsymbol{h}}_n}\|^2 = |r_{\widetilde{\boldsymbol{h}}_n}|^2\|\boldsymbol{U}_{\widetilde{\boldsymbol{h}}_n}\|^2 = r^2_{\widetilde{\boldsymbol{h}}_n} = d_R(\boldsymbol{S}, \widetilde{\boldsymbol{S}}_n)^2$ where $d_R(\cdot,\cdot)$ denotes the

Table 4.1: The constants a_1 and a_2 as they appear in (4.8) for the Cayley and matrix Fisher distributions on $SO(3)$. The symbol $I_i(\cdot)$ is the modified Bessel function of the first kind with order i, $\Gamma(x) = (x-1)!$ the Gamma function and $F_D(x) = \exp(-x^2)\int_0^x \exp(t^2)dt$ is the Dawson function.

Distribution	a_1	a_2
Cayley	$\frac{2\kappa+1}{6(\kappa+2)}$	$\frac{\kappa\sqrt{2}\Gamma(\kappa+2)}{3\sqrt{\pi}\Gamma(\kappa+2.5)}$
matrix-Fisher	$\frac{I_1(2\kappa)}{12\kappa[I_0(2\kappa)-I_1(2\kappa)]}$	$\frac{e^{2\kappa}[6\sqrt{\kappa}-(3+8\kappa)F_D(2\sqrt{\kappa})]}{24\sqrt{2}\pi\kappa^{1.5}[I_0(2\kappa)-I_1(2\kappa)]}$

geodesic distance as defined in (2.4). Therefore a $100(1-\alpha)$ large sample confidence region (denoted by the superscript L) for the true central orientation $\boldsymbol{S}$ based on $\widetilde{\boldsymbol{S}}_n$ contains all rotations $\boldsymbol{R} \in SO(3)$ in the set

$$\mathcal{C}_\alpha^L = \left\{ \boldsymbol{R} \in SO(3) : \frac{2na_2^2}{a_1} d_R(\boldsymbol{R}, \widetilde{\boldsymbol{S}}_n)^2 < \chi^2_{3,1-\alpha} \right\}. \tag{4.11}$$

That is, all rotations within a fixed geodesic distance of the median $\widetilde{\boldsymbol{S}}_n$ are contained in the confidence region for $\boldsymbol{S}$.

As demonstrated in Chapter 3, bootstrap calibrated confidence regions tend to have closer to nominal finite sample coverage rates. Therefore, we propose a bootstrap method that can be used to calibrate the confidence region (4.11).

Theorem 4.3. *Assume $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ are a sample of i.i.d. observations from a directionally symmetric location model with central orientation $\boldsymbol{S}$ according to (4.1). Let $\widetilde{\boldsymbol{S}}_n$ denote the mean as defined in (4.6) and let $\widetilde{\boldsymbol{S}}_n^*$ denote its bootstrap version. Then, as $n \longrightarrow \infty$,*

$$\sup_{x\in\mathbb{R}} \left| P\left(2na_2^2\|\widetilde{\boldsymbol{h}}_n\|^2/a_1 \le x\right) - P_*\left(2na_2^2\|\widetilde{\boldsymbol{h}}_n^*\|^2/a_1 \le x\right) \right| \longrightarrow 0 \quad \textit{with probability 1,}$$

where P_* denotes bootstrap probability conditional on the data.

From the above bootstrap consistency result, a Monte Carlo bootstrap procedure can be given as follows for calibrating a confidence region for the central location parameter $\boldsymbol{S}$.

Monte Carlo Implementation:

1. For the given sample $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n \in SO(3)$, compute the median $\widetilde{\boldsymbol{S}}_n$.

2. Generate a bootstrap sample $\boldsymbol{R}_1^*, \ldots, \boldsymbol{R}_n^*$ by randomly sampling n rotation matrices with replacement from $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$.

3. Compute the median $\widehat{\boldsymbol{S}}_n^*$ of the bootstrap data set and form the test quantity $2n(\hat{a}_{2_n}^*)^2\|\widetilde{\boldsymbol{h}}_n\|^2/\hat{a}_{1_n}^*$ where $\hat{a}_{1_n}^*$ and $\hat{a}_{2_n}^*$ are computed from the bootstrap sample by replacing $\boldsymbol{R}_i^*$ and $\widetilde{\boldsymbol{S}}_n^*$ for $\boldsymbol{R}_i$ and $\widehat{\boldsymbol{S}}_n$, respectively.

4. Repeat steps 2 and 3, $j = 1, \ldots, m$ times to obtain m values of the quantity $2n(\hat{a}_{2_n}^*)^2\|\widetilde{\boldsymbol{h}}_n\|^2/\hat{a}_{1_n}^*$.

5. Define $\widehat{q}_{1-\alpha}$ to be the $100(1-\alpha)\%$ sample percentile of the bootstrap realizations of $2n(\hat{a}_{2_n}^*)^2\|\widetilde{\boldsymbol{h}}_n\|^2/\hat{a}_{1_n}^*$.

6. An approximate $100(1-\alpha)\%$ bootstrap confidence region (denoted B) for the true central orientation $\boldsymbol{S}$ based on the direct approach contains all rotations $\boldsymbol{S} \in SO(3)$ in the set

$$\mathcal{C}_\alpha^B = \left\{ \boldsymbol{R} \in SO(3) : \frac{2n\hat{a}_{2_n}^2}{\hat{a}_{1_n}} d_R(\boldsymbol{R}, \widetilde{\boldsymbol{S}}_n)^2 < \hat{q}_{1-\alpha} \right\}. \tag{4.12}$$

4.4.3 Efficiency and Sensitivity

Next we evaluate the large sample behavior of the median relative to the mean. First we consider the median's efficiency. The *asymptotic relative efficiency* (ARE) of the median relative to the mean is given by the ratio of their asymptotic variances, i.e. $\text{ARE}(\widetilde{\boldsymbol{S}}_n, \widehat{\boldsymbol{S}}_n) = \text{Var}(\widehat{\boldsymbol{S}}_n)/\text{Var}(\widetilde{\boldsymbol{S}}_n)$. An ARE less than one indicates that the mean is more efficient than the median, and therefore a larger sample size is required for the median to achieve the mean's level of uncertainty. For the form of the means asymptotic variance see Chapter 3.

In Figure 4.1 the ARE of the median relative to the mean is illustrated for a range of concentrations κ. For both the Cayley and matrix Fisher distributions, the ARE converges to $8/(3\pi) \approx 0.85$ as $\kappa \to \infty$. An ARE of $8/(3\pi)$ implies that for the same sample size, the median has an $100(\sqrt{3\pi/8} - 1)\% \approx 8.5\%$ larger standard error than the mean as $\kappa \to \infty$. Brown (1983) demonstrated the ARE of the spatial median in $\mathbb{R}^3$ relative to the MLE is $8/(3\pi)$ for

the isotropic normal distribution. This matches our result since the Cayley and matrix Fisher distributions both converge to a local normal distribution as $\kappa \to \infty$.

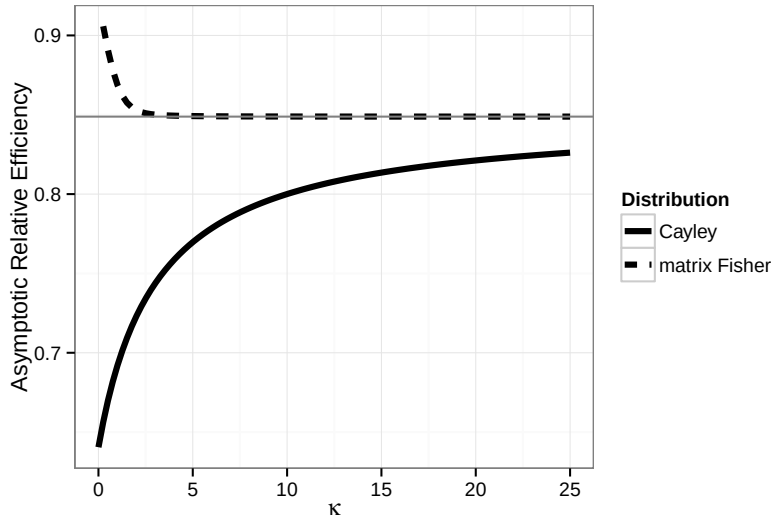


Figure 4.1: Asymptotic relative efficiency as a function of concentration κ for the Cayley and matrix Fisher distributions. In both distributions as κ goes to infinity, the ARE converges to $8/(3\pi)$.

Though the median is less efficient than the mean, it is robust to outliers while the mean is not. We next derive the median's *influence function* (IF), from which measures of robustness can be derived. Since $\widetilde{\boldsymbol{S}}_n$ is an M-estimator, an expression for the influence function is given by Hampel et al. (2011), page 230, which is given in the next proposition. For a proof see the Appendix.

Proposition 4.4. *Assume $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ are a sample of i.i.d. observations from a UARS distribution F on $SO(3)$ with central orientation $\boldsymbol{S} \in SO(3)$ and concentration $\kappa > 0$. Let $\widetilde{\boldsymbol{S}}_n$ be the median as defined in (4.6) and assume $\boldsymbol{R}_i \neq \widetilde{\boldsymbol{S}}_n$ for all i. The influence function of $\widetilde{\boldsymbol{S}}_n$ is given by*

$$IF\left(\boldsymbol{R}; \widetilde{\boldsymbol{S}}_n, F\right) = \frac{\sin(r)}{2a_2\sqrt{1-\cos(r)}}\boldsymbol{U}$$

where $a_2 = \mathrm{E}\left\{[1+3\cos(r)]/[12\sqrt{1-\cos(r)}]\right\}$ *and* $\exp\left[\boldsymbol{\Phi}\left(r\boldsymbol{U}\right)\right] = \boldsymbol{R}$.

The maximum bias that can be introduced by a small change in the sample, such as an outlier, is called the gross error sensitivity. Since the parameter space of interest $SO(3)$ is

bounded, Ko and Guttorp (1988) and others have argued in favor of replacing the notion of gross error sensitivity with *standardized gross error sensitivity* (SGES) when considering the robustness of an estimator in a bounded parameter space. The SGES of the estimator T is defined as

$$\gamma^*(T,F,V)^2 = \sup_{\boldsymbol{R}\in SO(3)} \left[\mathrm{IF}\left(\boldsymbol{R};T,F\right)^\top V\left(F\right)^{-1}\mathrm{IF}\left(\boldsymbol{R};T,F\right)\right]$$

where $IF(\boldsymbol{R};T,F)$ is the influence function of the estimator T for data from the distribution F and $V\left(F\right)$ is a measure of spread of the distribution F satisfying $V(F)>0$ (Ko and Guttorp, 1988; Hampel et al., 2011). The choice of spread $V(F)$ has been shown to be important (Laha and Mahesh, 2011). Common choices are the Fisher information matrix (Hampel et al., 2011), Kullback-Leibler discrepancy (He and Simpson, 1992) or the estimator's asymptotic variance (Ko and Guttorp, 1988). An estimator with a finite SGES for all permissible F is said to be *SB-robust* with respect to the measure of spread $V(F)$. For the remainder of this manuscript we will consider SGES with respect to the Fisher information matrix, denoted $\mathcal{I}$. In the next two propositions we demonstrate that for the Cayley and matrix Fisher distributions the median is SB-robust with respect to the Fisher information matrix while the mean is not. First we report the Fisher information matrices for the distributions of interest.

For the Cayley distribution the Fisher Information matrix is given by Equation (13) with $p=3$ of León et al. (2006),

$$\mathcal{I}_C\left(\boldsymbol{S}\right) = \frac{\kappa^2}{2\kappa-1}\boldsymbol{I}_{3\times3}. \tag{4.13}$$

Because the mean is the maximum likelihood estimator for the matrix Fisher distribution then the asymptotic variance of the mean from Chapter 3 is the Fisher information matrix. Using the quantities in Table 3.1,

$$\mathcal{I}_F\left(\boldsymbol{S}\right) = \frac{2}{3}\left[\frac{(\kappa+1)I_1(2\kappa)-\kappa I_0(2\kappa)}{I_0(2\kappa)-I_1(2\kappa)}\right]\boldsymbol{I}_{3\times3}. \tag{4.14}$$

Proposition 4.5. *The median is SB-robust for the Cayley and matrix Fisher distributions with respect to the Fisher information.*

Proof. First we consider the Cayley distribution. From (4.13) the squared SGES for the median is

$$\gamma^*\left(\widetilde{\boldsymbol{S}}_n,F,\mathcal{I}\right)^2 = \frac{9\pi\Gamma(\kappa+2.5)^2}{4(2\kappa-1)\Gamma(\kappa+2)^2}.$$

It can be shown that $\gamma^*(\widetilde{\boldsymbol{S}}_n, F, \mathcal{I}) \to 3\sqrt{2\pi}/4 \approx 1.88$ as $\kappa \to \infty$. Since γ^* is only defined when $\mathcal{I}_C(\boldsymbol{S})^{-1} > 0$ we don't consider the diffuse limit, i.e. $\kappa \to 0$. It follows the median is SB-robust for the Cayley family of distributions.

From (4.14) the squared SGES of the median for the matrix Fisher family of distributions is given by

$$\gamma^*\left(\widetilde{\boldsymbol{S}}_n, F, \mathcal{I}\right)^2 = \frac{1}{3a_2^2}\left[\frac{(\kappa+1)I_1(2\kappa) - \kappa I_0(2\kappa)}{I_0(2\kappa) - I_1(2\kappa)}\right]$$

where d is given in Table 4.1. Considering both the diffuse and concentrated cases: $\gamma^*(\widetilde{\boldsymbol{S}}_n, F, \mathcal{I}) \to 15\pi\sqrt{3}/(16\sqrt{2}) \approx 3.61$ as $\kappa \to 0$ and $\gamma^*(\widetilde{\boldsymbol{S}}_n, F, \mathcal{I}) \to 3\sqrt{2\pi}/4$ as $\kappa \to \infty$. Therefore the median is SB-robust with respect to the matrix Fisher family of distributions. □

Proposition 4.6. *The mean is not SB-robust for the Cayley or matrix Fisher distributions with respect to the Fisher information.*

Proof. From Chapter 3, the influence function of the extrinsic mean is given by

$$\text{IF}\left(\boldsymbol{R}; \widehat{\boldsymbol{S}}_n, F\right) = \frac{3\sin(r)}{\text{E}[1 + 2\cos(r)]}\boldsymbol{U}$$

where

$$\text{E}[1 + 2\cos(r)] = \begin{cases} \frac{3\kappa}{\kappa+2} & \text{if } \boldsymbol{R} \sim \text{Cayley} \\ \frac{(\kappa+1)I_1(2\kappa) - \kappa I_0(2\kappa)}{\kappa[I_0(2\kappa) - I_1(2\kappa)]} & \text{if } \boldsymbol{R} \sim \text{matrix Fisher.} \end{cases}$$

First we consider the Cayley distribution: from (4.13) the squared SGES for the mean is

$$\gamma^*(\widehat{\boldsymbol{S}}_n, F, \mathcal{I})^2 = \frac{(\kappa+2)^2}{2\kappa - 1}.$$

Clearly $\gamma^*(\widehat{\boldsymbol{S}}_n, F, \mathcal{I}) \to \infty$ as $\kappa \to \infty$. Again we don't consider the diffuse case due to the requirement that $\mathcal{I}_C(\boldsymbol{S})^{-1} > 0$. It follows the mean is not SB-robust for the Cayley distribution.

From (4.14) the SGES of the mean for the matrix Fisher family of distributions is given by

$$\gamma^*(\widehat{\boldsymbol{S}}_n, F, \mathcal{I})^2 = \frac{6\kappa^2\left[I_0(2\kappa) - I_1(2\kappa)\right]}{(\kappa+1)I_1(2\kappa) - \kappa I_0(2\kappa)}.$$

Considering both the diffuse and concentrated cases: $\gamma^*(\widehat{\boldsymbol{S}}_n, F, \mathcal{I}) \to \sqrt{6}$ as $\kappa \to 0$ but $\gamma^*(\widetilde{\boldsymbol{S}}_n, F, \mathcal{I}) \to \infty$ as $\kappa \to \infty$. Therefore the mean is not SB-robust for the matrix Fisher distribution. □

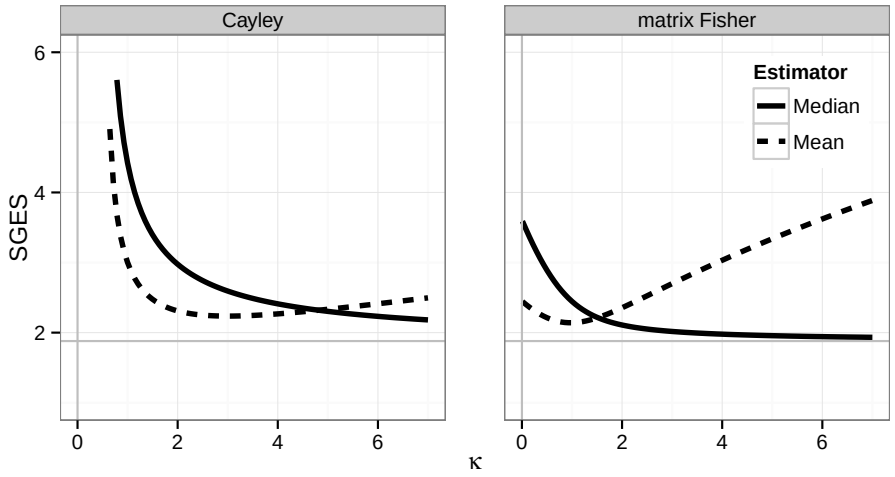


Figure 4.2: The mean and median SGES with respect to the Fisher information matrix as a function of κ for the Cayley and matrix Fisher distributions. A gray horizontal line is placed at $3\sqrt{2\pi}/4$ to indicate the limiting SGES for the median as $\kappa \to \infty$.

In a similar fashion it can be shown that the median is also SB-robust with respect to its asymptotic variance while the mean is not. The estimators' SGES for the Cayley and matrix Fisher distributions are plotted as a function of κ in Figure 4.2. Under both distributional assumptions, for small values of κ, the mean is less sensitive to outliers than the median relative to the variability of the data. As κ increases, however, the sensitivity of the mean relative to the data variability diverges to infinity while the median has a finite limit.

4.5 Simulation Study

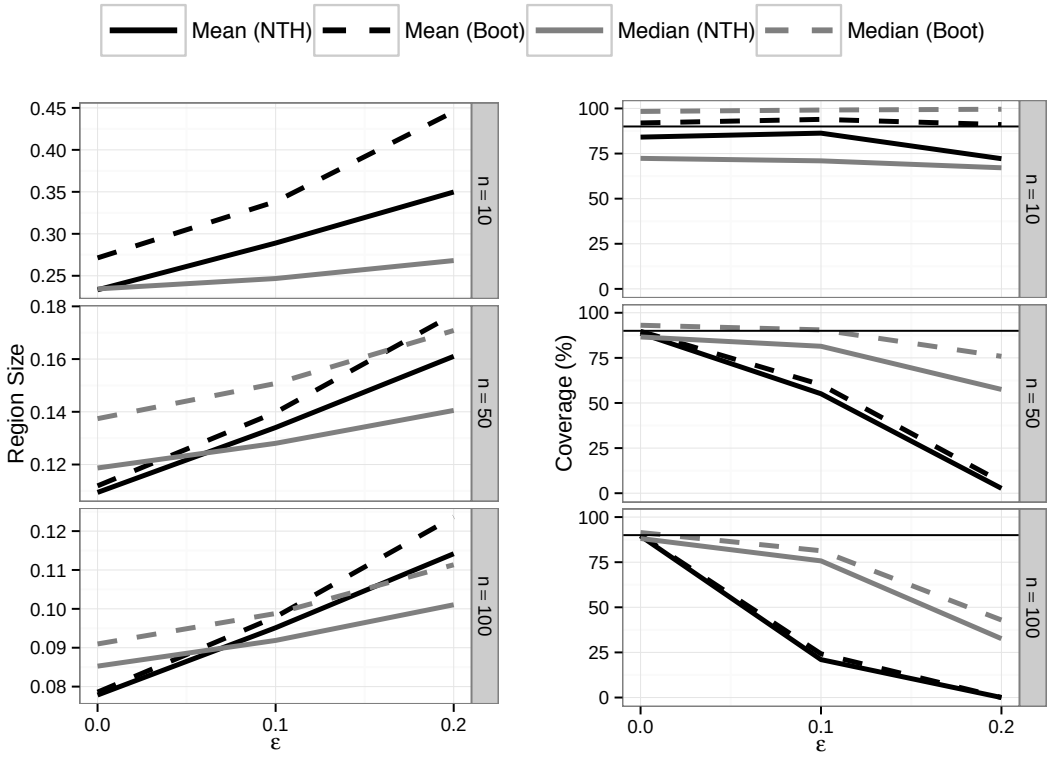
In this section we study the small sample behavior of the mean and median through simulation. In particular we evaluate how the confidence regions in Section 4.4.2 compare to those based on the mean in the presence of contamination. This confidence regions' size and coverage rate is reported to demonstrate the estimator bias and variance, respectively.

Ten-thousand samples of sizes $n = 10$, 50 and 100 were simulated from the contaminated distribution given by

$$F_\epsilon = (1-\epsilon)F(\boldsymbol{I}_{3\times 3}, \kappa) + \epsilon F(\boldsymbol{S}, \kappa)$$

where $\epsilon = 0.0$, 0.1 and 0.2, F is either the Cayley or matrix Fisher distribution, the concentra-

tion κ is set to 20, and the contamination central orientation parameter $\boldsymbol{S}$ is a rotation through $\pi/2$ radians. For all levels of contamination we consider the true central orientation to be $\boldsymbol{I}_{3\times 3}$.

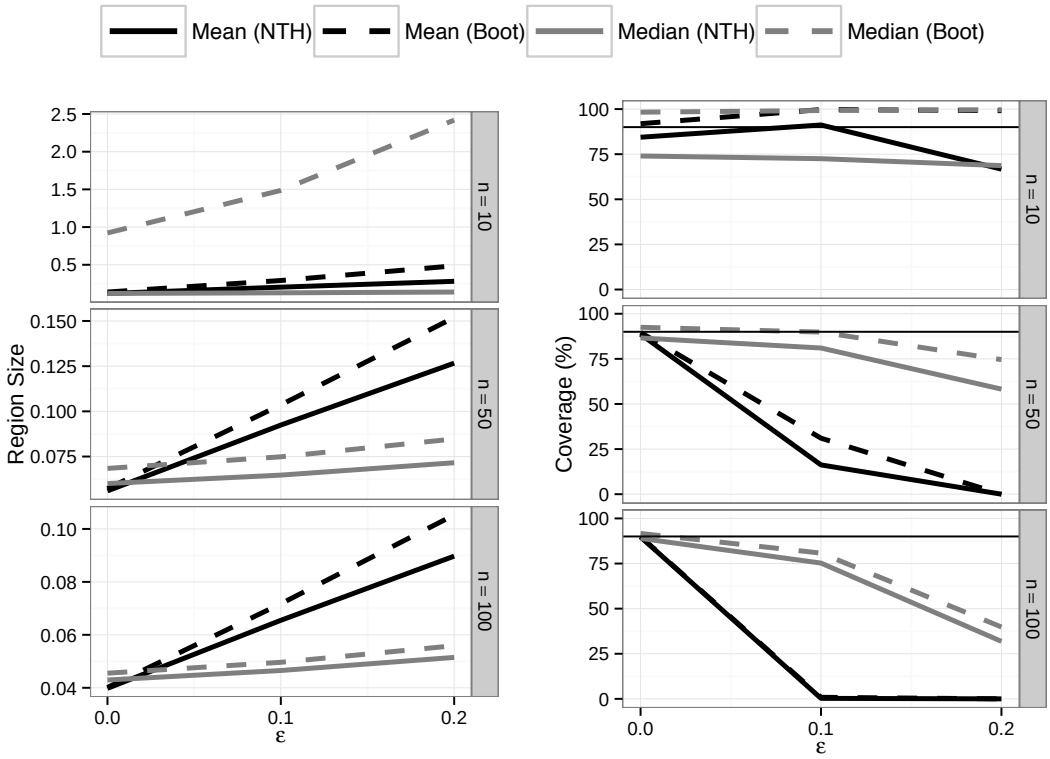


(a) Size (b) Coverage

Figure 4.4: Comparison of confidence region size (a) and coverage rate (b) for data from the Cayley distribution. The median bootstrap region sizes for $n = 10$ are 3.94, 6.29 and 9.87 for ϵ =0.0, 0.1 and 0.2, respectively.

The size of each confidence region $\mathcal{C}_\alpha$ is defined as the length of the longest chord in $SO(3)$ that connects the center of the region (located at the respective estimator) and a rotation $\boldsymbol{R} \in \mathcal{C}_\alpha$. This definition of region size can, in most cases, be interpreted as the region's radius. A region captures the true central orientation $\boldsymbol{S}$ if the radius of the region is greater than the distance between the estimator and the true central orientation $\boldsymbol{I}_{3\times 3}$. To illustrate, consider the normal theory region based on the projected median in (4.11). The squared radius of the region $\mathcal{C}_\alpha^L$ is given by

$$\mathrm{rad}^2(\mathcal{C}_\alpha^L) = \frac{\hat{a}_{1_n}\chi^2_{3,1-\alpha}}{2n\hat{a}^2_{2_n}}.$$



(a) Size (b) Coverage

Figure 4.6: Comparison of confidence region size (a) and coverage rate (b) for data from the Fisher distribution.

The region includes the true central orientation $\boldsymbol{S} \in \mathcal{C}_\alpha^L$ if

$$d_R(\widetilde{\boldsymbol{S}}_n, \boldsymbol{I}_{3\times 3}) < \text{rad}(\mathcal{C}_\alpha^L).$$

The regions' size as a function of contamination ϵ for the two estimators and two methods are given for the Cayley and matrix-Fisher distributions in Figures 4.4. The regions' coverage rate is illustrated similarly in Figure 4.6. For a tabular description of these results see the Appendix. Note that in Figure 4.3a, for $n = 10$ the median bootstrap regions results have been omitted because the values are too big to be compare to the other regions. According to Table 4.4 of the Appendix, for samples of size $n = 10$ the bootstrap regions based on the median have sizes 3.94, 6.29 and 9.87 for levels of contamination $\epsilon = 0.0$, 0.1 and 0.2, respectively. This illustrates two key ideas. First, the median bootstrap regions are unreliable for small sample sizes. Second, the interpretation of "radius" does not hold for regions of this size as the largest

geodesic distance possible in $SO(3)$ is of length π. These region sizes are correct, however, because these regions are for the vector $\widetilde{\boldsymbol{h}}_n$ while lies in an infinite space, only its projection into $SO(3)$ $\widetilde{\boldsymbol{S}}_n$ has a "radius" bounded by π.

In all scenarios the bootstrap regions are larger and have coverages closer to 100% than the normal theory regions based on the same estimator. As sample size increases the mean and median based regions become more disparate with the median based regions preferred in terms of both metrics. As a function of contamination ϵ, the regions based on the mean increase in size and decrease in coverage more quickly than those based on the median. This demonstrates that the mean is heavily biased by the contamination because the regions based on the mean are centered far from the true central orientation $\boldsymbol{I}_{3\times 3}$. This is especially true for the matrix Fisher distribution with $n = 100$ where the coverage rate for the mean region in 0 for 10% contamination while the median based regions remain close to 75%. The mean based regions are centered far away from the true central orientation and their size does not make up for the extreme bias.

4.6 Data Example

In material sciences rotation data are gathered via electron backscatter diffraction data (EBSD) to identify the cubic crystal orientation on the surface of metals. It is becoming increasingly common to use multiple scans of the same sample to determine the "true" crystal orientation at each measured location (Humbert et al., 1996; Cho et al., 2005; Bachmann et al., 2010). It has been shown that locations on the boundary between two grains can result in repeat scans that are highly variable due to proximity of different grains (Stanfill et al., 2013). In this section we illustrate that using the median to estimate between scan variability is more reliable then using the mean due to the robustness properties demonstrated in this manuscript.

We consider EBSD data obtained by scanning a fixed 12.5 μm $\times$ 10 μm nickel surface at individual locations spaced 0.2 μm apart. This scan was repeated up to 14 times for each location yielding a total of 3,449 observations (Bingham et al., 2009, 2010a). Figure 4.7 is a type of "grain-map" with each location shaded according to the estimated misorientation angle of the median of the repeated scans at each location. Sixteen locations believed to be on the

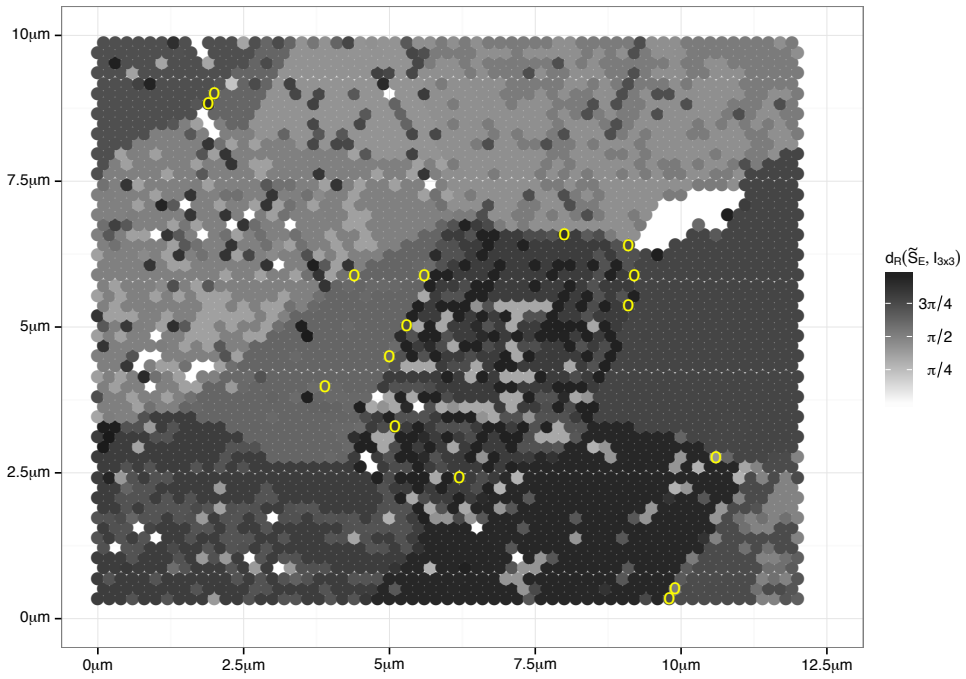


Figure 4.7: Grain map of the nickel data with locations of interest circled with yellow.

boundary of two grains are circled and the radius (in degrees) of the two 90% confidence regions centered at the mean and median are reported in Table 4.2.

The results in Table 4.2 are similar to those that were found in the simulation study. For samples that appear to come from a mixture of two distributions, normal theory regions based on the mean are uniformly larger than those based on the median. Further, bootstrap regions based on the mean are generally larger than the normal theory regions. Normal theory regions based on the median are the smallest in most cases. Figure 4.8 is an illustration of one of the locations, location 698, which was successfully scanned 13 times. Based on this figure and the raw data, there appears to be one large group consisting of eight scans and a smaller group consisting of five scans. Each axis of each scan at location 698 is projected onto a unit sphere along with the estimated mean and median with corresponding 90% bootstrap confidence regions. Normal theory regions are similar and are therefore omitted. The median estimate is positioned in the center of the large group and the corresponding confidence region is comparably small. The mean estimate is midway between the two groups of data and the

Table 4.2: Radius (in degrees) of 90% confidence regions centered at the respective estimators for locations with excessive deterioration or on grain boundaries.

	Large n Approx.		Bootstrap	
Location	Mean	Median	Mean	Median
50	3.700	0.402	4.955	22.120
111	3.527	0.294	4.257	15.338
698	12.298	0.352	20.432	7.543
758	8.039	0.298	12.246	0.352
901	10.397	0.682	10.891	0.869
1055	8.071	0.231	180.000	0.215
1291	8.192	0.268	12.183	0.248
1478	5.527	0.194	8.154	0.145
1661	5.033	0.231	7.415	0.415
1801	6.889	0.190	125.257	0.141
1959	6.622	0.233	9.719	0.207
1965	5.530	0.195	8.158	0.106
1983	6.898	0.194	180.000	0.118
2164	4.647	0.329	5.184	0.464
2219	6.617	0.491	7.606	100.921
2975	3.382	0.429	4.092	5.512
3036	3.411	0.254	4.745	0.233

confidence region is considerably larger. From this it is reasonable to assume the measure of uncertainty that results from the median is more reliable than that based on the mean.

Appendix

This section proves the propositions left unproven in the main manuscript as well as additional results from the simulation study.

Proofs

To prove $\widetilde{\boldsymbol{S}}_n$ is a consistent estimator for the central orientation $\boldsymbol{S}$, as stated in Proposition 4.1, we take an approach similar in spirit to the proof of Theorem 3.4 in Bhattacharya and Patrangenaru (2003), which established the extrinsic mean as a consistent estimator for $\boldsymbol{S}$. A sketch of the proof is as follows. The L_1 estimator over the space of all 3×3 matrices is a $\sqrt{n}$-consistent estimator for the central orientation parameter $\boldsymbol{S}$. Therefore, provided a continuous

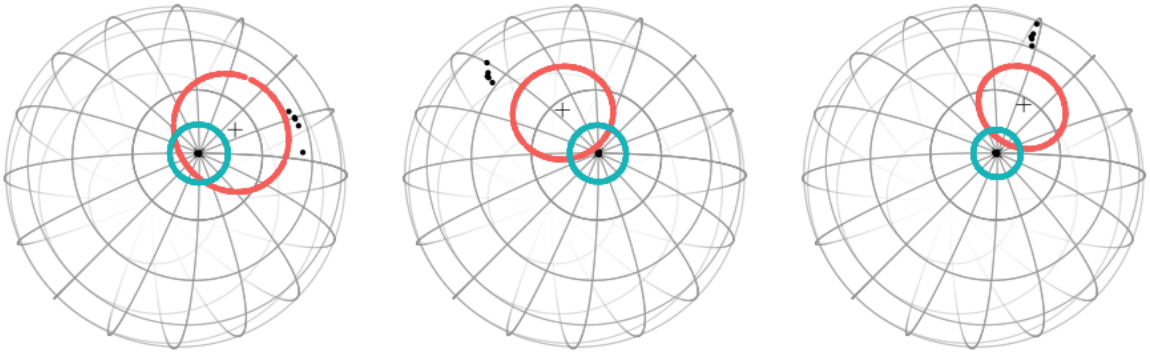


Figure 4.8: The x-, y- and z-axes of each scan at location 698 of the nickel dataset visualized with confidence regions based on the mean (red) and median (aqua).

function that maps all 3×3 matrices into $SO(3)$ exists (w.p. 1) then the continuous mapping theorem will give the L_1 estimator over for $SO(3)$ is $\sqrt{n}$-consistent for the central orientation.

Proof of Proposition 4.1.

Define the matrix $\boldsymbol{M}_n \in \mathbb{R}^{3\times3}$ such that $\boldsymbol{M}_n = \arg\min_{\boldsymbol{S}\in\mathcal{M}(3)} \sum_{i=1}^{n} \|\boldsymbol{R}_i - \boldsymbol{S}\|_F$. Let A_n denote the set on which $\boldsymbol{M}_n$ is focal, then by Theorem 3.2 of Bhattacharya and Patrangenaru (2003) $P(A_n) \to 0$ as $n \to \infty$. Further, the UARS distributions are directionally symmetric hence $\boldsymbol{M}_n$ is $\sqrt{n}$-consistent for $\mathrm{E}(\boldsymbol{R}_1) = b\boldsymbol{S}$ where $b = \frac{1}{3}[1 + 2\mathrm{E}(\cos r)]$ (see Brown 1983; Small 1990). Let $G : \mathbb{R}^{3\times3} \to SO(3)$ be the projection operator from the ambient space $\mathbb{R}^{3\times3}$ into $SO(3)$ as described in e.g. Moakher (2002) and note that G is a continuous function on A_n^c where $P(A_n^c) \to 1$ as $n \to \infty$. Therefore, by the continuous mapping theorem, $\widetilde{\boldsymbol{S}}_n = G(\boldsymbol{M}_n) \xrightarrow{p} G(b\boldsymbol{S}) = \boldsymbol{S}$ as $n \to \infty$, i.e. $\widetilde{\boldsymbol{S}}_n$ is $\sqrt{n}$-consistent for $\boldsymbol{S}$. □

To prove Proposition 4.2, we use the following steps.

1. Rewrite the loss function $\rho_1(\boldsymbol{h}, \boldsymbol{R}_i)$ as a linear function in $\boldsymbol{h}$.

2. Find the form of the derivative of the loss function with respect to $\boldsymbol{h}$, $\psi_1(\boldsymbol{R}_i, \boldsymbol{h}) = \partial\rho_1(\boldsymbol{R}_i, \boldsymbol{h})/\partial\boldsymbol{h}$.

3. Show the derivative function $\psi_1(\boldsymbol{R}_i, \boldsymbol{h})$ has a finite second moment for all $\boldsymbol{h}$.

4. Show the expectation of the derivative $\Psi_1(\boldsymbol{h}) = \mathrm{E}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h})]$ is zero if and only if it is evaluated at the true central orientation $\boldsymbol{h}_0$.

5. Finally show the derivative of the expectation $\Psi_1'(\boldsymbol{h}_0)$ exists and is non-zero $\|\Psi_1'(\boldsymbol{h}_0)\|_F \neq 0$.

Proof of Proposition 4.2.

Without loss of generality assume the central orientation is the identity $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$ with corresponding vector $\boldsymbol{h}_0 = \boldsymbol{0}_3$ satisfying $\exp[\boldsymbol{\Phi}(\boldsymbol{h}_0)] = \boldsymbol{S} = \boldsymbol{I}_{3\times 3}$. Define the L_1 estimator of the central orientation $\widetilde{\boldsymbol{S}}_n = \arg\min_{\boldsymbol{S}\in SO(3)} \sum_{i=1}^n \|\boldsymbol{R}_i - \boldsymbol{S}\|_F$ with the corresponding vector $\widetilde{\boldsymbol{h}}_n$ satisfying $\exp[\boldsymbol{\Phi}(\widetilde{\boldsymbol{h}}_n)] = \widetilde{\boldsymbol{S}}_n$.

Define the loss function $\rho_1(\boldsymbol{R}_i, \boldsymbol{h}) = \|\boldsymbol{R}_i - \exp[\Phi(\boldsymbol{h})]\|_F$, which can be written

$$\begin{aligned}
\rho_1(\boldsymbol{R}_i, \boldsymbol{h}) &= \|\boldsymbol{R}_i - \exp[\Phi(\boldsymbol{h})]\|_F \\
&= \left\{6 - 2\mathbf{tr}(\boldsymbol{R}_i^\top \exp[\Phi(\boldsymbol{h})])\right\}^{1/2} \\
&= \{6 - 2f_i(\boldsymbol{h})\}^{1/2}
\end{aligned}$$

where

$$\begin{aligned}
f_i(\boldsymbol{h}) &= \mathbf{tr}\left[\exp\left(\boldsymbol{\Phi}[\boldsymbol{h}]\right)^\top \boldsymbol{R}_i\right] \\
&= \mathbf{tr}\left[\exp\left(\boldsymbol{\Phi}[-\boldsymbol{h}]\right) \boldsymbol{R}_i\right] \\
&= 1 + 2\cos r_i + 2(\sin r_i)(\boldsymbol{h}^\top \boldsymbol{u}_i) \sum_{k=0}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^k}{(2k+1)!} \\
&\qquad + \left[(\boldsymbol{h}^\top \boldsymbol{h})(1+\cos r_i) - (1-\cos r_i)(\boldsymbol{h}^\top \boldsymbol{u}_i)^2\right] \sum_{k=1}^{\infty} \frac{(-1)^k (\boldsymbol{h}^\top \boldsymbol{h})^{k-1}}{(2k)!}
\end{aligned} \tag{4.15}$$

for all $\boldsymbol{h}$ and is continuously differentiable in $\boldsymbol{h}$.

The derivative of the loss function with respect to the vector $\boldsymbol{h}$, $\psi_1(\boldsymbol{R}_i, \boldsymbol{h}) = \partial\rho_1(\boldsymbol{R}_i, \boldsymbol{h})/\partial\boldsymbol{h}$

(Step 2):

$$\begin{aligned}\frac{\partial}{\partial h_j}\rho_1(\boldsymbol{R}_i,\boldsymbol{h}) &= \frac{\partial}{\partial h_j}\left\{6-2f_i(\boldsymbol{h})\right\}^{1/2}\\ &= \frac{1}{2\left\{6-2f_i(\boldsymbol{h})\right\}^{1/2}}\left[-2\frac{\partial}{\partial h_j}f_i(\boldsymbol{h})\right]\\ &= \frac{-1}{\left\{6-2f_i(\boldsymbol{h})\right\}^{1/2}}\Big[2u_{ij}\sin r_i+(1-\cos r_i)(\boldsymbol{h}^\top\boldsymbol{u}_i)u_{ij}\\ &\qquad -(1+\cos r_i)h_j+R^*_{ij}\Big] \end{aligned} \tag{4.16}$$

where

$$\begin{aligned}R^*_{ij} = (2\sin r_i)&\left[u_{ij}\sum_{k=1}^{\infty}\frac{(-1)^k(\boldsymbol{h}^\top\boldsymbol{h})^k}{(2k+1)!}+2h_j(\boldsymbol{h}^\top\boldsymbol{u}_i)\sum_{k=1}^{\infty}\frac{(-1)^k k(\boldsymbol{h}^\top\boldsymbol{h})^{k-1}}{(2k+1)!}\right]\\ &+\left[\left\{(1+\cos r_i)2h_j-(1-\cos r_i)(\boldsymbol{h}^\top\boldsymbol{u}_i)2u_{ij}\right\}\sum_{k=2}^{\infty}\frac{(-1)^k(\boldsymbol{h}^\top\boldsymbol{h})^{k-1}}{(2k)!}\right]\\ &+\left[\left\{(\boldsymbol{h}^\top\boldsymbol{h})(1+\cos r_i)\right\}2h_j\sum_{k=2}^{\infty}\frac{(-1)^k(k-1)(\boldsymbol{h}^\top\boldsymbol{h})^{k-2}}{(2k)!}\right]\\ &-\left[\left\{(1-\cos r_i)(\boldsymbol{h}^\top\boldsymbol{u}_i)^2\right\}2h_j\sum_{k=2}^{\infty}\frac{(-1)^k(k-1)(\boldsymbol{h}^\top\boldsymbol{h})^{k-2}}{(2k)!}\right]. \end{aligned} \tag{4.17}$$

We know $\widetilde{\boldsymbol{S}}_n \to \boldsymbol{I}_{3\times3} = \boldsymbol{S}$ a.s. P and $\widetilde{\boldsymbol{h}}_n \to \boldsymbol{0}_3$ a.s. P (by SLLN) so $\widetilde{\boldsymbol{h}}_n \xrightarrow{p} \boldsymbol{h}_0 = \boldsymbol{0}_3$. Let A_n be the event "$\|\widetilde{\boldsymbol{h}}_n\| < 1/4$" and note $P(A_n) \to 1$ as $n \to \infty$. Let $B(1/4) \equiv \{\boldsymbol{h} \in \mathbb{R}^3 : \|\boldsymbol{h}\| \le 1/4$ denote the closed ball then $|R_{ij}| < c\|\boldsymbol{h}\|^2$ uniformly for $\boldsymbol{h} \in B(1/4)$ for some $c \in (0,\infty)$, $j = 1,2,3$. It follows that for all i and j, $|R^*_{ij}| = 0$ when $\boldsymbol{h} = \boldsymbol{h}_0$. Next

$$\begin{aligned}\psi_1(\boldsymbol{R}_i,\boldsymbol{h}) &= \frac{\partial}{\partial\boldsymbol{h}}\rho_1(\boldsymbol{R}_i,\boldsymbol{h})\\ &= \frac{-1}{\left\{6-2f_i(\boldsymbol{h})\right\}^{1/2}}\left[2\sin r_i\boldsymbol{u}_i+(1-\cos r_i)\boldsymbol{u}_i\boldsymbol{u}_i{}^\top\boldsymbol{h}-(1+\cos r_i)\boldsymbol{h}+\boldsymbol{R}^*_i\right]\\ &= \frac{-1}{\left\{6-2f_i(\boldsymbol{h})\right\}^{1/2}}\left[2\sin r_i\boldsymbol{u}_i+\boldsymbol{C}_i\boldsymbol{h}+\boldsymbol{R}^*_i\right]\\ &= \frac{1}{\left\{6-2f_i(\boldsymbol{h})\right\}^{1/2}}\left[\boldsymbol{C}_i\boldsymbol{h}-2\sin r_i\boldsymbol{u}_i-\boldsymbol{R}^*_i\right] \end{aligned} \tag{4.18}$$

where

$$\boldsymbol{C}_i = (1+\cos r_i)\boldsymbol{I}_{3\times3}-(1-\cos r_i)\boldsymbol{u}_i\boldsymbol{u}_i{}^\top$$

and

$$\mathrm{E}\left(\boldsymbol{C}_i\right) = \mathrm{E}\left[(1+\cos r_i)\,\boldsymbol{I}_{3\times3}-(1-\cos r_i)\,\boldsymbol{u}_i\boldsymbol{u}_i{}^\top\right] = \frac{2}{3}[1+2\mathrm{E}\cos(r)]\boldsymbol{I}_{3\times3}$$

because r_i is independent of $\boldsymbol{u}_i$ and $\boldsymbol{u}_i$ is distributed uniformly on the $\mathbb{R}^3$-sphere and therefore $\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_i{}^\top) = \boldsymbol{I}_{3\times 3}/3$.

Next we show the second moment of the derivative $\mathrm{E}\|\psi_1(\boldsymbol{R}_i, \boldsymbol{h})\|_F^2$ is finite and continuous in $\boldsymbol{h}$ (Step 3).

$$\begin{aligned}
\mathrm{E}\|\psi_1(\boldsymbol{R}_i, \boldsymbol{h})\|_F^2 &= \mathrm{E}\left\| \frac{\boldsymbol{C}_i\boldsymbol{h} - 2\sin(r_i)\boldsymbol{u}_i - \boldsymbol{R}_i^*}{\{6 - 2f_i(\boldsymbol{h})\}^{1/2}} \right\|_F^2 \\
&= \mathrm{E}\left[\frac{\|\boldsymbol{C}_i\boldsymbol{h} - 2\sin(r_i)\boldsymbol{u}_i - \boldsymbol{R}_i^*\|_F^2}{6 - 2f_i(\boldsymbol{h})} \right] \\
&= \mathrm{E}\left[\frac{\boldsymbol{h}^\top\boldsymbol{C}_i\boldsymbol{C}_i\boldsymbol{h} - 4\sin(r_i)\boldsymbol{u}_i{}^\top\boldsymbol{C}_i\boldsymbol{h} + 4\sin^2(r_i) + \boldsymbol{R}_i^{**}}{6 - 2f_i(\boldsymbol{h})} \right] \qquad (4.19)
\end{aligned}$$

where

$$\boldsymbol{R}_i^{**} = \boldsymbol{R}_i^{*\top}\boldsymbol{R}_i^* + 4\sin(r_i)\boldsymbol{u}_i{}^\top\boldsymbol{R}_i^* - 2\boldsymbol{h}^\top\boldsymbol{C}_i\boldsymbol{R}_i^*.$$

To evaluate (4.19) at $\boldsymbol{h}_0$ recall $\mathbf{tr}(\boldsymbol{R}_i^\top) = \mathbf{tr}(\boldsymbol{R}_i) = 1 + 2\cos r_i$ then

$$6 - 2f_i(\boldsymbol{h}_0) = 6 - 2\mathbf{tr}\left(\boldsymbol{R}^\top\boldsymbol{I}_{3\times 3}\right) = 6 - 2(1 + 2\cos r_i) = 4(1 - \cos r_i). \qquad (4.20)$$

Further $\boldsymbol{R}_i^* = \boldsymbol{0}_3$ when $\boldsymbol{h} = \boldsymbol{h}_0$ which implies $\boldsymbol{R}_i^{**} = \boldsymbol{0}_3$ when $\boldsymbol{h} = \boldsymbol{h}_0$. Thus

$$\begin{aligned}
\mathrm{E}\|\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)\|_F^2 &= \mathrm{E}\left[\frac{\sin^2(r_i)}{1 - \cos(r_i)} \right] \\
&= \mathrm{E}\left[1 + \cos\left(r_i\right)\right].
\end{aligned}$$

It follows (4.19) is finite for $\boldsymbol{h}$ in a neighborhood of $\boldsymbol{h}_0$ with probability 1 because the numerator is bounded and $P(\exp[\Phi(\boldsymbol{h})] = \boldsymbol{R}_i) = 0$ for all i. Further $\mathrm{E}\|\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)\|_F^2 = \mathrm{E}[1+\cos(r_i)]$ which is well defined and therefore not a discontinuity.

Moving to Step 4, we verify the expected value of (4.18), denoted $\Psi_1(\boldsymbol{h}) = \mathrm{E}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h})]$ has norm zero if and only if $\boldsymbol{h} = \boldsymbol{h}_0$, i.e. $\|\Psi_1(\boldsymbol{h})\| = 0$ if and only if it is evaluated at the true central orientation $\boldsymbol{h} = \boldsymbol{0}_3$. Taking expectation of (4.18):

$$\begin{aligned}
\Psi_1(\boldsymbol{h}) &= \mathrm{E}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h})] \\
&= \mathrm{E}\left[\frac{\boldsymbol{C}_i\boldsymbol{h} - 2\sin(r_i)\boldsymbol{u}_i - \boldsymbol{R}_i^*}{[6 - 2f_i(\boldsymbol{h})]^{1/2}} \right] \\
&= \mathrm{E}\left[\frac{\boldsymbol{C}_i}{[6 - 2f_i(\boldsymbol{h})]^{1/2}} \right]\boldsymbol{h} - \mathrm{E}\left[\frac{2\sin(r_i)\boldsymbol{u}_i + \boldsymbol{R}_i^*}{[6 - 2f_i(\boldsymbol{h})]^{1/2}} \right]. \qquad (4.21)
\end{aligned}$$

From (4.20), evaluating (4.21) at the true central orientation $\boldsymbol{h}_0 = \mathbf{0}_3$ gives

$$\Psi_1(\boldsymbol{h}_0) = -\mathrm{E}\left[\frac{2\sin(r_i)}{[6-2f_i(\boldsymbol{h}_0)]^{1/2}}\boldsymbol{u}_i\right] = -\mathrm{E}\left[\frac{\sin(r_i)}{\sqrt{1-\cos(r_i)}}\boldsymbol{u}_i\right] = \mathbf{0}_3 \tag{4.22}$$

because r_i is distributed symmetrically about 0 on the interval $[-\pi, \pi)$ and $\sin(r_i)/\sqrt{1-\cos(r_i)}$ is an odd function. To show the converse recall $\mathrm{E}\sin(r) = 0$ and $\mathrm{E}(\boldsymbol{C}_i) = 2[1+2\mathrm{E}\cos(r)]\boldsymbol{I}_{3\times 3}/3$. Assume $\Psi_1(\boldsymbol{h}) = \mathbf{0}_3$ then

$$\begin{aligned}
&\left\|\mathrm{E}\left[\frac{\boldsymbol{C}_i\boldsymbol{h} - 2\sin(r_i)\boldsymbol{u}_i - \boldsymbol{R}_i^*}{[6-2f_i(\boldsymbol{h})]^{1/2}}\right]\right\| && = 0 \\
\implies &\|\mathrm{E}[\boldsymbol{C}_i\boldsymbol{h} - 2\sin(r_i)\boldsymbol{u}_i - \boldsymbol{R}_i^*]\| && = 0 \\
\implies &\|\mathrm{E}(\boldsymbol{C}_i)\boldsymbol{h} - \mathrm{E}(\boldsymbol{R}_i^*)\| && = 0 \\
\implies &\boldsymbol{h}^\top\mathrm{E}(\boldsymbol{C}_i)^\top\mathrm{E}(\boldsymbol{C}_i)\boldsymbol{h} - 2\mathrm{E}(\boldsymbol{R}_i^*)^\top\mathrm{E}(\boldsymbol{C}_i)\boldsymbol{h} + \mathrm{E}(\boldsymbol{R}_i^*)^\top\mathrm{E}(\boldsymbol{R}_i^*) && = 0 \\
\implies &\frac{4}{9}\left[1+2\mathrm{E}\cos(r)\right]^2\boldsymbol{h}^\top\boldsymbol{h} - \frac{4}{3}\left[1+2\mathrm{E}\cos(r)\right]\mathrm{E}(\boldsymbol{R}_i^*)^\top\boldsymbol{h} + \mathrm{E}(\boldsymbol{R}_i^*)^\top\mathrm{E}(\boldsymbol{R}_i^*) && = 0 \\
\implies &\frac{4}{3}\left[1+2\mathrm{E}\cos(r)\right]\left\{\frac{1}{3}\left[1+2\mathrm{E}\cos(r)\right]\boldsymbol{h}^\top - \mathrm{E}(\boldsymbol{R}_i^*)^\top\right\}\boldsymbol{h} + \mathrm{E}(\boldsymbol{R}_i^*)^\top\mathrm{E}(\boldsymbol{R}_i^*) && = 0
\end{aligned} \tag{4.23}$$

It can be shown that

$$\mathrm{E}(\boldsymbol{R}_i^*) = \frac{4}{3}[1+2\mathrm{E}\cos(r)]\left[\sum_{k=2}^{\infty}\frac{(-1)^k(\boldsymbol{h}^\top\boldsymbol{h})^{k-1}}{(2k)!} + \|\boldsymbol{h}\|^2\sum_{k=2}^{\infty}\frac{(-1)^k(k-1)(\boldsymbol{h}^\top\boldsymbol{h})^{k-1}}{(2k)!}\right]\boldsymbol{h}.$$

Since r is symmetric about zero, $\mathrm{E}\cos(r) > 0$ and therefore $\mathrm{E}(\boldsymbol{R}_i^*)^\top\mathrm{E}(\boldsymbol{R}_i^*) = 0$ if and only if $\boldsymbol{h} = \mathbf{0}_3$. It follows that (4.23) implies $\boldsymbol{h} = \mathbf{0}_3$. Putting this together with (4.22) gives that $\|\Psi_1'(\boldsymbol{h})\| = 0$ if and only if $\boldsymbol{h} = \boldsymbol{h}_0$.

Finally we show the derivative of the expectation $\Psi_1'(\boldsymbol{h}_0)$ exists and is non-zero $\|\Psi_1'(\boldsymbol{h}_0)\| \neq 0$ (Step 5). Notice that the derivative of the loss function $\psi_1(\boldsymbol{R}_i, \boldsymbol{h})$ is continuous and bounded in $\boldsymbol{h}$, therefore derivative and expectation interchange. That is,

$$\begin{aligned}
\Psi_1'(\boldsymbol{h}_0) &= \frac{\partial}{\partial\boldsymbol{h}}\mathrm{E}[-\psi_1(\boldsymbol{R}_i, \boldsymbol{h})]\Big|_{\boldsymbol{h}=\boldsymbol{h}_0} \\
&= \mathrm{E}\left[-\frac{\partial}{\partial\boldsymbol{h}}\psi_1(\boldsymbol{R}_i, \boldsymbol{h})\Big|_{\boldsymbol{h}=\boldsymbol{h}_0}\right].
\end{aligned} \tag{4.24}$$

First we find the derivative inside the expectation,

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{h}}\psi_1(\boldsymbol{R}_i,\boldsymbol{h})\Big|_{\boldsymbol{h}=\boldsymbol{h}_0} &= \frac{\partial^2}{\partial \boldsymbol{h}\partial \boldsymbol{h}^\top}\rho_1(\boldsymbol{R}_i,\boldsymbol{h})\Big|_{\boldsymbol{h}=\boldsymbol{h}_0} \\
&= \frac{\partial}{\partial \boldsymbol{h}}\left[\frac{\boldsymbol{C}_i\boldsymbol{h}^\top - 2\sin(r_i)\boldsymbol{u}_i{}^\top - \boldsymbol{R}_i^{*\top}}{[6-2f_i(\boldsymbol{h})]^{1/2}}\right]\Bigg|_{\boldsymbol{h}=\boldsymbol{h}_0} \qquad \text{(from (4.18))} \\
&= \frac{1}{[6-2f_i(\boldsymbol{h})]}\left\{[6-2f_i(\boldsymbol{h})]^{1/2}\left[\boldsymbol{C}_i - \frac{\partial}{\partial \boldsymbol{h}}\boldsymbol{R}_i^{*\top}\right]\right. \\
&\qquad \left. + \frac{\frac{\partial}{\partial \boldsymbol{h}}f_i(\boldsymbol{h})}{[6-2f_i(\boldsymbol{h})]^{1/2}}\left[\boldsymbol{C}_i\boldsymbol{h}^\top - 2\sin r_i\boldsymbol{u}_i{}^\top - \boldsymbol{R}_i^{*\top}\right]\right\}\Bigg|_{\boldsymbol{h}=\boldsymbol{h}_0} \\
&= \frac{\boldsymbol{C}_i - \frac{\partial}{\partial \boldsymbol{h}}\boldsymbol{R}_i^{*\top}\big|_{\boldsymbol{h}=\boldsymbol{h}_0}}{[6-2f_i(\boldsymbol{h}_0)]^{1/2}} + \frac{\left[\frac{\partial}{\partial \boldsymbol{h}}f_i(\boldsymbol{h}_0)\right]\left[-2\sin r_i\boldsymbol{u}_i{}^\top\right]}{[6-2f_i(\boldsymbol{h}_0)]^{3/2}} \\
&= \frac{1}{2\sqrt{1-\cos r_i}}\boldsymbol{C}_i + \frac{[-2\sin r_i\boldsymbol{u}_i]\left[-2\sin r_i\boldsymbol{u}_i{}^\top\right]}{8\,[1-\cos r_i]^{3/2}} \qquad \text{(from (4.18))} \\
&= \frac{1}{2\sqrt{1-\cos r_i}}\boldsymbol{C}_i + \frac{\sin^2 r_i}{2\,[1-\cos r_i]^{3/2}}\boldsymbol{u}_i\boldsymbol{u}_i{}^\top.
\end{aligned}$$

From (4.17), $\partial R^*_{ij}/\partial h_{j'}$ will contain at least one element of $\boldsymbol{h}$ for all $j' = 1,2,3$, therefore $\partial \boldsymbol{R}_i^{*\top}/\partial \boldsymbol{h} = \boldsymbol{0}_{3\times 3}$ when evaluated at $\boldsymbol{h} = \boldsymbol{h}_0$. Now taking the expectation,

$$\begin{aligned}
\Psi_1'(\boldsymbol{h}_0) &= \mathrm{E}\left[-\frac{\partial}{\partial \boldsymbol{h}}\psi_1(\boldsymbol{R}_i,\boldsymbol{h})\Big|_{\boldsymbol{h}=\boldsymbol{h}_0}\right] \\
&= \mathrm{E}\left[\frac{1+2\cos(r_i)-3\cos^2(r_i)}{6[1-\cos(r_i)]^{3/2}}\right]\boldsymbol{I}_{3\times 3} \\
&= \mathrm{E}\left[\frac{1+3\cos(r_i)}{6\sqrt{1-\cos(r_i)}}\right]\boldsymbol{I}_{3\times 3} \\
&= 2a_2\boldsymbol{I}_{3\times 3} \qquad (4.25)
\end{aligned}$$

where $a_2 = \mathrm{E}\left\{[1+3\cos(r)]\,/12\sqrt{1-\cos(r)}\right\}$. The assumption $\boldsymbol{R}_i \neq \widetilde{\boldsymbol{S}}_n$ for all i guarantees $\cos(r_i) \neq 1$ which guarantees d is well defined. Therefore, $\Psi_1'(\boldsymbol{h}_0)$ exists and $\|\Psi_1'(\boldsymbol{h}_0)\|_F^2 = 2a_2\sqrt{3}\boldsymbol{I}_{3\times 3} \neq 0$.

Therefore all the conditions necessary for $\sqrt{n}\,\widetilde{\boldsymbol{h}}_n$ to be asymptotically normal have been sat-

isfied. Next we derive the asymptotic variance of $\sqrt{n}\,\widetilde{\boldsymbol{h}}_n$. First we must derive $\text{Var}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)]$:

$$\begin{aligned}
\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)^\top &= \frac{\sin^2(r_i)}{1-\cos(r_i)}\boldsymbol{u}_i\boldsymbol{u}_i{}^\top \\
\implies \text{Var}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)] &= \text{E}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)^\top] - \boldsymbol{0}_3\boldsymbol{0}_3^\top \\
&= \frac{1}{3}\text{E}\left[\frac{\sin^2(r_i)}{1-\cos(r_i)}\right]\boldsymbol{I}_{3\times 3} \\
&= \frac{1}{3}\text{E}[1+\cos(r_i)]\boldsymbol{I}_{3\times 3} \\
&= 2a_1\boldsymbol{I}_{3\times 3} \qquad (4.26)
\end{aligned}$$

where $a_1 = \text{E}[1+\cos(r)]/6$. By the multivariate i.i.d. version of Theorem 5.13 from Shao (2003) on page 367

$$\sqrt{n}\,\widetilde{\boldsymbol{h}}_n \xrightarrow{\mathcal{L}} MVN_3\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right).$$

where

$$\begin{aligned}
\boldsymbol{\Sigma} &= [\Psi_1'(\boldsymbol{h}_0)]^{-1}\text{Var}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h})][\Psi_1'(\boldsymbol{h}_0)]^{-1} \\
&= (2a_2\boldsymbol{I}_{3\times 3})^{-1}(2a_1\boldsymbol{I}_{3\times 3})(2a_2\boldsymbol{I}_{3\times 3})^{-1} \\
&= \frac{a_1}{2a_2^2}\boldsymbol{I}_{3\times 3}.
\end{aligned}$$

□

Proof of Proposition 4.4.

Without loss of generality assume the central orientation $\boldsymbol{S} = \boldsymbol{I}_{3\times 3}$. In the Proof of Proposition 4.2 it was established that $\widetilde{\boldsymbol{S}}_n$ is an M-estimator, therefore according to Hampel et al. (2011), page 230, it has influence function

$$\text{IF}_1(\boldsymbol{R}_i, F) = [\Psi_1'(\boldsymbol{h}_0)]^{-1}\left[\psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0)\right]$$

where $\Psi_1'(\boldsymbol{h}_0) = \frac{\partial}{\partial \boldsymbol{h}}\text{E}[\psi_1(\boldsymbol{R}_i, \boldsymbol{h})]\big|_{\boldsymbol{h}=\boldsymbol{h}_0}$. From (4.18),

$$\begin{aligned}
\psi_1(\boldsymbol{R}_i, \boldsymbol{h}) &= \frac{2\sin(r_i)\boldsymbol{U}_i - \{[1-\cos(r_i)]\boldsymbol{U}_i\boldsymbol{U}_i^\top - [1+\cos(r_i)]\boldsymbol{I}_{3\times 3}\}\boldsymbol{h} - \boldsymbol{R}_i^*}{\|\boldsymbol{R}_i - \exp[\Phi(\boldsymbol{h})]\|} \\
\implies \psi_1(\boldsymbol{R}_i, \boldsymbol{h}_0) &= \frac{2\sin(r_i)}{\|\boldsymbol{R}_i - \boldsymbol{I}_{3\times 3}\|}\boldsymbol{u}_i \\
&= \frac{\sin(r_i)}{\sqrt{1-\cos(r_i)}}\boldsymbol{u}_i.
\end{aligned}$$

From (4.25), $\Psi_1'(\boldsymbol{h}_0) = 2a_2\boldsymbol{I}_{3\times 3}$ where $a_2 = \mathrm{E}\left\{[1+3\cos(r)] \,/\, \left[12\sqrt{1-\cos(r)}\right]\right\}$.

Putting it together

$$\mathrm{IF}_1(\boldsymbol{R}_i, F) = \frac{\sin(r_i)}{2a_2\sqrt{1-\cos(r_i)}}\boldsymbol{u}_i.$$

By assumption, r_i is distributed symmetrically about 0 on the interval $[-\pi, \pi)$ independent of $\boldsymbol{u}_i$ which is distributed uniformly on the unit sphere. Therefore $\mathrm{E}[\sin(r)/\sqrt{1-\cos(r)}] = 0$, which implies $\mathrm{E}[\mathrm{IF}_1(\boldsymbol{R}_i, F)] = 0$. Recall $\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_i{}^\top) = \boldsymbol{I}_{3\times 3}/3$, therefore

$$\begin{aligned}
\mathrm{E}[\mathrm{IF}_1(\boldsymbol{R}_i, F)\mathrm{IF}_1(\boldsymbol{R}_i, F)^\top] &= \mathrm{E}\left\{\frac{\sin^2(r)}{4a_2^2[1-\cos(r)]}\boldsymbol{u}_i\boldsymbol{u}_i{}^\top\right\} \\
&= \frac{1}{12a_2^2}\mathrm{E}\left[\frac{\sin^2(r)}{1-\cos(r)}\right]\boldsymbol{I}_{3\times 3} \\
&= \frac{1}{12a_2^2}\mathrm{E}\left[1+\cos(r)\right]\boldsymbol{I}_{3\times 3} \\
&= \frac{a_1}{2a_2^2}\boldsymbol{I}_{3\times 3}
\end{aligned}$$

where $a_1 = \mathrm{E}[1+\cos(r)]/6$. □

Proof of Theorem 4.3.

From Propositions 4.1 and 4.4, $\widetilde{\boldsymbol{S}}_n$ is an M-estimators with a continuous and bounded influence function. By Theorem 5.7 of Shao (2003), it follows that $\widetilde{\boldsymbol{S}}_n$ is ϱ_∞-Hadamard differentiable (see page 347). Further, the influence function of $\widetilde{\boldsymbol{S}}_n$ has a zero first and finite second moment. Therefore by Theorem 5.20 of Shao (2003) $\varrho_\infty(F_{boot}, F_n) \xrightarrow{p} 0$ as $n \to \infty$ (see page 383) and the result follows. □

Additional Results

Tables 4.3 and 4.4 are numerical summaries for the simulation study. The coverage rates of the theoretical and bootstrap regions based on the mean and median are in Table 4.3 and the region sizes are in Table 4.4.

Table 4.3: Coverage rates for the normal theory and bootstrap confidence regions based on the mean and median. The nominal coverage rate is 90%.

			Theory		Bootstrap	
Distribution	n	ϵ	Mean	Median	Mean	Median
Cayley	10	0.0	0.841	0.723	0.920	0.983
		0.1	0.863	0.710	0.939	0.991
		0.2	0.722	0.671	0.910	0.996
	50	0.0	0.886	0.866	0.897	0.931
		0.1	0.551	0.814	0.601	0.905
		0.2	0.027	0.575	0.062	0.758
	100	0.0	0.895	0.881	0.899	0.914
		0.1	0.210	0.757	0.242	0.814
		0.2	0.000	0.326	0.000	0.430
matrix-Fisher	10	0.0	0.844	0.740	0.919	0.983
		0.1	0.912	0.725	0.997	0.992
		0.2	0.667	0.687	0.991	0.997
	50	0.0	0.886	0.866	0.896	0.925
		0.1	0.162	0.810	0.310	0.898
		0.2	0.000	0.582	0.000	0.746
	100	0.0	0.897	0.889	0.899	0.917
		0.1	0.003	0.752	0.009	0.808
		0.2	0.000	0.319	0.000	0.398

Table 4.4: Confidence region size for the normal theory and bootstrap confidence regions based on the mean and median.

			Theory		Bootstrap	
Distribution	n	ϵ	Mean	Median	Mean	Median
Cayley	10	0.0	0.233	0.234	0.271	3.937
	10	0.1	0.289	0.247	0.338	6.286
	10	0.2	0.350	0.268	0.445	9.868
	50	0.0	0.109	0.119	0.112	0.137
	50	0.1	0.134	0.128	0.140	0.151
	50	0.2	0.161	0.141	0.177	0.171
	100	0.0	0.078	0.085	0.079	0.091
	100	0.1	0.095	0.092	0.098	0.099
	100	0.2	0.114	0.101	0.124	0.111
matrix-Fisher	10	0.0	0.119	0.119	0.138	0.922
	10	0.1	0.203	0.128	0.291	1.486
	10	0.2	0.280	0.138	0.486	2.419
	50	0.0	0.056	0.060	0.057	0.068
	50	0.1	0.092	0.065	0.104	0.075
	50	0.2	0.127	0.072	0.152	0.085
	100	0.0	0.040	0.043	0.040	0.046
	100	0.1	0.065	0.047	0.072	0.050
	100	0.2	0.090	0.051	0.105	0.056

CHAPTER 5. ROTATIONS: AN R PACKAGE FOR SO(3) DATA

A paper accepted by *The R Journal*

Bryan Stanfill, Heike Hofmann, Ulrike Genschel

Abstract

In this article we introduce the **rotations** package which provides users with the ability to simulate, analyze and visualize three-dimensional rotation data. More specifically it includes four commonly used distributions from which to simulate data, four estimators of the central orientation, six confidence region estimation procedures and two approaches to visualizing rotation data. All of these features are available for two different parameterizations of rotations: three-by-three matrices and quaternions. In addition, two datasets are included that illustrate the use of rotation data in practice.

5.1 Introduction

Data in the form of three-dimensional rotations have applications in many scientific areas, such as bio-medical engineering, computer vision, and geological and materials sciences where such data represent the positions of objects within a three-dimensional reference frame. For example, Humbert et al. (1996), Bingham et al. (2009) and Bachmann et al. (2010) apply rotation data to study the orientation of cubic crystals on the surfaces of metal. Rancourt et al. (2000) use rotations to represent variations in human movement while performing a task.

A common goal shared in the analysis of rotation data across all fields is to estimate the main or central orientation for a sample of rotations. More formally, let $SO(3)$ denote the rotation group, which consists of all real-valued 3×3 matrices $\boldsymbol{R}$ with determinant equal to $+1$. Then observations $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n \in SO(3)$ can be conceptualized as a random sample from a

location model

$$\mathbf{R}_i = \boldsymbol{S}\boldsymbol{E}_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where $\boldsymbol{S} \in SO(3)$ is the *fixed* parameter of interest indicating the central orientation, and $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n \in SO(3)$ denote i.i.d. *random* rotations which symmetrically perturb $\boldsymbol{S}$. Model (5.1) is a rotation-matrix analog of a location model for scalar data $Y_i = \mu + e_i$, where $\mu \in \mathbb{R}$ denotes a mean and $e_i \in \mathbb{R}$ denotes an additive error symmetrically distributed around zero.

Assuming the perturbations $\boldsymbol{E}_i$ symmetrically perturb $\boldsymbol{S}$ implies that the observations $\boldsymbol{R}_i$ have no preferred direction relative to $\boldsymbol{S}$ and that $E(\boldsymbol{R}_i) = c\boldsymbol{S}$ for some $c \in \mathbb{R}^+$ for all i. Also note that under the symmetry assumption, (5.1) could be equivalently specified as $\boldsymbol{R}_i = \boldsymbol{E}_i\boldsymbol{S}$, though the form given in (5.1) is the most common form in the literature (see Bingham et al. 2009 for details).

While there is a multitude of packages and functions available in R to estimate the mean in a location model, the toolbox for rotational data is limited. The **orientlib** (Murdoch, 2003) package includes the definition of an orientation class along with a few methods to summarize and visualize rotation data. A strength of the **orientlib** package is its thorough exploration of rotation representations, but the estimation and visualization techniques are lacking and no methods for inference are available. The **onion** (Hankin, 2011) package includes functions for rotation algebra but only the quaternion form is available and data analysis is not possible. The **uarsbayes** (Qiu, 2013) package includes functions for data generation and Bayes inference but this package is currently not publicly available. Packages for circular and spherical data, e.g. **circular** (Agostinelli and Lund, 2013) and **SpherWave** (Oh and Kim, 2013), can possibly be used but their extension to rotation data is not straightforward.

The **rotations** (Stanfill et al., 2014a) package fills this void by providing users with the tools necessary to simulate rotations from (5.1) with four distribution choices for the perturbation matrices $\boldsymbol{E}_i$. Estimation and inference for $\boldsymbol{S}$ in (5.1) is available along with two visualization techniques. The remainder of this manuscript introduces rotation data more fully and discusses the ways they are handled by the **rotations** package. For the latest on this package as well as a full list of available functions, see `help(package = "rotations")`.

5.2 Rotation Parameterizations

Several parameterizations of rotations exist. We consider two of the most commonly used: orthogonal 3×3 matrices with determinant one and four-dimensional unit vectors called *quaternions*. The **rotations** package allows for both parameterizations as input as well as transforming one into the other. We will briefly discuss each:

5.2.1 Matrix Form

Rotations in three-dimensions can be represented by 3×3 orthogonal matrices with determinant one. Matrices with these characteristics form a group called the *special orthogonal group*, or *rotation group*, denoted $SO(3)$. Every element in $SO(3)$ is associated with a skew-symmetric matrix $\boldsymbol{\Phi}(\boldsymbol{W})$ where

$$\boldsymbol{\Phi}(\boldsymbol{W}) = \begin{pmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{pmatrix}$$

and $\boldsymbol{W} \in \mathbb{R}^3$. Applying the exponential operator to the matrix $\boldsymbol{\Phi}(\boldsymbol{W})$ results in the rotation $\boldsymbol{R}$

$$\boldsymbol{R} = \exp\left[\boldsymbol{\Phi}(\boldsymbol{W})\right] = \sum_{k=0}^{\infty} \frac{\left[\boldsymbol{\Phi}(\boldsymbol{W})\right]^k}{k!}. \tag{5.2}$$

Since $\boldsymbol{\Phi}(\boldsymbol{W})$ is skew-symmetric, it can be shown that (5.2) reduces to

$$\boldsymbol{R} = \cos(r)\boldsymbol{I}_{3\times3} + \sin(r)\boldsymbol{\Phi}(\boldsymbol{U}) + [1-\cos(r)]\boldsymbol{U}\boldsymbol{U}^\top, \tag{5.3}$$

where $r = \|\boldsymbol{W}\|$, $\boldsymbol{U} = \boldsymbol{W}/\|\boldsymbol{W}\|$. In the material sciences literature r and $\boldsymbol{U} \in \mathbb{R}^3$ are termed the *misorientation angle* and *misorientation axis*, respectively.

Given a rotation matrix $\boldsymbol{R}$ one can find the associated skew-symmetric matrix $\boldsymbol{\Phi}(\boldsymbol{W})$ by applying the logarithm operator defined by

$$\mathrm{Log}(\boldsymbol{R}) = \begin{cases} \boldsymbol{0} & \text{if } \theta = 0 \\ \frac{r}{2\sin r}\left(\boldsymbol{R} - \boldsymbol{R}^\top\right) & \text{otherwise,} \end{cases} \tag{5.4}$$

where $r \in [-\pi, \pi)$ satisfies $\mathbf{tr}(\boldsymbol{R}) = 1 + 2\cos r$ and $\mathbf{tr}(\cdot)$ denotes the trace of a matrix. For more on the correspondence between $SO(3)$ and skew-symmetric matrices see Stanfill et al. (2013).

The **rotations** package defines the `S3` class `"SO3"`, which internally stores a sample of n rotations as a $n \times 9$ matrix. If $n = 1$ then an object of class `"SO3"` is printed as a 3×3 matrix but for $n > 1$ the $n \times 9$ matrix is printed. Objects can be coerced into, or tested for the class `"SO3"` with the `as.SO3` and `is.SO3` functions, respectively. Any object passed to `is.SO3` is tested for three characteristics: dimensionality, orthogonality and determinant one.

The `as.SO3` function coerces the input into the class `"SO3"`. There are three types of input supported by the `as.SO3` function. Given a singe angle r and axis $\boldsymbol{U}$, `as.SO3` will form a rotation matrix according to (5.3). Equivalently one could supply a three-dimensional vector $\boldsymbol{W}$, then the length of that vector will be taken to be the angle of rotation $r = \|\boldsymbol{W}\|$ and the axis is taken to be the unit-vector in the direction of $\boldsymbol{W}$, i.e. $\boldsymbol{U} = \boldsymbol{W}/\|\boldsymbol{W}\|$. One can also supply a rotation $\boldsymbol{Q}$ in the quaternion representation. The `as.SO3` function will return the matrix equivalent of $\boldsymbol{Q}$. For all input types the function `as.SO3` returns an $n \times 9$ matrix of class `"SO3"` where each row corresponds to a rotation matrix. Below we illustrate the use of the `as.SO3` function by constructing the 3×3 matrix associated with a 90° rotation about the y-axis, i.e. $r = \pi/2$ and $\boldsymbol{U} = (0, 1, 0)$. In this example and all that follow, we have rounded the output to three digits for compactness.

```
> r <- pi/2
> U <- c(0, 1, 0)
> W <- U * r
> R <- as.SO3(W)
> R
     [,1] [,2] [,3]
[1,]    0    0    1
[2,]    0    1    0
[3,]   -1    0    0
> identical(R, as.SO3(U, r))
[1] TRUE
```

Given a rotation matrix $\boldsymbol{R}$, the functions `mis.angle` and `mis.axis` will determine the

misorientation angle and axis of an object with class `"SO3"` as illustrated in the next example.

```
> mis.angle(R) * 2/pi
[1] 1
> mis.axis(R)
     [,1] [,2] [,3]
[1,]    0    1    0
```

5.2.2 Quaternion Form

Quaternions are unit vectors in $\mathbb{R}^4$ that are commonly written as

$$\boldsymbol{Q} = x_1 + x_2 i + x_3 j + x_4 k, \tag{5.5}$$

where $x_l \in [-1, 1]$ for $l = 1, 2, 3, 4$ and $i^2 = j^2 = k^2 = ijk = -1$. We can write $\boldsymbol{Q} = (s, \boldsymbol{V})$ as tuple of the scalar s for coefficient $\mathbf{1}$ and vector $\boldsymbol{V}$ for the remaining coefficients, i.e. $s = x_1$ and $\boldsymbol{V} = (x_2, x_3, x_4)$.

A rotation around axis $\boldsymbol{U}$ by angle r translates to $\boldsymbol{Q} = (s, \boldsymbol{V})$ with

$$s = \cos(r/2), \quad \boldsymbol{V} = \boldsymbol{U} \sin(r/2).$$

Note that rotations in quaternion form are over-parametrized: $\boldsymbol{Q}$ and $-\boldsymbol{Q}$ represent equivalent rotations. This ambiguity has no impact on the distributional models, parameter estimation or inference methods to follow. Hence, for consistency, the **rotations** package only generates quaternions satisfying $x_1 \geq 0$. Data provided by the user does not need to satisfy this condition however.

The `S3` class `"Q4"` is defined for the quaternion representation of rotations. All the functionality of the `"SO3"` class also exists for the `"Q4"` class, e.g. `is.Q4` and `as.Q4` will test for and coerce to class `"Q4"`, respectively. Internally, a sample of n quaternions is stored in the form of a $n \times 4$ matrix with each row a unit vector. Single quaternions are printed according to the representation in (5.5) (see example below) while a sample of size n is printed as a $n \times 4$ matrix with column names `Real`, `i`, `j` and `k` to distinguish between the four components.

The following code creates the same rotation from the previous section in the form of a quaternion with the `as.Q4` function. This function works much the same way as the `as.SO3` function in terms of possible inputs but returns a vector of length four of the class `"Q4"`.

```
> as.Q4(U, r)
0.707 + 0 * i + 0.707 * j + 0 * k
> as.Q4(as.SO3(U, r))
0.707 + 0 * i + 0.707 * j + 0 * k
```

5.3 Data Generation

If the rotation $\boldsymbol{E}_i \in SO(3)$ from (5.1) has an axis $\boldsymbol{U}$ that is uniformly distributed on the unit sphere and an angle r that is independently distributed about zero according to some symmetric distribution function then $\boldsymbol{E}_i$ is said to belong to the *uniform-axis random spin*, or *UARS*, class of distributions. From Bingham et al. (2009) the density for $\boldsymbol{E}_i$ is given by

$$f\left(\boldsymbol{E}_i|\kappa\right) = \frac{4\pi}{3 - \mathbf{tr}\left(\boldsymbol{E}_i\right)} C\left(\operatorname{acos}\left\{\frac{\mathbf{tr}(\boldsymbol{E}_i) - 1}{2}\right\}\middle|\, \kappa\right), \tag{5.6}$$

where $C(\cdot|\kappa)$ is the distribution function associated with the angle of rotation r with concentration parameter κ. Members of the UARS family of distributions are differentiated based on the angular distribution $C(\cdot|\kappa)$.

The **rotations** package gives the user access to four members of the UARS class. Each member is differentiated by the distribution function for r: the uniform, the matrix Fisher (Langevin, 1905; Downs, 1972; Khatri and Mardia, 1977; Jupp and Mardia, 1979), the Cayley (Schaeben, 1997; León et al., 2006) and the circular-von Mises distribution (Bingham et al., 2009). Note: probability distribution functions on $SO(3)$ such as (5.6) are defined with respect to the Haar measure, which we denote by λ. That is, the expectation of a random rotation $\boldsymbol{R} \in SO(3)$ with corresponding misorientation angle r is given by $E\left(\boldsymbol{R}\right) = \int_\Omega \boldsymbol{R} f\left(\boldsymbol{R}|\kappa\right) \mathrm{d}\lambda$ where $\Omega = SO(3)$, $\mathrm{d}\lambda = [1 - \cos(r)]\mathrm{d}r/(2\pi)$ and $\mathrm{d}r$ is the derivative of r with respect to the Lebesgue measure. Because the Haar measure acts as the uniform measure on $SO(3)$ and $\lambda\left(\Omega\right) = 1$, then the angular distribution $C(r) = [1 - \cos(r)]/(2\pi)$ is referred to as the uniform

distribution for misorientation angles r and has been included in the **rotations** package under the name `.haar` (see Table 5.1).

The spread of the Cayley, matrix Fisher and circular-von Mises distributions is controlled by the concentration parameter κ. Concentration is a distribution specific quantity and is not comparable across different distributions. To make comparisons across distributions possible we also allow for specification of the circular variance, which is defined as $\nu = 1 - E[\cos(r)]$ where $E[\cos(r)]$ is often referred to as the *mean resultant length* (Fisher, 1996). The form of each angular distribution along with the circular variance as a function of the concentration parameter is given in Table 5.1.

Table 5.1: Circular densities and circular variance ν; $I_i(\cdot)$ represents the modified Bessel function of order i and $\Gamma(\cdot)$ is the gamma function.

Name	**Density** $C(r\|\kappa)$	**Circular variance** ν	**Function**
Uniform	$\frac{1-\cos(r)}{2\pi}$	$\frac{3}{2}$	`.haar`
Cayley	$\frac{\Gamma(\kappa+2)(1+\cos r)^{\kappa}(1-\cos r)}{2^{(\kappa+1)}\sqrt{\pi}\Gamma(\kappa+1/2)}$	$\frac{3}{\kappa+2}$	`.cayley`
matrix Fisher	$\frac{[1-\cos(r)]\exp[2\kappa\cos(r)]}{2\pi[\mathrm{I}_0(2\kappa)-\mathrm{I}_1(2\kappa)]}$	$\frac{3\mathrm{I}_0(2\kappa)-4\mathrm{I}_1(2\kappa)+\mathrm{I}_2(2\kappa)}{2[\mathrm{I}_0(2\kappa)-\mathrm{I}_1(2\kappa)]}$	`.fisher`
circular-von Mises	$\frac{\exp[\kappa\cos(r)]}{2\pi\mathrm{I}_0(\kappa)}$	$\frac{\mathrm{I}_0(\kappa)-\mathrm{I}_1(\kappa)}{\mathrm{I}_0(\kappa)}$	`.vmises`

For a given concentration `d`, `p` and `r` take the same meaning as for the more familiar distributions such as `dnorm`. To simulate a sample of $SO(3)$ data, the `ruars` function takes arguments `n`, `rangle`, and `kappa` to specify the sample size, angular distribution and concentration as shown below. Alternatively, one can specify the circular variance ν. Circular variance is used in the event that both circular variance and concentration are provided. The `space` argument determines the parameterization to form. When a sample of rotations is printed then a $n \times 9$ matrix is printed with column titles that specify which element of the matrix each column corresponds to. For example, the $\boldsymbol{R}_{\{1,1\}}$ element of a rotation matrix is printed under the column heading `R11` as illustrated below.

```
> Rs <- ruars(n = 20, rangle = rcayley, kappa = 1, space = "SO3")
> Qs <- ruars(n = 20, rangle = rcayley, kappa = 1, space = "Q4")
```

```
> Rs <- ruars(n = 20, rangle = rcayley, nu = 1, space = "SO3")
> Qs <- ruars(n = 20, rangle = rcayley, nu = 1, space = "Q4")
> head(Rs, 3)
        R11    R21   R31   R12    R22    R32    R13    R23   R33
[1,] -0.425 -0.850 0.310 0.475 -0.501 -0.723  0.770 -0.160 0.617
[2,] -0.564 -0.733 0.379 0.745 -0.256  0.615 -0.354  0.630 0.691
[3,]  0.087 -0.716 0.692 0.117  0.698  0.707 -0.989  0.019 0.145
```

5.4 Data Analysis

In this section we present functions in the **rotations** package to compute point estimates and confidence regions for the central orientation $\boldsymbol{S}$.

5.4.1 Estimation of Central Orientation

Given a sample of n observations $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ generated according to (5.1), the **rotations** package offers four built-in ways to estimate the central orientation $\boldsymbol{S}$. These estimators are either Riemannian- or Euclidean-based in geometry and use either the L_1- or L_2- norm, i.e. they are median- or mean-type. We briefly discuss how the choice of geometry affects estimation of $\boldsymbol{S}$.

The choice of geometry results in two different metrics to measure the distance between rotation matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2 \in SO(3)$. The Euclidean distance, d_E, between two rotations is defined by

$$d_E\left(\boldsymbol{R}_1, \boldsymbol{R}_2\right) = \left\|\boldsymbol{R}_1 - \boldsymbol{R}_2\right\|_F,$$

where $\|\boldsymbol{A}\|_F = \sqrt{\mathbf{tr}(\boldsymbol{A}^\top \boldsymbol{A})}$ denotes the Frobenius norm. The Euclidean distance between two rotation matrices corresponds to the length of the shortest path in $\mathbb{R}^{3\times 3}$ that connects them and is therefore an *extrinsic* distance metric.

Estimators based on the Euclidean distance form the class of *projected* estimators. The name is derived from the method used to compute these estimators. That is, each estimator in this class is the projection of the the generic 3×3 matrix that minimizes the loss function

into $SO(3)$. For an object with class `"SO3"` the `median` or `mean` function with argument `type = "projected"` will return a 3×3 matrix in $SO(3)$ that minimizes the first- or second-order loss function, respectively.

By staying in the Riemannian space $SO(3)$ the natural distance metric becomes the Riemannian (or geodesic) distance, d_R, which for two rotations $\boldsymbol{R}_1, \boldsymbol{R}_2 \in SO(3)$ is defined as

$$d_R\left(\boldsymbol{R}_1, \boldsymbol{R}_2\right) = \frac{1}{\sqrt{2}} \left\| \text{Log}\left(\boldsymbol{R}_1^\top \boldsymbol{R}_2\right) \right\|_F = |r|,$$

where $\text{Log}(\boldsymbol{R})$ denotes the logarithm of $\boldsymbol{R}$ defined in (5.4) and $r \in [-\pi, \pi)$ is the misorientation angle of $\boldsymbol{R}_1^\top \boldsymbol{R}_2$. The Riemannian distance corresponds to the length of the shortest path that connects $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ *within* the space $SO(3)$ and is therefore an *intrinsic* distance metric. For this reason, the Riemannian distance is often considered the more natural metric on $SO(3)$. As demonstrated in Stanfill et al. (2013), the Euclidean and Riemannian distances are related by $d_E(\boldsymbol{R}_1, \boldsymbol{R}_2) = 2\sqrt{2}\sin\left[d_R(\boldsymbol{R}_1, \boldsymbol{R}_2)/2\right]$.

Estimators based on the Riemannian distance metric are called *geometric* estimators because they preserve the geometry of $SO(3)$. These can be computed using the `mean` and `median` functions with the argument `type = "geometric"`. Table 5.2 summarizes the four estimators including their formal definition and how they can be computed.

Table 5.2: A summary of the estimators included in the **rotations** package. `Rs` is a sample of n rotations with class `"SO3"` or `"Q4"`.

Estimator name	**Definition**	**Code**
Projected Mean	$\widehat{\boldsymbol{S}}_n = \underset{\boldsymbol{S} \in SO(3)}{\arg\min} \sum_{i=1}^{n} d_E^2(\boldsymbol{S}, \boldsymbol{R}_i)$	`mean(Rs, type = "projected")`
Projected Median	$\widetilde{\boldsymbol{S}}_n = \underset{\boldsymbol{S} \in SO(3)}{\arg\min} \sum_{i=1}^{n} d_E(\boldsymbol{S}, \boldsymbol{R}_i)$	`median(Rs, type = "projected")`
Geometric Mean	$\widehat{\boldsymbol{S}}_R = \underset{\boldsymbol{S} \in SO(3)}{\arg\min} \sum_{i=1}^{n} d_R^2(\boldsymbol{S}, \boldsymbol{R}_i)$	`mean(Rs, type = "geometric")`
Geometric Median	$\widetilde{\boldsymbol{S}}_R = \underset{\boldsymbol{S} \in SO(3)}{\arg\min} \sum_{i=1}^{n} d_R(\boldsymbol{S}, \boldsymbol{R}_i)$	`median(Rs, type = "geometric")`

The estimators in Table 5.2 find estimates based on minimization of L_1- and L_2-norms in the chosen geometry. The function `gradient.search` provides the option to optimize for

any other arbitrary minimization criterion. As the name suggests, the minimization is done along the gradient of the minimization function in the rotation space. Starting from an initial, user-specified rotation, the algorithm finds a (local) minimum by stepping iteratively in the direction of the steepest descent. Step size is regulated internally by adjusting for curvature of the minimization function.

We highlight this process in the example below. The function `L1.error` is defined to minimize the intrinsic L_1-norm, the result from the optimization should therefore agree with the geometric median of the sample. In fact, the difference between the two results is at the same level as the minimal difference (`minerr`) used for convergence of the gradient search. What is gained in flexibility of the optimization is, of course, paid for in terms of speed: the built-in median function is faster by far than the gradient search.

Also illustrated in the example below is the `rot.dist` function, which computes the distance between two objects of class `"SO3"`, e.g. `R1` and `R2`. The argument `method` specifies which type of distance to compute: the `"extrinsic"` option will return the Euclidean distance and the `"intrinsic"` option will return the Riemannian distance. If `R1` is an $n \times 9$ matrix representing a sample of rotations, then `rot.dist` will return a vector of length n where the ith element represents the specified distance between `R2` and the ith row of `R1`.

```
> # error function definition
> L1.error <- function(sample, Shat) {
+    sum(rot.dist(sample, Shat, method = "intrinsic", p = 1))
+ }
> cayley.sample <- ruars(n = 50, rangle = rcayley, nu = 1, space = "SO3")
> # gradient based optimization
> system.time(SL1 <- gradient.search(cayley.sample, L1.error))
   user  system elapsed
  3.464   0.007   3.473
> # in-built function
> system.time(S <- median(cayley.sample, type = "geometric"))
```

```
   user  system elapsed
  0.004   0.000   0.005
> rot.dist(S, SL1$Shat)
[1] 1.492e-05
```

5.4.2 Confidence Regions

Asymptotic results for the distribution of the projected mean $\widehat{\boldsymbol{S}}_n$ and median $\widetilde{\boldsymbol{S}}_n$ can be used to construct confidence regions for the central orientation $\boldsymbol{S}$. In the literature two approaches are available to justify the limiting distribution of the vector in $\mathbb{R}^3$ associated with the centered estimator through (5.2). More specifically, the vector $\sqrt{n}\widehat{\boldsymbol{h}}$ has been shown to have a trivariate normal distribution where $\widehat{\boldsymbol{h}} \in \mathbb{R}^3$ satisfies

$$\exp\left[\boldsymbol{\Phi}\left(\widehat{\boldsymbol{h}}\right)\right] = \boldsymbol{S}^\top \widehat{\boldsymbol{S}}_n.$$

The first approach transforms a result from directional statistics while the second uses M-estimation theory in $SO(3)$ directly. A summary of these methods is given next.

In the context of directional statistics, Prentice (1984) used results found in Tyler (1981) and the fact that $\widehat{\boldsymbol{S}}_n$ is a function of the spectral decomposition of $\overline{\boldsymbol{R}} = \sum_{i=1}^n \boldsymbol{R}_i/n$ in order to justify a multivariate normal limiting distribution for the scaled vector $\sqrt{n}\,\widehat{\boldsymbol{h}}$. Unsatisfied with the coverage rate achieved by Prentice (1986), Fisher et al. (1996) proposed a pivotal bootstrap procedure that results in coverage rates closer to the nominal level for small samples. A transformation from unit vectors in $\mathbb{R}^d$ to rotation matrices is required in order to apply the results of Prentice (1984) and Fisher et al. (1996) to $SO(3)$, therefore they are called *transformation-based*. The projected median $\widetilde{\boldsymbol{S}}_n$ cannot be expressed as a function of the sample spectral decomposition, therefore this approach cannot be used to create confidence regions based on $\widetilde{\boldsymbol{S}}_n$.

It has also been shown that both estimators $\widehat{\boldsymbol{S}}_n$ and $\widetilde{\boldsymbol{S}}_n$ are M-estimators, which motivates a direct approach to confidence region estimation in $SO(3)$ (Chang and Rivest, 2001). In Stanfill et al. (2014b), a pivotal bootstrap method based on the direct approach was implemented to improve coverage rates in small samples. Because the results in Chang and Rivest (2001) and

Stanfill et al. (2014b) deal with $SO(3)$ data directly, this approach is called *direct*.

The six possible confidence regions that result from these two methods are available through the wrapper function `region`. They are differentiated based on the `method`, `type` and `estimator` arguments. Set `estimator = "mean"` or `estimator = "median"` to estimate a region based on $\widehat{\boldsymbol{S}}_n$ or $\widetilde{\boldsymbol{S}}_n$, respectively. For $\widehat{\boldsymbol{S}}_n$ one can choose `method = "transformation"` for the transformation-based methods or `method = "direct"` for the direct method. Since the transformation-based methods cannot be applied to $\widetilde{\boldsymbol{S}}_n$ an error is returned if `estimator = "median"` and `method = "transformation"` are combined. A bootstrap version of the specified method is implemented if `type = "bootstrap"` or the normal limiting distribution can be chosen with `type = "asymptotic"`. If a bootstrap type region is specified one can additionally specify the bootstrap sample size with the `m` argument, which is set to 300 by default. Regardless of the method and type chosen a single value is returned on the interval $(0, \pi]$. This value corresponds to the radius of the confidence region centered at each of the axes of the specified estimator.

In the example code below a sample of $n = 50$ rotations are drawn from the Cayley-UARS($\boldsymbol{I}_{3\times3}, \kappa = 10$) distribution then the four types of confidence regions based on the direct approach are constructed. For a graphical representation of this dataset along with an interpretation of the confidence regions see Figure 5.1b.

```
> Rs <- ruars(50, rcayley, kappa = 10)
> region(Rs, method = "direct", type = "asymptotic",
+        estimator = "mean", alp = 0.05)
[1] 0.189
> region(Rs, method = "direct", type = "bootstrap", estimator = "mean",
+        alp = 0.05, m = 300)
[1] 0.201
> region(Rs, method = "direct", type = "asymptotic",
+        estimator = "median", alp = 0.05)
[1] 0.201
> region(Rs, method = "direct", type = "bootstrap", estimator = "median",
```

```
+        alp = 0.05, m = 300)
[1] 0.249
```

5.5 Visualizations

The **rotations** package offers two methods to visualize rotation data in three-dimensions. Because rotation matrices are orthogonal, each column of a rotation matrix has length one and is perpendicular to the other axes. Therefore each column of a rotation matrix can be illustrated as a point on the surface of a unit sphere, which represents the position of the x-, y- or z-axis for that rotation matrix. Since each sphere represents one of the three axes, three spheres are required to fully visualize a sample of rotations. Though the use of separate spheres to represent each axis can be seen as a disadvantage, the proposed visualization method makes the idea of a central orientation and a confidence region interpretable.

An existing function that can be used to illustrate rotation data is the `boat3d` function included in the **orientlib** package. Given a sample of rotations, the `boat3d` function produces either a static or interactive three-dimensional boat to represent the provided data. If only one rotation is of interest, the `boat3d` function is superior to the proposed method because it conveniently illustrates rotational data in a single image. If multiple rotations are provided, however, the `boat3d` function will produce separate side-by-side boats, which can be hard to interpret. In addition, the illustration of a estimated central orientation or a confidence region in $SO(3)$ with the `boat3d` function is not presently possible.

The **rotations** package can be used to produce high-quality static plots within the framework of the **ggplot2** package (Wickham, 2009). Static plots are specifically designed for datasets that are highly concentrated and for use in presentations or publications. Alternatively, the **rotations** package can produce interactive plots using functions included in the **sphereplot** package (Robotham, 2013). Interactive plots are designed so that the user can explore a dataset and visualize a diffuse sample.

Calling the `plot` function with a `"SO3"` or `"Q4"` object will result in an interactive or static sphere, differentiated by setting the argument `interactive` to `TRUE` or `FALSE`, respectively. The `center` argument defines the center of the plot and is usually set to the identity rotation

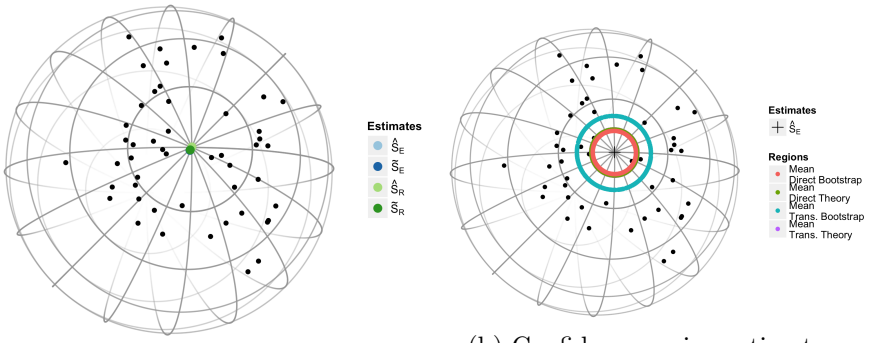
`id.SO3` or an estimate of the central orientation, e.g. `mean(Rs)`. The user can specify which columns to visualize with the `col` argument with options 1, 2 and 3 representing the x-, y- and z- axes, respectively. For static plots, multiple axes can be displayed simultaneously by supplying a vector to `col`; only one column will be displayed at a time for interactive plots. Also available to static plots is the argument `to_range`, which when set to `TRUE` will display the portion of the sphere where the observations are present.

All four estimates of the central orientation can be plotted along with a sample of rotations. Setting the argument `estimates_show = "all"` will display all four simultaneously. If only a few estimates are of interest then any combination of `"proj.mean"`, `"proj.median"`, `"geom.mean"` or `"geom.median"` are valid inputs. The estimators are indicated by color and a legend is provided, see Figure 5.1a. Finally, the `mean_regions` and `median_regions` options allow the user to draw a circle on the surface of the sphere representing the confidence region for that axis, centered at $\widehat{\boldsymbol{S}}_n$ and $\widetilde{\boldsymbol{S}}_n$ respectively. If estimators are plotted along with the different regions in static plots then shapes represent the estimators and colors represent the region methods, see Figure 5.1b, while regions and estimators are always distinguished by colors for the interactive plots. Given the sample of rotations generated in a previous example, the example below illustrates how to produce static plots using the `plot` function for objects of class `"SO3"` and Figure 5.2 illustrates the results of these commands.

```
> plot(Rs, center = mean(Rs), col = 1, show_estimates = "all",
+      interactive = FALSE)
> plot(Rs, center = mean(Rs), col = 1, show_estimates = "proj.mean",
+      mean_regions = "all", alp = .05, interactive = FALSE)
```

5.6 Datasets

Datasets `drill` and `nickel` are included in the **rotations** package to illustrate how the two representations of orientation data discussed here are used in practice. The `drill` dataset was collected to assess variation in human movement while performing a task (Rancourt, 1995). Eight subjects drilled into a metal plate while being monitored by infrared cameras. Quater-



(a) Point estimates

(b) Confidence region estimates

Figure 5.2: The x-axis of a random sample from the Cayley-UARS distribution with $\kappa = 1$, $n = 50$. All for point estimates are displayed in (a) and all three region methods along with the projected mean are in (b).

nions are used to represent the orientation of each subjects' wrist, elbow and shoulder in one of six positions. For some subjects several replicates are available. See Rancourt et al. (2000) for one approach to analyzing these data. In the example below we load the `drill` dataset, coerce the observations for subject one's wrist into a form usable by the **rotations** package via `as.Q4`, then estimate the central orientation with the projected mean.

```
> data(drill)
> head(drill)
  Subject Joint Position Replicate    Q1      Q2      Q3     Q4
1       1 Wrist        1         1 0.944  -0.192  -0.156  0.217
2       1 Wrist        1         2 0.974  -0.120  -0.111  0.158
3       1 Wrist        1         3 0.965  -0.133  -0.141  0.177
4       1 Wrist        1         4 0.956  -0.134  -0.115  0.233
5       1 Wrist        1         5 0.953  -0.199  -0.061  0.222
6       1 Wrist        2         1 0.963  -0.159  -0.127  0.177

> Subj1Wrist<-subset(drill, Subject == '1' & Joint == 'Wrist')
> Subj1Wdata <- as.Q4(Subj1Wrist[, 5:8])
> mean(Subj1Wdata)
```

```
0.987 - 0.070 * i - 0.134 * j + 0.049 * k
```

In the `nickel` dataset, rotation matrices are used to represent the orientation of cubic crystals on the surface of a nickel sample measured with Electron Backscatter Diffraction. Each `location` on the surface of the nickel is identified by the `xpos` and `ypos` columns while the `rep` column identifies which of the fourteen replicate scans that measurement corresponds to. The last nine columns, denoted `v1-v9`, represent the elements of the rotation matrix at that location in vector form. See Bingham et al. (2009, 2010a) and Stanfill et al. (2013) for more details. In the example below we estimate the central orientation at location one.

```
> data(nickel)
> head(nickel[, 1:6])
  xpos  ypos location rep     V1    V2
1    0 0.346        1   1 -0.648 0.686
2    0 0.346        1   2 -0.645 0.688
3    0 0.346        1   3 -0.645 0.688
4    0 0.346        1   4 -0.646 0.688
5    0 0.346        1   5 -0.646 0.686
6    0 0.346        1   6 -0.644 0.690
> Location1<-subset(nickel, location == 1)
> Loc1data<-as.SO3(Location1[, 5:13])
> mean(Loc1data)
       [,1]   [,2]   [,3]
[1,] -0.645 -0.286 -0.708
[2,]  0.687 -0.623 -0.374
[3,] -0.334 -0.728  0.599
```

5.7 Summary

In this manuscript we introduced the **rotations** package and demonstrated how it can be used to generate, analyze and visualize rotation data. The **rotations** package is compatible with

the quaternion specific **onion** package by applying its `as.quaternion` function to a transposed `"Q4"` object. Connecting to the **onion** package gives the user access to a wide range of algebraic functions unique to quaternions. Also compatible with the **rotations** package is the **orientlib** package, which includes additional parameterizations of rotations. To translate rotation matrices generated by the **rotations** package into a form usable by the **orientlib** package, first coerce a `"SO3"` object into a matrix of the same dimension, i.e. $n \times 9$, then apply the `rotvector` function provided by the **orientlib** package. Quaternions are defined in the **orientlib** package by $\boldsymbol{Q} = x_1 i + x_2 j + x_3 k + x_4$, cf. (5.5), which may lead to confusion when translating quaternions between the **orientlib** package and either of the **onion** or **rotations** packages. Below is a demonstration of how quaternions and rotation matrices generated by the **rotations** package can be translated into a form usable by the **onion** and **orientlib** packages, respectively. See `help(package = "onion")` and `help(package = "orientlib")` for more on these packages.

```
> Qs <- ruars(20, rcayley, space = 'Q4')
> Rs <- as.SO3(Qs)
> suppressMessages(require(onion))
> onionQs <- as.quaternion(t(Qs))
> suppressMessages(require(orientlib))
> orientRs <- rotvector(matrix(Rs, ncol = 9))
```

Computational speed of the **rotations** package has been enhanced through use of the **Rcpp** and **RcppArmadillo** packages (Eddelbuettel, 2013; Eddelbuettel and Sanderson, 2014). In future versions of the package we plan to extend the parameterization and estimator sections to include robust estimators currently being developed by the authors.

5.8 Acknowledgements

We would like to thank the reviewers for their comments and suggestions. The **rotations** package and this article have benefited greatly from their time and effort.

CHAPTER 6. GENERAL CONCLUSIONS

6.1 General Discussion

The analysis of object movement and orientation in three-dimensions has received an increase in attention, but the methods used to analyze these data are limited. In this dissertation we discussed point and confidence region estimation for the central orientation for the location model in $SO(3)$.

In Chapter 2 we explored a rage of literature from both approaches to estimating the central orientation in the $SO(3)$ location model. The existing intrinsic mean, intrinsic median and extrinsic mean are introduced and the existing methods for the computation in practice are reviewed. We introduce the extrinsic median for the first time and present an algorithm to compute it in practice. The results of our simulation study suggest the choice of extrinsic or intrinsic estimation depends upon the tail behavior of the data model. In general the extrinsic median is shown to perform well in all scenarios examined.Chapters 3 and 4 focus on the mean and median extrinsic estimators, respectively.

In Chapter 3, M-estimator theory is used to show the extrinsic mean is asymptotically normal. Though the asymptotic normality of the extrinsic mean is not novel, the method in which the limiting distribution is derived is and it is shown to lead to an estimator of the limiting variance that has better small sample properties in a simulation study. In addition, a pivotal bootstrap procedure is proposed that is proven to achieve the nominal coverage rate asymptotically. According to our simulation study, confidence regions based on the pivotal bootstrap have the closet to nominal coverage rates without being overly liberal. In a data example the confidence region methods proposed are used to verify a long held, though unconfirmed claim that ESBD data can be measure with 1° level of precision.

Though the extrinsic mean has many nice properties, it is also negatively influenced by extreme observations. As an alternative estimator, the extrinsic median is investigated in detail in Chapter 4. In particular, the median is shown to be a consistent estimator of the central orientation parameter, asymptotically normal and SB-robust for the family of Cayley and matrix Fisher distributions. Meanwhile the extrinsic mean is shown to not be SB-robust for the same family of distributions. The asymptotic normality of the extrinsic median is used to find robust confidence regions for the central orientation and the pivotal bootstrap region from Chapter 3 is extended to the median. In a simulation study, confidence regions based on the mean and median are computed for data generated from a contaminated distribution. As a function the contamination, the regions based on the mean increase in size dramatically while the coverage rate approaches zero. For the regions based on the median, however, the rate of increase in the region size is less and the coverage rates decrease at a slower rate as well. The EBSD data is revisited, this time we illustrate how uncertainty between scans can be greatly over estimated at locations on the boundary of two grains.

Finally, Chapter 5 details the `rotations` package which is an user-friendly collection of the code used in this dissertation. Also included are two datasets, the EBSD data analyzed throughout this dissertation and the drill dataset of Rancourt (1995). Tools for data generation, analysis and inference are available including both Bayesian and frequentist ideas. The visualization technique is also demonstrated.

6.2 Future Research

In the process of writing this dissertation questions arose that may prove fruitful for future research.

1. As demonstrated in the simulation study in Chapter 4, the confidence regions based on the median can have sizes larger than the $SO(3)$ space. The fact that a region can be larger than the entire parameter space is an artifact of extrinsic approach and could be seen as a reason to look for other region computation methods. That is, using theoretical results on the manifold itself could lead to more interpretable and always appropriate

confidence regions.

2. The assumption of rotational symmetry has been used in all parts of this dissertation. Allowing for non-symmetric distributions, such as the class of preferred axis-random spin distributions of Bingham et al. (2012), is a natural extension of this work. Though point estimation will likely be unchanged, the construction of confidence regions will need to be updated significantly.

3. No rigorous argument has been made as to when the extrinsic or intrinsic approach should be used for $SO(3)$ data analysis. Currently, the intrinsic approach is cite by its authors to be "more natural" and therefore appropriate. On the other hand, under several common distributions on $SO(3)$ the extrinsic approach leads to the maximum likelihood estimator which is desirable from a statistical point of view. We would like to take an approach similar to Chapter 2 and compare the four estimators from a theoretical standpoint. The goal of this project is to produce recommendations on when the extrinsic or intrinsic approach should be used.

4. A comparison of intrinsic and extrinsic estimators when extreme observations are present is also of interest. Towards this end, consider the extrinsic and intrinsic mean. From Chapter 4 the influence function of the extrinsic mean evaluated at the observation $\boldsymbol{R}_i \in SO(3)$ was shown to be proportional to the sine of the angle between $\boldsymbol{R}_i$ and the true central orientation $\boldsymbol{S}$. Therefore the extrinsic mean is most heavily influence by observations perpendicular (rotated through $\pi/2$ radians) to it and not influence at all by observations on the exact opposite pole. It can be shown, however, that the intrinsic mean's influence function is proportional to the value of the misorientation angle directly. Therefore, the observations furthest away from the intrinsic mean have the greatest impact on its performance. This observation can lead to recommendations on when the extrinsic approach should be used versus the intrinsic, but a rigorous study of this fundamental difference in estimators is warranted.

5. The influence functions proposed in Chapter 4 can be used to identify an influential point

in $SO(3)$ with respect to an estimator, but a formal definition of "outlier" in $SO(3)$ has yet to be proposed. For parameter space structures similar to $SO(3)$, such as the circle and sphere, the idea of an outlier has been explored, but their extension to $SO(3)$ data is non-trivial. Fletcher et al. (2009) considered outliers in $SO(3)$, though to produce outliers, observations were randomly rotated through $\pi/2$ radians with little justification.

6. Once an outlier in $SO(3)$ is identified, data analysis methods that can accommodate that outlier need to be developed. In Chapter 4 the extrinsic median was shown to be SB-robust, but it is also inefficient. The class of robust means, such as the trimmed and winsorized mean, could result in a more efficient estimator than the median that also accommodates for extreme observations. The multidimensional Huber estimator of Hampel et al. (2011) is also a promising.

BIBLIOGRAPHY

Agostinelli, C. and Lund, U. (2013). *circular: Circular Statistics*. R package version 0.4-7.

Arun, K., Huang, T., and Blostein, S. (1987). Least-squares fitting of two 3-D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700.

Bachmann, F., Hielscher, R., Jupp, P., Pantleon, W., Schaeben, H., and Wegert, E. (2010). Inferential statistics of electron backscatter diffraction data from within individual crystalline grains. *Journal of Applied Crystallography*, 43(6):1338–1355.

Bajaj, C. (1988). The algebraic degree of geometric optimization problems. *Discrete & Computational Geometry*, 3(1):177–191.

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics*, 31(1):1–29.

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *The Annals of Statistics*, 33(3):1225–1259.

Bingham, M. A., Lograsso, B. K., and Laabs, F. C. (2010a). A statistical analysis of the variation in measured crystal orientations obtained through electron backscatter diffraction. *Ultramicroscopy*, 110(10):1312–1319.

Bingham, M. A., Nordman, D. J., and Vardeman, S. B. (2009). Modeling and inference for measured crystal orientations and a tractable class of symmetric distributions for rotations in three dimensions. *Journal of the American Statistical Association*, 104(488):1385–1397.

Bingham, M. A., Nordman, D. J., and Vardeman, S. B. (2010b). Finite-sample investigation

of likelihood and Bayes inference for the symmetric von Mises–Fisher distribution. *Computational Statistics & Data Analysis*, 54(5):1317–1327.

Bingham, M. A., Nordman, D. J., and Vardeman, S. B. (2012). Bayes inference for a tractable new class of non-symmetric distributions for 3-dimensional rotations. *Journal of Agricultural, Biological, and Environmental Statistics*, 17(4):527–543.

Brown, B. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(1):25–30.

Bunge, H. (1982). *Texture Analysis in Material Science*. Butterworth, London.

Chan, Y. and He, X. (1993). On median-type estimators of direction for the von Mises-Fisher distribution. *Biometrika*, 80(4):869–875.

Chang, T. and Rivest, L.-P. (2001). M-estimation for location and regression parameters in group models: A case study using Stiefel manifolds. *The Annals of Statistics*, 29(3):784–814.

Chikuse, Y. (2003). *Statistics on Special Manifolds*. Springer Verlag.

Cho, J., Rollett, A., and Oh, K. (2005). Determination of a mean orientation in electron backscatter diffraction measurements. *Metallurgical and Materials Transactions A*, 36(12):3427–3438.

Dai, Y., Trumpf, J., Li, H., Barnes, N., and Hartley, R. (2010). Rotation averaging with application to camera-rig calibration. *Computer Vision–ACCV 2009*, pages 335–346.

Davis, A. (1977). Asymptotic theory for principal component analysis: non-normal case. *Australian Journal of Statistics*, 19(3):206–212.

Demirel, M. C., El-Dasher, B. S., Adams, B. L., and Rollett, A. D. (2000). Studies on the accuracy of electron backscatter diffraction measurements. *Electron Backscatter Diffraction in Materials Science*, pages 407–418.

Downs, T. (1972). Orientation statistics. *Biometrika*, 59(3):665–676.

Ducharme, G. and Milasevic, P. (1987). Spatial median and directional data. *Biometrika*, 74(1):212–215.

Durocher, S. and Kirkpatrick, D. (2009). The projection median of a set of points. *Computational Geometry*, 42(5):364–375.

Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer-Verlag, New York. ISBN 978-1-4614-6867-7.

Eddelbuettel, D. and Sanderson, C. (2014). Rcpparmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063.

Fisher, N. (1985). Spherical medians. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(2):342–348.

Fisher, N. I. (1996). *Statistical Analysis of Circular Data*. Cambridge University Press.

Fisher, N. I., Hall, P., Jing, B.-Y., and Wood, A. T. (1996). Improved pivotal methods for constructing confidence regions with directional data. *Journal of the American Statistical Association*, 91(435):1062–1070.

Fisher, R. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 217(1130):295–305.

Fletcher, P., Venkatasubramanian, S., and Joshi, S. (2008). Robust statistics on Riemannian manifolds via the geometric median. In *2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.

Fletcher, P., Venkatasubramanian, S., and Joshi, S. (2009). The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1):S143–S152.

Fletcher, P. T., Lu, C., and Joshi, S. (2003). Statistics of shape via principal geodesic analysis on lie groups. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 95–101. IEEE.

Gower, J. (1974). The mediancentre. *Applied Statistics*, 23(3):466–470.

Hadani, R. and Singer, A. (2011). Representation theoretic patterns in three-dimensional cryo-electron microscopy II–the class averaging problem. *Foundations of Computational Mathematics*, 11(5):589–616.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics*, 14(4):1453–1462.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2011). *Robust Statistics: The Approach Based on Influence Functions*. Wiley & Sons, New York.

Hankin, R. K. S. (2011). *onion: octonions and quaternions*. R package version 1.2-4.

Hartley, R., Aftab, K., and Trumpf, J. (2011). L1 rotation averaging using the Weiszfeld algorithm. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3041–3048. IEEE.

He, X. (1992). Robust statistics of directional data: a survey. *Nonparametric Statistics and Related Topics*, pages 87–96.

He, X. and Simpson, D. G. (1992). Robust direction estimation. *The Annals of Statistics*, 20(1):351–369.

Hielscher, R., Schaeben, H., and H., S. (2010). Orientation distribution within a single hematite crystal. *Mathematical Geosciences*, 42:359–375.

Horn, B., Hilden, H., and Negahdaripour, S. (1988). Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America*, 5(7):1127–1135.

Humbert, M., Gey, N., Muller, J., and Esling, C. (1996). Determination of a mean orientation from a cloud of orientations. Application to electron back-scattering pattern measurements. *Journal of Applied Crystallography*, 29(6):662–666.

Jupp, P. and Mardia, K. (1979). Maximum likelihood estimators for the matrix von Mises-Fisher and Bingham distributions. *The Annals of Statistics*, 7(3):599–606.

Jupp, P. and Mardia, K. (1989). A unified view of the theory of directional statistics. *International Statistical Review*, 57(3):261–294.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541.

Khatri, C. and Mardia, K. (1977). The von Mises-Fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):95–106.

Ko, D. and Chang, T. (1993). Robust M-estimators on spheres. *Journal of Multivariate Analysis*, 45(1):104–136.

Ko, D. and Guttorp, P. (1988). Robustness of estimators for directional data. *The Annals of Statistics*, 16(2):609–618.

Laha, A. K. and Mahesh, K. (2011). SB-robustness of directional mean for circular distributions. *Journal of Statistical Planning and Inference*, 141(3):1269–1276.

Langevin, P. (1905). Magnetism and the theory of the electron. *Annales de Chimie et de Physique*, 5:70.

Lenth, R. V. (1981). Robust measures of location for directional data. *Technometrics*, 23(1):77–81.

León, C., Massé, J., and Rivest, L. (2006). A statistical model for random rotations. *Journal of Multivariate Analysis*, 97(2):412–430.

Liu, R. Y., Singh, K., et al. (1992). Ordering directional data: Concepts of data depth on circles and spheres. *The Annals of Statistics*, 20(3):1468–1484.

Manton, J. (2004). A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups. In *8th Conference on Control, Automation, Robotics and Vision, (ICARCV)*, volume 3, pages 2211–2216. IEEE.

Mardia, K. (1972). *Statistics of Directional Data*. London: Academic Press.

Mardia, K. and Jupp, P. (2000). *Directional Statistics*. Wiley Chichester.

Matthies, S., Muller, J., and Vinel, G. (1988). On the normal distribution in the orientation space. *Textures and Microstructures*, 10(1):77–96.

Moakher, M. (2002). Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16.

Murdoch, D. (2003). orientlib: An R package for orientation data. *Journal of Statistical Software*, 8(19):1–11.

Oh, H. and Kim, D. (2013). *SpherWave: Spherical Wavelets and SW-based Spatially Adaptive Methods*. R package version 1.2.2.

Otieno, B. S. (2002). *An Alternative Estimate of Preferred Direction for Circular Data*. PhD thesis, Virginia Polytechnic Institute and State University.

Pierrynowski, M. and Ball, K. (2009). Oppugning the assumptions of spatial averaging of segment and joint orientations. *Journal of Biomechanics*, 42(3):375–378.

Preisig, P. and Kragic, D. (2006). Robust statistics for 3d object tracking. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, pages 2403–2408. IEEE.

Prentice, M. (1984). A distribution-free method of interval estimation for unsigned directional data. *Biometrika*, 71(1):147–154.

Prentice, M. (1986). Orientation statistics without parametric assumptions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):214–222.

Qiu, Y. (2013). *Isotropic Distributions for 3-Dimensional Rotations and One-sample Bayes Inference*. PhD thesis, Iowa State University.

Rancourt, D. (1995). *Arm Posture and Hand Mechanical Impedance in the Control of a Hand-held Power Drill*. Dissertation, MIT.

Rancourt, D., Rivest, L.-P., and Asselin, J. (2000). Using orientation statistics to investigate variations in human kinematics. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 49(1):81–94.

Randle, V. (2003). *Microtexture Determination and its Applications*. London: Maney for The Institute of Materials, Minerals and Mining.

Robotham, A. (2013). *sphereplot: Spherical Plotting*. R package version 1.5.

Savyolova, T. and Nikolayev, D. (1995). Normal distribution on the rotation group SO(3). *Textures and Microstructures*, 29(3):201–233.

Schaeben, H. (1997). A simple standard orientation density function: The hyperspherical de la Vallée Poussin kernel. *Physica Status Solidi (B)*, 200(2):367–376.

Schwartz, M. and Rozumalski, A. (2005). A new method for estimating joint parameters from motion data. *Journal of Biomechanics*, 38(1):107–116.

Shao, J. (2003). *Mathematical Statistics*. Springer Texts in Statistics. Springer.

Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review*, 58(3):263–277.

Stanfill, B., Genschel, U., and Hofmann, H. (2013). Point estimation of the central orientation of random rotations. *Technometrics*, 55(4):524–535.

Stanfill, B., Hofmann, H., and Genschel, U. (2014a). *rotations: Tools for Working with Rotation Data*. R package version 1.2.

Stanfill, B., Nordman, D., Hofmann, H., Genschel, U., and Zhang, J. (2014b). Nonparametric confidence regions for the central orientation of random rotations. Unpublished manuscript.

Tyler, D. E. (1981). Asymptotic inference for eigenvectors. *The Annals of Statistics*, 9(4):725–736.

Umeyama, S. (1991). Least squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:376–380.

Watson, G. S. (1983). *Statistics on Spheres*, volume 6.

Wehrly, T. E. and Shine, E. P. (1981). Influence curves of estimators for directional data. *Biometrika*, 68(1):334–335.

Weiszfeld, E. (1937). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematics Journal*, 43:355–386.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Wilson, A. and Spanos, G. (2001). Application of orientation imaging microscopy to study phase transformations in steels. *Materials Characterization*, 46(5):407–418.