Likelihood and Bayes inference for a class of distributions on orientations in 3 dimensions

Melissa Ann Bingham

Iowa State University
Likelihood and Bayes inference for a class of distributions on orientations in 3 dimensions

by

Melissa Ann Bingham

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:
Stephen Vardeman, Co-major Professor
Daniel Nordman, Co-major Professor
Alicia Carriquiry
Huaqing Wu
Frederick Lorenz

Iowa State University
Ames, Iowa
2009

Copyright © Melissa Ann Bingham, 2009. All rights reserved.
TABLE OF CONTENTS

ABSTRACT ............................................................................................................. v

GENERAL INTRODUCTION ................................................................................. 1
  1 Introduction ..................................................................................................... 1
  2 Dissertation Organization ............................................................................... 2

MODELING AND INFERENCE FOR MEASURED CRYSTAL ORIENTATIONS AND
A TRACTABLE CLASS OF SYMMETRIC DISTRIBUTIONS FOR ROTATIONS IN 3 DIMENSIONS ................................................................. 3

Abstract .............................................................................................................. 3
  1 Introduction ..................................................................................................... 4
  2 Uniform-Axis-Random-Spin distributions ...................................................... 11
    2.1 A concrete model for the generation of 3-dimensional orientation data .. 11
    2.2 The UARS(S, κ) density ........................................................................... 13
  3 One-sample quasi-likelihood and inference for the von Mises version of the UARS
distributions ......................................................................................................... 16
    3.1 Parameterization of S ............................................................................... 16
    3.2 Maximum quasi-likelihood estimation (MQL) .......................................... 18
    3.3 Asymptotic results ................................................................................... 19
    3.4 Confidence regions .................................................................................. 22
  4 Simulation study and asymptotic power comparison ........................................ 22
  5 Application of quasi-likelihood methodology for the von Mises version of the
UARS distributions to the crystal orientation data .............................................. 26
  6 Conclusion ...................................................................................................... 29
BAYES ONE-SAMPLE AND ONE-WAY RANDOM EFFECTS ANALYSES FOR 3-D ORIENTATIONS WITH APPLICATION TO MATERIALS SCIENCE

Abstract

1 Introduction

2 One-sample Bayes analyses for the vM-UARS(\mathbf{S}, \kappa) distribution

2.1 One-sample model

2.2 Simulation study

3 One-way random effects Bayes analyses for the vM-UARS(\mathbf{S}, \kappa) distribution

4 Application to EBSD data

5 Conclusion

References

FINITE-SAMPLE INVESTIGATION OF LIKELIHOOD AND BAYES INFERENDE FOR THE SYMMETRIC VON MISES-FISHER DISTRIBUTION

Abstract

1 Introduction

2 The symmetric matrix Fisher distribution

2.1 Model density

2.2 Likelihood function and large-sample distributional results
A STATISTICAL ANALYSIS OF VARIATION IN MEASURED CRYSTAL ORIENTATIONS OBTAINED THROUGH ELECTRON BACKSCATTER DIFFRACTION

Abstract

1 Introduction

2 EBSD background

3 The UARS class of distributions

4 Bayes analyses for a hierarchical model with vM-UARS components

5 Application to EBSD data

6 Conclusion

References

GENERAL CONCLUSION

ACKNOWLEDGEMENTS
This dissertation concerns variation in 3-dimensional orientations (represented by $3 \times 3$ orthogonal matrices with positive determinant). The Uniform-Axis-Random-Spin (UARS) class of distributions for 3-dimensional orientations is identified and likelihood and Bayes inference for the class are developed, resulting in new practical statistical methods for modeling orientation data. Two members of the UARS class, the von Mises version and the symmetric matrix von Mises-Fisher distribution, are considered in detail. The methodology developed is used in materials science and human kinematics applications.
GENERAL INTRODUCTION

1 Introduction

Orientation data arise in many different areas, and while modeling random orientations has received attention in the statistical literature, the methods studied often result from general “special manifold” considerations and may be hard to use in practice. Many existing distributions for orientations in 3 dimensions also have limitations including parameters that are not easily interpreted, complicated inference, and difficulty of simulation. Further, the existing distributions do not provide flexibility in modeling and may not be suitable for all data (as in the case of the crystal orientation data explored throughout much of this dissertation).

This dissertation identifies and develops inference methods for a useful class of distributions on orientations in 3 dimensions (as represented by $3 \times 3$ orthogonal matrices with positive determinant) that overcomes many of the shortcomings of existing tools. A direct modeling approach is taken, beginning with a simple mechanism for generating random 3-dimensional orientations (as opposed to beginning from more general considerations). The resulting Uniform-Axis-Random-Spin (UARS) class has directly interpretable parameters, relatively simple theory, and provides needed flexibility in modeling 3-dimensional rotations. Likelihood and Bayes inference for the UARS class are considered, producing new practically useful statistical methodology. Particular attention is paid to two members of the UARS class, refereed to in this dissertation as the von Mises version and the symmetric matrix von Mises-Fisher distribution, and applications are made to materials science and human kinematics problems.
2 Dissertation Organization

This dissertation is organized as a collection of journal-submission-ready papers. The first paper, “Modeling and Inference for Measured Crystal Orientations and a Tractable Class of Symmetric Distributions for Rotations in 3 Dimensions,” introduces the UARS class and examines quasi-likelihood inference for the von Mises version of the UARS distributions (vM-UARS). The second paper, “Bayes One-Sample and One-Way Random Effects Analyses for 3-D Orientations with Application to Materials Science,” examines Bayes methods for the vM-UARS distribution in the one-sample case and makes comparisons to the quasi-likelihood inference explored in the first paper. The Bayes methods are extended to cover one-way random effects scenarios. Both papers include applications to a materials science problem. In the third paper, “Likelihood and Bayes Inference for the Symmetric von Mises-Fisher Distribution,” the symmetric von Mises-Fisher distribution is treated as a member of the UARS class. Likelihood and Bayes inference are developed and application is made to a human kinematics problem. The last paper, “A Bayes Statistical Analysis of the Variation in Crystal Orientations Obtained through Electron Backscatter Diffraction,” further extends the Bayes analyses for the vM-UARS distribution by developing Bayes analyses for a hierarchical model that provides important measurement precision assessments in a materials science problem.
MODELING AND INference FOR MEASURED CRYSTAL ORIENTATIONS AND A TRACTABLE CLASS OF SYMMETRIC DISTRIBUTIONS FOR ROTATIONS IN 3 DIMENSIONS

A paper submitted to Journal of the American Statistical Association

Melissa A. Bingham¹, Daniel J. Nordman, and Stephen B. Vardeman

Iowa State University

Abstract

Electron Backscatter Diffraction (EBSD) is a technique used in materials science to study the microtexture of metals, producing data which measure orientations of crystals in a specimen. We examine the precision of such data based on a useful class of distributions on orientations in 3 dimensions (as represented by $3 \times 3$ orthogonal matrices with positive determinant). While such modeling has received attention in the statistical literature, the approach taken has typically been from general “special manifold” considerations and the resulting methodology may not be easily accessible to non-specialists. We take a more direct modeling approach, beginning from a simple intuitively appealing mechanism for generating random orientations specifically in 3-space. The resulting class of distributions has many desirable properties, including directly interpretable parameters and relatively simple theory. We investigate the basic properties of the entire class and one-sample quasi-likelihood-based inference for one member

¹Supported by NSF grant DMS #0502347 EMSW21-RTG awarded to the Department of Statistics, Iowa State University, and by the Ames Laboratory through U.S. Department of Justice COPS Program grant #2005CKWX0466 and interagency agreement #2002-LP-R-083 by the National Institute of Justice, through the Midwest Forensics Resource Center. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under contract #DE-AC02-07CH11358.
of the model class, producing new statistical methodology that proves practically useful in the analysis of EBSD data.

Key words: Directional data, Electron Backscatter Diffraction, Euler angles, Haar measure, von Mises distribution, orthogonal matrix, quasi-likelihood ratio test, UARS distribution, Wald test

1 Introduction

Our work is motivated by a materials science application that involves quantifying the precision of Electron Backscatter Diffraction (EBSD) data produced in studies of the microtexture of metals. Using a Scanning Electron Microscope (SEM), EBSD data result when a stationary beam of electrons is diffracted by atomic lattice planes in a target metal, creating an image called a Kikuchi diffraction pattern on a focal plane of sensors. These collected EBSD patterns are then matched to theoretical patterns based on the known crystal structure of the metal and machine geometry, which in turn indicates the orientation of cubic crystals in the metal (because different orientations produce angular differences in band location, band intensity, and band width in the diffraction pattern); see Randle (2003).

In the application motivating our work, a TSL MSC-2200 (EBSD) camera and Field Emission AmRay 1845 SEM are used along with OIM Version 4.2 Analytical Suite software to produce diffraction patterns and then fitted crystal orientations relative to some reference coordinate system (defined, for example, by the geometry of the machine or the macro geometry of a metal sample being studied) at points scanned on a 2-dimensional regular grid laid over a “flat” surface of a metal specimen. Data for two metals will be examined here. The first metal, high-iron-concentration nickel, was chosen for study because it produces ideal EBSD patterns. The nickel specimen had a surface area $40 \mu m \times 40 \mu m$ and the same region was scanned 14 times with at least 4000 measurements per run. Each observation represents the orientation of a cubic crystal at some scanned position on the metal surface (and can be described by a $3 \times 3$ orthogonal rotation matrix). Seven scans taken on a fixed region of an aluminum sample will also be considered, although in less detail.
Since EBSD is a commonly used methodology, it is important to investigate the consistency of measurements it produces. Although the issue of EBSD measurement precision has received some attention in the materials science literature, methods of data collection and inference for experimental precision are not yet widely agreed upon. Often referenced as a source for EBSD precision, Demirel, El-Dasher, Adams, and Rollet (2000) examine variation in orientation measurements from a single scan on a silicon single crystal wafer in terms of the spread in “misorientation angles” between an unspecified reference coordinate system and measured crystal orientations. (For two orientations specified by $3 \times 3$ orthogonal rotation matrices $M_1$ and $M_2$, the misorientation axis and angle are such that when $M_2$ is rotated around the misorientation axis by the (positive) misorientation angle, $M_1$ results (see Randle, 2003, sec. 6.4.1).) The misorientation angles of Demirel et al. (2000) are then centered by subtraction of their mean, resulting in a spread of roughly $-1^\circ$ to $1^\circ$, and a “$1^\circ$” precision is reported. A more natural analysis might be based on misorientation angles from a “central orientation.” Wilson and Spanos (2001) used this kind of approach with observations from a single scan on a single crystal gallium arsenide sample and found misorientation angles from an orientation chosen to “minimize average misorientation” to be roughly between $0^\circ$ and $1^\circ$. These authors then somewhat inexplicably report a “$0.5^\circ$” precision based on these angles. In the event that measured orientations are fairly consistent and substantially different from the orientation of the reference coordinate system, the spread in absolute values of angles produced by the somewhat ad hoc methods of Demirel et al. (2000) can be expected to be roughly comparable to the spread in (inherently positive) misorientation angles produced using the more natural approach of Wilson and Spanos (2001). With that in mind, based on both the silicon work of Demirel et al. (2000) and the gallium arsenide study of Wilson and Spanos (2001), a “$1^\circ$ misorientation angle precision” is representative of the commonly held perception of EBSD in materials science.

The precision estimates available in the current literature are based on single scans of single crystals/grains (i.e. scans at neighboring locations on homogeneous specimens). Our intention here is to begin development of methods that will eventually allow us to coherently
quantify multiple components of variation in more complicated cases, where multiple scans are taken on the same specimen and specimens are potentially composed of multiple crystals. Of first present interest is the basic single-site repeatability of data obtained through EBSD (a matter that seems to be thus far unaddressed in the materials science literature) — we wish to examine the variation in orientation measurements obtained when EBSD is used repeatedly at a single location on the same metal specimen. Then, quantifying the variation of measured orientations thought to be from within a single grain (in the style of existing work) is a second problem of subject matter interest. Probability models for rotation matrices are potentially useful for describing the variation in scanning results and quantifying the nature of within-grain variation.

In the statistical literature, the most commonly referenced distribution for rotation matrices is the matrix (von Mises) Fisher distribution (sometimes referred to as the Langevin distribution) introduced by Downs (1972). Important advances have since been made by Khatri and Mardia (1977) and Jupp and Mardia (1979), with further work done by Prentice (1986), Mardia and Jupp (1999), Rancourt, Rivest, and Asselin (2000), Chikuse (2003), and Rivest, Baillargeon, and Pierrynowski (2008). But practical limitations remain. Parameters of the distribution are not easily interpreted, inference is not simple, and simulation from this distribution is not obvious. These considerations motivate León, Massé, and Rivest (2006) to introduce the so-called Cayley distributions. Although the Cayley distributions provide improvements over the matrix Fisher distribution in these areas, some of these problems (particularly of interpretability of parameters and ease of use of existing methodology) remain substantial. What is more, much of the EBSD data we have seen are not well-described by these existing models. We have therefore found it useful to abstract a property known to be shared by many symmetric distributions on 3-dimensional orientations and to use this as the basis for defining a very broad class of models for rotation matrices with

i) an intuitively appealing constructive definition, and

ii) directly interpretable parameters

that facilitates the development and application of statistical methods. Here we will provide
some general development and show how a newly identified element of this model class can be used to address the instrument repeatability problem in materials science.

The following is a loose initial description of the modeling idea for rotation matrices (more geometric details can be found in Section 2.1). For $\Omega$ the set of $3 \times 3$ orthogonal matrices with positive determinant (and that thus preserve the right hand rule) and $S \in \Omega$ representing some central location or principal direction, view a random orientation $O \in \Omega$ as of the form $O = SM$ for a random perturbation $M \in \Omega$ of the $3 \times 3$ identity matrix. As it turns out, for the basic symmetric matrix Fisher and Cayley distributions for $O$, the matrix $M \equiv M(U, r)$ can be thought of as arising from a random rotation of the 3-dimensional axes through angle $r \in (-\pi, \pi]$ about a random axis prescribed by unit vector $U$ uniformly distributed over the unit sphere in 3-space. The variable $|r|$ corresponds to the materials science literature’s “misorientation angle” between $S$ and $O$ (see Randle, 2003). The angle $r$ is distributed independently of the random axis and has a marginal distribution symmetric on $(-\pi, \pi]$ which is indirectly and restrictively inherited from the matrix Fisher or Cayley model directly specified for $O$ (see León et al., 2006, sec. 5.2). It seems however, that more flexible modeling of orientation data might grow out of allowing other distributions for the angle $r$.

Figure 1 is a plot of theoretical quantiles for the distribution of $|r|$ versus empirical $|\hat{r}|$ quantiles after fitting the matrix Fisher and Cayley models to a small sample of EBSD data representing multiple measurements at the same location on the nickel specimen. Figure 1 suggests that the fits to the distributions from the existing literature are not good and that other forms for the distribution of $r$ might profitably be explored. For example, by selecting a von Mises circular distribution for $r$ (symmetric about 0) and plotting resulting quantiles of $|r|$ versus empirical quantiles of $|\hat{r}|$ for the same data represented in Figure 1, one obtains the substantially more pleasant Figure 2.

In this paper, we consider two small data sets representing a part of the nickel EBSD data. The first consists of the fourteen repeat measurements from one location on the scans represented in Figures 1 and 2 and will be used to study the repeatability of the EBSD data. The second data set, consisting of measurements for 70 different locations on a single scan
Figure 1: Q-Q plots for the fitted misorientation angles obtained from fourteen repeat EBSD observations, using the matrix Fisher (left) and Cayley (right) models on $\Omega$.

Figure 2: Q-Q plot for the fitted misorientation angles obtained from fourteen repeat EBSD observations, using the von Mises circular distribution on $r$. 

that all appeared to be within the same grain, will be used to study the within grain variation.

Table 1 presents the fourteen repeat measurements, represented in Euler angle form. (For more on the Euler angle representation of a $3 \times 3$ orthogonal rotation matrix, see Section 3.1.) In Section 5, two small data sets from the aluminum scans will also be considered. Seven repeat measurements from the same location on the scans and 50 locations appearing to be within a single grain on one scan will be considered.

Table 1: Repeat EBSD data for a single location on a nickel specimen (in Euler angle form)

<table>
<thead>
<tr>
<th>Observation</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.857001</td>
<td>0.9280639</td>
<td>4.220050</td>
</tr>
<tr>
<td>2</td>
<td>5.862054</td>
<td>0.9357848</td>
<td>4.216083</td>
</tr>
<tr>
<td>3</td>
<td>5.861929</td>
<td>0.9377740</td>
<td>4.219057</td>
</tr>
<tr>
<td>4</td>
<td>5.864537</td>
<td>0.9373403</td>
<td>4.215629</td>
</tr>
<tr>
<td>5</td>
<td>5.861906</td>
<td>0.9340686</td>
<td>4.217158</td>
</tr>
<tr>
<td>6</td>
<td>5.861424</td>
<td>0.9357253</td>
<td>4.217151</td>
</tr>
<tr>
<td>7</td>
<td>5.862159</td>
<td>0.9369018</td>
<td>4.215104</td>
</tr>
<tr>
<td>8</td>
<td>5.866320</td>
<td>0.9340180</td>
<td>4.216715</td>
</tr>
<tr>
<td>9</td>
<td>5.861765</td>
<td>0.9358679</td>
<td>4.216700</td>
</tr>
<tr>
<td>10</td>
<td>5.856148</td>
<td>0.9347951</td>
<td>4.221809</td>
</tr>
<tr>
<td>11</td>
<td>5.866616</td>
<td>0.9342294</td>
<td>4.212785</td>
</tr>
<tr>
<td>12</td>
<td>5.860286</td>
<td>0.9363732</td>
<td>4.217755</td>
</tr>
<tr>
<td>13</td>
<td>5.862718</td>
<td>0.9353843</td>
<td>4.216685</td>
</tr>
<tr>
<td>14</td>
<td>5.862907</td>
<td>0.9358418</td>
<td>4.215240</td>
</tr>
</tbody>
</table>

To be more precise about what is presented in Figures 1 and 2 and to prepare for a parallel treatment of the second nickel data set, we begin by defining an estimate of the principal orientation $\mathbf{S}$ for a sample. Suppose that $\overline{o} = \sum_{i=1}^{n} o_i / n$ for data orientations $o_1, \ldots, o_n \in \Omega$. The value of $\mathbf{S}$ that maximizes $\text{tr}(\mathbf{S}^T \overline{o})$ (for $\text{tr}(\mathbf{A})$ the trace of $\mathbf{A}$) is commonly used as an estimate for $\mathbf{S}$. It is the moment estimator for the modal rotation of the Cayley distribution in 3 dimensions (León et al., 2006, p. 421), and is the mean direction for the matrix Fisher distribution in 3 dimensions (Khatri and Mardia, 1977, p. 96).

Once $\mathbf{S}$ has been estimated, we find the misorientation angle $|r|$ required to obtain each observation from the fitted $\mathbf{S}$. We can then fit a choice of circular distribution for $r \in (-\pi, \pi]$ to the sample misorientation angles using maximum likelihood. For each data set, the von Mises circular distribution (see Mardia and Jupp, 2000) and the distributions on $r$ that give
the Cayley and symmetric matrix Fisher distributions were fit. Figures 1 and 2 gave the sample quantiles plotted against the theoretical quantiles for the first nickel data set (repeat scans at a single location). Figure 3 contrasts the fits for the between-location data.

Figure 3: Q-Q plots for the misorientation angles $|r|$ obtained from the 70 observations within a single grain of the nickel specimen, for fitted (a) von Mises (b) matrix Fisher and (c) Cayley distributions

The EBSD nickel data sets show that there are cases where previously developed models are not appropriate and that there is a need for greater flexibility in modeling 3-dimensional rotations. In the next section, we more formally develop what we will call the UARS class of distributions introduced above. Most existing work in this area begins by defining distributions on somewhat abstract “special manifolds” and develops results for 3-dimensional rotations as special cases. Our approach is more direct. We begin from a very concrete and easily described mechanism specifically for generating random rotations in 3-space. After defining the UARS class of distributions, we examine its properties and those of quasi-likelihood-based inference for one member of the model class, where the angle $r$ is assumed to follow a von Mises circular distribution. We then apply our results to the EBSD data. In addition to the nickel data (Figures 2 and 3), the von Mises circular distribution also adequately describes the misorientation angles obtained from two small aluminum data sets; see the supplementary on-line Appendix for Q-Q plots.
2 Uniform-Axis-Random-Spin distributions

2.1 A concrete model for the generation of 3-dimensional orientation data

To study the variation in a set of “measurements” belonging to Ω, we begin by supposing that \( o_i \in \Omega \) for \( i = 1, \ldots, n \) are data like those giving fitted orientations of cubic crystals known to share a common principal orientation \( S \in \Omega \). We wish to model the deviations of \( o_1, o_2, \ldots, o_n \) from \( S \).

We begin with a physical description of a model of data generation for the case \( S = I_{3 \times 3} \) (the \( 3 \times 3 \) identity matrix). To create a random rotation \( O \in \Omega \) we first generate a point uniformly on the unit sphere. We can represent this point in terms of polar coordinates \( \theta \in [0, \pi] \) and \( \phi \in [0, 2\pi] \) as the unit vector

\[
U = (u, v, w)^T = \left( \sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta \right)^T.
\]  

(1)

Secondly, we independently generate an angle \( r \in (-\pi, \pi] \) from a circular distribution symmetric about 0. In order to allow the spread of this circular distribution to depend on a parameter, we will write \( r \sim \text{Circ}(\kappa) \), where the \( \text{Circ}(\kappa) \) distribution has density \( C(r|\kappa) \). The last step is to rotate all of 3-space (including the unit sphere and the vectors \((1,0,0), (0,1,0), \) and \( (0,0,1) \) specifying the original coordinate axes) about the vector in (1) by the angle \( r \), giving \( O \in \Omega \) with columns that specify the rotated locations of \((1,0,0), (0,1,0), \) and \( (0,0,1) \) respectively.

An observation \( O \) generated as above can be expressed in terms of the elements of \( U \) in (1) and \( r \) as \( O = M(U,r) \), where

\[
M(U,r) \equiv UU^T + (I_{3 \times 3} - UU^T) \cos r + \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \sin r.
\]  

(2)

In this model, small \( |r| \) produces an observation \( O \) that differs less from \( I_{3 \times 3} \) than will one with large \( |r| \). (For example, consider the extreme case \( r = 0 \) which will produce an observation \( O = I_{3 \times 3} \) regardless of \( U \).) Therefore, the parameter \( \kappa \) of the circular distribution governs the “spread” of orientation data generated according to this model. So we will say that for \( U \)
uniform on the sphere independent of \( r \sim \text{Circ}(\kappa) \), \( \mathbf{O} \) as in (2) has a Uniform-Axis-Random-Spin distribution with parameters \( \mathbf{I}_{3 \times 3} \) and \( \kappa \), i.e. \( \mathbf{O} \sim \text{UARS}(\mathbf{I}_{3 \times 3}, \kappa) \).

Such distributions for \( \mathbf{O} \) can alternatively be expressed in terms of quaternions. If \( \mathbf{x} = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \) is a unit vector, then there is a mapping \( \mu : \mathbb{R}^4 \to \Omega \) with \( \mathbf{O} = \mu(\mathbf{x}) \) such that there is a one-to-one correspondence between \( \mathbf{O} \) and \( \pm\mathbf{x} \); see Prentice (1986) who uses axial distributions on the quaternions to define distributions on \( \Omega \). The quaternion representation of \( \mathbf{O} \) as in (2) is \( \mathbf{x} = (\sin(r/2)u, \sin(r/2)v, \sin(r/2)w, \cos(r/2)) \). In this form, a rotationally symmetric distribution for \( \mathbf{x} \) about the axis \((0, 0, 0, 1)\), as defined by Watson (1983, sec. 3.4), is possible through a distribution on \( \cos(r/2) \). The UARS distributions can thus also be viewed as a family of rotationally symmetric distributions on the quaternions. While this representation in terms of quaternions is elegant, we have not found it to simplify any of our analyses. The rotation-axis-and-angle motivation and representation that we use here both has an appealing concrete interpretation and leads to easily understood and effective calculations. Further, the concept of a misorientation angle \( |r| \) and misorientation axis \( \mathbf{U} \) is familiar to materials scientists and thus has a connection to our motivating application.

A UARS(\( \mathbf{I}_{3 \times 3}, \kappa \)) distribution describes “directionally symmetric” perturbations of \( \mathbf{I}_{3 \times 3} \). Consider now modeling directionally symmetric perturbations of some other orientation, described by the orthogonal matrix \( \mathbf{S} \). (This will allow modeling of measured crystal orientations where the true orientation is \( \mathbf{S} \), described by a principal orientation of the distribution at \( \mathbf{S} \).)

With \( \mathbf{U} \) uniform on the sphere independent of \( r \sim \text{Circ}(\kappa) \) the distribution of

\[
\mathbf{O} = \mathbf{S} \cdot M(\mathbf{U}, r)
\]

has principal direction at \( \mathbf{S} \). We will call the distribution of \( \mathbf{O} \) constructed as in (3) a Uniform-Axis-Random-Spin distribution with parameters \( \mathbf{S} \in \Omega \) and \( \kappa \) and write \( \mathbf{O} \sim \text{UARS}(\mathbf{S}, \kappa) \). This two-component parameterization is natural and useful in interpreting the effects of variability specified by \( \kappa \) and central location at \( \mathbf{S} \) in modeling. We have also enumerated some nice probabilistic (symmetry) properties of these models in the Appendix, such as the fact that (3) is distributionally equivalent to the definition \( \mathbf{O} = M(\mathbf{U}, r) \cdot \mathbf{S} \).

As mentioned in Section 1, distributions for matrix data studied in the existing literature,
like the Fisher and Cayley distributions, have often grown out of density derivations beginning with manifolds (see Khatri and Mardia, 1977), and specializing these calculations to rotation matrices often yields a characterization in terms of an independent pair \((U, r)\) where \(r\) is forced to have a particular distributional form. In this sense, distributions on rotation matrices defined in terms of an axis and spin are not unique to the present work. Our point here is that we have found beginning from this constructive definition made especially for 3-dimensional rotations, where one may freely model the \(r\)-spin distribution, to be an intuitively appealing, tractable and highly effective approach to modeling and statistical inference.

### 2.2 The UARS\((S, \kappa)\) density

In the modeling construction above, the first step requires specifying a density \(C(r|\kappa)\) for \(r \in (-\pi, \pi]\), symmetric about 0 and having concentration parameter \(\kappa\). It then follows that a matrix density for \(O \sim \text{UARS}(S, \kappa)\), with respect to the Haar measure which acts as a “uniform distribution” on the collection \(\Omega\) of rotation matrices, is given by

\[
f(O|S, \kappa) = \frac{4\pi}{3 - tr(S^T O)} C\left(\arccos[2^{-1}(tr(S^T O) - 1)]|\kappa\right), \quad O \in \Omega.
\]

See the supplementary on-line Appendix for details on the derivation of this density.

The UARS class of distributions allows unrestricted choice of distribution for \(r\) (or \(|r|\)). But unless the function \(\frac{C(r|\kappa)}{1 - \cos r}\) has a finite limit at \(r = 0\), the density (4) will be unbounded at \(O = S\). Thus, some of the natural choices for the distribution on \(r\), such as the von Mises circular distribution explored next, result in densities with singularities at \(S\). Other choices for the distribution on \(r\) that are perhaps less natural, but result in bounded matrix densities (4) are the “scaled” Beta\((3, \kappa)\) distribution and the truncated Maxwell-Boltzmann distribution. The Haar measure itself corresponds to a Lebesgue density for \(r\) of \((1 - \cos |r|)/\pi, r \in (-\pi, \pi]\) (Miles, 1965), for which (4) is bounded. León et al. (2006, sec. 5.2) give the density for \(r\) corresponding to the Cayley distribution (introduced in the same paper) and for the symmetric matrix Fisher distribution (introduced by Downs (1972)), for which (4) is again bounded.

For modeling the EBSD data, we consider UARS\((S, \kappa)\) matrix models where \(r\) has a von
Mises circular density given by

\[ C(r|\kappa) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(r)] , \quad r \in (-\pi, \pi) \]  

(5)

where \( I_0(\kappa) = (2\pi)^{-1} \int_{-\pi}^{\pi} \exp[\kappa \cos(r)] \, dr \) is the modified Bessel function of order zero. This density is unimodal, symmetric about 0, and as \( \kappa \to \infty \), the distribution becomes approximately Normal with mean 0 and variance \( \frac{1}{\kappa} \). For more on the von Mises distribution see Mardia and Jupp (2000).

Though the von Mises version of (4) given by

\[ f(o|S,\kappa) = \frac{2}{3-tr(S^T o)} [I_0(\kappa)]^{-1} \exp\left[ \frac{\kappa}{2} (tr(S^T o) - 1) \right] , \quad o \in \Omega, \]  

(6)

is unbounded, the choice of von Mises as a circular distribution for \( r \) is a natural one (and appropriately describes our crystallographic data as seen in Figures 2 and 3). We denote the von Mises version of the UARS(\( S, \kappa \)) distribution by vM-UARS(\( S, \kappa \)).

Because the density (6) gives an unbounded likelihood, in Section 3.2 we will discuss a type of quasi-likelihood inference for the vM-UARS distributions. The quasi-likelihood we will use results from treating

\[ f(o|S,\kappa) = [I_0(\kappa)]^{-1} \exp\left[ \frac{\kappa}{2} (tr(S^T o) - 1) \right] , \quad o \in \Omega, \]  

(7)

as a “quasi-density” on \( \Omega \) in replacement of (6). Notice that (7) is not the symmetric matrix Fisher density for \( O \) by virtue of the fact that it is not properly normalized. (The Fisher density for \( O \) corresponds to (4) with a density \( C(r|\kappa) \) given by multiplying the circular von Mises density (5) with concentration parameter \( 2\kappa \) by \( 1 - \cos r \) and normalizing.)

To get some idea of what the vM-UARS(\( S, \kappa \)) distribution is like, Figures 4 and 5 illustrate the vM-UARS(\( I_{3\times3}, \kappa \)) distributions with \( \kappa \) values of 5 and 10. In Figure 4 the contours shown on the spheres outline regions enclosing increasing amounts of probability associated with the placement of the (randomly) rotated coordinate axes for the two vM-UARS(\( I_{3\times3}, \kappa \)) distributions. If the contour closest to each axis is considered the first contour, then \((10 \times i)\%\) of realizations keep all 3 perpendicular axes within the region represented by the \( i^{th} \) contours about \( x, y, \) and \( z \) (simultaneously). As \( \kappa \) increases, probability accumulates more quickly.
as we move away from the principal direction $\mathbf{S} = \mathbf{I}_{3 \times 3}$. (If we had attempted to picture distributions with small $\kappa$, regions of high probability for the different axes would have seen significant overlap.) In Figure 5 each set of 3 perpendicular axes represents one orientation generated from the vM-UARS($\mathbf{I}_{3 \times 3}$, $\kappa$) distribution. As $\kappa$ increases, the orientations become less spread about the principal direction (represented by the axes at $x$, $y$, and $z$).

![Figure 4: Probability content contours for vM-UARS distributions with $\kappa = 5$ (left) and 10 (right) (the axes shown are those represented by principal direction $\mathbf{S}$)](image)

![Figure 5: Five random orientations generated from vM-UARS distributions with $\kappa = 5$ (left) and 10 (right) (the axes at $x$, $y$, and $z$ are those represented by principal direction $\mathbf{S}$)](image)

The value $\kappa$ controls the spread of the misorientation angle $|\mathbf{r}|$ between $\mathbf{S}$ and what is observed, and thus the corresponding “spread” of the vM-UARS($\mathbf{S}$, $\kappa$) distribution. Let $\Delta_1(\kappa)$
be the median of the distribution of $|r|$ so that $(-\Delta_1(\kappa), \Delta_1(\kappa))$ captures 50% of the von Mises(\kappa) probability. Then let $\Delta_2(\kappa)$ be the median of the distribution of the maximum angle between an $S$-rotated and a vM-UARS($S, \kappa$)-rotated coordinate axis. Table 2 illustrates the relationship between $\Delta_1(\kappa)$ and $\Delta_2(\kappa)$ for various choices of $\kappa$. The values of $\Delta_1(\kappa)$ were computed using numerical integration and each value of $\Delta_2(\kappa)$ presented is based on a sample of 100,000 vM-UARS realizations. Obviously, there is a close correspondence between $\Delta_1(\kappa)$ and $\Delta_2(\kappa)$, and $\kappa$ directly controls the concentrations of the von Mises and vM-UARS distributions. ($\Delta_2(\kappa)$ is slightly smaller than $\Delta_1(\kappa)$ as the maximum rotation of a coordinate axis is for each realization no more than the misorientation angle between the realization and $S$.)

Table 2: Values of the medians $\Delta_1(\kappa)$ and $\Delta_2(\kappa)$ for various choices of $\kappa$

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>$\Delta_1(\kappa)$</th>
<th>$\Delta_2(\kappa)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.80977</td>
<td>0.77526</td>
</tr>
<tr>
<td>5</td>
<td>0.31170</td>
<td>0.30218</td>
</tr>
<tr>
<td>10</td>
<td>0.21657</td>
<td>0.20923</td>
</tr>
<tr>
<td>20</td>
<td>0.15194</td>
<td>0.14697</td>
</tr>
<tr>
<td>50</td>
<td>0.09567</td>
<td>0.09232</td>
</tr>
<tr>
<td>100</td>
<td>0.06755</td>
<td>0.06519</td>
</tr>
<tr>
<td>500</td>
<td>0.03017</td>
<td>0.02909</td>
</tr>
</tbody>
</table>

3 One-sample quasi-likelihood and inference for the von Mises version of the UARS distributions

3.1 Parameterization of $S$

In what follows it will be useful to parameterize the principal direction $S = g(\alpha, \beta, \gamma)$ as a function of Euler angles $\alpha \in [0, 2\pi]$, $\beta \in [0, \pi]$, and $\gamma \in [0, 2\pi]$. (The Euler angle parameterization of a rotation matrix is familiar to material scientists and when using EBSD, the orientations are typically output in Euler angle form; additionally, the parameterization also allows simple evaluation of a quasi-likelihood function.) Euler angles can be used to specify the orientation of an object in 3-dimensional Euclidean space relative to some reference
coordinate system by subjecting the object to a sequence of three rotations, which are, in order,

1) a (counterclockwise) rotation of $\alpha$ radians about the $z$-axis at $(0,0,1)$,

2) a (counterclockwise) rotation of $\beta$ radians about the $x$-axis at $(1,0,0)$, and

3) a (counterclockwise) rotation of $\gamma$ radians again about the $z$-axis.

(Other orders and choices of axes are obviously possible.) Each of these rotations can describe a $3 \times 3$ rotation matrix $S = g(\alpha, \beta, \gamma)$, where $g : [0, 2\pi] \times [0, \pi] \times [0, 2\pi] \to \Omega$ is defined by

$$
g(\alpha, \beta, \gamma) = \begin{pmatrix}
\cos \gamma & \sin \gamma & 0 \\
-\sin \gamma & \cos \gamma & 0 \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \beta & \sin \beta \\
0 & -\sin \beta & \cos \beta
\end{pmatrix} \begin{pmatrix}
\cos \alpha & \sin \alpha & 0 \\
-\sin \alpha & \cos \alpha & 0 \\
0 & 0 & 1
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\cos \alpha \cos \gamma - \sin \alpha \sin \gamma \cos \beta & \sin \alpha \cos \gamma + \cos \alpha \sin \gamma \cos \beta & \sin \gamma \sin \beta \\
-\cos \alpha \sin \gamma - \sin \alpha \cos \gamma \cos \beta & -\sin \alpha \sin \gamma + \cos \alpha \cos \gamma \cos \beta & \cos \gamma \sin \beta \\
\sin \alpha \sin \beta & -\cos \alpha \sin \beta & \cos \beta
\end{pmatrix}.
$$

The function $g$ is onto $\Omega$ and is one-to-one except in cases where $\beta = 0$ or $\beta = \pi$ (see Morawiec, 2004). So strictly speaking, $g$ is not invertible. But in what follows we will treat it as if it were. Nearly all elements of $\Omega$ have unique representations in terms of the three Euler angles just introduced, and it is the elements of $\Omega$ that are fundamental (as opposed to the three angles used here). In a rare case where an orientation $S$ is needed without a unique representation in Euler angles, we may rotate all observations by an arbitrary rotation matrix $R$ (e.g., a counterclockwise rotation by $\frac{\pi}{2}$ radians about $(1,0,0)$) and do inference for the UARS($RS, \kappa$) distribution. Estimates and confidence regions for $RS$ can be obtained and then rotated by $RT$ to give estimates and confidence regions for $S$. 
With \( S = g(\alpha, \beta, \gamma) \), we write diagonal elements of \( S^T o \) as

\[
o_{11}(\alpha, \beta, \gamma) = (\cos \alpha \cos \gamma - \sin \alpha \sin \gamma \cos \beta) o_{11} + (\cos \alpha \sin \gamma - \sin \alpha \cos \gamma \cos \beta) o_{21} + (\sin \alpha \sin \beta) o_{31},
\]
\[
o_{22}(\alpha, \beta, \gamma) = (\sin \alpha \cos \gamma + \cos \alpha \sin \gamma \cos \beta) o_{12} + (\sin \alpha \sin \gamma + \cos \alpha \cos \gamma \cos \beta) o_{22} + (\cos \alpha \sin \beta) o_{32},
\]
\[
o_{33}(\alpha, \beta, \gamma) = (\sin \gamma \sin \beta) o_{13} + (\cos \gamma \sin \beta) o_{23} + (\cos \beta) o_{33}.
\]  

(9)

where “\( o_{ij} \)” denotes components of \( o \).

### 3.2 Maximum quasi-likelihood estimation (MQL)

The singularity of (6) at \( o = S \) means that the one-sample likelihood for the vM-UARS(\( S, \kappa \)) inference problem has singularities at \( S \) equal to each observation \( o_i \). So in place of inference based on the likelihood we consider quasi-likelihood-based inference based on quasi-density (7) for the one-sample problem for the vM-UARS(\( S, \kappa \)) distribution using the parameterization of \( S \) introduced in Section 3.1.

Suppose observations \( o_1, \ldots, o_n \in \Omega \) are iid from the vM-UARS(\( S, \kappa \)) distribution. Then, we have quasi-likelihood function

\[
L_n(\kappa, (\alpha, \beta, \gamma)) = [I_0(\kappa)]^{-n} \prod_{i=1}^{n} \exp \left[ \frac{\kappa}{2} (o_{11,i}(\alpha, \beta, \gamma) + o_{22,i}(\alpha, \beta, \gamma) + o_{33,i}(\alpha, \beta, \gamma) - 1) \right]
\]

(10)

where, as in (9), \( o_{11,i}(\alpha, \beta, \gamma), o_{22,i}(\alpha, \beta, \gamma), o_{33,i}(\alpha, \beta, \gamma) \) denote the diagonal entries of \( S^T o_i \), \( i = 1, \ldots, n \), with \( S = g(\alpha, \beta, \gamma) \). The one-sample log-quasi-likelihood for the vM-UARS(\( S, \kappa \)) distribution is then

\[
l_n(\kappa, (\alpha, \beta, \gamma)) = \frac{\kappa}{2} \sum_{i=1}^{n} (o_{11,i}(\alpha, \beta, \gamma) + o_{22,i}(\alpha, \beta, \gamma) + o_{33,i}(\alpha, \beta, \gamma) - 1) - n \log(I_0(\kappa))
\]

\[
= \frac{\kappa n}{2} \text{tr}(S^T \bar{o}) - \frac{\kappa n}{2} - n \log(I_0(\kappa)),
\]

(11)

using the sample mean \( \bar{o} = \sum_{i=1}^{n} o_i/n \). Thus, the maximum quasi-likelihood (MQL) estimates for \( (\alpha, \beta, \gamma) \), say \( (\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \), are such that \( \hat{S} = g(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \) maximizes \( \text{tr}(S^T \bar{o}) \). Therefore, the MQL estimate \( \hat{S} \) for the vM-UARS(\( S, \kappa \)) distribution is the moment estimator mentioned in
Section 1. Chang and Rivest (2001) also discuss M-Estimation for the location parameter $S$ by minimizing objective functions that include those that are functions of $tr(S^T o_i)$, while Rivest and Chang (2006) consider regression-type estimators of the location parameter.

3.3 Asymptotic results

For later inference, we develop some asymptotic results for one-sample quasi-likelihood inference; all proofs appear in the supplementary on-line Appendix. Our first result regards the asymptotic distribution of the maximum quasi-likelihood (MQL) estimator.

**Proposition 1.** Suppose $O_1, \ldots, O_n$ are iid vM-UARS($S$, $\kappa$). Let $\hat{\theta}_n^T$ be the MQL estimator of $\theta^T = (\kappa, (\alpha, \beta, \gamma))$ and suppose that the true value of $\theta$ is $\theta_0$. Then, under $\theta_0$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \text{MVN}(0, H_1^{-1}(\theta_0) I_1(\theta_0) H_1^{-1}(\theta_0)),$$

as $n \to \infty$, where

$$H_1(\theta) \equiv \begin{pmatrix} I_2(\kappa) + \frac{1}{2} I_1(\kappa) - \left( \frac{I_1(\kappa)}{I_0(\kappa)} \right)^2 & 0 & 0 & 0 \\ 0 & \kappa D(\kappa) & 0 & \kappa D(\kappa) \cos \beta \\ 0 & 0 & \kappa D(\kappa) & 0 \\ 0 & \kappa D(\kappa) \cos \beta & 0 & \kappa D(\kappa) \end{pmatrix},$$

and

$$I_1(\theta) \equiv \begin{pmatrix} I_2(\kappa) + \frac{1}{2} I_1(\kappa) - \left( \frac{I_1(\kappa)}{I_0(\kappa)} \right)^2 & 0 & 0 & 0 \\ 0 & \kappa I_1(\kappa) + 1 I_0(\kappa) & 0 & \kappa I_1(\kappa) \cos \beta + \frac{2 I_1(\kappa)}{I_0(\kappa)} \\ 0 & 0 & \kappa I_1(\kappa) + 1 I_0(\kappa) & 0 \\ 0 & \kappa I_1(\kappa) \cos \beta + \frac{2 I_1(\kappa)}{I_0(\kappa)} & 0 & \kappa I_1(\kappa) + \frac{3 I_1(\kappa)}{3 I_0(\kappa)} \end{pmatrix},$$

for $D(\kappa) = \frac{1}{3} + \frac{2}{3} I_1(\kappa)$ and $I_i(\kappa)$ the modified Bessel function of order $i$.

See the Appendix for expressions of $H_1(\theta)$ and $I_1(\theta)$ in terms of expected derivatives for the log-quasi-likelihood.

We proceed to look at limiting distributions associated with two different methods of hypothesis testing. For Propositions 2 and 3 we derive the asymptotic null distributions of
quasi-likelihood ratio test (Q-LRT) statistics and Wald type test statistics, while Propositions 4 and 5 concern non-null distributions for these statistics.

**Proposition 2.** (Asymptotic Null Distributions of Quasi-Likelihood Ratio Test (Q-LRT) Statistics) Suppose that we partition \( \theta_{4 \times 1} = (\kappa, (\alpha, \beta, \gamma)) \) into \( \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \) where \( \theta_1 \) is \( r \times 1 \) and \( \theta_2 \) is \( (4-r) \times 1 \). Consider the hypothesis

\[ H_0 : \theta_1 = \theta_{01} \]

where the true value of \( \theta \), \( \theta_0 = \begin{pmatrix} \theta_{01} \\ \theta_{02} \end{pmatrix} \), satisfies \( H_0 \), and define \( \lambda_n = \frac{L_n(\theta^*_n)}{L_n(\hat{\theta}_n)} \), where \( L_n \) is as in (10), \( \hat{\theta}_n \) is the MQL estimate over \( \Theta \), and \( \theta^*_n \) is the MQL estimate over \( \Theta_0 \) (the part of the parameter space where \( H_0 \) holds). Then, as \( n \to \infty \),

\[ -2 \log(\lambda_n) \xrightarrow{d} \begin{cases} \chi^2_1 + B(\kappa_0)\chi^2_3 & \text{for } \theta_1 = \theta, \\ B(\kappa_0)\chi^2_3 & \text{for } \theta_1^T = (\alpha, \beta, \gamma), \\ \chi^2_1 & \text{for } \theta_1 = \kappa \end{cases} \]  

(14)

where \( B(\kappa_0) = \frac{I_1(\kappa_0)}{I_0(\kappa_0) + 2I_1(\kappa_0)} \in (0, \frac{1}{3}] \) and \( \chi^2_1 \) and \( \chi^2_3 \) are independent chi-squared variables.

**Proposition 3.** (Asymptotic Null Distribution of the Overall Wald Test Statistic) Consider the hypothesis \( H_0 : \theta = \theta_0 \). Define

\[ T_n = n(\hat{\theta}_n - \theta_0)^T H_1(\theta_0)I^{-1}_1(\theta_0)H_1(\theta_0)(\hat{\theta}_n - \theta_0) \]

for \( H_1(\theta) \) and \( I_1(\theta) \) as in (12) and (13), respectively. Then, under \( H_0 \) (under \( \Theta_0 \)), as \( n \to \infty \),

\[ T_n \xrightarrow{d} \chi^2_1. \]

In a straightforward manner it also holds that if we wish to test \( H_0 : (\alpha, \beta, \gamma)^T = (\alpha_0, \beta_0, \gamma_0)^T \) and \( I_1^{-1}(\theta_0) \) accordingly, then under \( H_0 \) the Wald statistic \( T_n \xrightarrow{d} \chi^2_3 \) as \( n \to \infty \). Similarly, for the test of \( H_0 : \kappa = \kappa_0 \), under \( H_0 \) the Wald statistic \( T_n \xrightarrow{d} \chi^2_1 \).

Next, we will expand on the previous two propositions by finding limiting distributions of the overall Q-LRT statistic and overall Wald test statistic under local alternative hypotheses.
**Proposition 4.** (Limiting Non-Null Distribution for the Overall Quasi-Likelihood Ratio Statistic) For $\theta^T = (\kappa, (\alpha, \beta, \gamma))$, suppose $\theta_0$ denotes the true parameter value and

$$
\hat{\theta}_n = \theta_0 + H_1^{-1}(\theta_0)\mathcal{I}_1(\theta_0) \frac{1}{\sqrt{n}} \frac{\delta}{\sqrt{n}}
$$

(15)

for $\delta^T = (\delta_1, \delta_2, \delta_3, \delta_4) \in \mathbb{R}^4$. Let $\lambda_n = \frac{L_n(\hat{\theta}_n)}{L_n(\hat{\theta}_n)}$ denote the quasi-likelihood ratio along the sequence $\hat{\theta}_n$, where $\hat{\theta}_n$ is the MQL estimate of $\theta$. Then under the sequence of nulls $\hat{\theta}_n$, as $n \to \infty$,

$$
-2 \log(\lambda_n) \xrightarrow{d} \chi^2_1(\delta_\kappa^2) + B(\kappa_0)\chi^2_3(\delta_\alpha^2 + \delta_\beta^2 + \delta_\gamma^2),
$$

for $B(\kappa_0) = \frac{I_1(\kappa_0)}{I_0(\kappa_0) + 2I_1(\kappa_0)}$ where $\chi^2_1(\delta_\kappa^2)$ and $\chi^2_3(\delta_\alpha^2 + \delta_\beta^2 + \delta_\gamma^2)$ are independent chi-squared random variables with noncentrality parameters $\delta_\kappa^2$ and $\delta_\alpha^2 + \delta_\beta^2 + \delta_\gamma^2$.

Note that if we rewrite Proposition 4 by replacing (15) with $\hat{\theta}_n = \theta_0 + \frac{\delta}{\sqrt{n}}$, we find that under the sequence of nulls $\hat{\theta}_n$,

$$
-2 \log(\lambda_n) \xrightarrow{d} \chi^2_1(\eta_1) + B(\kappa_0)\chi^2_3(\eta_2),
$$

where $\chi^2_1(\eta_1)$ and $\chi^2_3(\eta_2)$ are independent chi-squared random variables with noncentrality parameters

$$
\eta_1 = \delta_\kappa^2 \left[ \frac{I_2(\kappa_0) - \frac{1}{\kappa_0}I_1(\kappa_0)}{I_0(\kappa_0)} - \left( \frac{I_1(\kappa_0)}{I_0(\kappa_0)} \right)^2 \right]
$$

and

$$
\eta_2 = \frac{I_1(\kappa_0)}{3\kappa_0I_0(\kappa_0)} \left( 1 + 2\frac{I_1(\kappa_0)}{I_0(\kappa_0)} \right)^2 \left[ \delta_\alpha^2 + \delta_\beta^2 + \delta_\gamma^2 + 2\delta_\alpha\delta_\gamma \cos \beta_0 \right].
$$

With $\hat{S}_n = g(\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n)$ and $S_0 = g(\alpha_0, \beta_0, \gamma_0)$, where $g$ is as in (8), suppose $\xi$ is the misorientation angle between $\hat{S}_n$ and $S_0$. Then, $n(2 - 2\cos \xi_n) \xrightarrow{p} \delta_\alpha^2 + \delta_\beta^2 + \delta_\gamma^2 + 2\delta_\alpha\delta_\gamma \cos \beta_0$, so the part of the noncentrality parameter $\eta_2$ not depending on $\kappa_0$ is obtained as the limit of a function of this misorientation angle.

**Proposition 5.** (Limiting Non-Null Distribution for the Overall Wald Test Statistic) For $\theta^T = (\kappa, (\alpha, \beta, \gamma))$, suppose $\theta_0$ denotes the true parameter value and

$$
\hat{\theta}_n = \theta_0 + H_1^{-1}(\theta_0)\mathcal{I}_1(\theta_0) \frac{1}{\sqrt{n}} \frac{\delta}{\sqrt{n}}
$$
for $\delta^T = (\delta_k, \delta_{\alpha}, \delta_{\beta}, \delta_{\gamma}) \in \mathbb{R}^4$. Then under the sequence of nulls $\tilde{\theta}_n$, as $n \to \infty$,

$$T_n \overset{d}{\to} \chi^2_4(\delta_k^2 + \delta_{\alpha}^2 + \delta_{\beta}^2 + \delta_{\gamma}^2).$$

### 3.4 Confidence regions

We can obtain confidence regions for $S$ or $\kappa$ by inversion of either quasi-likelihood ratio or Wald tests. This results in confidence intervals for $\kappa$. For $S$, inversion of tests of $H_0 : (\alpha, \beta, \gamma) = (\alpha_0, \beta_0, \gamma_0)$ produces the set of triples $(\alpha_0, \beta_0, \gamma_0)$ for which the test statistic is small. Applying the function $g$ from (8) to all such triples gives a set of orientations comprising the confidence set for $S$. Letting $\hat{S}$ denote the MQL estimate of $S$, the angle between each of the 3 perpendicular axes representing an orientation in the confidence set and the corresponding axis of $\hat{S}$ can be found. Let $\rho$ be the maximum of all such angles (found numerically). We consider a confidence region for $S$ (potentially slightly more conservative than that resulting from inversion of the tests) as the set of all orientations for which the maximum angle between an $S$ coordinate axis and an $\hat{S}$ coordinate axis is less than $\rho$. We can represent this confidence region by 3 cones of constant angle $\rho$ around axes representing $\hat{S}$ and the notion of “size” of the confidence region for $S$ can be reduced to thinking about the size of $\rho$. In Section 4 we give confidence intervals for $\kappa$ and figures representing confidence regions for $S$ based on data from the vM-UARS($S, \kappa)$ distribution.

### 4 Simulation study and asymptotic power comparison

To investigate the relevance of the asymptotic results of Section 3.3 to statistical practice, we simulated data from the vM-UARS($S, \kappa)$ distribution using different choices for the parameter $\kappa$ for various sample sizes. The values used for $\kappa$ were 2, 8, and 20 and sample sizes were $n = 10, 30, 100, \text{ and } 300$. The values used for $\alpha, \beta, \text{ and } \gamma$ in $S = g(\alpha, \beta, \gamma)$, where $g$ is as in (8), were fixed at respectively 2.3, 1.1, and 5.9 throughout the simulations (though simple symmetry arguments show that this detail is immaterial).

We simulated 1000 samples for each $(n, \kappa)$ combination and obtained values for the Q-LRT statistic and the Wald test statistic, for testing for the entire parameter vector. The empirical
cumulative distribution function for each test statistic was then plotted along with the limiting
distribution. Figure 6 contains the plots for \( \kappa = 2 \) and 8 (similar plots resulted for \( \kappa = 20 \)).

The asymptotic cutoff for the Q-LRT for the entire vector with \( \kappa = 8 \), and at an \( \alpha \)-level
of 0.05, is (based on Proposition 2) approximately 5.123. From the empirical cumulative
distribution functions represented in Figure 6 we find that this value corresponds to \( \alpha \)-levels
of approximately 0.0565, 0.055, 0.053, and 0.052 for \( n = 10, 30, 100, \) and 300, respectively. So,
when using the asymptotic cutoff with small sample sizes, the actual levels are more liberal
than the desired \( \alpha = 0.05 \). The asymptotic cutoff for the Wald test for the entire vector, at an
\( \alpha \)-level of 0.05, is (based on Proposition 3) approximately 9.488. From the empirical cumulative
distribution functions, we find that this value corresponds to \( \alpha \)-levels of approximately 0.173,
0.093, 0.072, and 0.057 for \( n = 10, 30, 100 \) and 300, respectively, when \( \kappa = 8 \). Thus, the
convergence in the case of the Wald test is not as fast as for the Q-LRT. Similar results hold
for \( \kappa = 2 \) and 20, and when testing for only \( \kappa \). When testing for only the set \( (\alpha, \beta, \gamma) \) the
results for both the Wald test and the Q-LRT are similar to those of the Q-LRT of the entire
parameter vector.

It is also instructive to consider MQL estimates for one sample of size \( n = 30 \) simulated
from the vM-UARS(\( S, \kappa \)) distribution under each of the 3 values of \( \kappa \) (where again where
\( S = g(2.3,1.1,5.9) \)). MQL estimates obtained from each of these three samples are presented
in Table 3.

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>( \hat{\kappa} )</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\beta} )</th>
<th>( \hat{\gamma} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.524342</td>
<td>2.275942</td>
<td>1.052625</td>
<td>5.931470</td>
</tr>
<tr>
<td>8</td>
<td>8.066815</td>
<td>2.290575</td>
<td>1.077529</td>
<td>5.876191</td>
</tr>
<tr>
<td>20</td>
<td>18.297676</td>
<td>2.305525</td>
<td>1.136357</td>
<td>5.892312</td>
</tr>
</tbody>
</table>

To graphically portray such data and the corresponding MQL estimate of \( S \), we display the
data on a 3-dimensional sphere. Figure 7 shows the 30 realizations from the vM-UARS(\( S, 8 \))
distribution as points on a sphere corresponding to the positions of the \( x, y, \) and \( z \) unit vectors
after rotation. The orientation given by the MQL estimate of \( S \) is portrayed by the axes at \( x, y, \) and \( z \). We see from Figure 7 that the MQL estimate of \( S \) is “in the center” of the data.
Figure 6: Limiting and empirical (estimated small $n$) null cumulative distribution functions for the Q-LRT (top) and Wald (bottom) statistics for the full parameter vector, $\kappa = 2$ (right) and 8 (left)
We can also give confidence regions based on inversion of both quasi-likelihood ratio and Wald tests. 95% confidence limits for the parameter \( \kappa \) under each of these methods are presented in Table 4. We represent the 95\% quasi-likelihood ratio confidence region for \( S \) in Figure 8 using the method described in Section 3.4. Here, the axes at \( x, y, \) and \( z \) represent the orientation of the MQL estimate for \( S \), as given by \( \hat{S} = g(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \). We note that in the limiting distribution given in (14), the quantity \( B(\kappa_0) = \frac{I_1(\kappa_0)}{I_0(\kappa_0)+2I_1(\kappa_0)} \) is bounded above by \( \frac{1}{3} \). So, when making the 95\% quasi-likelihood ratio confidence region for \( (\alpha, \beta, \gamma) \) we used the upper 5\% point of the “\( \frac{1}{3} \chi_3^2 \)” distribution as the cutoff between small and large values of \( -2 \log(\lambda_{30}) \).

By comparing the plots in Figure 8, we see that for larger values of \( \kappa \), since the data are more concentrated, we get smaller conic regions. The angle between each axis and the edge of the conic region is 0.19278 for \( \kappa = 2 \), 0.09862 for \( \kappa = 8 \), and 0.07174 for \( \kappa = 20 \). Although we do not present a corresponding figure here, in the case of a 95\% Wald confidence region, these angles are 0.18199, 0.09709, and 0.07025, respectively.

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>Quasi-likelihood ratio</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(1.571072, 3.796212)</td>
<td>(1.733614, 4.735426)</td>
</tr>
<tr>
<td>8</td>
<td>(4.777156, 12.69225)</td>
<td>(5.470979, 16.05493)</td>
</tr>
<tr>
<td>20</td>
<td>(10.64507, 29.02341)</td>
<td>(12.24265, 36.77996)</td>
</tr>
</tbody>
</table>
Using Propositions 4 and 5 we can compare the asymptotic power of the quasi-likelihood ratio and Wald theory tests (for the entire parameter vector) under various choices of $\kappa_0$ and $\delta$. Table 5 gives the limiting power of each test (at a limiting $\alpha$-level of 0.05) for some combinations of these values. In checking various combinations of $\kappa_0$ and $\delta$, the Q-LRT generally tended to have more asymptotic power than the Wald test in cases where the largest deviation between true and null-hypothesized parameters was located in the spread parameter $\kappa_0$; when the greatest deviations were located in the Euler angle parameters, the Wald test had better large sample power. Each test has its own strengths. The explanation is that the limit law of the quasi-likelihood ratio statistic is a convolution of two chi-squared variables where deviations between hypothothesized and true parameters are split into different noncentrality parameters, depending on the location of the deviations (i.e., whether these lie in the concentration parameter $\kappa$ or in the location/Euler angle parameters). In contrast, the chi-squared limit in the Wald test has one noncentrality parameter that combines all deviations between true and hypothesized parameters.

5 Application of quasi-likelihood methodology for the von Mises version of the UARS distributions to the crystal orientation data

As promised in Section 1, we examine the repeatability of measurements obtained using EBSD. As our matrix models describe rotational symmetry, we calculated the value of Pren-
Table 5: Asymptotic power of the quasi-likelihood ratio and Wald tests (for the entire vector) under various choices of $\kappa_0$ and $\delta$

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\kappa_0 = 1$</th>
<th>$\kappa_0 = 4$</th>
<th>$\kappa_0 = 16$</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>(.1, 1, 1, 1, 1)</td>
<td>0.0504</td>
<td>0.0507</td>
<td>0.0510</td>
<td>0.0520</td>
</tr>
<tr>
<td>(1, 1, 1, 1, 1)</td>
<td>0.1665</td>
<td>0.1619</td>
<td>0.1617</td>
<td>0.1073</td>
</tr>
<tr>
<td>(2, 1, 1, 1, 1)</td>
<td>0.5041</td>
<td>0.4834</td>
<td>0.4786</td>
<td>0.3224</td>
</tr>
<tr>
<td>(.1, 1, 1, 1)</td>
<td>0.0984</td>
<td>0.1218</td>
<td>0.1277</td>
<td>0.2449</td>
</tr>
<tr>
<td>(.1, 2, 2, 2)</td>
<td>0.4109</td>
<td>0.5500</td>
<td>0.5623</td>
<td>0.8028</td>
</tr>
<tr>
<td>(1, 1, 1, 1)</td>
<td>0.2278</td>
<td>0.2603</td>
<td>0.2622</td>
<td>0.3201</td>
</tr>
<tr>
<td>(2, 1, 1, 1)</td>
<td>0.5814</td>
<td>0.5912</td>
<td>0.5934</td>
<td>0.5400</td>
</tr>
<tr>
<td>(2, 2, 2, 2)</td>
<td>0.7997</td>
<td>0.8506</td>
<td>0.8530</td>
<td>0.9119</td>
</tr>
</tbody>
</table>


tice’s R statistic for each data set. Small values were found, suggesting that an assumption of spherical symmetry is reasonable in all cases (see Prentice, 1984). We then fit the von Mises version of the Uniform-Axis-Random-Spin distributions to the 14 repeat nickel observations in Table 1 and seven repeat aluminum observations via MQL. Because the data were so highly concentrated around their principal directions, the MQL estimates of the parameters $\kappa$ were extremely large. Thus, we used the normal approximation to the von Mises distribution as given in Section 2.2 for computing the quasi-likelihood. Estimates and 95% confidence intervals for $\kappa$ are provided in Table 6. The estimates for the location $S$ (represented as Euler angles $(\alpha, \beta, \gamma)$) are $(5.8620, 0.9352, 4.2170)$ and $(5.3552, 1.4566, 0.0580)$ for the nickel and aluminum data, respectively. (We will not present a figure giving the confidence region for $S$ as we did in the previous section. The region is so small that it visually appears as a “corner” positioned at the estimated principal direction instead of as a set of a conic regions).

Table 6: MQL estimates and 95% confidence limits for the parameter $\kappa$ for the repeat nickel and aluminum data sets

<table>
<thead>
<tr>
<th></th>
<th>MQL estimate</th>
<th>Q-LRT interval</th>
<th>Wald interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nickel</td>
<td>86373.7613</td>
<td>(37113.27, 167042.84)</td>
<td>(49617.48, 333217.06)</td>
</tr>
<tr>
<td>Aluminum</td>
<td>58829.1138</td>
<td>(994.2512, 70864.1789)</td>
<td>(28730.26, 14120080)</td>
</tr>
</tbody>
</table>

The intervals for $\kappa$ given in Table 6 contain only values that represent highly concentrated distributions on $\Omega$. To put the estimated values of $\kappa$ into the context of EBSD precision, we consider the normal approximation to the von Mises circular distribution. Under this
approximation, with \( \hat{\kappa} \) the MQL estimate for a set of data, the spin angles \( r \) effectively follow a normal distribution with mean 0 and variance \( 1/\hat{\kappa} \). Thus, the angles \( r \) for the repeated nickel data are approximately normal with standard deviation 0.003402 radians, or 0.19495°. So, the corresponding fitted distribution on the misorientation angles \( |r| \) places 99% probability in the interval \((0°, 0.5022°)\), giving an EBSD precision of 0.5022°. For the repeated aluminum data, the angles \( r \) are approximately normal with standard deviation 0.004123 radians, or 0.23623°, and the fitted distribution on the misorientation angles places 99% probability in the interval \((0°, 0.6085°)\) for an EBSD precision of 0.6085°. Therefore, both metals produce single-site estimated EBSD precisions better than the literature-standard 1°. But again note that EBSD precisions given in the literature are based on a single scan across a homogeneous specimen and not repeat readings at a given location. Our present analysis provides more optimistic precisions than those in the literature when pure repeatability is at issue.

Next, we consider measurements from a single scan that appear to be within the same grain. With 70 different locations for the nickel data we arrive at a MQL estimate for \( S \) given by the Euler angles \((5.8544, 0.9297, 4.2224)\). With 50 different locations for the aluminum data, we arrive at an estimate of \((5.3521, 1.4506, 0.0585)\). We expect the observations to be highly concentrated about the principal directions. Estimates and 95% confidence intervals for \( \kappa \) are provided in Table 7.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MQL estimate</th>
<th>Q-LRT interval</th>
<th>Wald interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nickel</td>
<td>19018.6767</td>
<td>(13393.45, 26033.30)</td>
<td>(14285.92, 28440.91)</td>
</tr>
<tr>
<td>Aluminum</td>
<td>30713.9345</td>
<td>(20193.15, 44375.14)</td>
<td>(22064.79, 50515.57)</td>
</tr>
</tbody>
</table>

Based on the estimated values of \( \kappa \), the angles \( r \) for the nickel data are approximately normal with standard deviation 0.007251 radians, or 0.41546°, and the angles \( r \) for the aluminum data are approximately normal with standard deviation 0.005706 radians, or 0.32693°. Fitted distributions on the misorientation angles \( |r| \) place 99% probability in the intervals \((0°, 1.0701°)\) and \((0°, 0.8421°)\), respectively. Thus, for the metals considered here, we obtain precision figures comparable to the commonly quoted 1° when considering observations representing
different locations in a single grain. We emphasize, however, that our methods provide a coherent fitting and inference framework for this problem, where existing published work on EBSD precision is entirely descriptive and to some degree ad hoc in terms of how sets of measurements are converted to precision statements. The fact that pure repeatability variation (single site precision) for EBSD is smaller than between location variation even in a single crystal/grain is probably traceable to small effects of both 1) slight inhomogeneity of actual material properties site to site (for example, due to preparation differences across a specimen) and 2) slight inhomogeneity of equipment behavior site to site (related, for example, to physical distortions, discreteness of pixel locations on the focal plane, slightly different geometries of beam paths, etc.).

6 Conclusion

The UARS class of distributions and the inference tools developed in this paper are extremely attractive and provide alternatives for modeling 3-dimensional rotations beyond the distributions previously studied in the literature. This has proved valuable in our application, allowing us to rationally and systematically quantify precision of EBSD measurements. The UARS models are tractable and have directly interpretable (“location” and “spread”) parameters ($S$ and $\kappa$). The motivating construction shows how to easily simulate from a UARS distribution, a fact that is proving to be extremely valuable in Bayes MCMC analyses (see Bingham, Vardeman, and Nordman, 2009 and Bingham, Nordman, and Vardeman, 2009). Our development of quasi-likelihood theory for the von Mises case is important in its own right and can serve as a template for parallel developments with other circular distributions (including those that lead to bounded densities, allowing for explicit likelihood theory) to provide a rich variety of alternative practical one-sample models for random orientations (see Bingham, Nordman, and Vardeman (2009) for likelihood and Bayes results for the symmetric matrix Fisher model based on the methods of the present paper). We can provide useful “one-way random effects” methods (see again Bingham, Vardeman, and Nordman (2009) and Bingham, Nordman, and Vardeman (2009)) and fully expect to provide simple “time series”
and clustering methods for 3-dimensional rotations based on this class of distributions.

While much good and difficult work as been done on the problem of modeling randomness
of rotations (Downs 1972; Khatri and Mardia 1977; Jupp and Mardia 1979; Prentice 1986;
Mardia and Jupp 1999; Rancourt et al. 2000; Chang and Rivest 2001; Chikuse 2003; León
et al. 2006; Rivest et al. 2008) our contention here is that focus on the UARS class provides
heretofore unavailable flexibility and tractability that can open myriad possibilities in the
analysis of 3-dimensional orientation data arising from various applications.

Acknowledgment

The authors wish to thank Dr. Barbara Lograsso of Michigan Technological University
for introducing us to the EBSD problem and Fran Laabs of Ames Laboratory, Iowa State
University, for collecting the EBSD data. We further thank Iowa State University Professor
Max Morris for an early suggestion that led us to identify and study the UARS class.

Appendix: Properties of the UARS Distributions

We establish some properties of UARS distributions; proofs of these can be found in a
supplementary on-line Appendix. For simplicity, some properties are stated only for the case
of principle direction $I_{3 \times 3}$. These may be easily extended to general $S$ using the fact that
$O \sim UARS(I_{3 \times 3}, \kappa)$ implies $S \cdot O \sim UARS(S, \kappa)$. Further, note that while Properties 7 - 11
below are those of the von Mises version of the UARS distributions, Properties 1 - 6 hold for
any choice of circular distribution $r \sim \text{Circ}(\kappa)$ (symmetric about 0 or not). Additionally, the
last three properties concern the score function from the quasi-density and are relevant to the
inference method described in Section 3.

Property 1. If $O \sim UARS(I_{3 \times 3}, \kappa)$, then $O^T \sim UARS(I_{3 \times 3}, \kappa)$.

Property 2. If $O \sim UARS(I_{3 \times 3}, \kappa)$, then $S \cdot O \cdot S^T \sim UARS(I_{3 \times 3}, \kappa)$ for any $3 \times 3$ rotation
matrix $S$.

Property 3. If $O \sim UARS(I_{3 \times 3}, \kappa)$, then $S \cdot O$ and $O \cdot S \sim UARS(S, \kappa)$. 
Property 4. If \( O \sim \text{UARS}(S, \kappa) \), then \( O^T \sim \text{UARS}(S^T, \kappa) \).

Property 5. If \( O \sim \text{UARS}(S, \kappa) \), then \( S^T \cdot O \) and \( O \cdot S^T \sim \text{UARS}(I_{3 \times 3}, \kappa) \).

Property 6. Suppose \( O \sim \text{UARS}(I_{3 \times 3}, \kappa) \) and \( O = (X \ Y \ Z) \), where \( X, Y, \) and \( Z \) are the three columns of \( O \). Let \( P_X \) be the spherical distribution of \( X \) about \( (1,0,0)^T \), \( P_Y \) be the spherical distribution of \( Y \) about \( (0,1,0)^T \), and \( P_Z \) be the spherical distribution of \( Z \) about \( (0,0,1)^T \). Then \( P_X = P_Y = P_Z \).

Property 7. If \( O \sim \text{vM-UARS}(I_{3 \times 3}, \kappa) \), then

\[
E(O) = \left( \frac{1}{3} + \frac{2}{3} \frac{I_1(\kappa)}{I_0(\kappa)} \right) I_{3 \times 3}, \quad E(O^2) = \left( \frac{1}{3} + \frac{2}{3} \frac{I_2(\kappa)}{I_0(\kappa)} \right) I_{3 \times 3},
\]

where \( I_i(\kappa) \) is the modified Bessel function of order \( i \).

Property 8. Suppose \( O = (X \ Y \ Z) \sim \text{vM-UARS}(I_{3 \times 3}, \kappa) \). Let \( x \) represent the cosine of the angle between \( X \) and \( (1,0,0)^T \), \( y \) represent the cosine of the angle between \( Y \) and \( (0,1,0)^T \), and \( z \) represent the cosine of the angle between \( Z \) and \( (0,0,1)^T \). Then,

\[
\text{Corr}(x,y) = \text{Corr}(x,z) = \text{Corr}(y,z) = \frac{1}{15} + \frac{8}{15} \frac{I_1(\kappa)}{I_0(\kappa)} + \frac{2}{5} \frac{I_2(\kappa)+\frac{1}{2} I_1(\kappa)}{I_0(\kappa)} - \left( \frac{1}{3} + \frac{2}{3} \frac{I_1(\kappa)}{I_0(\kappa)} \right)^2,
\]

Property 9. Suppose that \( O \sim \text{vM-UARS}(S, \kappa) \). If \( l(\kappa, (\alpha, \beta, \gamma)) = l_1(\kappa, (\alpha, \beta, \gamma)) \) as in (11) and

\[
l'(\kappa, (\alpha, \beta, \gamma)) = \left( \frac{\partial l}{\partial \kappa}, \frac{\partial l}{\partial \alpha}, \frac{\partial l}{\partial \beta}, \frac{\partial l}{\partial \gamma} \right)^T,
\]

then \( E(l'(\kappa, (\alpha, \beta, \gamma))) = 0 \).

Property 10. Suppose that \( O \sim \text{vM-UARS}(S, \kappa) \) and that

\[
l''(\kappa, (\alpha, \beta, \gamma)) = \begin{pmatrix}
\frac{\partial^2 l}{\partial \kappa^2} & \frac{\partial^2 l}{\partial \kappa \partial \alpha} & \frac{\partial^2 l}{\partial \kappa \partial \beta} & \frac{\partial^2 l}{\partial \kappa \partial \gamma} \\
\frac{\partial^2 l}{\partial \alpha \partial \kappa} & \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} & \frac{\partial^2 l}{\partial \alpha \partial \gamma} \\
\frac{\partial^2 l}{\partial \beta \partial \kappa} & \frac{\partial^2 l}{\partial \beta \partial \alpha} & \frac{\partial^2 l}{\partial \beta^2} & \frac{\partial^2 l}{\partial \beta \partial \gamma} \\
\frac{\partial^2 l}{\partial \gamma \partial \kappa} & \frac{\partial^2 l}{\partial \gamma \partial \alpha} & \frac{\partial^2 l}{\partial \gamma \partial \beta} & \frac{\partial^2 l}{\partial \gamma^2}
\end{pmatrix}
\]

for \( l(\kappa, (\alpha, \beta, \gamma)) \) as in Property 9. Then, \( H_1(\kappa, (\alpha, \beta, \gamma)) = -E(l''(\kappa, (\alpha, \beta, \gamma))) \) as given in (12).
Property 11. Suppose that $O \sim vM-UARS(S, \kappa)$ and that

$$l'(\kappa, (\alpha, \beta, \gamma)) = \left( \frac{\partial l}{\partial \kappa}, \frac{\partial l}{\partial \alpha}, \frac{\partial l}{\partial \beta}, \frac{\partial l}{\partial \gamma} \right)^T$$

for $l(\kappa, (\alpha, \beta, \gamma))$ as in Property 9. Then, $I_1(\kappa, (\alpha, \beta, \gamma)) = \text{Var}(l'(\kappa, (\alpha, \beta, \gamma)))$ is as given in (13).

References


**Supplementary Material**

Supplementary material to follow consists of Appendices A – D providing the development of the UARS density, proofs of the main results in the Appendix and Section 3.3 of the main manuscript, and Q-Q plots of EBSD aluminum data similar to those presented for the nickel data in Section 1 of the main manuscript.

**Appendix A: Development of the UARS density**

Suppose that $\mathbf{O} \sim UARS(\mathbf{S}, \kappa)$ as in (3). Note that we may obtain an observation from $UARS(\mathbf{S}, \kappa)$ equivalently as

$$\mathbf{O} = \mathbf{S} \cdot M(U, |r|). \quad (A.1)$$

The distributional equivalence of $\mathbf{S} \cdot M(U, r)$ and $\mathbf{S} \cdot M(U, |r|)$ follows from the fact that $M(U, r) = M(-U, -r)$ in (2) and that $U$ is uniformly distributed and independent of $r$ (see also Miles, 1965). Hence, we may develop a $UARS(\mathbf{S}, \kappa)$ density by considering the joint...
distribution of \((|r|, U)\). In this framework, every potential realization of \(O\) as in (A.1), for which \(O \neq S\), corresponds to a unique realization of \(U\) and \(|r| > 0\). In the case \(O = S\), \(r = 0\) and \(M(U, |r|) = I_{3 \times 3}\) hold.

To derive a density for a \(UARS(S, \kappa)\) distribution with respect to the invariant Haar measure, \(H\), which acts as a “uniform distribution” on \(\Omega\) (see Miles, 1965; Downs, 1972), first note that the uniform distribution for \(U\) on the unit sphere can be specified by allowing \(\phi\) and \(\tau \equiv \cos \theta\) in (1) to be independently and uniformly distributed on \([0, 2\pi]\) and \([-1, 1]\), respectively. If \(r \in (-\pi, \pi]\) is also independently distributed with a density \(C(r|\kappa)\) symmetric about 0, then \(|r|\) has density \(2 \cdot C(|r|/\kappa), |r| \in [0, \pi]\). Thus, a distribution \(F^*\) on \([-1, 1] \times [0, 2\pi] \times [0, \pi]\) for \((\tau, \phi, |r|)\) has a joint density \(dF^*/d\lambda = C(|r|/\kappa)/(2\pi)\), where \(\lambda\) stands for 3-dimensional Lebesgue measure. By definition, this distribution induces a corresponding distribution, \(F\), for \(O = M(U, |r|)\) in \(\Omega\) that is \(UARS(I_{3 \times 3}, \kappa)\). Recall that the mapping \(O = M(U, |r|)\) is one-to-one (except for \(r = 0\), an event of probability 0) so that \(M^{-1}(O) = (U, |r|) \equiv (\tau, \phi, |r|)\) is essentially well-defined and the trace of \(O\) yields

\[
|r| = \arccos\left(2^{-1}(tr(O) - 1)\right). \tag{A.2}
\]

In the particular case where \(|r|\) has a Lebesgue density \((1 - \cos |r|)/\pi\), we obtain a distribution \(H^*\) on \([-1, 1] \times [0, 2\pi] \times [0, \pi]\) for \((\tau, \phi, |r|)\), which induces the Haar measure \(H\) on \(\Omega\) via (A.1); see Miles (1965). Therefore, substituting (A.2) into

\[
\frac{dF^*}{dH^*} = \frac{dF^*}{d\lambda} \frac{d\lambda}{dH^*} = \frac{2\pi C(|r|/\kappa)}{1 - \cos |r|}, \quad \tau \in [-1, 1], \ \phi \in [0, 2\pi], \ |r| \in [0, \pi] \tag{A.3}
\]

gives the density \(f = dF/dH\) of the \(UARS(I_{3 \times 3}, \kappa)\) distribution on \(\Omega\) with respect to \(H\) as

\[
f(o|\kappa) = \frac{4\pi}{3 - tr(o)} C\left(\arccos\left(2^{-1}(tr(o) - 1)\right)\right), \quad o \in \Omega.
\]

That is, with probability space \((\Omega, \mathcal{F}, F)\) and \(A \in \mathcal{F}\),

\[
P_F(O \in A) = P_F^\tau((\tau, \phi, |r|) \in M^{-1}(A)) = \int_{(\tau,\phi,|r|)\in M^{-1}(A)} \frac{dF^*}{dH^*}dH^* = \int_{o \in A} f(o)dH
\]

where \(f(o) = dF^*(M^{-1}(o))/dH^*\) and \(P_H(A) = P_{H^*}(M^{-1}(A))\).

More generally, a density for \(O \sim UARS(S, \kappa)\) is given by

\[
f(o|S, \kappa) = \frac{4\pi}{3 - tr(S^T o)} C\left(\arccos\left(2^{-1}(tr(S^T o) - 1)\right)\right), \quad o \in \Omega,
\]
with respect to $H$, which follows from $S^T \cdot O \sim \text{UARS}(I_{3 \times 3}, \kappa)$ and the invariance of the Haar measure (i.e., $P_H(A) = P_H(S^T A)$).

Appendix B: Proofs of UARS properties from the main paper’s appendix

In what follows, let $U = (u, v, w)^T$ be uniformly distributed on the sphere, independently of $r \sim \text{Circ}(\kappa)$.

Proof of Property 1. Recall that $O = M(U, r) \sim \text{UARS}(I_{3 \times 3}, \kappa)$ and note that $O^T = M(-U, r)$ by (2). Since $U$ is uniformly distributed on the sphere and is independent of $r$, so is $-U$. Thus, $O^T = M(-U, r) \overset{d}{=} M(U, r) \sim \text{UARS}(I_{3 \times 3}, \kappa)$. □

Proof of Property 2. We note the matrix identity

$$S \cdot M(U, r) \cdot S^T = M(S \cdot U, r)$$

can be proven by considering $S = g(\alpha, \beta, \gamma)$ as in (8). Since $U$ is uniform on the sphere and is independent of $r$, so is $S \cdot U$ and

$$O = M(U, r) \overset{d}{=} M(S \cdot U, r) = S \cdot O \cdot S^T.$$  \hspace{1cm} (B.1)

Proof of Property 3. By definition $S \cdot O \sim \text{UARS}(S, \kappa)$. By (B.1), $O \overset{d}{=} S \cdot O \cdot S^T$. Multiplying by $S$ on the right gives $O \cdot S \overset{d}{=} S \cdot O \sim \text{UARS}(S, \kappa)$. □

Proof of Property 4. $S^T \cdot O \sim \text{UARS}(I_{3 \times 3}, \kappa)$ so that $O^T \cdot S \sim \text{UARS}(I_{3 \times 3}, \kappa)$ by Property 1. Then $O^T = (O^T \cdot S) \cdot S^T \sim \text{UARS}(S^T, \kappa)$ by Property 3. □

Proof of Property 5. By definition $S^T \cdot O \sim \text{UARS}(I_{3 \times 3}, \kappa)$. By (B.1), $S \cdot (S^T \cdot O) \cdot S^T \overset{d}{=} S^T \cdot O$. Thus $O \cdot S^T \sim \text{UARS}(I_{3 \times 3}, \kappa)$. □
Proof of Property 6. As in (B.1), \( \mathbf{O} = M(\mathbf{U}, r) \overset{d}{=} M(\mathbf{QU}, r) \) for any \( 3 \times 3 \) orthogonal rotation matrix \( \mathbf{Q} \). Let
\[
\mathbf{Q} = \begin{pmatrix}
0 & -1 & 0 \\
-1 & 0 & 0 \\
0 & 0 & -1
\end{pmatrix}.
\]
Then \( \mathbf{QU} = (-v, -u, -w)^T \), and
\[
M(\mathbf{QU}, r) = \begin{pmatrix}
v^2 + (u^2 + w^2) \cos r & uv(1 - \cos r) + w \sin r & vw(1 - \cos r) - u \sin r \\
v(1 - \cos r) - w \sin r & u^2 + (v^2 + w^2) \cos r & u(1 - \cos r) + v \sin r \\
vw(1 - \cos r) + u \sin r & u(1 - \cos r) - v \sin r & w^2 + (u^2 + v^2) \cos r
\end{pmatrix}.
\]
Now, to compare \( P_X \) with \( P_Y \) we must rotate \( Y \) to also be about the vector \((1, 0, 0)^T\). We can do this by taking \( \mathbf{RY} \) where
\[
\mathbf{R} = \begin{pmatrix}
0 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 1
\end{pmatrix}.
\]
Thus, \( \mathbf{RY} = (v^2 + (u^2 + w^2) \cos r, uv(1 - \cos r) - w \sin r, vw(1 - \cos r) + u \sin r)^T \) is distributed about \((1, 0, 0)^T\). But, \( \mathbf{RY} \) is the same as the first column of \( M(\mathbf{QU}, r) \overset{d}{=} M(\mathbf{U}, r) = \mathbf{O} \). Thus, \( \mathbf{RY} \overset{d}{=} \mathbf{X} \) and \( P_X = P_Y \). The fact that \( P_Y = P_Z \) can be established in a similar manner by using
\[
\mathbf{Q} = \begin{pmatrix}
-1 & 0 & 0 \\
0 & 0 & -1 \\
0 & -1 & 0
\end{pmatrix} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{pmatrix}.
\]

Proof of Property 7. We write \( \mathbf{O} = M(\mathbf{U}, r) \) for \( \mathbf{U} \) uniformly distributed on the sphere. Expressing \( \mathbf{U} \) in terms of \( \theta \) and \( \phi \) as in (1), and using the fact that \( \theta, \phi, \) and \( r \) are independently distributed, we have \( E(\mathbf{uv}) = E(\mathbf{uw}) = E(\mathbf{vw}) = 0 \), \( E(u^2) = E(v^2) = E(w^2) = \frac{1}{3} \), \( E(\sin r) = 0 \), and \( E(\cos r) = \frac{I_1(\kappa)}{I_0(\kappa)} \). The first expectation follows.
Now, for \( j_1, j_2, j_3, j_4 \in \{1, 2, 3\} \), we find

\[
E(O_{j_1j_2}^2) = \begin{cases} 
\frac{1}{5} + \frac{4}{15} \cdot \frac{I_1(\kappa)}{I_0(\kappa)} + \frac{8}{15} \cdot \frac{I_2(\kappa) + \frac{1}{\kappa} I_1(\kappa)}{I_0(\kappa)} & \text{for } j_1 = j_2 \\
\frac{1}{15} + \frac{I_2(\kappa) + (\frac{1}{\kappa} - 2)I_1(\kappa)}{15I_0(\kappa)} + \frac{I_1(\kappa)}{3\kappa I_0(\kappa)} & \text{for } j_1 \neq j_2
\end{cases}
\]

\[
E(O_{12} O_{21}) = E(O_{23} O_{32}) = E(O_{13} O_{31}) = \frac{1}{15} + \frac{I_2(\kappa) + (\frac{1}{\kappa} - 2)I_1(\kappa)}{15I_0(\kappa)} - \frac{I_1(\kappa)}{3\kappa I_0(\kappa)},
\]

and

\[
E(O_{11} O_{22}) = E(O_{11} O_{33}) = E(O_{22} O_{33}) = \frac{1}{15} + \frac{8}{15} \cdot \frac{I_1(\kappa)}{I_0(\kappa)} + \frac{2}{5} \cdot \frac{I_2(\kappa) + \frac{1}{\kappa} I_1(\kappa)}{I_0(\kappa)},
\]

while all other terms of the form \( E(O_{j_1j_2} O_{j_3j_4}) \) are zero.

Let \( P = O^2 \). Then

\[
E(P_{11}) = E(O_{11}^2 + O_{12} O_{21} + O_{13} O_{31}) = \frac{1}{3} + \frac{2}{3} \left( \frac{I_2(\kappa) + \frac{1}{\kappa} I_1(\kappa)}{I_0(\kappa)} - \frac{I_1(\kappa)}{\kappa I_0(\kappa)} \right) = \frac{1}{3} + \frac{2 I_2(\kappa)}{3 I_0(\kappa)}.
\]

Finding all other entries of \( P \) in a similar manner gives \( E(P_{22}) = E(P_{33}) = E(P_{11}) \) and \( E(P_{ij}) = 0 \) for all \( i, j \in \{1, 2, 3\}, i \neq j \). The second expectation follows.

\[\square\]

**Proof of Property 8.** First, we have that \( x = O_{11}, y = O_{22}, \) and \( z = O_{33} \). From the proof of Property 7,

\[
E(O_{11} O_{22}) = E(O_{11} O_{33}) = E(O_{22} O_{33}) = \frac{1}{15} + \frac{8}{15} \cdot \frac{I_1(\kappa)}{I_0(\kappa)} + \frac{2}{5} \cdot \frac{I_2(\kappa) + \frac{1}{\kappa} I_1(\kappa)}{I_0(\kappa)}
\]

and

\[
E(O_{11}^2) = E(O_{22}^2) = E(O_{33}^2) = \frac{1}{5} + \frac{4}{15} \cdot \frac{I_1(\kappa)}{I_0(\kappa)} + \frac{8}{15} \cdot \frac{I_2(\kappa) + \frac{1}{\kappa} I_1(\kappa)}{I_0(\kappa)}.
\]

Also, by Property 7,

\[
E(O_{11}) = E(O_{22}) = E(O_{33}) = \frac{1}{3} + \frac{2 I_1(\kappa)}{3 I_0(\kappa)}.
\]

Therefore, the result follows. \[\square\]
Proof of Property 9. Since \( \mathbf{O} \sim \text{vM-UARS}(\mathbf{S}, \kappa) \), we can write \( \mathbf{O} = \mathbf{S} \cdot \mathbf{P} \) where \( \mathbf{P} \sim \text{vM-UARS}(\mathbf{I}_{3 \times 3}, \kappa) \) and \( \mathbf{S} = g(\alpha, \beta, \gamma) \) as in (8). Now, by Property 7,

\[
E(\mathbf{O}) = E(\mathbf{S} \cdot \mathbf{P}) = \mathbf{S} \cdot E(\mathbf{P}) = \left( \frac{1}{3} + \frac{2}{3} \frac{I_1(\kappa)}{I_0(\kappa)} \right) \cdot \mathbf{S}.
\]

Then,

\[
E \left( \frac{\partial l}{\partial \kappa} \right) = E \left( \frac{1}{2} \left[ P_{11} + P_{22} + P_{33} - 1 \right] - \frac{I_1(\kappa)}{I_0(\kappa)} \right)
\]

\[
= \frac{1}{2} \left[ 3 \left( \frac{1}{3} + \frac{2}{3} \frac{I_1(\kappa)}{I_0(\kappa)} \right) - 1 \right] - \frac{I_1(\kappa)}{I_0(\kappa)} = 0
\]

and

\[
E \left( \frac{\partial l}{\partial \alpha} \right) = \frac{\kappa}{2} E \left( \frac{\partial P_{11}}{\partial \alpha} + \frac{\partial P_{22}}{\partial \alpha} + \frac{\partial P_{33}}{\partial \alpha} \right)
\]

where

\[
P_{11} = s_{11} \cdot O_{11} + s_{21} \cdot O_{21} + s_{31} \cdot O_{31}
\]

\[
P_{22} = s_{12} \cdot O_{12} + s_{22} \cdot O_{22} + s_{32} \cdot O_{32}
\]

\[
P_{33} = s_{13} \cdot O_{13} + s_{23} \cdot O_{23} + s_{33} \cdot O_{33}
\]

and \( s_{ij} \) are the elements of \( \mathbf{S} \). By calculating each of the partial derivatives, we find that

\[
E \left( \frac{\partial l}{\partial \alpha} \right) = \frac{\kappa}{2} E \left( -s_{12} \cdot O_{11} - s_{22} \cdot O_{21} - s_{32} \cdot O_{31} + s_{11} \cdot O_{12} + s_{21} \cdot O_{22} + s_{31} \cdot O_{32} \right)
\]

\[
= \frac{\kappa}{2} \left( \frac{1}{3} + \frac{2}{3} \frac{I_1(\kappa)}{I_0(\kappa)} \right) (-s_{12}s_{11} - s_{22}s_{21} - s_{32}s_{31} + s_{11}s_{12} + s_{21}s_{22} + s_{31}s_{32}) = 0.
\]

In a similar manner it can be shown that \( E \left( \frac{\partial l}{\partial \beta} \right) = 0 \) and \( E \left( \frac{\partial l}{\partial \gamma} \right) = 0 \).

Proof of Property 10. Again, since \( \mathbf{O} \sim \text{vM-UARS}(\mathbf{S}, \kappa) \), we can write \( \mathbf{O} = \mathbf{S} \cdot \mathbf{P} \) where \( \mathbf{P} \sim \text{vM-UARS}(\mathbf{I}_{3 \times 3}, \kappa) \). As in the proof of Property 9, \( E(\mathbf{O}) = D(\kappa) \cdot \mathbf{S} \). Now, we must calculate \( l''(\kappa, (\alpha, \beta, \gamma)) \). First, we note that

\[
\frac{\partial^2 l}{\partial \kappa^2} = - \frac{I_2(\kappa) + \frac{1}{\kappa} I_1(\kappa)}{I_0(\kappa)^2} + \left( \frac{I_1(\kappa)}{I_0(\kappa)} \right)^2.
\]

Second, let \( \eta, \psi \in \{\alpha, \beta, \gamma\} \). Then,

\[
\frac{\partial^2 l}{\partial \kappa \partial \eta} = \frac{1}{2} \left( \frac{\partial P_{11}}{\partial \eta} + \frac{\partial P_{22}}{\partial \eta} + \frac{\partial P_{33}}{\partial \eta} \right).
\]
and
\[
\frac{\partial^2 l}{\partial \eta \partial \psi} = \frac{\kappa}{2} \left( \frac{\partial^2 P_{11}}{\partial \eta \partial \psi} + \frac{\partial^2 P_{22}}{\partial \eta \partial \psi} + \frac{\partial^2 P_{33}}{\partial \eta \partial \psi} \right),
\]
where \( P_{11}, P_{22}, \) and \( P_{33} \) are as in (B.2).

We will provide the details of finding \( E \left( \frac{\partial^2 l}{\partial \alpha \partial \gamma} \right) \) and leave the other terms to the reader.

By (B.3),
\[
E \left( \frac{\partial^2 l}{\partial \alpha \partial \gamma} \right) = \frac{\kappa}{2} \left[ E \left( \frac{\partial^2 P_{11}}{\partial \alpha \partial \gamma} \right) + E \left( \frac{\partial^2 P_{22}}{\partial \alpha \partial \gamma} \right) + E \left( \frac{\partial^2 P_{33}}{\partial \alpha \partial \gamma} \right) \right].
\]

Now,
\[
E \left( \frac{\partial^2 P_{11}}{\partial \alpha \partial \gamma} \right) = -s_{22} \cdot E(O_{11}) + s_{12} \cdot E(O_{21}) + 0 \cdot E(O_{31}) = (-s_{22}s_{11} + s_{12}s_{21}) D(\kappa),
\]
\[
E \left( \frac{\partial^2 P_{22}}{\partial \alpha \partial \gamma} \right) = s_{21} \cdot E(O_{12}) - s_{11} \cdot E(O_{22}) + 0 \cdot E(O_{32}) = (s_{21}s_{12} - s_{11}s_{22}) D(\kappa),
\]
and
\[
E \left( \frac{\partial^2 P_{33}}{\partial \alpha \partial \gamma} \right) = 0.
\]
Thus,
\[
E \left( \frac{\partial^2 l}{\partial \alpha \partial \gamma} \right) = \frac{\kappa}{2} \left[ 2(s_{12}s_{21} - s_{11}s_{22}) D(\kappa) \right] = \kappa(s_{12}s_{21} - s_{11}s_{22}) D(\kappa).
\]

Now,
\[
s_{12}s_{21} - s_{11}s_{22} = (\sin \alpha \cos \gamma + \cos \alpha \sin \gamma \cos \beta)(- \cos \alpha \sin \gamma - \sin \alpha \cos \gamma \cos \beta)
\]
\[
- (\cos \alpha \cos \gamma - \sin \alpha \sin \gamma \cos \beta)(- \sin \alpha \sin \gamma + \cos \alpha \cos \gamma \cos \beta)
\]
\[
= -(\cos^2 \alpha \sin^2 \gamma + \sin^2 \alpha \cos^2 \gamma + \sin^2 \alpha \sin^2 \gamma + \cos^2 \alpha \cos^2 \gamma) \cos \beta
\]
\[
= - \cos \beta
\]
Therefore,
\[
- E \left( \frac{\partial^2 l}{\partial \alpha \partial \gamma} \right) = \kappa \cos \beta \cdot D(\kappa). \]

Proof of Property 11. By Property 9, \( \text{Var}(l'(\kappa, (\alpha, \beta, \gamma))) = E(l'(\kappa, (\alpha, \beta, \gamma))^2) \). Again, since \( O \sim \nuM-UARS(S, \kappa) \), we can write \( O = S \cdot P \) where \( P \sim \nuM-UARS(I_{3 \times 3}, \kappa) \). Recall that we can express \( P \) as \( P = M(U, r) \) and in the proof of Property 7 we found \( E(P_{j_1j_2}P_{j_3j_4}) \) for all \( j_1, j_2, j_3, j_4 \in \{1, 2, 3\} \).

We present the details required for finding just one term of \( E(l'(\kappa, (\alpha, \beta, \gamma))^2) \). Others may be found similarly. We will calculate \( E \left( \left( \frac{\partial l}{\partial \alpha} \right)^2 \right) \).
First,
\[
E \left( \left( \frac{\partial l}{\partial \alpha} \right)^2 \right) = \kappa^2 \frac{4}{4} E \left( \left( \frac{\partial P_{11}}{\partial \alpha} + \frac{\partial P_{22}}{\partial \alpha} + \frac{\partial P_{33}}{\partial \alpha} \right)^2 \right),
\]
where \( P_{11}, P_{22}, \) and \( P_{33} \) are as in (B.2). Using the proof of Property 9,
\[
E \left( \left( \frac{\partial l}{\partial \alpha} \right)^2 \right) = \kappa^2 \frac{4}{4} E \left( -s_{12} \cdot O_{11} - s_{22} \cdot O_{21} - s_{32} \cdot O_{31} + s_{11} \cdot O_{12} + s_{21} \cdot O_{22} + s_{31} \cdot O_{32} \right)^2.
\]
Expanding, we have
\[
E \left( \left( \frac{\partial l}{\partial \alpha} \right)^2 \right) = \kappa^2 \frac{4}{4} \left( s_{12}^2 E(O_{11}^2) + s_{22}^2 E(O_{21}^2) + s_{32}^2 E(O_{31}^2) \right.
\]
\[
+ s_{11}^2 E(O_{12}^2) + s_{21}^2 E(O_{22}^2) + s_{31}^2 E(O_{32}^2) \]
\[
+ 2s_{12}s_{22} E(O_{11}O_{21}) + 2s_{12}s_{32} E(O_{11}O_{31}) - 2s_{12}s_{11} E(O_{11}O_{12}) \]
\[
- 2s_{12}s_{21} E(O_{11}O_{22}) - 2s_{12}s_{31} E(O_{11}O_{32}) + 2s_{22}s_{32} E(O_{21}O_{32}) \]
\[
- 2s_{22}s_{11} E(O_{21}O_{12}) - 2s_{22}s_{21} E(O_{21}O_{22}) - 2s_{22}s_{31} E(O_{21}O_{32}) \]
\[
- 2s_{32}s_{11} E(O_{31}O_{12}) - 2s_{32}s_{21} E(O_{31}O_{22}) - 2s_{32}s_{31} E(O_{31}O_{32}) \]
\[
+ 2s_{11}s_{21} E(O_{12}O_{22}) + 2s_{11}s_{31} E(O_{12}O_{32}) \]
\[
\left. + 2s_{21}s_{31} E(O_{22}O_{32}) \right) \quad \text{(B.4)}
\]
Now, for any \( j, k \in \{1, 2, 3\}, \)
\[
E(O_{j1}O_{k2}) = E((s_{j1}P_{11} + s_{j2}P_{21} + s_{j3}P_{31})(s_{k1}P_{12} + s_{k2}P_{22} + s_{k3}P_{32}))
\]
\[
= s_{j2}s_{k1}E(P_{21}P_{12}) + s_{j1}s_{k2}E(P_{11}P_{22}),
\]
\[
E(O_{j1}O_{k1}) = s_{j1}s_{k1}E(P_{11}^2) + s_{j2}s_{k2}E(P_{21}^2) + s_{j3}s_{k3}E(P_{31}^2)
\]
\[
E(O_{j2}O_{k2}) = s_{j1}s_{k1}E(P_{12}^2) + s_{j2}s_{k2}E(P_{22}^2) + s_{j3}s_{k3}E(P_{32}^2) \quad \text{(B.5)}
\]
By placing the quantities from (B.5) into (B.4) and simplifying, we find

\[
E\left(\left(\frac{\partial l}{\partial \alpha}\right)^2\right) = \frac{\kappa^2}{4} (E(P^2_{11})(s_{11}^2 + s_{21}^2 + s_{31}^2))^2 + E(P^2_{21})(s_{12}^2 + s_{22}^2 + s_{32}^2)^2 + E(P^2_{12})(s_{11}^2 + s_{21}^2 + s_{31}^2)^2 + E(P^2_{22})(s_{12}^2 + s_{22}^2 + s_{32}^2)^2
\]

\[
+2E(P_{21}P_{22})(s_{12}^2 + s_{22}^2 + s_{32}^2)(s_{11}^2 + s_{21}^2 + s_{31}^2) + 2E(P_{12}P_{22})(s_{12}^2 + s_{22}^2 + s_{32}^2)
\]

\[
= \frac{\kappa^2}{4} (E(P^2_{21}) + E(P^2_{12}) - 2E(P_{21}P_{12})) = \frac{\kappa^2}{4} \left( \frac{4I_1(\kappa)}{3\kappa I_0(\kappa)} \right) = \frac{\kappa I_1(\kappa)}{3I_0(\kappa)}.
\]

\[\Box\]

Appendix C: Proofs of propositions from Section 3.3

Proof of Proposition 1. By Taylor expansion we have

\[
0 = l'_n(\hat{\theta}_n) \approx l'_n(\theta_0) + l''_n(\theta_0)(\hat{\theta}_n - \theta_0)
\]

Since \(\frac{1}{\sqrt{n}}l'_n(\theta_0) \xrightarrow{d} \text{MVN}(0, \mathcal{I}_1(\theta_0))\) and \(-\frac{1}{n}l''_n(\theta_0) \xrightarrow{p} H_1(\theta_0), \) (positive definite), we have

\[
 \sqrt{n}(\hat{\theta}_n - \theta_0) \approx \left[\frac{1}{\sqrt{n}}l''_n(\theta_0)\right]^{-1} \frac{1}{\sqrt{n}}l'_n(\theta_0) \xrightarrow{d} \text{MVN}(0, H^{-1}_1(\theta_0)\mathcal{I}_1(\theta_0)H^{-1}_1(\theta_0)).
\]

\[\Box\]

Proof of Proposition 2. First we write \(-2\log(\lambda_n) = 2\left[l_n(\hat{\theta}_n) - l_n(\theta^*_n)\right]\) where \(\hat{\theta}_n\) is the maximum quasi-likelihood estimate over \(\Theta\) and \(\theta^*_n\) is the maximum quasi-likelihood estimate over \(\Theta_0\). By a Taylor expansion of \(l_n(\theta^*_n)\) about \(\hat{\theta}_n\), we get

\[
l_n(\theta^*_n) \approx l_n(\hat{\theta}_n) + l'_n(\hat{\theta}_n)(\theta^*_n - \hat{\theta}_n) + \frac{1}{2}(\theta^*_n - \hat{\theta}_n)^Tl''_n(\hat{\theta}_n)(\theta^*_n - \hat{\theta}_n)
\]
Now, \(-\frac{1}{n} l''_n(\hat{\theta}_n) \overset{p}{\to} H_1(\theta_0)\) and, since \(\hat{\theta}_n\) is the maximum quasi-likelihood estimate over \(\Theta\), \(l'_n(\hat{\theta}_n) = 0\). So,

\[
2 \left[ l_n(\hat{\theta}_n) - l_n(\theta_n^*) \right] \approx -2 l'_n(\hat{\theta}_n)(\theta_n^* - \hat{\theta}_n) - (\theta_n^* - \hat{\theta}_n)^T l''_n(\hat{\theta}_n)(\theta_n^* - \hat{\theta}_n)
\]

\[
\approx n(\theta_n^* - \hat{\theta}_n)^T \left[ -\frac{1}{n} l''_n(\hat{\theta}_n) \right] (\theta_n^* - \hat{\theta}_n)
\]

\[
\approx n(\theta_n^* - \hat{\theta}_n)^T H_1(\theta_0)(\theta_n^* - \hat{\theta}_n).
\]

(C.2)

Next, we expand \(l'_n(\theta_n^*)\) about \(\hat{\theta}_n\) giving

\[
\frac{1}{\sqrt{n}} l'_n(\theta_n^*) \approx \frac{1}{\sqrt{n}} l'_n(\hat{\theta}_n) + \frac{1}{n} l''_n(\hat{\theta}_n) \sqrt{n}(\theta_n^* - \hat{\theta}_n),
\]

or

\[
\sqrt{n}(\theta_n^* - \hat{\theta}_n) \approx \left[ \frac{1}{n} l''_n(\hat{\theta}_n) \right]^{-1} \frac{1}{\sqrt{n}} l'_n(\theta_n^*) \approx -H_1^{-1}(\theta_0) \frac{1}{\sqrt{n}} l'_n(\theta_n^*).
\]

By (C.2)

\[
-2 \log(\lambda_n) \approx \frac{1}{\sqrt{n}} l'_n(\theta_n^*) H_1^{-1}(\theta_0) \frac{1}{\sqrt{n}} l'_n(\theta_n^*).
\]

(C.3)

Now, we expand \(l'_n(\theta_n^*)\) about \(\theta_0\), yielding

\[
\frac{1}{\sqrt{n}} l'_n(\theta_n^*) \approx \frac{1}{\sqrt{n}} l'_n(\theta_0) + \frac{1}{n} l''_n(\theta_0) \sqrt{n}(\theta_n^* - \theta_0)
\]

\[
\approx \frac{1}{\sqrt{n}} l'_n(\theta_0) - H_1(\theta_0) \sqrt{n}(\theta_n^* - \theta_0).
\]

(C.4)

Now, partition \(H_1(\theta_0)\) into \(H_1(\theta_0) = \begin{bmatrix} G_1 & G_2 \\ G_2^T & G_3 \end{bmatrix}\) where \(G_1\) is \(r \times r\) and let \(J = \begin{bmatrix} 0 & 0 \\ 0 & G_3^{-1} \end{bmatrix}\).

Since \(\theta_n^*\) is the maximum quasi-likelihood estimate over \(\Theta_0\), we have \(\theta_n^* = \begin{pmatrix} \theta_{01} \\ \theta_2^* \end{pmatrix}\) where \(\theta_2^*\) maximizes the quasi-likelihood under the restriction that \(\theta_1 = \theta_{01}\). Thus, the last \(4-r\) components of \(l'_n(\theta_n^*)\) are 0, so that \(J \cdot l'_n(\theta_n^*) = 0\). Then, by (C.4),

\[
\frac{1}{\sqrt{n}} J \cdot l'_n(\theta_n^*) \approx \frac{1}{\sqrt{n}} J \cdot l'_n(\theta_0) - J H_1(\theta_0) \sqrt{n}(\theta_n^* - \theta_0),
\]

implying that

\[
\frac{1}{\sqrt{n}} J \cdot l'_n(\theta_0) \approx J H_1(\theta_0) \sqrt{n}(\theta_n^* - \theta_0) = \sqrt{n}(\theta_n^* - \theta_0),
\]
which follows from \((\theta_n^* - \theta_0) = \begin{pmatrix} 0 \\ \theta_n^* - \theta_{02} \end{pmatrix}\) and \(JH_1(\theta_0) = \begin{bmatrix} 0 & 0 \\ G_3^{-1} G_2^T & I_{3 \times 3} \end{bmatrix}\) so that

\[
JH_1(\theta_0)\sqrt{n}(\theta_n^* - \theta_0) = \sqrt{n} \begin{pmatrix} 0 \\ \theta_n^* - \theta_{02} \end{pmatrix} = \sqrt{n}(\theta_n^* - \theta_0). \text{ Thus, by (C.4),}
\]

\[
\frac{1}{\sqrt{n}} J_n'((\theta_n^* - \theta_0)) \approx \frac{1}{\sqrt{n}} J_n'((\theta_0)) - H_1(\theta_0)J \frac{1}{\sqrt{n}} J_n'((\theta_0)) = [I_{3 \times 3} - H_1(\theta_0)J] \frac{1}{\sqrt{n}} J_n'((\theta_0)). \quad (C.5)
\]

By the Central Limit Theorem, \(\frac{1}{\sqrt{n}} J_n'((\theta_0)) \xrightarrow{d} Y\) where \(Y \sim \text{MVN}(0, I_1(\theta_0))\). Therefore, by (C.3) and (C.5), we have

\[
-2 \log(\lambda_n) \xrightarrow{d} Y^T[I_{3 \times 3} - H_1(\theta_0)J]^{-1}I_1(\theta_0)[I_{3 \times 3} - H_1(\theta_0)J]Y - 2JH_1(\theta_0)JY
\]

Now,

\[
JH(\theta_0)J = \begin{bmatrix} 0 & 0 \\ 0 & G_3^{-1} \end{bmatrix} \begin{bmatrix} G_1 & G_2 \\ G_2^T & G_3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & G_3^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ G_3^{-1} \end{bmatrix} = J,
\]

giving \(-2 \log(\lambda_n) \xrightarrow{d} Y^T[I_1(\theta_0)^{-1} - J]Y\). Let \(Z = I_1(\theta_0)^{-\frac{1}{2}}Y\). Then, \(Z \sim \text{MVN}(0, I_{3 \times 3})\) and

\[
-2 \log(\lambda_n) \xrightarrow{d} Z^T I_1(\theta_0)^{-\frac{1}{2}}[H_1^{-1}(\theta_0) - J]I_1(\theta_0)^{-\frac{1}{2}}Z
\]

\[
= Z^T P(\theta_0)Z, \text{ where } P(\theta_0) = I_1(\theta_0)^{-\frac{1}{2}}[H_1^{-1}(\theta_0) - J]I_1(\theta_0)^{-\frac{1}{2}}.
\]

Now, it can be calculated that

\[
P(\theta_0) = \begin{cases} \text{diag}(1, B(\kappa_0), B(\kappa_0), B(\kappa_0)) \text{ for } \theta_1 = \theta \\ \text{diag}(0, B(\kappa_0), B(\kappa_0), B(\kappa_0)) \text{ for } \theta_1^T = (\alpha, \beta, \gamma) \\ \text{diag}(1, 0, 0, 0) \text{ for } \theta_1 = \kappa \end{cases}
\]

Therefore,

\[
-2 \log(\lambda_n) \xrightarrow{d} \begin{cases} \chi_1^2 + B(\kappa_0)\chi_3^2 \text{ for } \theta_1 = \theta \\ B(\kappa_0)\chi_3^2 \text{ for } \theta_1^T = (\alpha, \beta, \gamma) \\ \chi_1^2 \text{ for } \theta_1 = \kappa \end{cases}
\]

\(\square\)
Proof of Proposition 3. This follows directly from Proposition 1 that states
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \text{MVN}(0, H^{-1}_1(\theta_0)I_1(\theta_0)H^{-1}_1(\theta_0)). \]
\[ \square \]

Proof of Proposition 4. By expanding \( l'_n(\theta_0) \) about \( \hat{\theta}_n \), we have
\[ \frac{1}{\sqrt{n}}l'_n(\theta_0) \approx \frac{1}{\sqrt{n}}l'_n(\hat{\theta}_n) + \frac{1}{n}l''_n(\hat{\theta}_n)\sqrt{n}(\theta_0 - \hat{\theta}_n) \]
\[ \xrightarrow{d} \text{MVN}(0, I_1(\theta_0)) + I_1(\theta_0)^{\frac{1}{2}}\delta \]

Let \( Y \sim \text{MVN}(I_1(\theta_0)^{\frac{1}{2}}\delta, I_1(\theta_0)) \). By modifying the proof of Proposition 2, we have
\[ -2\log(\lambda_n) \xrightarrow{d} Y^T[H^{-1}_1(\theta_0) - J]Y \]
\[ = Z^TI_1(\theta_0)^{\frac{1}{2}}[H^{-1}_1(\theta_0) - J]I_1(\theta_0)^{\frac{1}{2}}Z \text{ where } Z = I_1(\theta_0)^{-\frac{1}{2}}Y \]
\[ = Z^TP(\theta_0)Z, \]
where \( P(\theta_0) = \text{diag}(1, B(\kappa_0), B(\kappa_0), B(\kappa_0)) \) and
\[ Z^T = (Z_1, Z_2, Z_3, Z_4) \sim \text{MVN}(\delta, I_{4 \times 4}). \]

Now
\[ Z^TP(\theta_0)Z = Z_1^2 + B(\kappa_0)(Z_2^2 + Z_3^2 + Z_4^2) \]
and \( Z_1^2 \sim \chi^2(\delta^2_\kappa) \) and \( Z_2^2 + Z_3^2 + Z_4^2 \sim \chi^2(\delta^2_\alpha + \delta^2_\beta + \delta^2_\gamma) \) are independent. \[ \square \]

Proof of Proposition 5. As in the proof of Proposition 4, \( \frac{1}{\sqrt{n}}l'_n(\theta_0) \xrightarrow{d} Y \), where
\[ Y \sim \text{MVN}(H^{-1}_1(\theta_0)I_1(\theta_0)^{\frac{1}{2}}\delta, I_1(\theta_0)), \]
so that
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \text{MVN}(H^{-1}_1(\theta_0)I_1(\theta_0)^{\frac{1}{2}}\delta, H^{-1}_1(\theta_0)I_1(\theta_0)H^{-1}_1(\theta_0)) \]
and hence,
\[ T_n \xrightarrow{d} Z^T Z \text{ where } Z \sim \text{MVN}(\delta, I_{4 \times 4}) \]
\[ \sim \chi^2(\delta^2_\kappa + \delta^2_\alpha + \delta^2_\beta + \delta^2_\gamma). \]
\[ \square \]
Appendix D: Q-Q plots for aluminum EBSD data

Q-Q plots for the fitted misorientation angles obtained from EBSD observations, using the von Mises circular distribution, are presented in Figure D.1 for two sets of aluminum data. The data used here are seven repeat measurements from a single location on the aluminum scans and 50 measurements appearing to be within a single grain from a single scan, respectively. These plots demonstrate that the von Mises distribution appropriately describes the aluminum data. (Figures 2 and 3 of the main manuscript showed goodness-of-fit for nickel EBSD data.)

Figure D.1: Q-Q plot for the fitted misorientation angles obtained from two sets of aluminum EBSD observations, using the von Mises circular distribution on the misorientation angle
BAYES ONE-SAMPLE AND ONE-WAY RANDOM EFFECTS ANALYSES FOR 3-D ORIENTATIONS WITH APPLICATION TO MATERIALS SCIENCE

A paper submitted to Bayesian Analysis

Melissa A. Bingham\textsuperscript{1}, Stephen B. Vardeman, and Daniel J. Nordman

Iowa State University

Abstract

We consider Bayes inference for a class of distributions on random orientations in 3 dimensions described by $3 \times 3$ rotation matrices. Non-informative priors are identified and Metropolis-Hasting within Gibbs algorithms are used to generate samples from posterior distributions in one-sample and one-way random effects scenarios. A simulation study investigates the performance of Bayes analyses in the one-sample case and includes comparisons to quasi-likelihood inference. Bayes one-way random effect analyses of orientation matrix data are then developed and the Bayes methods are illustrated in a materials science application.

Key words: Bayes, credible intervals, Electron Backscatter Diffraction, Gibbs, Markov chain Monte Carlo, Metropolis-Hastings, one-way random effects model, orthogonal matrix, posterior density, prior distribution, UARS distribution

\textsuperscript{1}Supported by NSF grant DMS \#0502347 EMSW21-RTG awarded to the Department of Statistics, Iowa State University
1 Introduction

This paper presents Bayes methods of inference for 3-dimensional orientations. Random observations $O_1, \ldots, O_n$ in the form of $3 \times 3$ orthogonal rotation matrices that preserve the right hand rule are common in many fields, including vectorcardiography (Downs, 1972) and human kinetics (Rancourt, Rivest, and Asselin, 2000). Such data are also important in materials science applications and arise as the output of an Electron Backscatter Diffraction (EBSD) machine. EBSD is used to study the microtexture of crystalline materials, such as metals (see Randle, 2003). An EBSD camera, coupled with a Scanning Electron Microscope, produces diffraction patterns (when atomic planes within a target material diffract a stationary beam of electrons) which are converted to crystal orientations relative to some reference coordinate system. Each resulting observation, expressed as a $3 \times 3$ rotation matrix, gives the orientation of a cubic crystal at some scanned position on a metal surface. Bingham, Nordman, and Vardeman (2009b) describe a machine precision problem, where interest lies in quantifying the variation in orientation measurements obtained when EBSD is used repeatedly at a single location as well as the variation of measured orientations within a “grain” of scanned points on a metal specimen. (A grain is a group of observations which generally share a common orientation).

To prescribe a probability model for this problem, Bingham et. al. (2009b) identified the Uniform-Axis-Random-Spin (UARS) class of distributions on the set of $3 \times 3$ matrices, that we denote by $\Omega$. This UARS class is useful for modeling the deviation of random orientations $O_i \in \Omega$ from a common “central location” orientation of $S \in \Omega$, here referred to as the true or principal orientation. A random rotation $O \in \Omega$ from a UARS distribution with principal orientation $S$ can be written as $O = SP$, where

$$P = UU^T + (I_{3 \times 3} - UU^T) \cos r + \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \sin r \in \Omega$$

comes about by rotating $I_{3 \times 3}$ (the $3 \times 3$ identity matrix) about an axis identified by a point $U^T = (u_1, u_2, u_3) \in \mathbb{R}^3$ uniformly distributed on the unit sphere by a random angle $r \in (-\pi, \pi]$. 
The angle \( r \) is sometimes referred to as a “misorientation angle” (Randle, 2003), and is assumed to follow a circular distribution on \((-\pi, \pi]\) that is symmetric about 0 with spread depending on a concentration parameter \( \kappa \geq 0 \). Small realizations of \(|r|\) imply that a rotation \( O \) from a UARS model will deviate little from the location parameter \( S \). Since the parameter \( \kappa \) controls the spread of this circular distribution for \( r \), it consequently controls the variation of a corresponding UARS observation \( O \) from the true direction \( S \). Therefore, for a particular family of circular distributions for \( r \), the resulting UARS distributions on \( \Omega \) have “location” parameter \( S \in \Omega \) and “spread” parameter \( \kappa \in \mathbb{R} \) and can be denoted by \( \text{UARS}(S, \kappa) \).

In the literature, the most studied distribution for rotation matrices has been the matrix Fisher (or Langevin) distribution, introduced by Downs (1972) and further investigated by other authors (Khatri and Mardia, 1977; Prentice, 1986; Mardia and Jupp, 2000; Rancourt et al., 2000; Chikuse, 2003). Jupp and Mardia (1979) studied maximum likelihood estimation for this distribution, but the inference is not simple and the usual parameterization is not so easily interpreted, as that of the UARS models. For estimation of a location parameter \( S \), Chang and Rivest (2001) and Rivest and Chang (2006) have also discussed approaches involving M- and regression-type estimators. While some of these works have considered the large sample properties of maximum likelihood or moment estimators in some models, little consideration has been given to Bayes methods.

Motivated by the same materials science application, our goal is to investigate Bayes one-sample inference for the parameters of the \( \text{UARS}(S, \kappa) \) distribution and to further develop Bayes one-way random effects analyses for this model. The flexible two-parameter model for rotation matrices used here allows us to develop useful priors for several inference scenarios where the resulting Bayes estimators are extremely effective. For clarity in exposition and concreteness, we shall focus on one particular UARS model (called \( \text{vM-UARS}(S, \kappa) \) distributions), for which \( r \) follows the von Mises circular distribution with direction 0 and concentration parameter \( \kappa \geq 0 \) (denoted by \( \text{vM}(0, \kappa) \)). However, we point out that the Bayes methods introduced here can also be applied to other members of the UARS class (for example, see Bingham, Nordman, and Vardeman (2009a) for Bayes inference for the symmetric matrix von Mises-
Figure 1: Sample quantiles of misorientation angles for fifty measured crystal orientations against theoretical quantiles for spins using fitted von-Mises and Fisher distributions.

The vM(0, κ) distribution for r is unimodal on [−π, π], symmetric about 0, converges to the uniform distribution on the circle as κ → 0, and becomes approximately Normal with mean 0 and variance \( \frac{1}{\kappa} \) as κ → ∞; see Mardia and Jupp (2000).

To verify that the vM-UARS(S, κ) distribution provides an adequate model for the measured crystal orientations, fifty observations from within a single grain were fit using both the vM-UARS(S, κ) and matrix Fisher distributions (see Downs, 1972). For each observation the misorientation angle r was found after fitting the principal direction S with a standard moment estimator, described in Bingham et. al. (2009b). Figure 1 shows the sample quantiles plotted against the theoretical quantiles for each distribution. The quantile plots provide evidence that vM-UARS(S, κ) distribution describes the crystal data far better than the more standard matrix Fisher model.

In Section 2, we develop Bayes analyses for the one-sample case of the vM-UARS(S, κ) distribution. Priors are chosen for the parameters S and κ, the posterior distribution is given, and we outline a Markov chain Monte Carlo (MCMC) algorithm for generating samples from...
the posterior. A simulation study for the one-sample Bayes analysis is then presented. Section 3 outlines Bayes inference in one-way random effects models for rotation matrices, and Section 4 applies the Bayes methods in the analysis of EBSD measurements.

2 One-sample Bayes analyses for the vM-UARS(S, κ) distribution

2.1 One-sample model

We begin by choosing potentially non-informative prior distributions for the parameters of the vM-UARS(S, κ) distribution. For the location parameter S we use a prior distribution uniform on Ω. The invariant Haar measure acts as a “uniform distribution” on Ω (see Downs, 1972) so we adopt a prior distribution on S specified by density

\[ h_1(S) = 1, \quad S \in \Omega \]  

(1)

with respect to the Haar measure. This prior is equivalent to a UARS prior model for S where r has a density on [−π, π] proportional to \(1 - \cos r\) (Miles, 1965); see Downs (1972) for other characterizations of the Haar measure. For the spread parameter κ, we use a Jeffreys prior. Since κ is the concentration parameter from the von Mises circular distribution with density 

\[ v(r|κ) = [2πI_0(κ)]^{-1} \exp [κ \cos(r)], \quad r \in (-π, π), \]  

(3)

for \(I_i(κ)\) the modified Bessel function of order \(i\), we have

\[-E \left( \frac{d^2}{dκ^2} \log(v(r|κ)) \right) = \frac{I_0(κ)^2 - \frac{1}{κ}I_0(κ)I_1(κ) - I_1(κ)^2}{I_0(κ)^2}.\]

Therefore, the Jeffreys prior is specified by

\[ h_2(κ) = \sqrt{\frac{I_0(κ)^2 - \frac{1}{κ}I_0(κ)I_1(κ) - I_1(κ)^2}{I_0(κ)}}, \quad κ \in [0, \infty). \]

(2)

Now, the vM-UARS(S, κ) distribution has density

\[ f(o|S, κ) = \frac{2}{3 - tr(S^T o)} [I_0(κ)]^{-1} \exp \left[ \frac{κ}{2} (tr(S^T o) - 1) \right], \quad o \in Ω, \]  

(3)

with respect to the Haar measure (see Bingham et al., 2009b), giving likelihood

\[ L(S, κ) = \frac{2^n \exp \left[ \frac{κ}{2} \sum_{i=1}^{n} (tr(S^T o_i) - 1) \right]}{[I_0(κ)]^{n} \prod_{i=1}^{n} (3 - tr(S^T o_i))}. \]

(4)
Multiplying (1), (2), and (4) the posterior density is proportional to
\[
g(S, \kappa) = 2^p \exp \left[ \frac{\kappa}{2} \sum_{i=1}^{n} (tr(S^T o_i) - 1) \right] \frac{\sqrt{I_0(\kappa)^2 - \frac{1}{\kappa} I_0(\kappa) I_1(\kappa) - I_1(\kappa)^2}}{[I_0(\kappa)]^{n+1} \prod_{i=1}^{n} (3 - tr(S^T o_i))}.
\] (5)

To simulate values from the posterior distribution we use a Metropolis-Hastings within Gibbs algorithm. Suppose that after \(j - 1\) iterations of the algorithm one has parameters \(S^{j-1}\) and \(\kappa^{j-1}\). For the parameter \(S\), we obtain a candidate for \(S^j\) as \(S^j \sim vM-UARS(S^{j-1}, \rho)\), where \(\rho\) is a tuning parameter that can be adjusted to make the algorithm efficient. We note that this choice of proposal for \(S\) is symmetric in the sense that \(f(S'|S, \rho) = f(S|S', \rho)\) for \(f\) in (3). For the concentration parameter \(\kappa\), we take \(\log(\kappa^j) \sim N(\log(\kappa^{j-1}), \sigma^2)\), i.e. \(\kappa^j\) is log-normal with parameters \(\log(\kappa^{j-1})\) and (the tuning parameter) \(\sigma^2\). If \(t(x|\mu, \sigma^2)\) represents the log-normal density, then \(t(\kappa' | \log(\kappa'), \sigma^2) = \frac{\kappa}{\kappa'}\).

Thus, for observations \(o_1, \ldots, o_n \in \Omega\), beginning with some starting values \(S^0\) and \(\kappa^0\), our Metropolis-Hastings within Gibbs algorithm for \(j = 1, 2, \ldots\) is:

1. Generate \(S^{j*} \sim vM-UARS(S^{j-1}, \rho)\) as a proposal for \(S^j\).

2. Compute \(r_j^1 = \frac{g(S^{j*}, \kappa^{j-1})}{g(S^{j-1}, \kappa^{j-1})}\) for \(g\) in (5) and generate \(W_j^1 \sim \text{Bernoulli}(\min(1, r_j^1))\).

3. Take \(S^j = W_j^1 S^{j*} + (1 - W_j^1) S^{j-1}\).

4. Generate \(\log(\kappa^{j*}) \sim N(\log(\kappa^{j-1}), \sigma^2)\), with \(\kappa^{j*}\) as a candidate for \(\kappa^j\).

5. Compute \(r_j^2 = \frac{g(S^j, \kappa^{j*})}{g(S^j, \kappa^{j-1})}\kappa^{j*} \frac{\kappa^{j*}}{\kappa^{j-1}}\) for \(g\) in (5) and generate \(W_j^2 \sim \text{Bernoulli}(\min(1, r_j^2))\).

6. Take \(\kappa^j = W_j^2 \kappa^{j*} + (1 - W_j^2) \kappa^{j-1}\).

Next, we report a simulation study in which the above algorithm was used to do one-sample Bayes analyses for the vM-UARS(S, \(\kappa\)) distribution for various values of \(\kappa\) and sample size, \(n\). We compare coverage probabilities and sizes of confidence/credible regions obtained from the Bayes analyses to the ones produced by a quasi-likelihood method of Bingham et al. (2009b). This comparison is of interest because the quasi-likelihood method produces an estimator of the
location parameter \( \mathbf{S} \) which is common in the matrix literature (explained in the next section). Additionally, the quasi-likelihood method requires distributional approximations which are potentially poor in finite samples. We emphasize that while the Bayes analyses outlined here are for the vM-UARS(\( \mathbf{S} \), \( \kappa \)) distribution, they are easily modified for use with other members of the UARS class (see Bingham et al. (2009a) for one such example).

2.2 Simulation study

We simulated data from the vM-UARS(\( \mathbf{S} \), \( \kappa \)) distribution for various combinations of \( \kappa \) and \( n \) and then used the algorithm outlined above to generate samples from the posterior distribution. The values used for \( \kappa \) were 1, 5, 20, and 500 and sample sizes were \( n = 10, 30, \) and 100. The parameter \( \mathbf{S} \) was held constant. Although symmetry arguments show this detail to be irrelevant, the value used was

\[
\mathbf{S} = \begin{pmatrix}
-0.491493 & 0.804619 & -0.333202 \\
-0.562824 & -0.001501 & 0.826576 \\
0.664578 & 0.593790 & 0.453596
\end{pmatrix}.
\] (6)

For each \((n, \kappa)\) combination we simulated 1000 samples from the vM-UARS(\( \mathbf{S} \), \( \kappa \)) distribution. Based on each of these samples we generated a sample of size 20000 from the posterior distribution in (5) (taken after a burn-in of 5000 iterations) using the algorithm with starting values \( \mathbf{S}^0 \) and \( \kappa^0 \) set at the true parameters. (Inspection of various starting values indicated that the choice of starting value did not affect the output of posterior simulations after the indicated burn-in period.) Further, for each \((n, \kappa)\) combination the tuning parameters \( \rho \) and \( \sigma \) were set to give Metropolis-Hastings jumping rates near 40% (see Gelman, Carlin, Stern, and Rubin, 2004, sec. 11.10). The values of the tuning parameters used are given in Table 1.

For each vM-UARS(\( \mathbf{S} \), \( \kappa \)) sample simulated under a given \((n, \kappa)\) pair we obtained point and set estimates for \( \mathbf{S} \) and \( \kappa \) using both Bayes methods and the quasi-likelihood method of Bingham et al. (2009b). For the latter, with iid observations \( \mathbf{o}_1, \ldots, \mathbf{o}_n \in \Omega \) from the vM-UARS(\( \mathbf{S} \), \( \kappa \)) distribution, point estimates were obtained by maximizing (the quasi-likelihood)

\[
Q(\mathbf{S}, \kappa) = [I_0(\kappa)]^{-n} \prod_{i=1}^{n} \exp \left[ \frac{\kappa}{2} (o_{11,i} + o_{22,i} + o_{33,i} - 1) \right]
\]
Table 1: Values of tuning parameters $\rho$ and $\sigma$

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>$n = 10$</th>
<th>$n = 30$</th>
<th>$n = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td>$\sigma$</td>
<td>$\rho$</td>
</tr>
<tr>
<td>$\kappa = 1$</td>
<td>25</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>$\kappa = 5$</td>
<td>100</td>
<td>1.5</td>
<td>1000</td>
</tr>
<tr>
<td>$\kappa = 20$</td>
<td>1000</td>
<td>1.5</td>
<td>1000</td>
</tr>
<tr>
<td>$\kappa = 500$</td>
<td>1000</td>
<td>1</td>
<td>10000</td>
</tr>
</tbody>
</table>

where $o_{11,i}$, $o_{22,i}$, $o_{33,i}$ are the diagonal entries of $S^T o_i$, $i = 1, \ldots, n$. Bingham et al. (2009b) note that $\hat{S}$ (the maximum quasi-likelihood estimate for the parameter $S$) is the moment estimator for the modal rotation of the Cayley distribution in 3 dimensions (León et al., 2006, p. 421), and is the mean direction for the matrix von Mises-Fisher distribution in 3 dimensions (Khatri and Mardia, 1977, p. 96), frequently used in the literature as an estimate of “location” for 3 dimensional rotation data. Both the posterior density (5) and likelihood (4) have singularities in $S$ at each observation $o_i$ and cannot be maximized. To obtain a Bayes point estimate for $S$ we instead find the maximizer of a quasi-likelihood based on the 20000 values simulated from the posterior. A Bayes point estimate for $\kappa$ can be simply obtained by taking the mean of the $\kappa$ values from the simulated posterior sample.

Although point estimates may be easily obtained, our focus here will be on set estimation. Bingham et al. (2009b) derived the asymptotic distributions of a quasi-likelihood ratio test statistic and a Wald test statistic (see sec. 3.3, Propositions 2 and 3) for the vMUARS($S, \kappa$) distribution and developed confidence regions for $\kappa$ and $S$ through inversion of such tests. Our aim here is to compare the coverage probabilities and sizes of sets obtained using quasi-likelihood ratio and Wald methods to those of credible sets from Bayes analyses. For concreteness, we consider 95% credible sets. To get a posterior credible interval for $\kappa$, let $\Delta_\kappa = \{\kappa^1, \kappa^2, \ldots, \kappa^N\}$ represents the set of $N = 20000$ values for $\kappa$ simulated from the posterior. With $\kappa_{.025}$ = the .025 quantile of $\Delta_\kappa$ and $\kappa_{.975}$ = the .975 quantile of $\Delta_\kappa$, the interval $[\kappa_{.025}, \kappa_{.975}]$ represents a 95% credible interval for $\kappa$. To construct a credible set for $S$, suppose that $S_B$ represents the Bayes point estimate for $S$ discussed above and that $\Delta_S = \{S^1, S^2, \ldots, S^N\}$ represents the set of values for $S$ simulated from the posterior. For
Figure 2: Graphical display of a confidence or credible region for the parameter $S$, with $x$, $y$, and $z$ representing the orientation of a corresponding point estimate for $S$.

For $i = 1, \ldots, N$, we find the angle between each of the coordinate axes rotated from their reference direction by $S_B$ with the corresponding axis rotated by $S_i$ and let $\delta^i$ represent the maximum of these three angles. With $\delta_{.95} = \text{.95 quantile of } \{\delta^1, \ldots, \delta^N\}$, cones of constant angle $\delta_{.95}$ around the coordinate axis rotated by $S_B$ create a region representing 95% of sets of directions of coordinate axes rotated by values from $\Delta_S$. Bingham et al. (2009b) use conic regions centered at coordinate axes rotated by the maximum quasi-likelihood estimate for $S$ to construct Wald and quasi-likelihood ratio confidence regions. Thus, a confidence or credible region for $S$ can be graphically displayed as in Figure 2 and the notion of “size” of these regions can be reduced to the size of the angle between the center and edge of the cones.

After creating confidence and credible regions for each of the 1000 vM-UARS($S$, $\kappa$) samples generated for a given $(n, \kappa)$ combination, we checked if the regions for $\kappa$ and $S$ contained the true parameter values. Corresponding estimated coverage probabilities for the inference methods are given in Table 2. The probabilities in the table verify that, with the exception of the small $n$ Wald intervals for $\kappa$, all regions are holding their nominal coverage rates. The fact that the Wald intervals for $\kappa$ cover the true parameter value less often than nominal
when \( n \) is small is in accord with the results in Bingham et al. (2009b). It is observed there that convergence to the limiting distribution for the Wald test statistic is slower than the corresponding convergence of the quasi-likelihood ratio test statistic.

**Table 2: Coverage rates for \( S \) and \( \kappa \) under Bayes, quasi-likelihood ratio test, and Wald techniques for various choices of \((n, \kappa)\)**

<table>
<thead>
<tr>
<th>((n, \kappa))</th>
<th>Bayes (S)</th>
<th>(\kappa)</th>
<th>Q-LRT (S)</th>
<th>(\kappa)</th>
<th>Wald (S)</th>
<th>(\kappa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1)</td>
<td>0.947</td>
<td>0.972</td>
<td>0.970</td>
<td>0.948</td>
<td>0.946</td>
<td>0.887</td>
</tr>
<tr>
<td>(30, 1)</td>
<td>0.956</td>
<td>0.944</td>
<td>0.982</td>
<td>0.948</td>
<td>0.949</td>
<td>0.934</td>
</tr>
<tr>
<td>(100, 1)</td>
<td>0.952</td>
<td>0.955</td>
<td>0.989</td>
<td>0.955</td>
<td>0.946</td>
<td>0.941</td>
</tr>
<tr>
<td>(10, 5)</td>
<td>0.961</td>
<td>0.958</td>
<td>0.941</td>
<td>0.944</td>
<td>0.956</td>
<td>0.835</td>
</tr>
<tr>
<td>(30, 5)</td>
<td>0.942</td>
<td>0.949</td>
<td>0.956</td>
<td>0.949</td>
<td>0.957</td>
<td>0.911</td>
</tr>
<tr>
<td>(100, 5)</td>
<td>0.947</td>
<td>0.941</td>
<td>0.961</td>
<td>0.953</td>
<td>0.957</td>
<td>0.935</td>
</tr>
<tr>
<td>(300, 5)</td>
<td>0.940</td>
<td>0.954</td>
<td>0.958</td>
<td>0.958</td>
<td>0.954</td>
<td>0.958</td>
</tr>
<tr>
<td>(1000, 5)</td>
<td>0.936</td>
<td>0.948</td>
<td>0.958</td>
<td>0.949</td>
<td>0.951</td>
<td>0.944</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>0.938</td>
<td>0.945</td>
<td>0.937</td>
<td>0.940</td>
<td>0.965</td>
<td>0.825</td>
</tr>
<tr>
<td>(30, 20)</td>
<td>0.944</td>
<td>0.957</td>
<td>0.953</td>
<td>0.955</td>
<td>0.959</td>
<td>0.917</td>
</tr>
<tr>
<td>(100, 20)</td>
<td>0.948</td>
<td>0.946</td>
<td>0.940</td>
<td>0.946</td>
<td>0.940</td>
<td>0.934</td>
</tr>
<tr>
<td>(10, 500)</td>
<td>0.947</td>
<td>0.950</td>
<td>0.933</td>
<td>0.926</td>
<td>0.958</td>
<td>0.803</td>
</tr>
<tr>
<td>(30, 500)</td>
<td>0.955</td>
<td>0.966</td>
<td>0.948</td>
<td>0.952</td>
<td>0.953</td>
<td>0.886</td>
</tr>
<tr>
<td>(100, 500)</td>
<td>0.949</td>
<td>0.946</td>
<td>0.952</td>
<td>0.942</td>
<td>0.955</td>
<td>0.942</td>
</tr>
</tbody>
</table>

Recognizing that both Bayes and quasi-likelihood sets produce about the “right” coverage probabilities, we compare the “sizes” of the regions obtained from the different methods. For each \((n, \kappa)\) combination, we have 1000 confidence intervals for \(\kappa\) and 1000 confidence sets for \(S\) from inversion of the quasi-likelihood ratio tests and Wald tests. Similarly, we have 1000 posterior credible regions for each parameter. Table 3 gives the median widths for the 1000 intervals for \(\kappa\). From the table we see that the posterior credible intervals for \(\kappa\) outperform both types of quasi-likelihood intervals. While the median width of the quasi-likelihood ratio intervals is relatively close to the that of the posterior credible intervals, the median width for the Wald intervals is much larger than that of the credible intervals for small sample sizes. Thus, Bayes methods capture \(\kappa\) in narrower intervals, still while holding the desired coverage probabilities.

Again, the size of a confidence and credible region for \(S\) is characterized by the angle of
the conic regions. The median angles for the 1000 regions produced by each method are given in Table 4. The credible regions are perhaps even surprisingly smaller than both the quasi-likelihood ratio and Wald sets for $S$, and the Bayes median angles seem to decrease at rate $1/n$ while the quasi-likelihood median angles seem to decrease at the “usual” $1/\sqrt{n}$ rate. (The $(300,5)$ and $(1000,5)$ lines of Tables 2, 3, and 4 were added to further study this phenomenon and bear out this observation. For $\kappa = 5$, the Bayes median angles are approximately $1.46/n$ while the Q-LRT median angles are approximately $.749/\sqrt{n}$ and Wald median angles are approximately $.729/\sqrt{n}$.) Our conjecture regarding the origin of this phenomenon is that it can be traced to the non-regularity of the likelihood (and posterior) in this problem. The likelihood (4) has singularities at all observations (which drove us to initially consider quasi-likelihood). The quasi-likelihood is smooth and its asymptotics “standard.” But it is not unheard of to be able to get a better rate than $1/\sqrt{n}$ with an appropriate method in a non-regular problem. (For a more complete analysis of this kind of rate issue for Bayes methods, albeit in a simpler context, see Nordman, Vardeman, and Bingham (2009).)

So in sum, the Bayes methods for both $\kappa$ and $S$ in the vM-UARS$(S, \kappa)$ one-sample problem are preferable to the quasi-likelihood methods of Bingham et al. (2009b). This is weakly true for inference about $\kappa$ and strongly true for inference on $S$.

### 3 One-way random effects Bayes analyses for the vM-UARS$(S, \kappa)$ distribution

With numerical evidence indicating that Bayes methods perform well for the one-sample case, we next develop Bayes analyses for a 3-d rotation version of a one-way random effects model. This allows effective study of two sources of variation arising between and among groups of matrix observations. Suppose that for $i = 1, \ldots, r$ and $k = 1, \ldots, m_i$,

$$O_{ik} = P_i Q_{ik} \in \Omega$$

for $P_i \sim^{iid}$ vM-UARS$(S, \kappa)$ independent of $Q_{ik} \sim^{iid}$ vM-UARS$(I_{3\times3}, \tau)$. Thus, we have $r$ groups with $m_i$ (that may differ with $i$) observations per group, where $\kappa$ represents the between-group variation and $\tau$ represents the within-group variation (with larger values indicating less
Table 3: Median width of Bayes credible intervals, quasi-likelihood ratio intervals, and Wald intervals for $\kappa$ for various choices of $(n, \kappa)$

<table>
<thead>
<tr>
<th>$(n, \kappa)$</th>
<th>Bayes</th>
<th>Q-LRT</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1)</td>
<td>2.0520</td>
<td>2.2740</td>
<td>6.9911</td>
</tr>
<tr>
<td>(30, 1)</td>
<td>1.2044</td>
<td>1.2314</td>
<td>1.3717</td>
</tr>
<tr>
<td>(100, 1)</td>
<td>0.6603</td>
<td>0.6650</td>
<td>0.6831</td>
</tr>
<tr>
<td>(10, 5)</td>
<td>8.6458</td>
<td>9.9403</td>
<td>42.5220</td>
</tr>
<tr>
<td>(30, 5)</td>
<td>4.7448</td>
<td>4.9787</td>
<td>6.6954</td>
</tr>
<tr>
<td>(100, 5)</td>
<td>2.5659</td>
<td>2.5989</td>
<td>2.8181</td>
</tr>
<tr>
<td>(300, 5)</td>
<td>1.4887</td>
<td>1.4971</td>
<td>1.5371</td>
</tr>
<tr>
<td>(1000, 5)</td>
<td>0.8145</td>
<td>0.8172</td>
<td>0.8237</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>36.3178</td>
<td>42.0862</td>
<td>177.9051</td>
</tr>
<tr>
<td>(30, 20)</td>
<td>20.4424</td>
<td>21.3608</td>
<td>28.5171</td>
</tr>
<tr>
<td>(100, 20)</td>
<td>11.0215</td>
<td>11.2212</td>
<td>12.1299</td>
</tr>
<tr>
<td>(1000, 500)</td>
<td>944.8629</td>
<td>1075.3761</td>
<td>4543.5468</td>
</tr>
<tr>
<td>(1000, 500)</td>
<td>279.1206</td>
<td>283.2678</td>
<td>306.1883</td>
</tr>
</tbody>
</table>

Table 4: Median angle of Bayes credible regions, quasi-likelihood ratio regions, and Wald regions for $S$ for various choices of $(n, \kappa)$

<table>
<thead>
<tr>
<th>$(n, \kappa)$</th>
<th>Bayes</th>
<th>Q-LRT</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1)</td>
<td>0.3857</td>
<td>0.6215</td>
<td>0.5439</td>
</tr>
<tr>
<td>(30, 1)</td>
<td>0.1264</td>
<td>0.3641</td>
<td>0.3104</td>
</tr>
<tr>
<td>(100, 1)</td>
<td>0.0374</td>
<td>0.1960</td>
<td>0.1665</td>
</tr>
<tr>
<td>(10, 5)</td>
<td>0.1474</td>
<td>0.2513</td>
<td>0.2312</td>
</tr>
<tr>
<td>(30, 5)</td>
<td>0.0495</td>
<td>0.1334</td>
<td>0.1321</td>
</tr>
<tr>
<td>(100, 5)</td>
<td>0.0152</td>
<td>0.0723</td>
<td>0.0714</td>
</tr>
<tr>
<td>(300, 5)</td>
<td>0.0048</td>
<td>0.0432</td>
<td>0.0426</td>
</tr>
<tr>
<td>(1000, 5)</td>
<td>0.0014</td>
<td>0.0237</td>
<td>0.0233</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>0.0683</td>
<td>0.1177</td>
<td>0.1091</td>
</tr>
<tr>
<td>(30, 20)</td>
<td>0.0247</td>
<td>0.0654</td>
<td>0.0642</td>
</tr>
<tr>
<td>(100, 20)</td>
<td>0.0074</td>
<td>0.0352</td>
<td>0.0351</td>
</tr>
<tr>
<td>(1000, 500)</td>
<td>0.0138</td>
<td>0.0234</td>
<td>0.0215</td>
</tr>
<tr>
<td>(1000, 500)</td>
<td>0.0050</td>
<td>0.0130</td>
<td>0.0126</td>
</tr>
<tr>
<td>(1000, 500)</td>
<td>0.0015</td>
<td>0.0070</td>
<td>0.0070</td>
</tr>
</tbody>
</table>
variation). We will use Bayes methods to estimate these parameters as well as the overall “location” parameter $S$.

First, we can write

$$f(o_{ik}, p_i | S, \kappa, \tau) = f(o_{ik} | p_i, \tau)f(p_i | S, \kappa),$$

for $f$ in (3). Thus, for the one-way random effects case we have a joint density for the observable $o_{ik}$ and unobservable $p_i$

$$f(o_{11}, \ldots, o_{rm}, p_1, \ldots, p_r | S, \kappa, \tau) = \prod_{i=1}^r \left( \prod_{k=1}^{m_i} f(o_{ik} | p_i, \tau) \right) f(p_i | S, \kappa)$$

$$= \prod_{i=1}^r \left( 2^{m_i+1} \exp \left[ \frac{r}{2} \sum_{k=1}^{m_i} (tr(p_i^T o_{ik}) - 1) + \frac{s}{2} (tr(S_i^T o_{ik}) - 1) \right] \right).$$

(7)

Now, we again place a uniform (Haar) prior on $S$ and place independent Jeffreys priors on both $\kappa$ and $\tau$. We then have posterior density for the unobservable $p_i$ and parameters proportional to

$$G(S, \kappa, \tau, p_1, \ldots, p_r) = f(p_1, \ldots, p_r | S, \kappa, \tau)h_2(\tau)h_2(\kappa)$$

(8)

for $f$ given in (7) with $o_{ik}$ fixed and dependence upon them suppressed and $h_2$ in (2). For observations $o_{ik}, i = 1, \ldots, r$ and $k = 1, \ldots, m_i$, beginning with starting values $S^0, \kappa^0, \tau^0$, and $\{p_1^0, \ldots, p_r^0\}$ we implement a Metropolis-Hastings within Gibbs algorithm for $j = 1, 2, \ldots$ as follows:

1. Generate $S^{j*} \sim \text{vM-UARS}(S^{j-1}, \rho_1)$ as a proposal for $S^j$.

2. Compute $r_j^1 = \frac{G(S^{j*}, \kappa^{j-1}, \tau^{j-1}, p_1^{j-1}, \ldots, p_r^{j-1})}{G(S^{j-1}, \kappa^{j-1}, \tau^{j-1}, p_1^{j-1}, \ldots, p_r^{j-1})}$ for $G$ in (8) and generate $W_j^1 \sim \text{Bernoulli}(\text{min}(1, r_j^1))$.

3. Take $S^j = W_j^1 S^{j*} + (1-W_j^1) S^{j-1}$.

4. Generate $\log(\kappa^{j*}) \sim \text{N}(\log(\kappa^{j-1}), \sigma_1^2)$, with $\kappa^{j*}$ as a proposal for $\kappa^j$.

5. Compute $r_j^2 = \frac{G(S^j, \kappa^{j*}, \tau^{j-1}, p_1^{j-1}, \ldots, p_r^{j-1})}{G(S^j, \kappa^{j-1}, \tau^{j-1}, p_1^{j-1}, \ldots, p_r^{j-1})}$ for $G$ in (8) and generate $W_j^2 \sim \text{Bernoulli}(\text{min}(1, r_j^2))$. 


6. Take $\kappa^j = W_j^2 \kappa^{j*} + (1 - W_j^2) \kappa^{j-1}$.

7. Generate $\log(\tau^{j*}) \sim N(\log(\tau^{j-1}), \sigma_j^2)$, with $\tau^{j*}$ as a proposal for $\tau^j$.

8. Compute $r_j^3 = \frac{G(S^j, \kappa^j, \tau^{j*}, \mathbf{p}_1^{j-1}, \ldots, \mathbf{p}_r^{j-1}) \tau^j}{G(S^j, \kappa^j, \tau^{j-1}, \mathbf{p}_1^{j-1}, \ldots, \mathbf{p}_r^{j-1}) \tau^{j-1}}$ for $G$ in (8) and generate $W_j^3 \sim \text{Bernoulli}(\min(1, r_j^3))$.

9. Take $\tau^j = W_j^3 \tau^{j*} + (1 - W_j^3) \tau^{j-1}$.

10. For $k = 1, \ldots, r$

    (a) Generate $\mathbf{p}_k^{j*} \sim \text{vM-UARS}([\mathbf{p}_k^{j-1}, \rho_2])$ as a proposal for $\mathbf{p}_k^j$.

    (b) Compute $q_j^k = \frac{G(S^j, \kappa^j, \tau^j, \mathbf{p}_1^j, \ldots, \mathbf{p}_k^{j-1}, \mathbf{p}_k^{j*}, \mathbf{p}_{k+1}^{j-1}, \ldots, \mathbf{p}_r^{j-1})}{G(S^j, \kappa^j, \tau^{j-1}, \mathbf{p}_1^{j-1}, \ldots, \mathbf{p}_k^{j-1}, \mathbf{p}_{k+1}^{j-1}, \ldots, \mathbf{p}_r^{j-1})}$ for $G$ in (8) and generate $V_j^k \sim \text{Bernoulli}(\min(1, q_j^k))$.

    (c) Take $\mathbf{p}_k^j = V_j^k \mathbf{p}_k^{j*} + (1 - V_j^k) \mathbf{p}_k^{j-1}$.

Again, we emphasize that while here we concentrate on the von Mises case, this methodology is easily modified and applied to other members of the UARS class.

To check the efficacy of the one-way random effects method above, we simulated 1000 independent data sets where $m_i = 30$ for $i = 1, \ldots, 30$, $\tau = 20$, $\kappa = 20$, and $S$ is as in Section 2.2. (Because the Metropolis-Hastings algorithm for the one-way random effects case requires much more computing time than for the one-sample case, we could not perform an extensive simulation study.) Using each of the 1000 data sets, we generated a sample of size 8000 from the posterior density in (8) after a burn-in of 2000 iterations. As in the one-sample case, starting values had no effect on posterior samples for such a burn-in period. For each of the 1000 cases, we obtained Bayes point estimates and 95% posterior credible sets. The point estimates and credible sets for $S$ were computed in the same manner as for the one-sample situation, and estimates and intervals for $\kappa$ and $\tau$ were both computed as for $\kappa$ in the one-sample case.

When checking the 95% credible regions for each parameter, we found the true values were captured at rates of 0.939, 0.953, and 0.942 for $S$, $\kappa$, and $\tau$, respectively. So, the coverage rates are as expected. We also found the median size of the credible regions for each parameter. For $S$ the median conic angle was 0.0241, for $\kappa$ the median credible interval width was 20.5221,
and for $\tau$ the median credible interval width was 3.6522. Thus, we are able to capture the indicator of within group variation $\tau$ in smaller intervals than the indicator of between group variation $\kappa$. Now, we apply the methods from this section, as well as the one-sample methods from the previous section, to real EBSD data.

4 Application to EBSD data

Bingham et al. (2009b) analyzed data from a sample of TSL (an EBSD company) calibration standard high-Iron-concentration nickel. The nickel specimen had a surface area $40\mu m \times 40\mu m$ and the same area was scanned 14 times with at least 4000 measurements per run. Using the quasi-likelihood inference of Bingham et al. (2009b), we fit the $vM-UARS(S, \kappa)$ distribution to two sets of observations: data set 1 contains 14 repeat observations from one location in the scans and data set 2 contains 50 observations from different locations appearing to be within the same grain on a single scan. We also applied the one-sample Bayes methods from Section 2.1 to the same two data sets by simulating a sample of size 20000 from the posterior given in (5). The estimates obtained are presented in Table 5 and are similar for the two estimation approaches.

Table 5: Estimates of the parameters $\kappa$ and $S$ (given in Euler angle form (see Bingham et al., 2009b)) for the two EBSD data sets using maximum quasi-likelihood and Bayes inference

<table>
<thead>
<tr>
<th>Data set 1:</th>
<th>MQL</th>
<th>$\kappa$ (\alpha, \beta, \gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33599.2028</td>
<td>5.8532, 0.9291, 4.2265</td>
</tr>
<tr>
<td></td>
<td>30774.0214</td>
<td>5.8534, 0.9292, 4.2259</td>
</tr>
<tr>
<td>Data set 2:</td>
<td>MQL</td>
<td>2197.0395 (5.8599, 0.9252, 4.2191)</td>
</tr>
<tr>
<td></td>
<td>Bayes</td>
<td>2156.5243 (5.8567, 0.9259, 4.2225)</td>
</tr>
</tbody>
</table>

We also provide 95% confidence and credible intervals for $\kappa$ for both data sets. The parameter $\kappa$ is of interest in these applications because it measures the spread of the orientations about some principal direction. It is desirable that data obtained through EBSD show a high degree of repeatability, making large values for $\kappa$ desirable. Confidence intervals obtained using the inference of Bingham et al. (2009b) are in Table 6 along with the posterior credible intervals we obtained for $\kappa$ using the Bayesian approach of Section 2.1. For both data sets, the
width of the Bayes credible interval for $\kappa$ is less than that of the Wald and quasi-likelihood ratio intervals. All intervals contain only values for $\kappa$ that represent highly concentrated distributions on $\Omega$. This is evidence that the machine does an adequate job of taking repeat readings using EBSD. The intervals for data set 2 are located entirely to the left of those for data set 1. This is as expected, since the intervals for the second data set capture both single site variation and location-to-location variation between sites.

Table 6: 95% quasi-likelihood ratio test and Wald confidence intervals and 95% credible intervals for $\kappa$ for the two EBSD data sets

<table>
<thead>
<tr>
<th></th>
<th>Data set 1</th>
<th>Data set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q-LRT</td>
<td>(14436.9125, 64979.3152)</td>
<td>(1444.3103, 3173.9226)</td>
</tr>
<tr>
<td>Wald</td>
<td>(19301.1624, 129622.7164)</td>
<td>(1578.4121, 3613.3484)</td>
</tr>
<tr>
<td>Bayes</td>
<td>(11961.3872, 58327.3703)</td>
<td>(1410.8453, 3055.6280)</td>
</tr>
</tbody>
</table>

Using the Bayes analysis for one-way random effects developed in Section 3, we can further look simultaneously at these two types of variation. We used a sample of 14 repeat observations from each of 50 positions on the scans that all appear to be within a single grain. Thus, $m_i = 14$ for each $i = 1, \ldots, 50$. Using the algorithm outlined in Section 3, we obtained a sample of size 8000 from the posterior in (8). Based on this sample, Bayes estimates are $\kappa_B = 2651.9501$, $\tau_B = 5474.5402$, and

$$S_B = \begin{pmatrix} -0.6449 & -0.2866 & -0.7084 \\ 0.6874 & -0.6228 & -0.3738 \\ -0.3341 & -0.7280 & 0.5987 \end{pmatrix}.$$ 

The 95% posterior credible intervals are (1717.1060, 3797.2093) and (4898.7185, 6066.4056) for $\kappa$ and $\tau$, respectively, with conic regions of angle 0.0034 representing a 95% credible region for $S$. Notice that the interval for $\tau$ represents orientations with less variability than that for $\kappa$, so the between group variation is greater than the within group variation.

Because the estimates for $\kappa$ and $\tau$ are much larger here than the parameters used in the single simulation in Section 3, we also conducted a simulation using $\kappa_B$, $\tau_B$, and $S_B$ as the true parameter values with $m_i = 14$ for $i = 1, \ldots, 50$. Again, 1000 samples were generated using these parameter values and then the Metropolis-Hastings within Gibbs algorithm was used to
get a sample of size 8000 from the posterior for each of the 1000 samples. Median 95% credible region sizes were calculated as before, yielding median width of 2071.9742 for $\kappa$, 1146.6534 for $\tau$, and median angle of 0.0013 for $S$. Coverage rates were 0.933, 0.962, and 0.954 for $S$, $\kappa$, and $\tau$, respectively. This is at least anecdotal evidence that Bayes one-way random effects analyses also behave sensibly for large values of $\kappa$ and $\tau$.

5 Conclusion

As we have shown, one-sample Bayes analyses for the vM-UARS($S$, $\kappa$) distribution are effective an provide an alternative to the quasi-likelihood inference introduced in Bingham et al. (2009b). By using Bayes methods and MCMC we are also able to extend our analyses for the vM-UARS($S$, $\kappa$) distribution to a one-way random effects version, which provides a useful new form of inference for 3-d rotation data. Although only the von Mises version of the UARS class of distributions was examined here, the Bayes methods are easily extended to other members of the class (see, for example, Bingham et al., 2009a).

References


FINITE-SAMPLE INVESTIGATION OF LIKELIHOOD AND BAYES INFERENCE FOR THE SYMMETRIC VON MISES-FISHER DISTRIBUTION

A paper submitted to Computational Statistics and Data Analysis

Melissa A. Bingham, Daniel J. Nordman, and Stephen B. Vardeman
Iowa State University

Abstract

We consider likelihood and Bayes analyses for the symmetric matrix von Mises-Fisher distribution, which is a common model for 3-dimensional orientations (represented by $3 \times 3$ orthogonal matrices with positive determinant). We extend one-sample likelihood theory to a parameterization describing mean rotation through Euler angles, which has not been directly considered in the literature but has attractive features in computation and interpretation of rotation matrices. Bayesian methods with non-informative priors are also developed using Metropolis-Hasting within Gibbs algorithms and we illustrate how the Bayes framework extends inference from one-sample problems to more complicated one-way random effects models based on the symmetric matrix Fisher model. A simulation study compares the finite-sample performance of Bayes and likelihood analyses in the one-sample case and the methods are applied to a human kinematics example for illustration.

Key words: Bayes, credible regions, Euler angles, matrix von Mises-Fisher distribution, one-way random effects model, rotation matrix

\footnote{Supported by NSF grant DMS \#0502347 EMSW21-RTG awarded to the Department of Statistics, Iowa State University}
1 Introduction

This paper compares likelihood and Bayes inference for the symmetric von Mises-Fisher distribution on 3-dimensional rotation matrices. Such matrix-valued data may arise in material science investigations (e.g., crystal orientations in metals, Mackenzie, 1957; Bingham, Nordman, and Vardeman, 2009) as well as studies of human kinematics (Rancourt, Rivest, and Asselin, 2000), and the matrix Fisher is the most widely referenced distribution for modeling such observations (Downs, 1972). A great deal of literature exists particularly surrounding likelihood methodology for this distribution, but there has been considerably less consideration for Bayes analyses. For example, advances in maximum likelihood inference for the matrix Fisher distribution have been made by Khatri and Mardia (1977) and Jupp and Mardia (1979), with further work done by Prentice (1986), Mardia and Jupp (1999), and Chikuse (2003). Chang and Rivest (2001) also developed M-estimation connected to likelihood inference. In terms of Bayes inference for the matrix Fisher distribution, Chang and Bingham (1996) outlined an approach for stipulating informative priors while, in unpublished work, Camano-Garcia (2006) considered Gibbs samplers for general Langevin (or matrix Fisher) distributions.

However, despite these developments and the clear popularity of the matrix Fisher distribution, little appears to be known about the relative merits of finite-sample likelihood and Bayes inference in this distribution. For example, the Bayes development in Chang and Bingham (1996) targeted large-sample approximations of posterior distributions, often with informative priors amenable to such approximations. We aim instead to explore the finite-sample implementation of Bayes methods with non-informative priors, employing the common parametrization of the symmetric matrix Fisher used by Chang and Bingham (1996), in an effort to contrast the quality of Bayes inference to that of likelihood in the one-sample problem. As a contribution to likelihood inference, we study through simulation how closely likelihood test statistics match their limit laws for finite-sample inference in the symmetric Fisher model and additionally describe likelihood methods based on an Euler angle parametrization (of the parameter $S$ described below). Euler angles offer a simple route for computing likelihood functions with rotation matrix parameters but require some theoretical development that (to our
knowledge) has not appeared in the literature. Beyond the one-sample problem, we also build
and illustrate Bayes inference in one-way random effects models based on the symmetric matrix
Fisher distribution. Our intent in this is to stimulate greater interest in Bayes methods for
the Fisher model by demonstrating how such methods extend naturally from the one-sample
situation to more complex inference scenarios. The Bayes framework may be more tractable
than a purely likelihood approach for some problems and may open new inference possibilities
for the matrix Fisher distribution. This paper lays some groundwork in this direction and
provides an MCMC-basis for implementing work in Chang and Bingham (1996).

Although we focus on the matrix Fisher distribution in particular, the Bayes and likelihood
methodology presented here is not restricted solely to this distribution. The same inference
methods can be extended to a larger class of distributions for random rotations, which have
the same geometric “construction” and parametrization as the symmetric matrix Fisher, as we
next explain. Let \( \Omega \) represent the set of \( 3 \times 3 \) rotation matrices that preserve the right hand
rule. The symmetric Fisher distribution is characterized through a location parameter \( S \in \Omega \)
and a spread parameter \( \kappa > 0 \), where the model itself (denoted \( F(S, \kappa) \) here) describes the
deviation of random orientations \( O_1, \ldots, O_n \in \Omega \) from a common “true” orientation \( S \in \Omega \).
One random orientation \( O \in \Omega \) from a \( F(S, \kappa) \) model may be constructed as \( O = S \cdot P \), where

\[
P = UU^T + (I_{3 \times 3} - UU^T) \cos r + \begin{pmatrix}
0 & -u_3 & u_2 \\
u_3 & 0 & -u_1 \\
-u_2 & u_1 & 0
\end{pmatrix} \sin r \in \Omega
\]

(1)
is a random rotation built from two independent components: a unit vector \( U^T = (u_1, u_2, u_3) \in \mathbb{R}^3 \)
identified by a point uniformly distributed on the unit sphere and an angle \( r \in (-\pi, \pi] \)
distributed according to a circular distribution symmetric about 0 with spread depending on
\( \kappa \). The matrix \( P \) represents the positions of coordinate axes in \( \mathbb{R}^3 \) (denoted by the \( 3 \times 3 \)
identity matrix \( I_{3 \times 3} \)) after spinning the \( \mathbb{R}^3 \)-frame around the axis \( U \in \mathbb{R}^3 \) by the angle \( r \),
where a small \(|r|\) value entails a small rotational deviation \( P \) from \( I_3 \) (e.g., \( r = 0 \) implies
\( P = I_{3 \times 3} \)). Since \( \kappa \) controls the spread or concentration of \( r \) around 0, this parameter also
controls the variation of an \( F(S, \kappa) \) observation \( O \) from the location parameter \( S \in \Omega \); see
also Chang and Bingham (1996) and León, Massé and Rivest (2006, p. 425) for this \( F(S, \kappa) \) parametrization. For the symmetric matrix Fisher, the distribution of the “spin” \( r \) has a particular distributional form (described in Section 2). However, allowing other circular distributions for \( r \in (-\pi, \pi] \), similarly involving a spread parameter \( \kappa \), creates a wide class of distributions for rotation matrices identified by Bingham, Nordman, and Vardeman (2009) as Uniform-Axis-Random Spin (UARS) models. Our point is that this constructive definition of rotation matrices particularly facilitates MCMC-based Bayes inference, so that the Bayes methods illustrated here for \( F(S, \kappa) \) distribution may be carried over to inference in other UARS(\( S, \kappa \)) models for rotations. We shall provide more details and connections to previous work in Section 3.

Section 2 gives the distribution of symmetric matrix Fisher distribution and provides large-sample distributional results for one-sample likelihood inference, which accommodate an Euler angle representation of \( S \). Although these results are asymptotic in nature, simulation evidence indicates that resulting likelihood ratio tests can be accurate in samples as small as \( n = 10 \). Section 3 provides numerical studies comparing likelihood and Bayes inference for the one-sample problem (using non-informative priors and a Metropolis-Hastings within Gibbs algorithm). Section 4 examines the performance of Bayes methods for one-way random effects models with the symmetric matrix Fisher distribution, while Section 5 illustrates the matrix methods with human kinematics data. Section 6 provides some conclusions.

We end this section by mentioning a general parametrization of the matrix Fisher distribution. This involves a model density \( a(F) \exp(tr(F^T o)) \), \( o \in \Omega \) with respect to the Haar measure on \( \Omega \) where \( a(F) \) is a normalizing constant with \( F \) a \( 3 \times 3 \) matrix of full rank. The invariant Haar measure acts as a “uniform distribution” on \( \Omega \) (cf. Downs, 1972). The parameter \( F \) can be decomposed as \( F = KM \) where \( M \) is the “polar component” (sometimes called the “mean direction” as in Downs (1972)) and \( K \) is the “elliptic component” (Khatri and Mardia, 1977). If data \( o_1, \ldots, o_n \) come from the general matrix Fisher distribution, the maximum likelihood estimate of \( F \), or equivalently \( M \) and \( K \), can be obtained by considering the singular value decomposition of \( \bar{o} = \frac{1}{n} \sum_{i=1}^{n} o_i \). This decomposition yields \( \bar{o} = \hat{\Delta} D_g \hat{F} \), where \( D_g \) is a di-
agonal matrix with entries \( g = (g_1, g_2, g_3) \), so that the maximum likelihood estimate for \( M \) is \( \hat{M} = \hat{\Delta} \hat{\Gamma} \). Suppose \( \mu_i(\phi) = (\partial/\partial \phi_i)_0 F_1(3/2, 1/4 D^2_\phi) \) for \( \phi = (\phi_1, \phi_2, \phi_3) \) where \( 0 F_1 \) is the hypergeometric function with matrix argument. Then for \( \hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3) \) solving \( g_i = \mu_i(\hat{\phi}), i = 1, 2, 3 \), the maximum likelihood estimate for \( K \) is \( \hat{K} = \Delta D_{\phi} \hat{\Delta}^T \) (Khatri and Mardia, 1977). Thus, although possible using this technique, maximum likelihood estimation for the elliptic component is not simple and, following Chang and Bingham (1996), we will consider the symmetric version of matrix Fisher distribution for which \( K = \kappa \cdot I_{3 \times 3} \) and \( M = S \) in our notation.

2 The symmetric matrix Fisher distribution

2.1 Model density

In the random rotation matrix construction (1) described in Section 1, if we specify an arbitrary probability density \( C(r|\kappa) \) for the random angle \( r \in (-\pi, \pi] \) (which is symmetric around 0 and has spread parameter \( \kappa \)), then we induce a probability distribution on \( 3 \times 3 \) rotations \( O = S \cdot P \in \Omega \) with density

\[
f(o|S, \kappa) = \frac{4\pi}{3 - tr(S^T o)} C \left( \arccos[2^{-1}(tr(S^T o) - 1)] \right), \quad o \in \Omega,
\]

with respect to the Haar measure (i.e., the density for a UARS(\( S, \kappa \)) model on \( \Omega \) in the terminology of Bingham, Nordman, and Vardeman, 2009, sec. 2.2). An appropriate choice of \( C(r|\kappa) \) produces a symmetric Fisher distribution \( F(S, \kappa) \) on \( \Omega \) as follows.

A standard von Mises circular density for \( r \in (-\pi, \pi] \) with concentration parameter \( 2\kappa \) is proportional to \( \exp(2\kappa \cos r) \) (Mardia and Jupp, 2000) so that, multiplying by \( 1 - \cos r \) and normalizing, we arrive at a density

\[
C(r|\kappa) = \frac{(1 - \cos r) \exp(2\kappa \cos r)}{2\pi (I_0(2\kappa) - I_1(2\kappa))}, \quad r \in (-\pi, \pi],
\]

where \( I_i \) is the modified Bessel function of order \( i \) and \( \kappa \in (0, \infty) \). Substitution of (3) into (2) then yields the \( F(S, \kappa) \) density

\[
f(o|S, \kappa) = \frac{\exp(\kappa[tr(S^T o) - 1])}{I_0(2\kappa) - I_1(2\kappa)}, \quad o \in \Omega,
\]
with respect to the Haar measure (i.e., elliptic component \( K = \kappa \cdot I_{3 \times 3} \), mean direction \( M = S \)).

To understand how the spread component \( \kappa \) in the density (3) of \( r \) translates into “spread” for the \( F(S, \kappa) \) distribution, consider Table 1. Here, \( \Delta_1(\kappa) \) is the median of the distribution on \(|r|\) (where \(|r|\) is sometimes referred to as a “misorientation angle” and has density \( 2 \cdot C(|r|/\kappa) \)) and \( \Delta_2(\kappa) \) is the median of the distribution of the maximum angle between \( S = I_{3 \times 3} \) and \( F(I_{3 \times 3}, \kappa) \)-rotated coordinate axes (i.e., the maximum angle over 3 rotated axes). The values \( \Delta_1(\kappa) \) were computed using numerical integration and each value of \( \Delta_2(\kappa) \) is based on a sample of 100,000 Fisher realizations.

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta_1(\kappa) )</td>
<td>1.29604</td>
<td>0.49845</td>
<td>0.34798</td>
<td>0.24460</td>
<td>0.15416</td>
<td>0.10889</td>
<td>0.04864</td>
<td>0.03439</td>
</tr>
<tr>
<td>( \Delta_2(\kappa) )</td>
<td>1.24709</td>
<td>0.48193</td>
<td>0.33585</td>
<td>0.23691</td>
<td>0.14868</td>
<td>0.10507</td>
<td>0.04696</td>
<td>0.03323</td>
</tr>
</tbody>
</table>

2.2 Likelihood function and large-sample distributional results

For purposes of likelihood inference, we next parameterize \( S \) using Euler angles \( \alpha \), \( \beta \), and \( \gamma \) by setting

\[
S = S(\alpha, \beta, \gamma) = \begin{pmatrix}
\cos \alpha \cos \gamma - \sin \alpha \sin \gamma \cos \beta & \sin \alpha \cos \gamma + \cos \alpha \sin \gamma \cos \beta & \sin \gamma \sin \beta \\
-\cos \alpha \sin \gamma - \sin \alpha \cos \gamma \cos \beta & -\sin \alpha \sin \gamma + \cos \alpha \cos \gamma \cos \beta & \cos \gamma \sin \beta \\
\sin \alpha \sin \beta & -\cos \alpha \sin \beta & \cos \beta
\end{pmatrix}
\]

for \( S : [0, 2\pi] \times [0, \pi] \times [0, 2\pi] \to \Omega \). Other parameterizations for \( S \) are possible, such as quaternions or an angle/axis characterization (analogous to defining \( S \) through fixed values of a rotation \( r \) and axis \( U \in \mathbb{R}^3 \) in (1)). Theoretical results in Chang and Rivest (2001) and León et al. (2006, p. 424), for example, are framed in terms of the angle/axis representation of rotation matrices. Although admittedly less common in the statistical literature, we employ this Euler angle representation because it is commonly used and understood by subject matter specialists in many fields (e.g., material science, Randle, 2003) and offers non-specialists a computationally straightforward representation for evaluating the likelihood function over a parameter space of a familiar type. Of course, the value of \( S \) is the fundamental quantity,
not any particular parametrization of this rotation matrix. However, our use of Euler angles requires developing some likelihood theory around this matrix parametrization, which we next consider. We note additionally that parametrization of the location parameter $S$ is unnecessary in Bayes inference of Section 4.

Supposing iid observations $o_1, \ldots, o_n \in \Omega$ are from $F(S, \kappa)$ distribution (4), we have likelihood function

$$L_n(\kappa, (\alpha, \beta, \gamma)) = \left[ I_0(2\kappa) - I_1(2\kappa) \right]^{-n} \prod_{i=1}^{n} \exp \left( \kappa [tr(S(\alpha, \beta, \gamma)^T o_i) - 1] \right).$$

(6)

and one-sample log-likelihood as

$$l_n(\kappa, (\alpha, \beta, \gamma)) = \kappa \sum_{i=1}^{n} [tr(S(\alpha, \beta, \gamma)^T o_i) - 1] - n \log (I_0(2\kappa) - I_1(2\kappa)).$$

(7)

Let $\hat{\theta}_n$ denote the maximum likelihood estimator (MLE) of $\theta = (\kappa, \alpha, \beta, \gamma)$. The work of Jupp and Mardia (1977) indicates that MLE will uniquely exist with probability 1 (except for potentially small sample sizes) and the MLE of $S$ is known to be the minimizer of $\sum_{i=1}^{n} tr(A^T o_i)$ over $A \in \Omega$ (cf. Downs, 1972), from which the MLE values of Euler angles can be determined through $\hat{S} = S(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ in (5). Proposition 1 establishes the asymptotic distribution of likelihood statistics under our Euler angle parametrization of the symmetric matrix Fisher distribution, useful for calibrating joint confidence regions.

**Proposition 1.** Suppose that $O_1, \ldots, O_n$ are a random sample from $F(S, \kappa)$.

(a) **Limiting Distribution of MLE:** As $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0_{4 \times 1}, [I_1(\theta)]^{-1})$$

i.e., convergence to a multivariate normal based on information matrix

$$I_1(\kappa, (\alpha, \beta, \gamma)) = \begin{pmatrix} A(\kappa) & 0 & 0 & 0 \\ 0 & 2\kappa D(\kappa) & 0 & 2\kappa D(\kappa) \cos \beta \\ 0 & 0 & 2\kappa D(\kappa) & 0 \\ 0 & 2\kappa D(\kappa) \cos \beta & 0 & 2\kappa D(\kappa) \end{pmatrix},$$

(8)
for $D(\kappa) = \frac{1}{3} + \frac{2}{3} \left( 1 + \frac{1 - \kappa}{\kappa^2} \right) I_1(2\kappa) - I_0(\kappa)$ and $A(\kappa) = \frac{2}{\kappa} I_0(\kappa)^2 - \frac{2}{\kappa^2} I_0(\kappa) I_1(\kappa) + \left( \frac{1 - \kappa}{\kappa^2} \right) I_1(\kappa)^2$.

where $I_i(\kappa)$ is the modified Bessel function of order $i$.

(b) Distribution of Likelihood Ratio Test Statistics: Partition $\theta^T = (\kappa, \alpha, \beta, \gamma)$ into $(\theta_1^T, \theta_2^T)^T$, where $\theta_1$ is $p \times 1$ and $\theta_2$ is $(4 - p) \times 1$, and define the likelihood ratio test statistic

$$\lambda_n = \frac{L_n(\theta_n^*)}{L_n(\hat{\theta}_n)}$$

for $H_0: \theta_1 = \theta_{01}$ using (6) and the MLE $\theta_n^*$ under the $H_0$-restricted parameter space. If $H_0$ is true and $n \to \infty$,

$$-2 \log(\lambda_n) \stackrel{d}{\to} \begin{cases} \chi_4^2 & \text{for } \theta_1 = \theta \\ \chi_3^2 & \text{for } \theta_1^T = (\alpha, \beta, \gamma) \\ \chi_1^2 & \text{for } \theta_1 = \kappa \end{cases} \quad \text{(9)}$$

(c) Distribution of Overall Wald Test Statistic: Under the null hypothesis $H_0: \theta = \theta_0$, let $T_n(\theta_0) = n(\hat{\theta}_n - \theta_0)^T I_1(\theta_0)(\hat{\theta}_n - \theta_0)$. If $H_0$ is true and $n \to \infty$,

$$T_n(\theta_0) \stackrel{d}{\to} \chi_4^2.$$ 

It is also the case that if we wish to test $H_0: (\alpha, \beta, \gamma)^T = (\alpha_0, \beta_0, \gamma_0)^T$ and partition (8) accordingly, then under $H_0$ the Wald statistic $T_n(\alpha_0, \beta_0, \gamma_0) \stackrel{d}{\to} \chi_3^2$ as $n \to \infty$. Similarly, for the test of $H_0: \kappa = \kappa_0$, under $H_0$ the Wald statistic $T_n(\kappa_0) \stackrel{d}{\to} \chi_1^2$. Khatri and Mardia (1977) provide likelihood ratio tests in general formulations of the matrix Fisher distribution, which are more complicated than the symmetric case here and less straightforward to invert into confidence regions. In a different asymptotic framework, Chang and Rivest (2001) also derive limiting distributions for M-estimates of the location parameter that minimize expressions of the form $\sum_i \rho(tr(S^T o_i))$, where the function $\rho$ satisfies certain conditions; Euler angles are not used, but limit laws there depend on underlying population quantities requiring estimation.

2.3 Simulation studies of finite-sample performance

To investigate the small sample impact of the asymptotic results, we generated 1000 samples from the $F(S, \kappa)$ distribution for different choices of $n$ and $\kappa$. The values used for $\kappa$ were 1, 5, 20, and 500 with sample sizes of $n = 10, 30, 100$. The location parameter was held fixed at
\(S = S(2.3, 1.1, 5.9)\) in (5), although symmetry arguments show any particular selection to be irrelevant. For each sample we calculated the likelihood ratio and Wald test statistics for the entire parameter vector. The cumulative empirical distribution function for each test statistic was then plotted along with the limiting distribution function. Plots for the likelihood ratio test statistics are presented in Figure 1, while Figure 2 contains similar plots for the Wald statistic.

The asymptotic cutoff for the likelihood ratio and Wald tests for the entire vector at an \(\alpha\)-level of 0.05 is 9.488. Examining the case of \(\kappa = 5\), this value corresponds to \(\alpha\)-levels of approximately 0.065, 0.067, and 0.059 for \(n = 10, 30,\) and 100, respectively, for the likelihood ratio test based on the empirical cumulative distribution functions represented in Figure 1, and to levels of 0.140, 0.085, and 0.066 for the Wald test based on Figure 2. Thus, the actual levels are more liberal than the desired \(\alpha = .05\) in small samples and convergence is apparently much slower in the case of the Wald test. Similar results hold for the other choices of \(\kappa\) and when testing for only the parameter \(\kappa\). When testing for only \((\alpha, \beta, \gamma)\) (that is, only for \(S\)), actual type I error probabilities for both tests are in closer agreement and are similar to those of the likelihood ratio test for the entire parameter vector.

In the next section, we consider Bayes analyses for the one-sample case of the \(F(S, \kappa)\) distribution. Non-informative priors are discussed along with an approach for simulation from the posterior. A simulation study then compares Bayes methods to the likelihood methods of this section.

3 One-sample Bayes analyses for the \(F(S, \kappa)\) distribution

3.1 One-sample model

We start by choosing non-informative prior distributions for the parameters of the \(F(S, \kappa)\) distribution. For the spread parameter \(\kappa\) we use a Jeffreys prior. The parameter \(\kappa\) controls the spread of the distribution with density in (3) and

\[-E \left( \frac{d^2}{d\kappa^2} \log(C(r|\kappa)) \right) = A(\kappa)\]
Figure 1: Limiting and empirical (estimated small $n$) null cumulative distribution functions for the likelihood ratio test statistic for the full parameter vector, $\kappa = 1, 5, 20, \text{and} 500$ (left to right, top to bottom)
Figure 2: Limiting and empirical (estimated small $n$) null cumulative distribution functions for the Wald test statistic for the full parameter vector, $\kappa = 1, 5, 20,$ and $500$ (left to right, top to bottom)
where \(A(\kappa)\) is specified in Proposition 1. Therefore, the Jeffreys prior for \(\kappa\) has density proportional to
\[
h_2(\kappa) = \frac{\sqrt{2} I_0(\kappa)^2 - 2 \frac{2}{\kappa^2} I_0(\kappa) I_1(\kappa) + \left(\frac{1}{\kappa^2} - 2\right) I_1(\kappa)^2}{I_0(2\kappa) - I_1(2\kappa)}, \quad \kappa \in [0, \infty).
\] (10)

For the location parameter \(S\) we use a uniform prior distribution on \(\Omega\), resulting in a density
\[
h_1(S) = 1, \quad S \in \Omega
\] (11)
with respect to the Haar measure; see also Chang and Bingham (1996) who consider large-sample approximations to the posterior when the location parameter \(S\) has a matrix Fisher prior. Multiplying (6), (11), and (10) the posterior density is proportional to
\[
g(S, \kappa) = \exp\left[\kappa \sum_{i=1}^{n} (tr(S^T o_i) - 1) \right] \frac{\sqrt{2} I_0(\kappa)^2 - 2 \frac{2}{\kappa^2} I_0(\kappa) I_1(\kappa) + \left(\frac{1}{\kappa^2} - 2\right) I_1(\kappa)^2}{[I_0(2\kappa) - I_1(2\kappa)]^{n+1}}. \tag{12}
\]

To compare the one-sample Bayes methods for the \(F(S, \kappa)\) distribution against the likelihood methods of Section 2, we implement an algorithm prescribed by Bingham, Vardeman, and Nordman (2009, sec. 2.1) for simulating posterior values using a Metropolis-Hastings within Gibbs algorithm. Those authors consider Bayes analyses for a UARS model (see Section 1) where the random “spin” \(r\) has a von-Mises circular density \(C(r|\kappa)\) in (2); we denote this matrix model as \(vM-UARS\). That is, in the Section 1 construction of rotation matrices, the density for \(r\) in a \(vM-UARS\) model differs from the corresponding density (3) in the symmetric matrix Fisher by a factor proportional to \((1 - \cos r)\). As a result, unlike the matrix Fisher density (4), the \(vM-UARS\) matrix density is irregular and has an infinite spike. (This implies that standard maximum likelihood does not exist for this model and work in Bingham, Vardeman, and Nordman (2009, sec. 2.2) compares a Bayes analysis to a “quasi-likelihood” method.)

We may implement this basic algorithm in the present context by replacing instances of the \(vM-UARS\) density with the Fisher density (4). With observations \(o_1, \ldots, o_n \in \Omega\), tuning parameters \(\rho\) and \(\sigma^2\), and some starting values \(S^0\) and \(\kappa^0\), our Metropolis-Hastings within Gibbs algorithm for \(j = 1, 2, \ldots\) is as follows (see Bingham, Vardeman, and Nordman (2009, sec. 2.1) for further details).
1. Generate $S^j_\ast \sim F(S, \kappa^j - 1)$ as a proposal for $S^j$. Compute $r^j_1 = \frac{g(S^j_\ast, \kappa^j - 1)}{g(S^j, \kappa^j - 1)}$ for $g$ in (12). Generate $W^j_1 \sim \text{Bernoulli}(\min(1, r^1_j))$ and let $S^j = W^j_1 S^j_\ast + (1 - W^j_1) S^{j - 1}$.

2. Generate $\log(\kappa^j_\ast) \sim N(\log(\kappa^j - 1), \sigma^2)$, with $\kappa^j_\ast$ as a candidate for $\kappa^j$. Compute $r^2_j = \frac{g(S^j, \kappa^j_\ast)_{\kappa^j_\ast}}{g(S^j, \kappa^{j - 1})_{\kappa^{j - 1}}} = \frac{g(S^j, \kappa^j_\ast)_{\kappa^j}}{g(S^j, \kappa^{j - 1})_{\kappa^{j - 1}}}$ for $g$ in (12). Generate $W^2_j \sim \text{Bernoulli}(\min(1, r^2_j))$ and let $\kappa^j = W^2_j \kappa^j_\ast + (1 - W^2_j) \kappa^{j - 1}$.

3.2 Simulation study of likelihood vs. Bayes inference (one-sample case)

For various combinations of $n$ and $\kappa$ we simulated 1000 samples from the $F(S, \kappa)$ distribution and then used the Metropolis-Hastings within Gibbs algorithm to generate samples from the posterior distribution. Again, the parameter $S$ was held constant at $S = S(2.3, 1.1, 5.9)$ in (5), with values of 1, 5, 20, and 500 for $\kappa$ and sample sizes of $n = 10, 30, 100$.

Based on each of the 1000 samples generated for a given $(n, \kappa)$ pair, we simulated a sample of size 20000 from the posterior distribution in (12) (taken after a burn-in of 5000 iterations). Starting values $S^0$ and $\kappa^0$ were set at the true parameters (after verifying that the choice of starting value did not affect the posterior simulations after the indicated burn-in period) and for each $(n, \kappa)$ combination tuning parameters were set to give Metropolis-Hastings jumping rates near 40% (see Gelman, Carlin, Stern, and Rubin, 2004, sec. 11.10).

We obtained point and set estimates for $S$ and $\kappa$ using both Bayes and likelihood methods for each $F(S, \kappa)$ sample simulated under a given $(n, \kappa)$ pair. Maximum likelihood point estimates for $\kappa$ and $S$ were computed. Bayes point estimates for $S$ were obtained by maximizing the posterior density, but we note that in the present case this estimate coincides with the MLE for $S$, described in Section 2.2. The Bayes point estimate used for $\kappa$ was the approximate posterior mean.

Our primary aim here is to compare the coverage probabilities and sizes of credible sets obtained from the Bayes analyses to those obtained using the likelihood methods. By inversion of the likelihood ratio and Wald tests given in Proposition 1, we obtain confidence regions for $\kappa$ and $S$. This results in a confidence interval for $\kappa$ and a set of orientations comprising the confidence set for $S$. In a manner similar to Bingham, Nordman, and Vardeman (2009), we
Figure 3: Graphical display of a confidence or credible region for the parameter $S$, with $x$, $y$, and $z$ representing the orientation of a corresponding point estimate for $S$

find the angle between each of the 3 perpendicular axes representing an orientation in the confidence set of $S$ and the corresponding axis of the MLE $\hat{S}$. With $\rho$ the maximum of all such angles, a confidence region for $S$ is viewed as the set of 3 cones of constant angle $\rho$, centered at $\hat{S}$. We consider 95% credible sets, so to obtain a credible interval for $\kappa$ we use the .025 and .975 quantiles of the $N = 20000$ values for $\kappa$ simulated from the posterior as the lower and upper bounds. Suppose $S_B$ represents the Bayes point estimate for $S$. To construct a credible set for $S$ we use the method of Bingham, Vardeman, and Nordman (2009). We find the maximum of the three angles between a coordinate axis rotated from its reference direction by $S_B$ and the corresponding axis rotated by $S^i$, where for $i = 1, \ldots, 20000$, $S^i$ are the values for $S$ simulated from the posterior. With $\delta_{.95} = \text{the .95 quantile of these 20000 angles}$, sets of 3 simultaneous cones of constant angle $\delta_{.95}$ around the coordinate axis rotated by $S_B$ create a region capturing 95% of posterior values for $S$. Figure 3 gives a graphical representation of a confidence or credible region for $S$. The notion of “size” of the regions for $S$ is reduced to considering the size of the angle between the center and edge of the cones.

We obtained 1000 posterior credible sets for both $S$ and $\kappa$ for each $(n, \kappa)$ combination, as well as 1000 confidence sets for each parameter from inversion of the likelihood ratio and Wald
tests of Section 2. To calculate the coverage rates for the inference methods under each \((n, \kappa)\) combination, we checked if the credible and confidence regions for \(\kappa\) and \(S\) obtained from each \(F(S, \kappa)\) sample contained the true parameter values and made relative frequencies of coverage. These estimated coverage probabilities are given in Table 2. The Bayes coverage is quite good (actual coverage probabilities agree closely with nominal ones set through specification of approximate credible levels) for both location and spread parameters, even in small sample sizes. Particularly for the \(\kappa\) parameter and small sample sizes \(n\), both likelihood ratio and Wald methods cover the true parameter value less often than nominal (with the Wald intervals being most liberal which agrees with convergence results in Section 2.3). Since both Bayes and likelihood methods have similar coverage rates as the sample sizes increase, we proceed to compare the “sizes” of confidence/credible regions obtained from both methods.

Table 2: Coverage rates (%) for \(\kappa\) and \(S\) under Bayes, likelihood ratio test, and Wald techniques for various choices of \((n, \kappa)\)

<table>
<thead>
<tr>
<th>((n, \kappa))</th>
<th>(\kappa)</th>
<th>(S)</th>
<th>(\kappa)</th>
<th>(S)</th>
<th>(\kappa)</th>
<th>(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((10, 1))</td>
<td>93.4</td>
<td>96.4</td>
<td>92.1</td>
<td>95.3</td>
<td>89.1</td>
<td>95.4</td>
</tr>
<tr>
<td>((30, 1))</td>
<td>94.9</td>
<td>96.1</td>
<td>95.0</td>
<td>98.1</td>
<td>94.5</td>
<td>98.3</td>
</tr>
<tr>
<td>((100, 1))</td>
<td>94.9</td>
<td>95.8</td>
<td>94.9</td>
<td>96.9</td>
<td>95.0</td>
<td>96.7</td>
</tr>
<tr>
<td>((10, 5))</td>
<td>93.9</td>
<td>94.4</td>
<td>91.1</td>
<td>98.6</td>
<td>81.3</td>
<td>98.6</td>
</tr>
<tr>
<td>((30, 5))</td>
<td>95.3</td>
<td>94.7</td>
<td>94.4</td>
<td>95.6</td>
<td>92.7</td>
<td>97.3</td>
</tr>
<tr>
<td>((100, 5))</td>
<td>94.5</td>
<td>94.9</td>
<td>94.9</td>
<td>96.2</td>
<td>93.7</td>
<td>95.2</td>
</tr>
<tr>
<td>((10, 20))</td>
<td>94.5</td>
<td>94.6</td>
<td>90.8</td>
<td>98.0</td>
<td>81.9</td>
<td>97.8</td>
</tr>
<tr>
<td>((30, 20))</td>
<td>95.2</td>
<td>94.5</td>
<td>95.1</td>
<td>96.7</td>
<td>92.7</td>
<td>96.8</td>
</tr>
<tr>
<td>((100, 20))</td>
<td>95.0</td>
<td>94.9</td>
<td>94.2</td>
<td>95.8</td>
<td>93.8</td>
<td>96.2</td>
</tr>
<tr>
<td>((10, 500))</td>
<td>93.7</td>
<td>96.2</td>
<td>92.2</td>
<td>98.1</td>
<td>85.7</td>
<td>98.4</td>
</tr>
<tr>
<td>((30, 500))</td>
<td>95.6</td>
<td>94.7</td>
<td>95.8</td>
<td>97.0</td>
<td>94.9</td>
<td>96.5</td>
</tr>
<tr>
<td>((100, 500))</td>
<td>94.7</td>
<td>94.8</td>
<td>95.0</td>
<td>95.7</td>
<td>94.5</td>
<td>95.6</td>
</tr>
</tbody>
</table>

The median widths for the 1000 intervals for \(\kappa\) are given in Table 3. From the table, we see that posterior credible intervals for \(\kappa\) slightly outperform both types of likelihood intervals for \(\kappa \geq 5\) and are particularly superior to Wald intervals in small samples. When \(\kappa = 1\), the likelihood ratio intervals are slightly narrower than the credible intervals, but the two methods produce similar results as \(n\) increases.

We characterize the size of confidence and credible regions for \(S\) by the angle of the conic
Table 3: Median width of Bayes credible intervals, likelihood ratio intervals, and Wald intervals for $\kappa$ for various choices of $(n, \kappa)$

<table>
<thead>
<tr>
<th>$(n, \kappa)$</th>
<th>Bayes</th>
<th>LRT</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1)</td>
<td>1.2491</td>
<td>1.1621</td>
<td>1.4248</td>
</tr>
<tr>
<td>(30, 1)</td>
<td>0.6481</td>
<td>0.6428</td>
<td>0.6744</td>
</tr>
<tr>
<td>(100, 1)</td>
<td>0.3481</td>
<td>0.3475</td>
<td>0.3521</td>
</tr>
<tr>
<td>(10, 5)</td>
<td>5.3165</td>
<td>5.6748</td>
<td>7.5852</td>
</tr>
<tr>
<td>(30, 5)</td>
<td>2.8716</td>
<td>2.9355</td>
<td>3.2037</td>
</tr>
<tr>
<td>(100, 5)</td>
<td>1.5585</td>
<td>1.5676</td>
<td>1.6079</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>21.9420</td>
<td>23.4370</td>
<td>31.2846</td>
</tr>
<tr>
<td>(30, 20)</td>
<td>11.7951</td>
<td>12.0597</td>
<td>13.1545</td>
</tr>
<tr>
<td>(100, 20)</td>
<td>6.4067</td>
<td>6.4390</td>
<td>6.6036</td>
</tr>
<tr>
<td>(10, 500)</td>
<td>552.8769</td>
<td>591.3957</td>
<td>778.5897</td>
</tr>
<tr>
<td>(30, 500)</td>
<td>298.0555</td>
<td>306.2049</td>
<td>330.8834</td>
</tr>
<tr>
<td>(100, 500)</td>
<td>161.1862</td>
<td>162.3764</td>
<td>166.5256</td>
</tr>
</tbody>
</table>

Table 4: Median angle of Bayes credible regions, likelihood ratio regions, and Wald regions for $S$ for various choices of $(n, \kappa)$

<table>
<thead>
<tr>
<th>$(n, \kappa)$</th>
<th>Bayes</th>
<th>LRT</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1)</td>
<td>1.02560</td>
<td>0.91892</td>
<td>1.92793</td>
</tr>
<tr>
<td>(30, 1)</td>
<td>0.54276</td>
<td>0.54905</td>
<td>0.62062</td>
</tr>
<tr>
<td>(100, 1)</td>
<td>0.29240</td>
<td>0.29930</td>
<td>0.31732</td>
</tr>
<tr>
<td>(10, 5)</td>
<td>0.29699</td>
<td>0.29279</td>
<td>0.33814</td>
</tr>
<tr>
<td>(30, 5)</td>
<td>0.16778</td>
<td>0.17117</td>
<td>0.17958</td>
</tr>
<tr>
<td>(100, 5)</td>
<td>0.09081</td>
<td>0.09369</td>
<td>0.09550</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>0.14307</td>
<td>0.14084</td>
<td>0.15646</td>
</tr>
<tr>
<td>(30, 20)</td>
<td>0.08069</td>
<td>0.08228</td>
<td>0.08499</td>
</tr>
<tr>
<td>(100, 20)</td>
<td>0.04356</td>
<td>0.04505</td>
<td>0.04535</td>
</tr>
<tr>
<td>(10, 500)</td>
<td>0.02815</td>
<td>0.02783</td>
<td>0.03033</td>
</tr>
<tr>
<td>(30, 500)</td>
<td>0.01586</td>
<td>0.01622</td>
<td>0.01662</td>
</tr>
<tr>
<td>(100, 500)</td>
<td>0.00861</td>
<td>0.00891</td>
<td>0.00892</td>
</tr>
</tbody>
</table>

Now that we have verified the one-sample Bayes methods produce practically reliable re-
sults, Section 4 extends these to cover one-way random effects models with the symmetric matrix Fisher distribution. In Section 5, the Bayes methods are applied to an example in human kinematics.

4 One-way random effects Bayes analyses for the Fisher model

Bingham, Vardeman, and Nordman (2009) explore the use of Bayes analyses for a 3-d rotation version of a one-way random effects model based on the von Mises UARS model. In a manner similar to the presentation there, suppose that for $i = 1, \ldots, r$ and $k = 1, \ldots, m_i$,

$$O_{ik} = P_i Q_{ik} \in \Omega$$

for $P_i \sim iid F(S, \kappa)$ independent of $Q_{ik} \sim iid F(I_{3x3}, \tau)$. Here $\kappa$ quantifies the between-group variation and $\tau$ quantifies the within-group variation (with large values indicating small variation). These parameters and the location parameter $S$ can be estimated using Bayes methods.

Placing independent Jeffreys priors on both $\kappa$ and $\tau$, and uniform (Haar) prior on $S$, following the work of Bingham, Vardeman, and Nordman (2009), we have posterior density for the unobservable $p_i$ and parameters proportional to

$$G(S, \kappa, \tau, p_1, \ldots, p_r) = f(p_1, \ldots, p_r|S, \kappa, \tau) h_2(\tau) h_2(\kappa)$$

(13)

for

$$f(o_{11}, \ldots, o_{rm_r}, p_1, \ldots, p_r|S, \kappa, \tau) = \prod_{i=1}^{r} \left( \exp \left[ \tau \sum_{k=1}^{m_i} \left( tr(p_i^T o_{ik}) - 1 \right) + \kappa \left( tr(S^T p_i) - 1 \right) \right] \right) \left[ \frac{I_0(2\tau)}{I_1(2\tau)} \right]^{m_i} \left[ \frac{I_0(2\kappa)}{I_1(2\kappa)} \right]$$

with $o_{ik}$ fixed and dependence upon them suppressed, and $h_2$ as given in (10).

We use the Metropolis-Hastings within Gibbs algorithm for the one-way random effects case of Bingham, Vardeman, and Nordman (2009, sec. 3) to simulate values from this posterior, replacing instances of the vM-UARS distribution by the Fisher distribution. In Section 3, we verified the performance of Bayes methods over several level combinations of sample size and parameter values, but here we will restrict ourselves to one case (due to far longer computing times involved in the one-way random effects case and the huge variety of potentially relevant sample size and parameter combinations). We simulated 1000 independent data sets where
$m_i = 30$ for $i = 1, \ldots, 30$, $\tau = 20$, $\kappa = 20$, and $S = S(2.3, 1.1, 5.9)$ for $S$ in (5), and using each of the 1000 data sets we generated a sample of size 8000 from the posterior density in (13) after a burn-in of 2000 iterations. Bayes point estimates and 95% posterior credible sets were then obtained. Estimates and intervals for $\kappa$ and $\tau$ were both computed as for $\kappa$ in the one-sample case. The mean directions (see Section 1) of the 8000 values for $S$ simulated from the posterior were used as the point estimates for $S$ and credible sets for $S$ were computed in the same manner as for the one-sample situation.

In checking if the 95% credible intervals contained the true parameters values, we found that the coverage rates were again as desired. Rates were .951, .950, and .962 for $\kappa$, $\tau$, and $S$, respectively. Median sizes of credible regions were also found, with a median width of 11.87740 for $\kappa$, 2.15577 for $\tau$, and a median cone angle of 0.08032 for $S$. We are correctly able to capture the indicator of within group variation $\tau$ in smaller intervals than the indicator of between group variation $\kappa$. We now apply the methods discussed thus far to a human kinematics example.

### 5 Application to human kinematics

Rancourt et al. (2000) discuss an experiment in which an infra-red camera system was used to obtain orientations of the wrist, elbow, and shoulder of eight individuals when drilling into a vertically positioned metal plane at six different locations. Each subject performed the drilling at each location five times. Of interest here is the variation of the five repetitions as well as the variation between different subjects when performing the same drilling task. (Since the infra-red emitting markers were not placed at exactly the same orientation on each subject, the latter includes both subject-to-subject variation and variation in marker placement.) We will use a subset of the original data, namely the wrist orientations when drilling at location 4. (Because there were instances when one of the limb markers was out of the camera’s field of view, only three of the joint $\times$ location combinations have complete data for all five repetitions for each subject, which includes the location 4 data.) For the chosen combination of wrist and location 4, we verified that the assumption of spherical symmetry is reasonable by computing
Prentice’s R statistic for the eight data sets of size \( n = 5 \) (see Prentice, 1984); Mauchly’s sphericity test for comparing subjects also produced a small test statistic (see Rancourt et al., 2000). Hence, the symmetric Fisher is a plausible model.

We first applied the one-sample methods of Sections 2 and 3.1 to the repetitions for each subject independently. For each subject, this resulted in a point estimate for \( S \) (same value obtained by maximizing either the likelihood or the posterior), a maximum likelihood estimate for \( \kappa \), and a Bayes point estimate for \( \kappa \) (based on 20000 values simulated from the posterior). The estimates for all subjects are given in Table 5. Additionally, 95% confidence/credible intervals for \( \kappa \) were obtained as presented in Table 6. Location parameter \( S \) estimates vary by subject due to the differing marker placements. In terms of the local coordinate system for the wrist marker for each subject, estimates of \( S \) imply the rotation needed to bring the upper arm from a position to parallel to the body to a horizontal position for drilling; see also León et al. (2006). We also observe from the tables that even though the point estimates for \( S \) vary, the estimated values for the spread parameter \( \kappa \) are large for all subjects, indicating small variation among repetitions. To better characterize this variation, note that for large \( \kappa \) values (as in Table 5) the distribution (3) on \( r \) is approximately the Maxwell-Boltzmann distribution with parameter \( 1/\sqrt{2\kappa} \). So, for example, using the Bayes estimate \( \hat{\kappa} \), a 99% “prediction” interval for the misorientation angle of rotation, \(|r|\), is \((0^\circ, 6.04^\circ)\) for Subject 1 and \((0^\circ, 9.38^\circ)\) for Subject 2.

Table 5: Point estimates for \( S \) in the form of Euler angles and maximum likelihood and Bayes point estimates for \( \kappa \) for each of 8 subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Point estimate for ( S )</th>
<th>MLE for ( \kappa )</th>
<th>Bayes for ( \kappa )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1.7660, 0.3410, 4.9723)</td>
<td>642.7361</td>
<td>510.6493</td>
</tr>
<tr>
<td>2</td>
<td>(3.4494, 0.2359, 2.7195)</td>
<td>266.8360</td>
<td>211.4876</td>
</tr>
<tr>
<td>3</td>
<td>(2.2912, 0.5886, 3.9828)</td>
<td>344.4092</td>
<td>273.4118</td>
</tr>
<tr>
<td>4</td>
<td>(2.1723, 0.4354, 4.2163)</td>
<td>328.1112</td>
<td>261.2676</td>
</tr>
<tr>
<td>5</td>
<td>(1.8157, 0.4082, 4.7234)</td>
<td>588.9460</td>
<td>471.3010</td>
</tr>
<tr>
<td>6</td>
<td>(2.6089, 0.3111, 3.8741)</td>
<td>732.4504</td>
<td>582.8049</td>
</tr>
<tr>
<td>7</td>
<td>(5.2642, 0.0960, 0.7900)</td>
<td>320.2209</td>
<td>257.5881</td>
</tr>
<tr>
<td>8</td>
<td>(4.6643, 0.2583, 1.3429)</td>
<td>290.9681</td>
<td>232.7668</td>
</tr>
</tbody>
</table>

We examine the subject-to-subject variation by extending the analyses to the one-way ran-
Table 6: 95% likelihood ratio, Wald, and Bayes intervals for $\kappa$ for each of 8 subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>LRT</th>
<th>Wald</th>
<th>Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(285.3059, 1218.4857)</td>
<td>(374.6528, 2260.7559)</td>
<td>(181.4162, 984.0362)</td>
</tr>
<tr>
<td>2</td>
<td>(118.2568, 505.0517)</td>
<td>(155.2921, 937.0745)</td>
<td>(75.3067, 422.5639)</td>
</tr>
<tr>
<td>3</td>
<td>(152.7311, 652.2893)</td>
<td>(200.9962, 1212.8655)</td>
<td>(97.3861, 526.2531)</td>
</tr>
<tr>
<td>4</td>
<td>(145.5078, 621.4351)</td>
<td>(191.0722, 1152.9809)</td>
<td>(92.4487, 514.3660)</td>
</tr>
<tr>
<td>5</td>
<td>(257.3139, 1123.3561)</td>
<td>(302.0207, 1822.4741)</td>
<td>(174.1514, 919.6204)</td>
</tr>
<tr>
<td>6</td>
<td>(324.3013, 1386.4408)</td>
<td>(414.9447, 2503.8885)</td>
<td>(220.7126, 1120.7901)</td>
</tr>
<tr>
<td>7</td>
<td>(141.5005, 605.6558)</td>
<td>(178.9143, 1079.6171)</td>
<td>(96.0619, 506.3005)</td>
</tr>
</tbody>
</table>

dom effects case. The parameter $\kappa$ will quantify the variation in wrist orientation between subjects when drilling at location 4 on the metal plane, while $\tau$ quantifies the variation present in the repeat drillings where $m_i = 5$ for each $i = 1, \ldots, 8$. The Metropolis-Hastings within Gibbs algorithm was used to simulate 8000 values from the posterior in (13) and Bayes estimates and credible intervals were obtained as discussed in Section 4. The Bayes estimates are $\hat{\kappa} = 9.4673$, $\hat{\tau} = 306.8598$, and $\hat{S} = S(2.2659, 0.2136, 4.0702)$ for $S$ in (5). The 95% credible intervals are (4.7697, 16.3782) and (230.9254, 397.6437) for $\kappa$ and $\tau$, respectively, with conic regions of angle 0.0373 representing a 95% credible region for $S$. As expected, the interval for $\tau$ represents orientations with less variability than that for $\kappa$, so the between subject variation is greater than the within subject variation. To view these measured variations in a different manner, with $\hat{\kappa} = 9.4673$ the distribution on the misorientation angles places 99% probability in the interval $\left(0^\circ, 45.901^\circ\right)$, and with $\hat{\tau} = 306.8598$, 99% probability is placed the interval $\left(0^\circ, 7.798^\circ\right)$. So, while the within subject variation is small, there is moderate variation present across subjects which can be attributed to subject-to-subject variation and variation in marker placement.

6 Conclusion

For one-sample inference on the symmetric matrix Fisher distribution, we have examined both likelihood and non-informative Bayes inference and found that Bayes methods produce competitive point and interval estimates. In particular, the coverage rates of Bayes credible
regions are quite close to nominal levels in small samples and, as expected, the coverage rates
match those of likelihood regions for larger samples. However, under the symmetric Fisher
model, likelihood ratio tests do exhibit rather fast convergence to their chi-squared limits and
appear to surpass Wald statistics for calibrating confidence regions.

We have also illustrated that Bayes methods can be extended for handling one-way random
effects models, which represents a new inference scenario for the matrix Fisher distribution. It
appears that the Bayes approach may open other inference possibilities involving the Fisher
distribution, and work in Chang and Bingham (1996) suggests some possibilities for informative
prior development. Further development and investigation of Bayes tools may help advance
inference, not only in the Fisher model, but also in other UARS models for rotation matrices
which share a common “matrix construction” that is amenable to modern MCMC-Bayes

References

Measured Crystal Orientations and a Tractable Class of Symmetric Distributions for Ro-
tations in 3 Dimensions,” Preprint, Department of Statistics, Iowa State University.

Bingham, M. A., Vardeman, S. B., and Nordman, D. J. (2009) “Bayes One-Sample and One-
Way Random Effects Analyses for 3-D Orientations with Application to Materials Science,”
Preprint, Department of Statistics, Iowa State University.

in Statistics and Econometrics, eds. Berry, D. A., Chaloner, K. M., and Geweke, J. K.,
New York: John Wiley & Sons.

in Group Models: A Case Study Using Stiefel Manifolds,” The Annals of Statistics, 29,
784-814.

State University.


Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), Bayesian Data Analysis,
Boca Raton: Chapman & Hall.


A STATISTICAL ANALYSIS OF VARIATION IN MEASURED CRYSTAL ORIENTATIONS OBTAINED THROUGH ELECTRON BACKSCATTER DIFFRACTION

A paper submitted to Ultramicroscopy

Melissa A. Bingham¹, Barbara K. Lograsso, and Francis C. Laabs

Iowa State University, Michigan Technological University, and Ames Laboratory

Abstract

Electron Backscatter Diffraction (EBSD) techniques (sometimes known as Orientation Imaging Microscopy (OIM)) are used to determine the crystallography of individual metal grains. This paper examines the variability in measurements obtained by EBSD. Based on repeat scans of the same region on a standard nickel sample, three sources of variability are explored: variation in repeat measurements at a fixed location, variation among locations within a grain, and grain-to-grain variation. To quantify the importance of these three sources of variation, Bayes statistical methods are applied to a hierarchical model with the Uniform-Axis-Random-Spin (UARS) components of Bingham, Nordman, and Vardeman (2009).

Key words: Bayes, credible intervals, Electron Backscatter Diffraction, Orientation Imaging Microscopy, misorientation angle, UARS distribution, von Mises distribution

¹Supported by NSF grant DMS #0502347 EMSW21-RTG awarded to the Department of Statistics, Iowa State University, and by the Ames Laboratory through U.S. Department of Justice COPS Program grant #2005CKWX0466 and interagency agreement #2002-LP-R-083 by the National Institute of Justice, through the Midwest Forensics Resource Center. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under contract #DE-AC02-07CH11358.
1 Introduction

Most crystalline materials are composed of a collection of grains of varying sizes and orientations of their internal crystal lattices. This paper focuses on metals with a cubic crystal structure. A large mismatch in orientation between two crystal lattices at physically adjacent locations defines a crystal boundary (grain boundary) and can be described in terms of axis-angle pairs (see Humphries, 2001). That is, if cube A is positioned next to cube B and cube B is rotated with reference to cube A, there is an axis common to a crystallographic plane for both cubes. A rotation of cube B about this axis will bring it into coincidence with cube A. Due to the cubic symmetry of the structures considered here, there are twenty-four such rotation angles. The smallest angle which brings cube B into coincidence with cube A is known as the misorientation angle (see Mackenzie, 1957). We consider analyses in which data have already been preprocessed so that the rotation angle between any two orientations gives the misorientation angle (and there is no 24-fold ambiguity).

Since Electron Backscatter Diffraction (EBSD) is commonly used to determine the microstructure of metals, it is important to investigate the consistency of measurements taken using this methodology. In this study, fourteen repeat scans of the same region on a standard nickel sample were taken. Grains of similarly oriented crystals were identified and twenty locations were chosen from each of eight grains. To investigate the adequacy of the measurements, we look at variation in the corresponding measured crystal orientations. Of interest are variation in the repeat measurements for the fourteen scans, variation among crystal orientations for locations within a grain, and grain-to-grain variation. Brewer, Kotula, and Michael (2008) explored the use of multivariate statistical analysis to reduce a large number of patterns into statistical sets for improved indexing. Funderberger, Morawiec, Bouzy, and Lecomte (2003) determined that the accuracy of orientation measurements from TEM Kikuchi patterns can be better than 0.1 degree. Our study seeks to consider the repeatability of measurements not only within a grain but also scan-to-scan, which is possible through a class of probability models for describing orientation measurements. The Uniform-Axis-Random-Spin (UARS) class of distributions identified by Bingham, Nordman, and Vardeman (2009) can be used to model
the deviations of a set of orientations from some “true” or central orientation. A hierarchical model with UARS components allows us to simultaneously represent all three sources of variation, and Bayes analyses will be used to provide estimates for the sizes of contributions of these three sources, along with an estimate of the “true” or principal orientation. In the next section, we provide details about the EBSD measurement process.

2 EBSD background

This paper focuses on crystal orientation measurements made using EBSD, which gives point-to-point measurements across the surface of the metal. When an electron beam impinges on lattice planes of atoms within the metal crystals, the electron beam is diffracted in a spread that produces reflected Kikuchi lines, appearing as bands on a recording screen. The clarity of the generated diffraction patterns, and hence the ability to determine the orientation, is affected by both the degree of crystal deformation and by the surface quality, since the patterns are generated from the very surface (first few atomic planes) of the metal. A resulting pattern, shown in Figure 1, consists of intersecting Kikuchi lines (bands) whose intensity can be calculated using a kinematical electron diffraction model which is a function of the crystal structure and chemistry of the material and instrument parameters. Further, the spacing and intersections of the bands depend on the geometry of the crystal lattice (i.e., angles between atomic planes) thereby allowing the atomic planes to be indexed and the overall orientation of the crystal lattice to be determined. The intersections of Kikuchi bands are indexed by calculating the angular distances between intersections and comparing them to known angles for a cubic unit cell. The indexing is automated, allowing for the fast collection of crystal orientations. Collected EBSD patterns are analyzed by comparison to theoretical patterns based on the known cubic crystal structure of the metal. Angular differences in band location, band intensity, and differences in the width of bands allow for the unique identification of crystal orientation. For more on Kikuchi diffraction patterns, see Randle (2003) and Schwarzer (1997).

EBSD patterns are collected automatically at regularly spaced locations (pixels) across a pre-designated area. At each location, the specific orientation of the crystal is determined
Figure 1: A typical EBSD pattern showing variations in band intensity that result from the differences in the atomic arrangement associated with the specific lattice plane from the pattern collected and a confidence index number for each point is assigned based on how well the calculated diffraction pattern matches the experimental pattern. The value of the confidence index factor is between zero and one, with one being the highest confidence. This confidence index is dependent on the deformation and quality of polish, i.e. the overall surface quality at that location. The distance the beam moves between adjacent locations is known as the step size and for this study was set to be less than the size of the grain at .2 µm. The beam then step scans over the chosen area and orientation information is collected for all points on a rectangular grid within the area. The complete data set is then used for generating orientation maps using proprietary software. Line scans were also used in this study.

The orientation of a crystal relative to some reference coordinate system is commonly represented in the form of Euler angles. Euler angles specify a sequence of three rotations that bring a crystal into coincidence with the reference coordinate system. For Euler angles $\varphi_1$, $\phi$, and $\varphi_2$, where $\varphi_1 \in [0, 2\pi]$, $\phi \in [0, \pi]$, and $\varphi_2 \in [0, 2\pi]$, these three rotations, in order, are

- a rotation of $\varphi_1$ radians about the $z$-axis at $(0, 0, 1)$,
- a rotation of $\phi$ radians about the $x$-axis at $(1, 0, 0)$, and
- a rotation of $\varphi_2$ radians again about the $z$-axis.

Since rotations in $\mathbb{R}^3$ can be represented by a $3 \times 3$ orthogonal rotation matrix, the orientation of a crystal relative to a reference coordinate system can also be described by a $3 \times 3$ orthogonal rotation matrix.
Figure 2: Standard crystallographic triangle relating grain orientation color to the vector normal of the crystallographic planes

In this study, the data were represented using maps to show the crystal orientations (i.e., Euler angles) with a pseudo color microstructural image. This microstructural imaging based on orientation is referred to as Orientation Imaging Microscopy (OIM). The OIM software assigned a particular color to each crystal based on the orientation of the cubic structure of the crystal at each point relative to the surface normal. The color selected corresponds to the crystal orientation given in the standard crystallographic triangle shown in Figure 2. This figure is related to the cubic structure of the atomic arrangements. Note that the vertex labeled $<111>$ is the vector notation for the direction normal to the 111 plane and represents the cube diagonal. The cube face is represented by the notation $<001>$, and $<101>$ corresponds to the face diagonal. Therefore, each color was assigned based not only on how it is physically related to the surface of the metal, but to the underlying atomic arrangement of the cubic lattice. Figure 3 is a map of the microstructure observed on the nickel sample studied as determined by one of fourteen scans taken on the standard nickel specimen. In the next section we review the UARS class of distributions identified by Bingham, Nordman, and Vardeman (2009) that will later be applied to modeling crystal orientations using a hierarchical structure.

3 The UARS class of distributions

The UARS class of distributions is useful for modeling the deviation of random orientations $O_1, \ldots, O_n \in \Omega$ from a common “true” orientation of $S \in \Omega$, here referred to as the true or...
Figure 3: Map of color coded crystal orientations for a nickel sample

principal orientation, where $\Omega$ represents the set of $3 \times 3$ orthogonal rotation matrices that preserve the right hand rule. If a random orientation $O \in \Omega$ has a UARS distribution with principal direction $S$, we can write $O = S \cdot P$, where

$$P = UU^T + (I_{3 \times 3} - UU^T) \cos r + \begin{pmatrix}
0 & -u_3 & u_2 \\
u_3 & 0 & -u_1 \\
-u_2 & u_1 & 0
\end{pmatrix} \sin r \in \Omega$$

is obtained by rotating $I_{3 \times 3}$ (the $3 \times 3$ identity matrix) about the unit vector $U^T = (u_1, u_2, u_3) \in \mathbb{R}^3$ by a random angle $r \in (-\pi, \pi]$. The axis $U$ is identified by a point uniformly distributed on the unit sphere. By construction, the angle $r$ is the misorientation angle between the principal direction $S$ and the final orientation $O$ and we suppose $r$ comes from a distribution on $(-\pi, \pi]$ that is symmetric about 0, with spread depending on the parameter $\kappa$. Since $\kappa$ controls the spread of the distribution on the misorientation angle, $r$, it controls the variation of a corresponding UARS observation $O$ from the true direction $S$.

In this paper, we suppose $r$ is distributed according to the von Mises circular distribution with direction 0 and concentration parameter $\kappa \in [0, \infty)$, denoted by $\text{vM}(0, \kappa)$ (see Mardia and Jupp, 2000). The $\text{vM}(0, \kappa)$ distribution is unimodal on $(-\pi, \pi]$, symmetric about 0, converges to the uniform distribution on $(-\pi, \pi]$ as $\kappa \to 0$, and becomes approximately Normal with
mean 0 and variance $\frac{1}{\kappa}$ as $\kappa \to \infty$. The UARS distribution with $\text{vM}(0, \kappa)$ distribution on $r$ and principal direction $\mathbf{S}$ is denoted by $\text{vM-UARS}(\mathbf{S}, \kappa)$.

Large values of $\kappa$ represent distributions on the misorientation angle with small spread. Consequently, when fitting a $\text{vM-UARS}(\mathbf{S}, \kappa)$ distribution to a set of measured crystal orientations, a larger estimated value for $\kappa$ represents orientations with less variation. Figures 4 and 5 (taken from Bingham, Nordman, and Vardeman (2009)) illustrate the variation present in orientations generated from $\text{vM-UARS}(\mathbf{S}, \kappa)$ distributions with $\kappa$ values of 5 and 10. In Figure 4, each set of 3 perpendicular axes represents one orientation and we see that the variation in the orientations is less for the larger value of $\kappa$. The spheres in Figure 5 are marked with contours outlining regions that encompass increasing amounts of probability associated with the two $\text{vM-UARS}(\mathbf{S}, \kappa)$ distributions. With the contour closest to each axis considered to be the “first” contour, $(10 \times i)\%$ of realizations keep all 3 perpendicular axes within the area represented by the $i^{th}$ contours about $x$, $y$, and $z$ (simultaneously). Probability accumulates more quickly as orientations move away from the principal direction with increasing $\kappa$.

Figure 4: Five random orientations generated from $\text{vM-UARS}$ distributions with $\kappa = 5$ and 10, respectively (the axes at $x$, $y$, and $z$ are those represented by principal direction $\mathbf{S}$)

To examine the effect that $\kappa$ has on the spread of the distribution for the misorientation angle $r$ (i.e. $\text{vM}(0, \kappa)$) and thus the corresponding “spread” of the $\text{vM-UARS}(\mathbf{S}, \kappa)$ distribution, Bingham, Nordman, and Vardeman (2009) presented Table 1. Here, $\Delta_1(\kappa)$ is the median of
Figure 5: Probability content contours for vM-UARS distributions with $\kappa = 5$ and 10, respectively (the axes shown are those represented by principal direction $S$)

the distribution on $|r|$ so that $(-\Delta_1(\kappa), \Delta_1(\kappa))$ captures 50% of the vM(0, $\kappa$) probability. The median of the distribution of the maximum angle between an $S$-rotated and a vM-UARS($S$, $\kappa$)-rotated coordinate axis is represented by $\Delta_2(\kappa)$. From the table, we see there is a close relationship between $\Delta_1(\kappa)$ and $\Delta_2(\kappa)$, and $\kappa$ directly controls the concentrations of the von Mises and vM-UARS distributions. In the next section we introduce a hierarchical model with vM-UARS components. Bayes analyses are then used to estimate the parameters of this model as applied to crystal orientation measurements obtained using EBSD.

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>$\Delta_1(\kappa)$</th>
<th>$\Delta_2(\kappa)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.80977</td>
<td>0.77526</td>
</tr>
<tr>
<td>5</td>
<td>0.31170</td>
<td>0.30218</td>
</tr>
<tr>
<td>10</td>
<td>0.21657</td>
<td>0.20923</td>
</tr>
<tr>
<td>20</td>
<td>0.15194</td>
<td>0.14697</td>
</tr>
<tr>
<td>50</td>
<td>0.09567</td>
<td>0.09232</td>
</tr>
<tr>
<td>100</td>
<td>0.06755</td>
<td>0.06519</td>
</tr>
<tr>
<td>500</td>
<td>0.03017</td>
<td>0.02909</td>
</tr>
</tbody>
</table>

4 Bayes analyses for a hierarchical model with vM-UARS components

Suppose that for $i = 1, \ldots, t$, $j = 1, \ldots, m_i$, and $k = 1, \ldots, l_j$

$$O_{ijk} = P_i Q_{ij} R_{ijk} \in \Omega$$
for \( P_i \sim_{iid} \text{vM-UARS}(S, \nu) \) independent of \( Q_{ij} \sim_{iid} \text{vM-UARS}(I_{3 \times 3}, \tau) \), all independent of \( R_{ijk} \sim_{iid} \text{vM-UARS}(I_{3 \times 3}, \eta) \). In our context we have \( t \) different grains with \( m_i \) (that may differ with \( i \)) scanned locations per grain and \( l_j \) (that may differ with \( j \)) repeat scans per location. Then \( \nu \) quantifies the grain-to-grain variation, \( \tau \) quantifies the within grain variation, and \( \eta \) quantifies the variation across repeated scans (with larger values indicating less variation in all cases). We will use Bayes statistical methods to estimate these parameters as well as the overall “location” parameter \( S \).

First, we write

\[
g(o_{ijk}, p_i, q_{ij}|S, \nu, \tau, \eta) = f(o_{ijk}|q_{ij}, \eta)f(q_{ij}|p_i, \tau)f(p_i|S, \nu),
\]

where

\[
f(o|S, \kappa) = \frac{2}{3 - tr(S^T o)} [I_0(\kappa)]^{-1} \exp \left[ \frac{\kappa}{2} (tr(S^T o) - 1) \right], \quad o \in \Omega,
\]

is the density of the vM-UARS\((S, \kappa)\) distribution with respect to the Haar probability distribution on \( \Omega \) (see Bingham, Nordman, and Vardeman, 2009) where \( I_0 \) represents the modified Bessel function of order zero. (The invariant Haar distribution acts as a “uniform distribution” on \( \Omega \) (see Downs, 1972).) Thus, for the hierarchical model we have a joint density for the (observable) \( o_{ijk} \) and (unobservable) \( p_i \) and \( q_{ij} \) as

\[
g(o_{111}, \ldots, o_{mtm_t}, p_1, \ldots, p_t, q_{11}, \ldots, q_{tm}, S, \nu, \tau, \eta) = \prod_{i=1}^{t} \left\{ \prod_{j=1}^{m_i} \left\{ \prod_{k=1}^{l_j} \frac{1}{f(o_{ijk}|q_{ij}, \eta)f(q_{ij}|p_i, \tau)} f(p_i|S, \nu) \right\} \right\}.
\]

To perform a Bayes analysis (see Gelman, Carlin, Stern, and Rubin, 2004; Carlin and Louis, 2000) we begin by choosing “non-informative” prior distributions for the parameters \( S, \nu, \tau, \) and \( \eta \). For the variation parameters we use a Jeffreys prior with density proportional to

\[
h_1(\lambda) = \frac{\sqrt{I_0(\lambda)^2 - \frac{1}{4} I_0(\lambda)I_1(\lambda) - I_1(\lambda)^2}}{I_0(\lambda)}, \quad \lambda \in [0, \infty)
\]

(see Bingham, Vardeman, and Nordman, 2009), where \( I_i \) is the modified Bessel function of order \( i \). For the location parameter \( S \) we use a prior distribution uniform on \( \Omega \) which is
specified by the density

\[ h_2(S) = 1, \ S \in \Omega \]  

with respect to the Haar measure. We then have posterior density for the unobservable \( p_i \) and \( q_{ij} \) and parameters proportional to

\[ G(S, \nu, \tau, \eta, p_1, \ldots, p_t, q_{11}, \ldots, q_{tm}) = g(p_1, \ldots, p_t | S, q_{11}, \ldots, q_{tm}, \nu, \tau, \eta) h_1(\eta) h_1(\tau) h_1(\nu) h_2(S) \]  

for \( g \) given in (2) with \( o_{ijk} \) fixed and dependence upon them suppressed, \( h_1 \) in (3), and \( h_2 \) in (4).

To simulate values from the posterior distribution we use a Metropolis-Hastings-within-Gibbs simulation algorithm that is a generalization of a simpler “one-way random effects” algorithm introduced in Bingham, Vardeman, and Nordman (2009). (For more on Metropolis Hastings and Gibbs, see Gelman et. al. (2004).) Suppose that after \( b - 1 \) iterations of the algorithm one has parameters \( S^{b-1}, \nu^{b-1}, \tau^{b-1}, \) and \( \eta^{b-1} \), as well as \( \{p_1^{b-1}, \ldots, p_t^{b-1}\} \) and \( \{q_{11}^{b-1}, \ldots, q_{tm}^{b-1}\} \). For the parameter \( S \), we obtain a candidate for \( S^b \) as \( S^b \sim \nu \text{M-UARS}(S^{b-1}, \rho_1) \) (where \( \rho_1 \) is a tuning parameter that can be adjusted to make the algorithm efficient). We note that this choice of proposal for \( S \) is symmetric in the sense that \( f(S'|S, \rho_1) = f(S|S', \rho_1) \) for \( f \) in (1). We use a similar proposal distribution when updating the members of \( \{p_1, \ldots, p_t\} \) and \( \{q_{11}, \ldots, q_{tm}\} \). For the parameter \( \nu \), we take \( \log(\nu^{b+1}) \sim \text{N}(\log(\nu^{b-1}), \sigma_1^2) \), i.e. \( \nu^{b+1} \) is log-normal with parameters \( \log(\nu^{b-1}) \) and (the tuning parameter) \( \sigma_1^2 \). If \( t(x|\mu, \sigma_1^2) \) represents the log-normal density, then \( \frac{t(\nu'|\log(\nu), \sigma_1^2)}{t(\nu|\log(\nu'), \sigma_1^2)} = \frac{\nu}{\nu'} \). We use a similar proposal for \( \tau \) and \( \eta \).

Thus, for observations \( o_{ijk}, i = 1, \ldots, t, j = 1, \ldots, m_i, \) and \( k = 1, \ldots, l_j, \) beginning with starting values \( S^0, \nu^0, \tau^0, \eta^0, \{p_1^0, \ldots, p_t^0\}, \) and \( \{q_{11}^0, \ldots, q_{tm}^0\} \), and using tuning parameters \( \sigma_1^2, \sigma_2^2, \sigma_3^2, \rho_1, \rho_2, \) and \( \rho_3 \), we implement a Metropolis-Hastings-within-Gibbs algorithm for \( b = 1, 2, \ldots \) as follows:

1. Generate \( S^{b*} \sim \nu \text{M-UARS}(S^{b-1}, \rho_1) \) as a proposal for \( S^b \).

2. Compute \( d_b^1 = G(S^{b*}, \nu^{b-1}, \tau^{b-1}, \eta^{b-1}, p_1^{b-1}, \ldots, p_t^{b-1}, q_{11}^{b-1}, \ldots, q_{tm}^{b-1}) \)
   \( \frac{G(S^{b-1}, \nu^{b-1}, \tau^{b-1}, \eta^{b-1}, p_1^{b-1}, \ldots, p_t^{b-1}, q_{11}^{b-1}, \ldots, q_{tm}^{b-1})}{G(S^{b-1}, \nu^{b-1}, \tau^{b-1}, \eta^{b-1}, p_1^{b-1}, \ldots, p_t^{b-1}, q_{11}^{b-1}, \ldots, q_{tm}^{b-1})} \) for \( G \) in (6) and generate \( W_b^1 \sim \text{Bernoulli}(\min(1, d_b^1)) \).
3. Take $S^b = W^1_b S^{b*} + (1 - W^1_b) S^{b-1}$.

4. Generate $\log(\nu^{b*}) \sim N(\log(\nu^{b-1}), \sigma^2_2)$, with $\nu^{b*}$ as a proposal for $\nu^b$.

5. Compute $d^2_b = \frac{G(S^b, \nu^{b*}, \tau^{b-1}, \eta^{b-1}, \phi^{b-1}, \theta^{b-1}, \xi^{b-1}, \psi^{b-1})}{G(S^b, \nu^{b*}, \tau^{b-1}, \eta^{b-1}, \phi^{b-1}, \theta^{b-1}, \xi^{b-1}, \psi^{b-1})}$ for $G$ in (6) and generate $W^2_b \sim \text{Bernoulli}(\min(1, d^2_b))$.

6. Take $\nu^b = W^2_b \nu^{b*} + (1 - W^2_b) \nu^{b-1}$.

7. Generate $\log(\tau^{b*}) \sim N(\log(\tau^{b-1}), \sigma^2_2)$, with $\tau^{b*}$ as a proposal for $\tau^b$.

8. Compute $d^3_b = \frac{G(S^b, \nu^b, \tau^{b*}, \eta^{b*}, \phi^{b*}, \theta^{b*}, \xi^{b*}, \psi^{b*})}{G(S^b, \nu^b, \tau^{b*}, \eta^{b*}, \phi^{b*}, \theta^{b*}, \xi^{b*}, \psi^{b*})}$ for $G$ in (6) and generate $W^3_b \sim \text{Bernoulli}(\min(1, d^3_b))$.

9. Take $\tau^b = W^3_b \tau^{b*} + (1 - W^3_b) \tau^{b-1}$.

10. Generate $\log(\eta^{b*}) \sim N(\log(\eta^{b-1}), \sigma^2_2)$, with $\eta^{b*}$ as a proposal for $\eta^b$.

11. Compute $d^4_b = \frac{G(S^b, \nu^b, \tau^b, \eta^b, \phi^b, \theta^b, \xi^b, \psi^b)}{G(S^b, \nu^b, \tau^b, \eta^b, \phi^b, \theta^b, \xi^b, \psi^b)}$ for $G$ in (6) and generate $W^4_b \sim \text{Bernoulli}(\min(1, d^4_b))$.

12. Take $\eta^b = W^4_b \eta^{b*} + (1 - W^4_b) \eta^{b-1}$.

13. For $i = 1, \ldots, t$

   (a) Generate $p^b_i \sim \text{vM-UARS}(p^b_{i-1}, \rho_2)$ as a proposal for $p^b_i$.

   (b) Compute $a^*_b = \frac{G(S^b, \nu^b, \tau^b, \eta^b, \phi^b, \theta^b, \xi^b, \psi^b)}{G(S^b, \nu^b, \tau^b, \eta^b, \phi^b, \theta^b, \xi^b, \psi^b)}$ for $G$ in (6) and generate $V^*_b \sim \text{Bernoulli}(\min(1, a^*_b))$.

   (c) Take $p^b_i = V^*_b p^b_{i-1} + (1 - V^*_b) p^b_{i-1}$.

   (d) For $j = 1, \ldots, m_i$

      i. Generate $q^b_{ij} \sim \text{vM-UARS}(q^b_{i-1}, \rho_3)$ as a proposal for $q^b_{ij}$.

      ii. Compute $c^b_{ij} = \frac{G(S^b, \nu^b, \tau^b, \eta^b, \phi^b, \theta^b, \xi^b, \psi^b)}{G(S^b, \nu^b, \tau^b, \eta^b, \phi^b, \theta^b, \xi^b, \psi^b)}$ for $G$ in (6) and generate $U^b_{ij} \sim \text{Bernoulli}(\min(1, c^b_{ij}))$.

      iii. Take $q^b_{ij} = U^b_{ij} q^b_{ij} + (1 - U^b_{ij}) q^b_{ij-1}$. 
Suppose the above Metropolis-Hastings-within-Gibbs algorithm is used to simulate sets of values \( \{S^1, \nu^1, \tau^1, \eta^1\}, \ldots, \{S^N, \nu^N, \tau^N, \eta^N\} \) from the posterior after an appropriate burn-in (a number of iterations in which simulated values are not used because effects of the starting values are still visible). Bayes point estimates for the parameters \( \nu, \tau, \) and \( \eta \), say \( \hat{\nu}, \hat{\tau}, \) and \( \hat{\eta} \), are taken to be the mean of the corresponding simulated values, i.e. \( \hat{\nu} = \frac{1}{N} \sum_{J=1}^{N} \nu^J \), \( \hat{\tau} = \frac{1}{N} \sum_{J=1}^{N} \tau^J \), and \( \hat{\eta} = \frac{1}{N} \sum_{J=1}^{N} \eta^J \). Suppose now that \( \hat{S} = \hat{\Delta} \hat{D} \hat{\Gamma} \) is the singular value decomposition of \( \bar{S} = \frac{1}{N} \sum_{J=1}^{N} S^J \). Then we use \( \hat{S} = \hat{\Delta} \hat{\Gamma} \) as the Bayes point estimate for the location parameter \( S \). (This value maximizes the quasi-likelihood for the vM-UARS(\( S, \kappa \)) distribution studied by Bingham, Nordman, and Vardeman (2009).)

Credible sets can also be found from the simulated values. A Bayes credible set can be thought of as enclosing a certain amount of posterior probability. If \( \nu_\delta \) is \( \delta \) quantile of \( \{\nu^1, \nu^2, \ldots, \nu^N\} \), then a 95% posterior credible interval for \( \nu \) is \( (\nu_{0.025}, \nu_{0.975}) \). Similar 95% credible intervals are made for \( \tau \) and \( \eta \). We use the method of Bingham, Vardeman, and Nordman (2009) to obtain a credible set for \( S \). For \( J = 1, \ldots, N \), we find the angle between each of the coordinate axes rotated from their reference direction by \( \hat{S} \) with the corresponding axis rotated by \( S^J \) and let \( \omega^J \) represent the maximum of these three angles. Suppose \( \omega_{0.95} \) = the .95 quantile of \( \{\omega^1, \ldots, \omega^N\} \). Then, sets of 3 simultaneous cones of constant angle \( \omega_{0.95} \) around the coordinate axis rotated by \( \hat{S} \) create a region capturing 95% of posterior values for \( S \). Figure 6 (taken from Bingham, Vardeman, and Nordman (2009)) gives a graphical representation of a credible region for \( S \), and we can use the size of the angle between the center and edge of the cones as a measure of the “size” of the region for \( S \). In the following section, we apply the Bayes methods presented here to the set of measured crystal orientations obtained using EBSD.

5 Application to EBSD data

EBSD was used to obtain measured crystal orientations for a standard nickel specimen provided by Materials Analysis Division, Ametek Incorporated for TSL EDAX (see TexSEM Laboratories, Inc.). Grinding was used to remove the deformation layer introduced during
Figure 6: Graphical display of a confidence or credible region for the parameter $S$, with $x$, $y$, and $z$ representing the orientation of a corresponding point estimate for $S$

sectioning, and to generate a flat surface for examination. The grinding materials used ran through 240 grit to 1200 grit SiC paper. Each of the grinding stages was of 30-second duration and water was used as a lubricant to flush away ground off material. The coupons were mechanically polished using standard metallographic techniques and the nickel specimen was sectioned using low speed, low deformation precision sectioning with cooling to preserve the microstructure.

Finer polishing was used to remove the deformation introduced during the grinding steps. The polishing was done with alumina abrasives and then diamond abrasives on general-purpose cloths. Polishing ran through fives steps starting with 9.0 micron abrasive and ending with a 0.05 micron abrasive. Each of the solution-based polishing stages was of 10-minute duration. As a final polishing step, the samples were polished on the colloidal silica vibratory polisher for two hours. The samples were coated with a thin, amorphous, conductive carbon coating to eliminate charging effects.

The nickel sample was secured on a pedestal oriented at 70 degrees from the horizontal. Fourteen scans were performed on the sample over the same $40 \mu m \times 40 \mu m$ surface area. A feature on the microstructure was used to register the scans and all scans were performed using
standard EBSD operation of a TSL MSC-2200 camera with TSL OIM hardware and version 4.2 software and a Field Emission AmRay 1845 SEM. Individual orientations at locations were automatically measured using EBSD on an array with 0.2 \( \mu \text{m} \) step size, resulting in at least 4000 measured crystal orientations per scan.

Figure 3 gives a color coded map of the crystal orientations taken from one of the scans and clearly shows where different grains are located within the nickel sample. Eight of the grains were chosen and twenty locations were selected from within each grain. For each of these locations we have crystal orientation data from the fourteen repeat scans, so that in the notation of Section 4, \( t = 8 \), \( m_i = 20 \) for \( i = 1, \ldots, 8 \), and \( l_j = 14 \) for \( j = 1, \ldots, 20 \). A Bayes analysis for the hierarchical model with vM-UARS components was done using the algorithm outlined in Section 4 to obtain a sample of size 8000 from the posterior in (6) (after a burn-in of 2000 iterations). Bayes point estimates for the parameters were found to be \( \hat{\nu} = 2.885404 \), \( \hat{\tau} = 31673.96 \), \( \hat{\eta} = 42750.19 \), and

\[
\hat{S} = \begin{pmatrix}
-0.8381482 & 0.01391734 & -0.5452650 \\
0.3709414 & -0.71836068 & -0.5885239 \\
-0.3998876 & -0.69553162 & 0.5969302
\end{pmatrix}.
\]

The intervals \((0.92828, 5.96885)\), \((25077.82, 38642.39)\), and \((40233.68, 45435.11)\) represent 95\% credible intervals for the parameters \( \nu \), \( \tau \), and \( \eta \), respectively. A credible region for \( S \) can be displayed as in Figure 6 with cone angle 0.44815 radians (in some sense the “overall” orientation of crystals is known to be within .45 radians).

Notice that the estimates for \( \tau \) and \( \eta \) are large (as are the values contained in the intervals for \( \tau \) and \( \eta \)). These estimates and intervals for \( \tau \) and \( \eta \) represent highly concentrated distributions for the misorientation angles, and thus highly concentrated distributions on \( \Omega \). When a measurement technique results in such a large estimate for \( \eta \), it is considered to be highly repeatable. The large estimate for \( \tau \) provides evidence that the EBSD readings from within a grain also show a high degree of precision. With \( \Delta_1 \) as in Section 3, \( \Delta_1(\hat{\tau}) = 0.00378987 \) and \( \Delta_1(\hat{\eta}) = 0.00326217 \). The credible interval for \( \eta \) represents orientations with less variability than that for \( \tau \), so the variation across repeated scans is less than the within grain variation.
As expected, the variation across grains is the largest, with \( \Delta_1(\hat{\nu}) = 0.424007 \).

6 Conclusion

The statistical analysis performed here provides quantification of the precision of measurements taken by EBSD. These results are significant since EBSD is widely used to determine the microstructure of metals. The methodology used here can be extended to other materials besides the nickel used as the featured example to serve as a means of quantifying the repeatability of this measurement technology.

References


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


GENERAL CONCLUSION

This dissertation has identified and developed inference for the UARS class of distributions for 3-dimensional rotations. The first paper introduced the UARS class and explored one member of the class, the von Mises version of the UARS distributions (vM-UARS). Because the vM-UARS distribution has unbounded likelihood (and ordinary maximum likelihood is unavailable), one-sample quasi-likelihood-based inference was explored.

Bayes inference for the vM-UARS distribution was investigated in the second paper. Non-informative priors were identified and Metropolis-Hastings within Gibbs algorithms were used to generate samples from posterior distributions in one-sample and one-way random effects scenarios. A simulation study investigated the performance of Bayes analyses in the one-sample case and included comparisons to the quasi-likelihood inference presented in the first paper. The fourth paper extended these Bayes analyses for the vM-UARS distribution to include hierarchical models. In each of the above-mentioned papers, the methods developed were used in a materials science application to quantify the importance of various sources of variation present in crystal orientation measurements.

The matrix von Mises-Fisher distribution is the most widely studied distribution for 3-dimensional rotation data, and the symmetric version of this distribution is a member of the UARS class. The third paper examined one-sample likelihood theory for the symmetric matrix von Mises-Fisher distribution and developed Bayes methods by using Metropolis-Hastings within Gibbs algorithms in one-sample and one-way random effects models. A simulation study was used to compare Bayes analyses to likelihood inference in the one-sample case and the methods were applied to a human kinematics study examining the variation in limb position during a drilling task.
The UARS class of distributions developed in this dissertation provides alternatives in modeling 3-dimensional rotations and overcomes many of the limitations of existing distributions. The UARS class has many attractive properties including directly interpretable parameters and relatively simple theory. Orientations can also be easily simulated from a UARS distribution. The quasi-likelihood, likelihood, and Bayes inference approaches employed here for two members of the UARS class can be used to provide inference for other members of the class. Thus, the UARS class offers previously unavailable flexibility in modeling 3-dimensional orientation data (a fact that was useful for modeling the crystal orientation data explored in much of this dissertation). Additionally, the UARS class opens many possibilities for further work in the area. Future research directions include clustering (with spatial considerations) and time series methods for 3-dimensional data, and extending the UARS class (thus far describing a symmetric model for 3-dimensional orientations) to allow for non-symmetric data.
ACKNOWLEDGEMENTS

I would like to thank my advisors, Dr. Vardeman and Dr. Nordman, for their guidance and support over the past few years. I have learned so much from each of them about research, writing, and the world of academia. Their wisdom has helped me greatly as a graduate student and I will carry it with me as I embark on my journey as a professor.

I would also like to thank my wonderful husband, Bob, for standing by me every step of the way throughout my time as a student. He has made sacrifices so that I could accomplish my goals and has allowed me to make decisions impacting our lives. I thank him for being so willing to “come along for the ride,” as there is no other person I would rather have beside me.