involve

Bootstrap techniques for measures of center
for three-dimensional rotation data

L. Katie Will and Melissa A. Bingham

msp

Bootstrap techniques for measures of center for three-dimensional rotation data

L. Katie Will and Melissa A. Bingham

(Communicated by Mary C. Meyer)

Bootstrapping is a nonparametric statistical technique that can be used to estimate the sampling distribution of a statistic of interest. This paper focuses on implementation of bootstrapping in a new setting, where the data of interest are 3-dimensional rotations. Two measures of center, the mean rotation and spatial average, are considered, and bootstrap confidence regions for these measures are proposed. The developed techniques are then used in a materials science application, where precision is explored for measurements of crystal orientations obtained via electron backscatter diffraction.

1. Introduction

Three-dimensional rotation data is common in the field of materials science, where electron backscatter diffraction (EBSD) can be used to study the microtexture of metals, including the orientation of crystals within the metal. Using EBSD, a fixed beam of electrons is diffracted off of a metal sample, creating an image on a focal plane of sensors. These images reveal information about crystal structure and orientation in the metal [Randle 2003]. One area of interest in regards to EBSD measurements is precision. As Bingham, Nordman, and Vardeman [Bingham et al. 2009a] point out, methods used for quantifying EBSD precision in the materials science literature are not standard, with ad hoc precision estimates often reported (see, for example, [Demirel et al. 2000; Wilson and Spanos 2001]). This led Bingham et al. [2009a] to investigate the precision of measurements obtained via EBSD by developing new statistical distributions for 3-dimensional rotations. While the distributions developed by Bingham et al. [2009a] do allow for some flexibility in modeling, our intent here is development of nonparametric techniques, namely bootstrap confidence regions, that can be used without the need for any distributional assumptions. While bootstrapping techniques are commonly used in

one-dimensional nonparametric statistics, these techniques have not been applied to 3-dimensional rotation data.

Suppose that $O_1, \ldots, O_n \in SO(3)$ represent orientations at n scanned locations on a metal specimen as measured by EBSD, where $SO(3)$ denotes the set of all 3×3 orthogonal rotation matrices. When adjacent locations produce similar EBSD crystal orientations, those locations are considered to be part of the same grain. We are interested in estimating the central rotation of a set of orientations from within the same grain, since the true central orientation would represent the actual grain orientation, with random scatter in $O_1, \ldots, O_n$ due to measurement error in the EBSD process. We will investigate two different measures of center for 3-dimensional rotation data.

The mean rotation, M, is a commonly used measure of center [León et al. 2006; Bingham et al. 2009a; Khatri and Mardia 1977] that is defined to be the rotation that maximizes trace$(M^T \bar{O})$, where $\bar{O} = \frac{1}{n} \sum_{i=1}^{n} O_i$ for $O_1, \ldots, O_n \in SO(3)$. The mean rotation M can be found by using $M = VW$, where $\bar{O} = V \Sigma W$ is the singular value decomposition of $\bar{O}$. It is necessary to use these components from the singular value decomposition since $\bar{O}$ may not be an element of $SO(3)$, but M is.

The second measure of center considered is the spatial average of Ball and Greiner [2012]. The spatial average of a set of rotations $O_1, \ldots, O_n \in SO(3)$ is obtained through an iterative procedure that uses what is referred to as the axis-angle representation of a matrix. For a given matrix, the axis and angle are such that if you rotate the 3×3 identity matrix about the axis by the angle, you will arrive at the specified matrix. The steps to find the spatial average are outlined below, where the end result is the matrix S found in step (4) in the final iteration. The procedure begins by looking at just the first two matrices, O_1 and O_2. Starting at O_1, we rotate half of the way towards O_2, resulting in a matrix S. Then we consider the third matrix in the data set O_3 and rotate S one-third of the way towards this matrix, giving an updated matrix S. We then rotate S one-fourth of the way towards O_4, again updating S. This process continues until we have been through all n matrices in the data set. Note that for large samples both the mean rotation and spatial average converge to the population central matrix.

(1) Let $S = O_1$ and let $i = 2$.

(2) Compute $G = S^T O_i$.

(3) Let u be the axis of G, let θ be the angle of G, and let $p = \theta/i$.

(4) Compute $S = SP$, where P is the matrix form of (u, p), and let $i = i + 1$.

(5) If $i \leq n$, return to step (2).

In Section 2, development of bootstrap confidence regions for these measures of center for 3-dimensional rotations will be discussed. Accuracy of the bootstrap techniques will be explored through a simulation study in Section 3. Finally, the

bootstrap procedure will be applied to data from a nickel specimen to evaluate EBSD precision in Section 4.

2. Development of bootstrapping technique

Bootstrapping is a nonparametric statistical technique that uses resampling with replacement. It can be used to estimate the sampling distribution of almost any statistic, e.g., mean, proportion, variance. It is commonly used to find confidence regions for population parameters. To find a 95% confidence interval in one-dimension, a large number (say 1000) of samples of size n are drawn from the original sample of size n with replacement and the statistic of interest is computed for each bootstrap sample. This creates a sampling distribution for the statistic of interest. Under the bootstrap percentile method, a 95% confidence interval is then obtained by using the 2.5th and 97.5th percentiles as confidence bounds.

Although bootstrapping has been used to create confidence regions in a wide variety of settings, including analyzing directional data such as p-dimensional unit vectors [Fisher and Hall 1989], we introduce the concept of bootstrapping for measures of center in the 3-dimensional setting, where the data can be represented by 3×3 orthogonal rotation matrices. To estimate measures of center for 3-dimensional rotation data by bootstrapping, we sample with replacement from the original sample of n matrices 1000 times. Each sample is a bootstrap sample, for which we compute a measure of center (mean rotation or spatial average). We will refer to these 1000 matrices as bootstrap central matrices. To provide an estimate of center for the 3-dimensional rotation data, the mean rotation of the 1000 bootstrap central matrices can be computed. Since this matrix is analogous to what we would consider a "point estimate" when considering 1-dimensional data, we also refer to it as a point estimate here.

After obtaining the point estimate for our central rotation, we want to find a confidence region around this matrix. Because rotation matrices do not have a natural ordering, the percentile bootstrap method of using the 2.5th and 97.5th percentiles as confidence bounds does not translate directly to 3-dimensional rotation data. Instead, we will use a set of three cones centered at the point estimate to give a confidence region for the true central rotation in a similar fashion to Bingham et al. [2009a]. Figure 1 illustrates this concept for two different-sized sets of cones. To determine the size of the cones needed to give a 95% confidence region, we first think of each matrix as a set of three axes (x, y, and z) and consider the angles between the axes of the point estimate and the bootstrap central matrices. For each bootstrap central matrix, we find three angles. Each one is the angle between an axis and the corresponding axis of the point estimate. We then take the maximum of these three angles, so that moving a distance of that angle away from all three axes
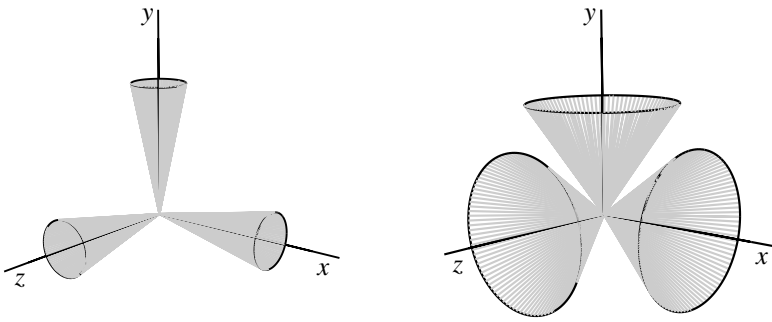
Figure 1. Plot of confidence cones around a point estimate (represented as three axes) with an angle of 0.2 radians (left) and 0.6 radians (right).

of the point estimate would contain all three axes of the bootstrap central matrix. Once these maximum angles are computed for all 1000 bootstrap central matrices, we take the 95th percentile and use this as the cone size. Since this set of three cones centered at the point estimate will capture 95% of the bootstrap central matrices, we think of it as a 95% bootstrap percentile region for the true central matrix.

3. A simulation study

To examine the accuracy of the bootstrap technique developed in Section 2, a simulation study was performed. Data sets were simulated from both the von Mises version of the uniform axis random spin (vM-UARS) distributions [Bingham et al. 2009a] and the matrix Fisher distribution [Khatri and Mardia 1977]. A vM-UARS or matrix Fisher distribution is characterized by a central rotation $S \in SO(3)$ and a spread parameter $\kappa \in (0, \infty)$. The spread parameter κ is best described as a concentration parameter since larger values of κ yield rotations with less variability.

For this simulation study we used κ values of 1, 5, 20, and 500 and sample size n of 10, 30, and 100. For each combination of κ and n both the mean rotation and spatial average were considered and the bootstrapping procedure was replicated 1000 times (i.e., 1000 different samples were drawn from each of the vM-UARS and matrix Fisher distributions) with 1000 bootstrap samples taken from the original sample each time. For each of the 1000 replications a 95% confidence region as a set of three cones was found. The coverage rates of the confidence cones were then found as the proportion of times out of 1000 that the true central rotation S was captured. Note that our choice of S for simulation purposes was arbitrary, as results are the same regardless of what true central rotation is used. Tables 1 and 2 show the coverage rates along with the median cone size, in radians, for each case.

The coverage rates fluctuate closely around 95%, which validates that the bootstrapping procedure is behaving as desired for the two distributions considered here.

(κ, n)	Mean rotation		Spatial average	
	Coverage rate	Median cone size	Coverage rate	Median cone size
(1, 10)	0.966	0.63441	0.944	0.72272
(1, 30)	0.962	0.32435	0.980	0.51034
(1, 100)	0.948	0.16934	0.979	0.26583
(5, 10)	0.944	0.22523	0.924	0.22695
(5, 30)	0.932	0.13152	0.947	0.13726
(5, 100)	0.954	0.07147	0.965	0.07526
(20, 10)	0.944	0.10876	0.943	0.10816
(20, 30)	0.946	0.06436	0.945	0.06500
(20, 100)	0.956	0.03512	0.959	0.03535
(500, 10)	0.944	0.02157	0.927	0.02176
(500, 30)	0.963	0.01269	0.945	0.01271
(500, 100)	0.949	0.00697	0.961	0.00702

Table 1. Coverage rates and median cone sizes (in radians) for estimating the center of the vM-UARS distribution using the mean rotation and the spatial average.

(κ, n)	Mean rotation		Spatial average	
	Coverage rate	Median cone size	Coverage rate	Median cone size
(1, 10)	0.957	1.05138	0.951	1.35648
(1, 30)	0.950	0.55232	0.995	0.88534
(1, 100)	0.942	0.29247	0.991	0.43596
(5, 10)	0.922	0.27921	0.909	0.27613
(5, 30)	0.945	0.16369	0.939	0.16325
(5, 100)	0.954	0.09025	0.947	0.09023
(20, 10)	0.919	0.13289	0.920	0.13365
(20, 30)	0.932	0.07896	0.952	0.07891
(20, 100)	0.950	0.04319	0.939	0.04318
(500, 10)	0.918	0.02629	0.925	0.02633
(500, 30)	0.940	0.01556	0.948	0.01544
(500, 100)	0.947	0.00851	0.948	0.00851

Table 2. Coverage rates and median cone sizes (in radians) for estimating the center of the matrix Fisher distribution using the mean rotation and the spatial average.

We also see that the values of κ and n impact the median confidence region sizes as expected, with larger κ (less spread) and larger n resulting in smaller regions. It is also important to compare the nonparametric bootstrap techniques developed here to existing parametric methods for the vM-UARS and matrix Fisher distributions. Bingham, Vardeman, and Nordman [Bingham et al. 2009b, Table 5, page 618] provide median cone sizes for the central rotation of the vM-UARS distributions obtained by maximum quasi-likelihood estimation using the same κ and n values considered here. Bingham, Nordman, and Vardeman [Bingham et al. 2010b, Table 5, page 1325] use maximum likelihood estimation to provide similar results for the matrix Fisher distribution. To compare the cone sizes of these works to the results given in Tables 1 and 2 presented here, we calculated the relative difference between the sizes as $d(a_p, a_b) = |a_p - a_b|/a_p$, where a_p is the angle from the parametric approach and a_b is the angle from the bootstrap approach. For the vM-UARS distribution, the largest relative difference was 0.1360 (for $\kappa = 1$ and $n = 100$). For the matrix Fisher distribution, the largest relative difference was 0.1441 (for $\kappa = 1$ and $n = 10$). Both of these differences are small, indicating that the bootstrap techniques developed here produce results that are equivalent to existing parametric approaches.

4. Application to electron backscatter diffraction data

Now that the bootstrapping technique developed in Section 2 has been shown to keep coverage rates as expected and perform similarly to existing parametric methods, we use it to investigate precision of measurements obtained through EBSD. A high-iron-concentration nickel specimen of size $40 \, \mu\text{m} \times 40 \, \mu\text{m}$ was scanned using EBSD. The scanning was done over a regularly spaced grid with $0.2 \, \mu\text{m}$ step size across the top of the specimen, resulting in 4121 crystal orientations. See [Bingham et al. 2010a] for more details regarding the machinery used and data collection process.

For two orientations P and Q, the misorientation angle between them is the smallest angle of rotation needed to get from P to Q when rotating about some axis. When using EBSD, orientations close in proximity are classified as composing a grain when the misorientation angle between them is small, so that a grain is thought of as a homogeneous piece of material that produces observations which generally share a common orientation. Figure 2 gives the grain map that resulted from using EBSD on the nickel specimen [Bingham et al. 2010a]. The grain map can be viewed as if one was looking down on the piece of nickel, so that the axes give the x- and y-locations on a rectangular grid. Each dot in the figure corresponds to a single measured orientation from the total 4121 orientations in this scan. Similar orientations are classified into grains, with each colored block on the map indicating a different grain. The similarity of color within grains makes some of the 4121 dots indistinguishable from others. Dots that clearly stand out represent locations on
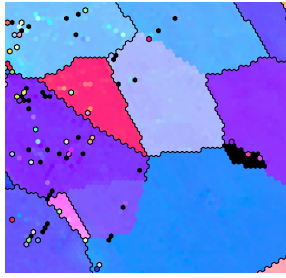
Figure 2. EBSD grain map for the nickel specimen, with grains represented by blocks of similar color.

Grain	Sample size (n)	Original cone size	Reported EBSD precision
1	49	0.0647°	0.4531°
2	31	0.1742°	0.9699°
3	21	0.0876°	0.4016°
4	44	0.0823°	0.5461°
5	22	0.1054°	0.4942°

Table 3. Size of 95% confidence regions for central rotations and reported precision of EBSD measurements (in degrees).

the scan with deformities. Although there are over ten grains present, we will use subsets of data from five of these grains in the analysis here.

For each of the five grains considered, we applied the bootstrapping technique to the 3×3 matrices representing crystal orientations. Using the mean rotation as our measure of center, 95% confidence regions for the central rotation were found. The sizes of the confidence cones are provided in Table 3, in degrees. Because confidence region sizes decrease at a rate of $1/\sqrt{n}$ (which can be verified by examining the cone sizes presented in Tables 1 and 2), before reporting the degree of precision we multiply each of the five cone sizes by $\sqrt{n}$. The reported precision estimates are also provided in Table 3. We find EBSD precision estimates comparable to the 1° reported by Demirel, El-Dasher, Adams, and Rollett [Demirel et al. 2000] and the 0.5° reported by Wilson and Spanos [2001] by using methods that are much more statistically sound than the methods employed in these works.

5. Conclusion

The study of precision for EBSD measurements considered here is just one of many applications that could benefit from the bootstrapping techniques developed. These bootstrapping techniques, while simple to implement, have been shown to perform as well as existing parametric approaches. Given the complexity of

existing parametric methods, they are likely not easily accessible to practitioners (such as materials scientists) who often collect 3-dimensional rotation data. Further, statistical methods that do not rely on distributional assumptions are important in the area of 3-dimensional rotation data since there are relatively few developed distributions for which parametric methods are even available [Bingham et al. 2009a; Khatri and Mardia 1977; León et al. 2006]. Therefore, the bootstrapping techniques presented here could play an important role in the field of statistics, as well as in areas of study where 3-dimensional rotations are commonly found.

References

[Ball and Greiner 2012] K. A. Ball and T. M. Greiner, "A procedure to refine joint kinematic assessments: functional alignment", *Comput. Meth. Biomech. Biomed. Engin.* **15**:5 (2012), 487–500.

[Bingham et al. 2009a] M. A. Bingham, D. J. Nordman, and S. B. Vardeman, "Modeling and inference for measured crystal orientations and a tractable class of symmetric distributions for rotations in three dimensions", *J. Amer. Statist. Assoc.* **104**:488 (2009), 1385–1397. MR 2596996 Zbl 1205.62215

[Bingham et al. 2009b] M. A. Bingham, S. B. Vardeman, and D. J. Nordman, "Bayes one-sample and one-way random effects analyses for 3-D orientations with application to materials science", *Bayesian Anal.* **4**:3 (2009), 607–629. MR 2551048

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

[Bingham et al. 2010b] M. A. Bingham, D. J. Nordman, and S. B. Vardeman, "Finite-sample investigation of likelihood and Bayes inference for the symmetric von Mises–Fisher distribution", *Comput. Statist. Data Anal.* **54**:5 (2010), 1317–1327. MR 2600834

[Demirel et al. 2000] M. C. Demirel, B. S. El-Dasher, B. L. Adams, and A. D. Rollett, "Studies on the accuracy of electron backscatter diffraction measurements", pp. 65–74 in *Electron backscatter diffraction in materials science*, edited by A. J. Schwartz et al., Springer, New York, 2000.

[Fisher and Hall 1989] N. I. Fisher and P. Hall, "Bootstrap confidence regions for directional data", *J. Amer. Statist. Assoc.* **84**:408 (1989), 996–1002. MR 1134489

[Khatri and Mardia 1977] C. G. Khatri and K. V. Mardia, "The von Mises–Fisher matrix distribution in orientation statistics", *J. Roy. Statist. Soc. Ser. B* **39**:1 (1977), 95–106. MR 0494687 Zbl 0356.62044

[León et al. 2006] C. A. León, J.-C. Massé, and L.-P. Rivest, "A statistical model for random rotations", *J. Multivariate Anal.* **97**:2 (2006), 412–430. MR 2234030 Zbl 1085.62066

[Randle 2003] V. Randle, *Microtexture determination and its applications*, 2nd ed., Maney for the Institute of Materials, Minerals and Mining, London, 2003.

[Wilson and Spanos 2001] A. W. Wilson and G. Spanos, "Application of orientation imaging microscopy to study phase transformations in steels", *Mater. Charact.* **46**:5 (2001), 407–418.

kwill@iastate.edu   *Department of Statistics, Iowa State University, Ames, IA 50011, United States*

mbingham@uwlax.edu   *Department of Mathematics and Statistics, University of Wisconsin-La Crosse, 1725 State Street, La Crosse, WI 54601, United States*

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

MANAGING EDITOR

Kenneth S. Berenhaut   Wake Forest University, USA

BOARD OF EDITORS

PRODUCTION

Silvio Levy, Scientific Editor

Cover: Alex Scorpan

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

involve