Supplementary Appendix: Bayesian Inference for a New Class of Distributions on Equivalence Classes of 3-D Orientations With Applications to Materials Science

Chuanlong Du, Daniel J. Nordman

Department of Statistics

Iowa State University

and

Stephen B. Vardeman

Department of Statistics

Department of Industrial and Manufacturing Systems Engineering

Iowa State University

This supplementary appendix consists of three parts. Appendix A first provides additional details about uniform-axis-random-spin (UARS) models for random rotations, as explained in Section 2 of the main manuscript. Appendix B supplies further numerical summaries for the simulations of Section 3.3 of the main manuscript. An EM algorithm approach for computing point estimators in models for unlabeled orientations is described in Appendix C. A final reference section provides citations appearing in this Appendix.

A UARS($\boldsymbol{S}, \kappa$) Models for Rotation Matrices

Let Ω be the collection of all 3×3 rotation matrices, i.e., $SO(3)$. Bingham, Nordman, and Vardeman (2009) and Hielscher, Schaeben, and Siemes (2010) identified a class of UARS models on Ω, which are useful for describing random rotations symmetrically distributed around a fixed, mean rotation (i.e., central location) parameter $\boldsymbol{S} \in \Omega$, and where the amount of variability in rotations can be directly controlled by a concentration parameter $\kappa > 0$ in the model. Rotations in the UARS model class have a single simple, geometric construction as follows. Suppose we have a unit vector $\boldsymbol{u} \in \mathbb{R}^3$, and we spin the axes of

the standard coordinate system (represented by the columns of the identity matrix $\boldsymbol{I}_{3\times 3}$) around the direction $\boldsymbol{u}$ (i.e., a signed axis) counter-clockwise through angle r. Upon this rotation, the columns of $\boldsymbol{I}_{3\times 3}$ move to positions given by a corresponding rotation matrix

$$\boldsymbol{M}(\boldsymbol{u}, r) = \boldsymbol{u}\boldsymbol{u}^T + (\boldsymbol{I}_{3\times 3} - \boldsymbol{u}\boldsymbol{u}^T)\cos r + \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \sin r, \tag{A.1}$$

which is Euler's angle-axis representation of a 3-D rotation; see Figure 1 of the main manuscript for illustration.

UARS models are a stochastic version of this angle-axis representation. If we now take $\boldsymbol{u}$ as uniformly distributed on the unit sphere and independent of a random angle $r \sim \text{Cir}(\kappa)$, where $\text{Cir}(\kappa)$ is a circular distribution on $(-\pi, \pi]$, symmetric around 0 and with a positive concentration parameter κ (with concentration increasing in κ), then we obtain a random rotation matrix denoted by $\boldsymbol{M}(\boldsymbol{u}, r)$ whose distribution we denote as $\text{UARS}(I, \kappa)$. For $\boldsymbol{S} \in \Omega$, $\boldsymbol{S} \cdot \boldsymbol{M}(\boldsymbol{u}, r) \sim \text{UARS}(\boldsymbol{S}, \kappa)$ gives a so-called UARS random rotation with location and concentration parameters respectively $\boldsymbol{S}$ and κ. Different choices of the circular distribution for r produce different models for symmetric random rotations. If the $\text{Cir}(\kappa)$ distribution for the angle r has a density $C(r|\kappa)$ (with respect to the usual Lebesgue measure), then a $\text{UARS}(\boldsymbol{S}, \kappa)$ rotation has a corresponding density

$$f(\boldsymbol{O}|\boldsymbol{S}, \kappa) = \frac{4\pi}{3 - \text{tr}(\boldsymbol{S}^T\boldsymbol{O})} C\left(\arccos(\frac{\text{tr}(\boldsymbol{S}^T\boldsymbol{O}) - 1}{2})|\kappa\right), \ \boldsymbol{O}, \boldsymbol{S} \in \Omega, \ \kappa > 0, \tag{A.2}$$

with respect to the "uniform distribution" on Ω (or Haar measure), which provides a dominating measure for defining densities on Ω (see Downs, 1972).

B Additional Numerical Summaries

Section 3.3 of the main manuscript describes a simulation study of Bayesian and likelihood-based methods for computing regions to estimate parameters κ and $[\boldsymbol{S}]$ in models for unlabeled orientations. Figure 7 there displays plots of coverage accuracies for all methods for various sample sizes n and concentration parameters κ. The following table reports these actual numerical values, as summarized in Figure 7.

(n, κ)	Bayes		LRT		Wald	
	κ	$\boldsymbol{S}$	κ	$\boldsymbol{S}$	κ	$\boldsymbol{S}$
(10,1)	98.8	91.0	35.2	82.6	17.8	43.2
(30,1)	99.4	92.1	39.4	87.0	11.6	53.2
(100,1)	99.9	94.1	43.1	84.7	12.6	59.1
(10, 2)	99.7	95.1	62.0	81.6	70.8	31.5
(30, 2)	99.7	97.0	83.4	90.3	83.0	50.4
(100, 2)	98.7	98.5	92.1	90.6	90.0	67.8
(10,3)	99.0	97.7	85.4	87.9	93.9	70.2
(30,3)	97.3	98.3	92.7	93.0	94.9	87.6
(100,3)	96.8	96.8	94.8	94.5	95.5	96.1
(10,4)	96.9	98.9	90.2	90.7	95.8	81.1
(30,4)	96.1	96.1	94.6	93.3	93.7	93.7
(100,4)	95.9	95.0	95.7	93.8	95.7	95.1
(10,5)	95.0	97.5	90.8	91.4	96.4	85.5
(30,5)	95.3	95.1	93.6	94.5	95.6	95.0
(100,5)	95.6	95.4	94.9	95.2	95.3	97.0
(10, 7)	93.5	95.1	90.8	92.4	95.7	86.1
(30, 7)	96.1	95.2	94.5	94.5	95.1	93.2
(100, 7)	95.5	94.2	95.2	94.2	95.4	93.7
(10, 10)	95.6	95.2	92.1	92.3	95.7	89.0
(30, 10)	94.4	95.3	92.8	95.3	94.3	94.5
(100, 10)	95.6	95.6	94.9	96.1	95.1	95.9
(10,20)	94.0	94.7	89.9	93.7	94.0	92.3
(30,20)	94.6	94.5	93.9	94.3	96.1	94.8
(100,20)	95.2	95.0	94.7	95.2	95.4	95.8

Table 2: Coverage rates (as percentages) for κ and $[\boldsymbol{S}]$ for nominally 95% Bayesian (Bayes) regions, inverted likelihood ratio test (LRT) regions and Wald regions, for some choices of (n, κ).

C An EM Algorithm for Unlabeled Orientations

This section provides an EM (Expectation-Maximization) algorithm (cf. Dempster, Laird and Rubin, 1977; Schafer, 1997; Casella and Berger, 2002) that can be applied to maximize the log-likelihood function for the concentration κ and location $[\boldsymbol{S}]$ parameters, given by

$$l(\kappa, [\boldsymbol{S}]) = -n \log\left(b(\kappa)\right) + \sum_{i=1}^{n} \log\left(\sum_{j=1}^{24} \exp\left\{\kappa \operatorname{tr}(\boldsymbol{S}^T \boldsymbol{O}_i^{(j)})\right\}\right) + c, \tag{C.1}$$

$b(\kappa) = e^{\kappa}\left(I_0(2\kappa) - I_1(2\kappa)\right)$, (cf. equation (6), Sec. 2.3 of the main manuscript) based on i.i.d. observed equivalence classes $[\boldsymbol{O}_1], \ldots, [\boldsymbol{O}_n]$ following a symmetric matrix Fisher von Mises distribution of Section 2.2 (denoted by SMF$([\boldsymbol{S}]; \kappa)$). (Recall in (C.1) that $\boldsymbol{O}_i^{(j)}$ runs through elements of the equivalence class $[\boldsymbol{O}_i]$ and $\boldsymbol{S}$ is any element of $[\boldsymbol{S}]$.) That is, rather than maximizing $l(\kappa, [\boldsymbol{S}])$ directly in (C.1), we may iteratively solve an alternative sequence of maximizations whose limiting solution provides the target maximum likelihood estimators. We thank a reviewer for suggesting this possibility, which can offer another computational approach for obtaining point estimators from maximum likelihood. With a straightforward modification to include the prior $\pi(\kappa)$, the same EM algorithm could also be applied for maximizing the posterior density $p(\kappa, [\boldsymbol{S}])$ from Section 3.2 to obtain Bayesian point estimators (i.e., maximum a posteriori (MAP) estimators).

To avoid confusion, we also note that the main methodology proposed in the manuscript for inference with unlabeled orientations is a MCMC-based Bayesian approach. Our intent with this is to go beyond point estimation to provide credible regions having good frequentist coverage accuracies and to also allow geometrically interpretable "cone" regions for the location parameter $[\boldsymbol{S}]$ (cf. Sec. 3.1) which are not available through maximum likelihood. Our approach also provides point estimators for κ and $[\boldsymbol{S}]$ parameters, though we achieve these through MCMC posterior samples in an alternative way than maximizing the posterior density directly. Hence, the following EM algorithm can be viewed as a complement to point estimation via maximum likelihood or MAP estimation.

We next state an EM algorithm for maximum likelihood, using the modeling conventions from Section 2. Rather than maximize the log-likelihood (C.1) directly for $([\boldsymbol{S}], \kappa)$ based on equivalence classes of orientations $[\boldsymbol{O}_1], \ldots, [\boldsymbol{O}_n]$ under the SMF$([\boldsymbol{S}]; \kappa)$ model, we may apply an EM algorithm by augmenting or completing these unlabeled orientation data

with labels. For any element $\boldsymbol{S} \in [\boldsymbol{S}]$, define $\boldsymbol{O}_i \sim \text{SMF}(\boldsymbol{S}; \kappa)$ for $i = 1, \ldots, n$, so that, upon mapping $\boldsymbol{O}_i$ to its 24-fold equivalence class or its unlabeled counterpart, we have $[\boldsymbol{O}_i] \sim \text{SMF}([\boldsymbol{S}]; \kappa)$ (see Sec. 2.1 for details). The space of (labeled) orientations Ω (or $SO(3)$) can be partitioned into 24 parts and each part, $j = 1, \ldots, 24$, contains exactly one element $\boldsymbol{O}_i^{(j)}$ of $[\boldsymbol{O}_i]$. The EM algorithm then proceeds by treating $[\boldsymbol{O}_1], \ldots, [\boldsymbol{O}_n]$ as "incomplete" data and formulating "complete" data as $[\boldsymbol{O}_1], \ldots, [\boldsymbol{O}_n], L_1, \ldots, L_n$ using labels L_i, $i = 1, \ldots, n$, where $L_i = j$ if $\boldsymbol{O}_i = \boldsymbol{O}_i^{(j)}$, the jth element of $[\boldsymbol{O}_i]$. In the following, an exact labeling of the 24 elements of an equivalence class $[\boldsymbol{O}]$ (see equation (1) of Sec. 2.1) will be unimportant and unnecessary.

If $L([\boldsymbol{S}], \kappa \big| [\boldsymbol{O}], \boldsymbol{L})$ denotes the likelihood from the joint density of $[\boldsymbol{O}] = ([\boldsymbol{O}_1], \ldots, [\boldsymbol{O}_n])$ and $\boldsymbol{L} = (L_1, \ldots, L_n)$ (with respect to the uniform distribution on $[\Omega]$), then, similarly to (C.1), we may write

$$\log L([\boldsymbol{S}], \kappa \big| [\boldsymbol{O}], \boldsymbol{L}) = -n \log\left(b(\kappa)\right) + \kappa \left(\sum_{i=1}^{n} \sum_{j=1}^{24} \mathbb{I}(L_i = j) \cdot \text{tr}(\boldsymbol{S}^T \boldsymbol{O}_i^{(j)}) \right) + c, \qquad \text{(C.2)}$$

where $b(\kappa) = e^{\kappa}\left(I_0(2\kappa) - I_1(2\kappa)\right)$, $\boldsymbol{S}$ is any element of $[\boldsymbol{S}]$, and $\mathbb{I}(\cdot)$ denotes the indicator function. We then formulate the EM algorithm as follows. From initial values $([\boldsymbol{S}]^{(0)}, \kappa^{(0)})$, we define a sequence $([\boldsymbol{S}]^{(m)}, \kappa^{(m)})$ according to

$$([\boldsymbol{S}]^{(m+1)}, \kappa^{(m+1)}) \equiv \text{argmax}_{[\boldsymbol{S}], \kappa} \, \text{E}\left[\log L([\boldsymbol{S}], \kappa | [\boldsymbol{O}], \boldsymbol{L}) \big| [\boldsymbol{O}], [\boldsymbol{S}]^{(m)}, \kappa^{(m)} \right] \qquad \text{(C.3)}$$

where the E-step in (C.3) is defined with respect to $\boldsymbol{L}$, based on (C.2), as

$$\begin{aligned} & \text{E}\left[\log L([\boldsymbol{S}], \kappa | [\boldsymbol{O}], \boldsymbol{L}) \big| [\boldsymbol{O}], [\boldsymbol{S}]^{(m)}, \kappa^{(m)} \right] \qquad \text{(C.4)} \\ = \; & -n \log\left(b(\kappa)\right) + \kappa \left(\sum_{i=1}^{n} \sum_{j=1}^{24} P\left(L_i = j \big| [\boldsymbol{O}], [\boldsymbol{S}]^{(m)}, \kappa^{(m)}\right) \cdot \text{tr}(\boldsymbol{S}^T \boldsymbol{O}_i^{(j)}) \right) + c, \end{aligned}$$

using that, for $i = 1, \ldots, n$ and $j = 1, \ldots, 24$,

$$P\left(L_i = j \big| [\boldsymbol{O}], [\boldsymbol{S}]^{(m)}, \kappa^{(m)}\right) = \exp\left\{ \kappa^{(m)} \, \text{tr}(\boldsymbol{S}_m^T \boldsymbol{O}_i^{(j)}) \right\} \Big/ \sum_{\ell=1}^{24} \exp\left\{ \kappa^{(m)} \, \text{tr}(\boldsymbol{S}_m^T \boldsymbol{O}_i^{(\ell)}) \right\}$$

for a given element $\boldsymbol{S}_m$ from $[\boldsymbol{S}]^{(m)}$.

From (C.3)-(C.4), the value of $\kappa > 0$ does not influence the maximization step for $[\boldsymbol{S}]$ so that the $(m+1)$th step maximizer $[\boldsymbol{S}]^{m+1}$ in (C.3) corresponds to $[\boldsymbol{S}]^{m+1} = [\boldsymbol{S}_{m+1}]$ where

$$\boldsymbol{S}_{m+1} = \text{argmax}_{\boldsymbol{S} \in \Omega} \, \text{tr}(\boldsymbol{S}^T \bar{\boldsymbol{O}}_{n,m}) \qquad \text{(C.5)}$$

for a weighted mean

$$\bar{\boldsymbol{O}}_{n,m} \equiv \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{24}\boldsymbol{O}_i^{(j)}\exp\left\{\kappa^{(m)}\,\mathrm{tr}(\boldsymbol{S}_m^T\boldsymbol{O}_i^{(j)})\right\}\Big/\sum_{\ell=1}^{24}\exp\left\{\kappa^{(m)}\,\mathrm{tr}(\boldsymbol{S}_m^T\boldsymbol{O}_i^{(\ell)})\right\}.$$

Hence, the value $\boldsymbol{S}_{m+1}$ corresponds to a type of moment estimator commonly found for location parameters with rotation data (i.e., a projected mean), which may be solved using Procrustes techniques (cf. Downs, 1972; Jupp and Mardia, 1979; León et al., 2006, p. 421). If $\bar{\boldsymbol{O}}_{n,m} = \boldsymbol{U}_{n,m}\mathrm{diag}(\lambda_{1,n,m}, \lambda_{2,n,m}, \lambda_{3,n,m})\boldsymbol{V}_{n,m}^T$ denotes a singular value decomposition of $\bar{\boldsymbol{O}}_{n,m}$ corresponding to eigenvalues $\lambda_{1,n,m} > \lambda_{2,n,m} > |\lambda_{3,n,m}| > 0$ and 3×3 rotation matrices $\boldsymbol{U}_{n,m}$ and $\boldsymbol{V}_{n,m}$, then $\boldsymbol{S}_{m+1} = \boldsymbol{U}_{n,m}\boldsymbol{V}_{n,m}^T$ maximizes the trace in (C.5). Given $[\boldsymbol{S}]^{m+1} = [\boldsymbol{S}_{m+1}]$, the maximizer κ^{m+1} in (C.3) can be found by taking derivatives with respect to κ in (C.4) and then finding the solution to

$$1 + \mathrm{tr}(\boldsymbol{S}_{m+1}^T\bar{\boldsymbol{O}}_{n,m}) = \frac{I_1(2\kappa)}{\kappa}\cdot\frac{1}{I_0(2\kappa) - I_1(2\kappa)},$$

based on derivatives $dI_0(x)/dx = I_1(x)$ and $dI_1(x)/dx = I_0(x) - x^{-1}I_1(x)$ of the modified Bessel function. The sequence $([\boldsymbol{S}]^{(m+1)}, \kappa^{(m+1)})$ in (C.3) is then iteratively determined until convergence, establishing the EM algorithm for unlabeled orientations. See Wu (1983) and McLachlan and Krishnan (1997) for more details on the convergence properties of the EM algorithm.

References

Bingham, M. A., Nordman, D. J., and Vardeman, S. B. (2009). Modeling and Inference for Measured Crystal Orientations and a Tractable Class of Symmetric Distributions for Rotations in Three Dimensions. *Journal of the American Statistical Association*, **104**, 1385-1397.

Casella G. and Berger, R. L. (2002). *Statistical Inference*, 2nd Ed. Duxbury, Pacific Grove, CA.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-22.

Downs, T. D. (1972). Orientation Statistics. *Biometrika*, **59**, 665-676.

Hielscher, R., Schaeben, H., and Siemes, H. (2010). Orientation Distribution within a Single Hematite Crystal. *Mathematical Geosciences*, **42**, 359-375.

McLachlan, G. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.

Wu, C. F. J. (1983). Invariance of Maximum Likelihood Estimators. *The Annals of Statistics*, **11**, 95-103.