
Article:
Kent, JT orcid.org/0000-0002-1861-8349, Ganeiber, AM and Mardia, KV orcid.org/0000-0003-0090-6235 (2018) A New Unified Approach for the Simulation of a Wide Class of Directional Distributions. Journal of Computational and Graphical Statistics, 27 (2). pp. 291-301. ISSN 1061-8600

https://doi.org/10.1080/10618600.2017.1390468

A new unified approach for the simulation of a wide class of directional distributions

John T. Kent
Department of Statistics, University of Leeds
and
Asaad M. Ganeiber
Department of Statistics, University of Benghazi
and
Kanti V. Mardia
Department of Statistics, University of Leeds

October 12, 2017

Abstract

The need for effective simulation methods for directional distributions has grown as they have become components in more sophisticated statistical models. A new acceptance-rejection method is proposed and investigated for the Bingham distribution on the sphere using the angular central Gaussian distribution as an envelope. It is shown that the proposed method has high efficiency and is also straightforward to use. Next, the simulation method is extended to the Fisher and Fisher-Bingham distributions on spheres and related manifolds. Together, these results provide a widely applicable and efficient methodology to simulate many of the standard models in directional data analysis. An R package `simdd`, available in the online supplementary material, implements these simulation methods.

Keywords: acceptance-rejection, angular central Gaussian distribution, Bingham distribution, bivariate von Mises sine distribution, matrix Fisher distribution, simulation efficiency.

1 Introduction

Directional data analysis is concerned with statistical analysis on various non-Euclidean manifolds, starting with circle and the sphere, and extending to related manifolds. Comprehensive monographs are available for statistical analysis in this setting; see, e.g., Fisher et al. (1987), Mardia & Jupp (2000), Chikuse (2003). However, the subject of simulation has received much less coverage, with the key contributions scattered through the literature.

The need for effective simulation methods has grown in recent years as directional distributions have become components in more sophisticated statistical models, which are studied using MCMC methods. For example, Green & Mardia (2006) used the matrix Fisher distribution for random 3×3 rotation matrices in a Bayesian model to align two unlabelled configurations of points in $\mathbb{R}^3$, with an application to a problem of protein alignment in bioinformatics.

In general there exist suitable direct methods of simulation, especially methods based acceptance rejection, for the simpler directional models. However, it is necessary to resort to cumbersome MCMC methods for the more complicated distributions. The purpose of this paper is to extend availability of acceptance rejection methods to a wider class of directional distributions. The starting point is a new acceptance rejection method for the Bingham distribution, which can then be used as a building block in a wider range of applications.

The paper is organized as follows. Following some background and preparation in Section 2, the new acceptance rejection simulation method for the Bingham distribution is proposed and analyzed in Section 3. Extensions and special cases are covered in Sections 4–7. Section 8 sets the results of this paper in context by reviewing the literature and summarizing the best available methods in different settings. Some uses of this simulation methodology are explored in Section 9. Earlier versions of this work appeared in Kent & Ganeiber (2012) and Kent et al. (2012).

The unit sphere $S_{q-1} = \{x \in \mathbb{R}^q : \ x^T x = 1\}, \ q \geq 2$, comprises the unit vectors in $\mathbb{R}^q$. The surface area of S_{q-1} is given by $\pi_q = 2\pi^{q/2}/\Gamma(q/2)$ and the differential element of surface area can be written as $[dx]$. Thus the uniform distribution on S_{q-1} can be written as $\pi_q^{-1}[dx]$. A more explicit formula can be given using polar coordinates. For example, the

circle S_1 can be parameterized by $\theta \in [0, 2\pi)$ with uniform measure $d\theta/(2\pi)$. The sphere S_2 can be parameterized by colatitude $\theta \in [0, \pi]$ and longitude $[0, 2\pi)$ with uniform measure

$$\sin\theta d\theta d\phi/(4\pi). \tag{1.1}$$

Strictly speaking a probability density on a manifold is a density with respect to an underlying measure. In Euclidean space $\mathbb{R}^p$ the underlying measure is usually taken to be Lebesgue measure dx without explicit comment. But on non-Euclidean manifolds more care is needed. This paper is concerned with spheres and related compact manifolds for which there is a natural underlying uniform measure with a finite total measure. To avoid repeated occurences of normalizing constants such as π_q and differential elements such as $[dx]$, all such probability densities will be expressed with respect to the uniform distribution. Thus we will write the density for the uniform distribution on S_2 as $f(x) = 1$ (with respect to itself) rather than as $f(x) = 1/(4\pi)$ (with respect to $[dx]$) or as $f(x) = \sin\theta$ (with respect to $d\theta d\phi$).

2 Background

Recall the acceptance-rejection method of simulation. Consider two densities,

$$f(x) = c_f f^*(x), \quad g(x) = c_g g^*(x) \tag{2.1}$$

where f^* and g^* are known functions, but where the normalizing constants may or may not have a known explicit form. Suppose it is possible to simulate easily from g, known as the envelope, and it is desired to simulate observations from f. The key requirement is that there is a known bound of the form

$$f^*(x) \le M^* g^*(x) \text{ for all } x \tag{2.2}$$

for some finite constant M^*. The acceptance-rejection algorithm proceeds as follows.

Step 1. Simulate $X \sim g$ independently of $W \sim \text{Unif}(0, 1)$.

Step 2. If $W < f^*(X)/\{M^* g^*(X)\}$, then accept X.

Step 3. Otherwise go back to step 1.

Comments

(a) If we set $M = c_f M^*/c_g$, then (2.2) can be expressed equivalently as $f(x) \leq Mg(x)$ for all x.

(b) The bound M satisfies $M \geq 1$. The number of trials needed from g is geometrically distributed with mean $M \geq 1$. The efficiency is defined by $1/M$. For high efficiency the bound M should be as close to 1 as possible.

(c) The algorithm can be used even if the normalizing constants do not have a known explicit form. However, to compute the efficiency analytically, it is necessary to know the normalizing constants.

(d) Suppose the density $g(x) = g(x; b)$ depends on a parameter b with corresponding bound $M^*(b)$ in (2.2). If the normalizing constant $c_g = c_g(b)$ has a known explicit form, then it is possible to maximize the efficiency with respect to b, even if c_f does not have a known explicit form.

When developing acceptance-rejection simulation methods for directional distributions, there are several issues to consider:

- the need for good efficiency for a wide range of concentration parameters for f, ranging from uniform to highly concentrated. In similar problems on $\mathbb{R}^p$, the task is simpler when distributions are closed under affine transformations; in such cases it is sufficient to consider just a single standardized form of the distribution for f.
- the challenge in finding a tractable envelope distribution.
- the presence of trigonometric factors in the base measure when expressed in polar coordinates, such as in (1.1).

The basic strategy in this paper is to bound certain exponential family densities using tractable densities with heavier tails. A simple example in Euclidean space is given by the multivariate normal density, which can be bounded using the multivariate Cauchy density.

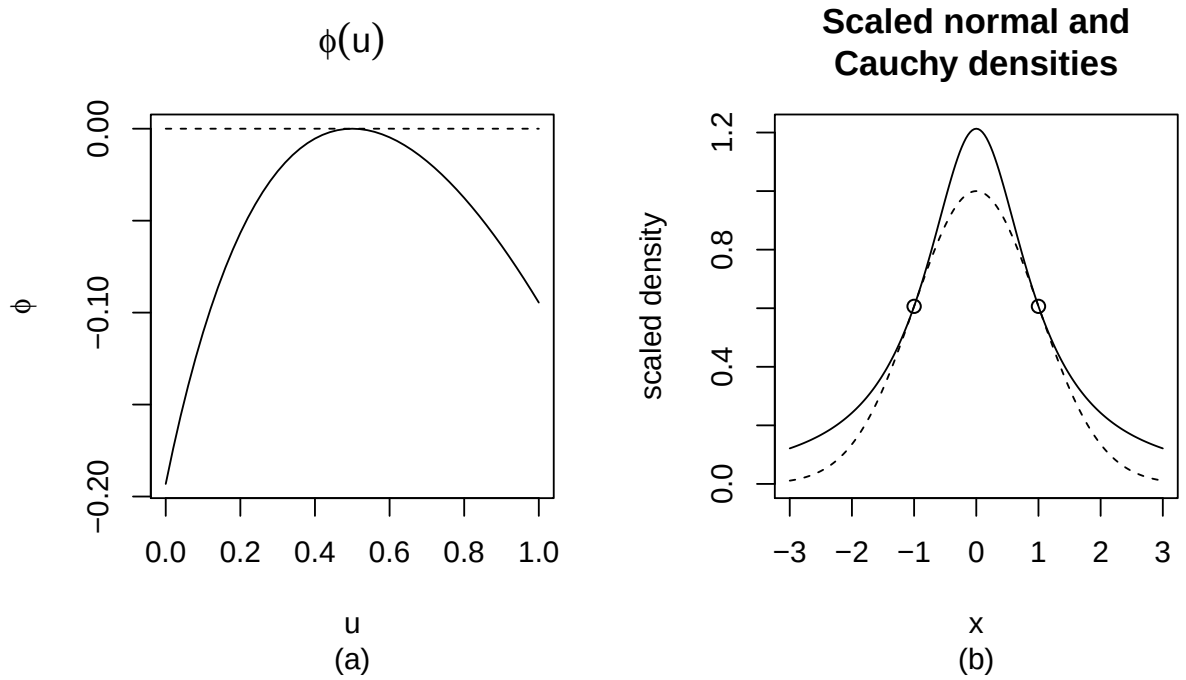

Figure 1: (a) The modified log function $\phi(u)$ in (2.3) (solid line) with $q = 2$. The upper bound of 0 (dashed line) is attained when $u = 1/2$. (b) Scaled Cauchy envelope (solid line) for a scaled normal density (dashed line) in $p = 1$ dimension. The two curves touch at $x = \pm 1$.

This example is important both the illustrate the general procedure and to set the scene for the Bingham distribution in the next section.

To develop a bound for the normal density, consider first a version of the log function, which has been modified to simplify a later inequality,

$$\phi(u) = \frac{q}{2}\log(1 + 2u/b) - u - \frac{q}{2}\log(1 + 2u_0/b) + u_0, \quad u \geq 0, \tag{2.3}$$

where $q \geq 2$ and $0 < b < q$ are fixed constants and $u_0 = (q - b)/2$. The last two terms on the righthand side of (2.3) are constants, chosen so that $\phi(u_0) = 0$. The value of u_0 is chosen so that the function $\frac{q}{2}\log(1 + 2u/b)$ has slope 1 at $u = u_0$; hence $\phi'(u_0) = 0$. Also note that $\phi''(u) < 0$ for $u \geq 0$ so that $\phi(u)$ is a concave function. Therefore, $\phi(u) \leq 0$ for all $u \geq 0$; see Figure 1(a). After exponentiating, the inequality $\phi(u) \leq 0$ can be re-arranged

Table 1: Efficiency $1/M$, where M is given in (2.7), of the A/R simulation method for the multivariate normal distribution in p dimensions using a multivariate Cauchy envelope. Here $q = p + 1$.

p	1	2	3	4	5	10	50	100
q	2	3	4	5	6	11	51	101
eff.	66%	52%	45%	40%	36%	26%	12%	9%

as

$$e^{-u} \leq e^{-(q-b)/2} \left(\frac{q/b}{1 + 2u/b} \right)^{q/2}. \tag{2.4}$$

The multivariate normal distribution $N_p(0, \Sigma)$ has density

$$f(x; \Sigma) = f(x) = c_f f^*(x), \quad f^*(x) = \exp(-\frac{1}{2} x^T \Sigma^{-1} x), \quad c_f = c_f(\Sigma) = |2\pi\Sigma|^{-1/2}, \tag{2.5}$$

for $x \in \mathbb{R}^p$. The multivariate Cauchy distribution $C_p(0, \Psi)$ has density

$$g(x; \Psi) = g(x) = c_g g^*(x), \quad g^*(x) = (1 + x^T \Psi^{-1} x)^{-q/2}, c_g = c_g(\Psi) = \frac{\Gamma(q/2)}{\pi^{q/2}} |\Psi|^{-1/2}, \tag{2.6}$$

(Mardia et al. 1979, p. 57), where here and elsewhere $q = p + 1$.

If we set $\Psi = b\Sigma$ so that the scatter matrix for the Cauchy is a scalar multiple of the covariance matrix for the normal, and if we set $u = \frac{1}{2} x^T \Sigma^{-1} x$, then the inequality (2.4) leads to a bound on the densities, $f(x; \Sigma) \leq M(b) g(x; \Psi)$, with

$$M(b) = 2^{-(q-1)/2} q^{q/2} e^{-q/2} \pi^{1/2} b^{-1/2} e^{b/2} / \Gamma(q/2).$$

Minimizing over $0 < b < q$ yields the optimal parameter $b = 1$ with optimal bound

$$M = M(1) = \sqrt{2\pi e} \left(\frac{q}{2e} \right)^{q/2} / \, \Gamma(q/2). \tag{2.7}$$

Figure 1(b) illustrates the comparison between the two densities. Table 1 gives a collection of efficiencies $1/M$ as a function of dimension p. For large p, $M \sim \sqrt{pe/2}$ by Stirling's formula.

Note that the efficiency declines slowly with the dimension, but is still high enough to be feasible even for dimension $p = 100$. Of course, this is just a toy example since there are better ways to simulate the normal distribution. However, it is important for the next section, both as a motivating example and as a limiting case.

3 The key building block: simulating the Bingham distribution

For the purposes of this paper, the starting point for the simulation of directional distributions is the Bingham distribution. In this section we describe the "BACG" acceptance rejection method to simulate the Bingham distribution using the angular central Gaussian distribution as an envelope. In later sections, we show how the simulation of the Bingham distribution can be used as a building block to simulate other directional distributions.

The Bingham distribution, $\text{Bing}_{q-1}(A)$ on S_{q-1}, $q \geq 2$, where the parameter matrix A is $q \times q$ symmetric, has density

$$f_{\text{Bing}}(x) = c_{\text{Bing}} f^*_{\text{Bing}}(x), \quad f^*_{\text{Bing}}(x) = \exp(-x^T A x). \tag{3.1}$$

Note the distribution is antipodally symmetric, $f(x) = f(-x)$. The normalizing constant $c_{\text{Bing}} = c_{\text{Bing}}(A)$ can be expressed as a hypergeometric function of matrix argument (Mardia & Jupp 2000, p. 182), but is not sufficiently tractable to be of interest here. The use of a minus sign in the exponent is unconventional but simplifies later formulae. Since A and $A + cI$ define the same distribution for any real constant c, we may assume without loss of generality that the eigenvalues of A satisfy

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_q. \tag{3.2}$$

In $q = 3$ dimensions, this distribution can exhibit isotropic bipolar behavior ($0 = \lambda_1 < \lambda_2 = \lambda_3$), girdle behavior ($0 = \lambda_1 = \lambda_2 < \lambda_3$), and intermediate behavior. Provided $\lambda_1 < \lambda_2$ the density is unimodal in terms of the axis $\pm x$ (or equivalently, bimodal in terms of the direction x), with the mode lying along the axis given by the eigenvector for the eigenvalue λ_1. See Section 5.1 for more about the distinction between a direction and an axis.

The angular central Gaussian distribution, $\text{ACG}(\Omega)$ on S_{q-1}, where the parameter matrix Ω is $q \times q$ symmetric positive definite, takes the form

$$f_{\text{ACG}}(x) = c_{\text{ACG}} f^*_{\text{ACG}}(x), \quad f^*_{\text{ACG}}(x) = \left(x^T \Omega x\right)^{-q/2}, \quad c_{\text{ACG}} = |\Omega|^{1/2}. \tag{3.3}$$

The angular central Gaussian distribution is simple to simulate. If $y \sim N_q(0, \Sigma)$, where Σ is positive definite, then $x = y/||y|| \sim \text{ACG}(\Omega)$ with $\Omega = \Sigma^{-1}$ (e.g., Mardia & Jupp 2000, p. 182).

Setting $u = x^T Ax$ in (2.4) and setting $\Omega = \Omega(b) = I + 2A/b,\ b > 0$, yields the following envelope inequality on the starred densities,

$$\begin{aligned} f^*_{\text{Bing}}(x) &= e^{-u} \\ &\le e^{-(q-b)/2} \left(\frac{q/b}{1 + 2x^T Ax/b} \right)^{q/2} \\ &= e^{-(q-b)/2} \left(\frac{q/b}{x^T \Omega x} \right)^{q/2} \\ &= e^{-(q-b)/2} (q/b)^{q/2} f^*_{\text{ACG}}(x), \end{aligned} \tag{3.4}$$

using the constraint $x^T x = 1$. The corresponding bound $M(b)$ takes the form

$$M(b) = c_{\text{Bing}} e^{-(q-b)/2} (q/b)^{q/2} |\Omega(b)|^{-1/2}. \tag{3.5}$$

Since $|\Omega(b)| = \prod_{i=1}^q (1 + 2\lambda_i/b)$, the function $\log M(b)$ and its first two derivatives take the form

$$\begin{aligned} \log M(b) &= \frac{1}{2} b - \frac{1}{2} \sum_{i=1}^q \log(b + 2\lambda_i) + \text{ const.}, \\ \{\log M(b)\}' &= \frac{1}{2} - \frac{1}{2} \sum_{i=1}^q (b + 2\lambda_i)^{-1}, \\ \{\log M(b)\}'' &= \frac{1}{2} \sum_{i=1}^q (b + 2\lambda_i)^{-2}, \end{aligned}$$

where prime denotes differentiation. Note $\{\log M(b)\}'' > 0$; hence $\log M(b)$ is convex for $b \in (0, \infty)$. Since $\{\log M(b)\}'$ increases monotonically from $-\infty$ to $1/2$ as for $b \in (0, \infty)$, it follows that the equation $\{\log M(b)\}' = 0$ has a unique solution b_0, say, which is therefore the unique minimum of $\log M(b)$. The equation $\{\log M(b_0)\}' = 0$ can be rewritten as

$$\sum_{i=1}^q \frac{1}{b_0 + 2\lambda_i} = 1. \tag{3.6}$$

It is not difficult to check that $1 \le b_0 \le q$.

Let $M(b_0)$ denote the optimal bound. Curiously, the same value of b_0 also appears in the saddlepoint approximation of Kume & Wood (2005) for the Bingham normalizing constant (where b here is the same as $-2t$ there), and leads to the approximate formula,

$$\hat{M}(b_0) = \frac{2\pi^{1/2}}{\Gamma(q/2)} \left(\frac{q}{2e} \right)^{q/2} Q^{1/2}, \quad Q = \sum_{i=1}^q \frac{1}{(b_0 + 2\lambda_i)^2}. \tag{3.7}$$

Caution is needed with this approximation because it is not exact in either the limiting case of uniformity or the limiting case of high concentration. For the parameter values in Table 2, the saddlepoint approximation overestimates the efficiency by between 4% and 9%.

However, it is possible to say exactly what happens in the limiting cases. If all the λ_i converge to 0, then b converges to q. Both the Bingham density and the ACG envelope converge to the uniform distribution and the efficiency converges to 1.

To deal with the high concentration case, it is simplest to replace A in (3.1) by βA and think of A as a fixed matrix as $\beta > 0$ gets large. Provided the $q-1$ largest eigenvalues of A are strictly positive, the ACG distribution (restricted to a hemisphere about the mode) converges to a $(q-1)$-dimensional multivariate Cauchy distribution, the Bingham distribution converges to a $(q-1)$-dimensional multivariate normal distribution, b_0 converges to 1 and the bound $M(b_0)$ converges to the bound (2.7), with $p = q-1$.

Empirically, it has been noticed that the limiting case is the worst possible case. For smaller values of the concentration matrix A, the efficiencies will be higher. The lefthand column of Table 2 illustrates the pattern for $q = 3$. The efficiency is never lower than 52%, the value from Table 1 for $p = 2$. This limiting value is attained in the concentrated bipolar case (when $\lambda_2 = \lambda_3$ is large). The girdle case ($\lambda_2 = 0$) has higher efficiencies. Similar conclusions are reached from the righthand side of Table 2 for $q = 4$. The efficiency is never lower than 45%, the value from Table 1 for $p = 3$. Each entry in this table has been constructed from one million simulations, so that the standard errors are negligible.

This general pattern persists for higher values of q. The efficiency lies between the entry in Table 1 under high concentration and 1 under uniformity. Other than the slow decline in efficiency under high concentration as q increases and questions of computer storage, there seems to be no upper bound to the feasible values of q. For example, we have found no problems for $q = 1000$.

4 The Fisher-Bingham model on S_{q-1}

Simulation of the Bingham distribution is important in its own right. However, it can also serve as a building block to simulate a wider class of directional distributions, both on the

Table 2: Efficiency of the BACG A/R simulation method on S_2 with $A = \text{diag}(0, \lambda_2, \lambda_3)$, and on S_3 with $A = \text{diag}(0, \lambda_2, \lambda_3, \lambda_4)$, for the Bingham distribution with an ACG envelope.

Efficiency on S_2			Efficiency on S_3							
λ_2	λ_3	Efficiency	λ_2	λ_3	λ_4	Efficiency	λ_2	λ_3	λ_4	Efficiency
0	0	100%	0	0	0	100%	0	100	100	69%
0	10	84%	0	0	10	89%	10	10	10	53%
10	10	58%	0	0	100	86%	10	10	100	50%
0	100	80%	0	10	10	75%	10	100	100	48%
100	100	53%	0	10	100	72%	100	100	100	45%

sphere and on related manifolds. For each of these manifolds there is a unique invariant measure which can be used to define a uniform distribution.

This section focuses on the Fisher-Bingham distribution on S_{q-1} with density

$$f^*_{\text{FB}}(x) = \exp(\kappa \mu_0^T x - x^T A x) = \exp\{\kappa(\mu_0^T x - x^T A^* x)\}, \tag{4.1}$$

where $\kappa > 0$, $\mu_0 \in S_{q-1}$ and $A(q \times q)$ is symmetric. Without loss of generality the smallest eigenvalue of A can be taken equal to 0. In the second form, κ has been factored out of the exponent, with $A^* = A/\kappa$; this form will be useful when considering efficiency in the limiting case $\kappa \to \infty$ with A^* held fixed.

The full FB family is too general to be of much interest statistically; practical interest is centered on various special cases of the *aligned* Fisher-Bingham family, for which μ_0 is an eigenvector of A. For this paper we are interested in distributions with a unique mode at $x = \mu_0$, which from the Appendix occurs if and only if $I + 2A^*$ is positive semi-definite. When studying simulation efficiency, it is also important to distinguish the *nonsingular* case (where $I + 2A^*$ is positive definite) from the *singular* case (where $I + 2A^*$ has some zero eigenvalues).

Important examples of unimodal aligned models include the following, with some simulated patterns given in Figure 2. For theoretical purposes, the aligned models are easiest to describe if the coordinate system is rotated so that $\mu_0 = [1, 0, \ldots, 0]^T$ lies on the first coordinate axis and $A = \text{diag}(\lambda_1, \ldots, \lambda_q)$ is diagonal.

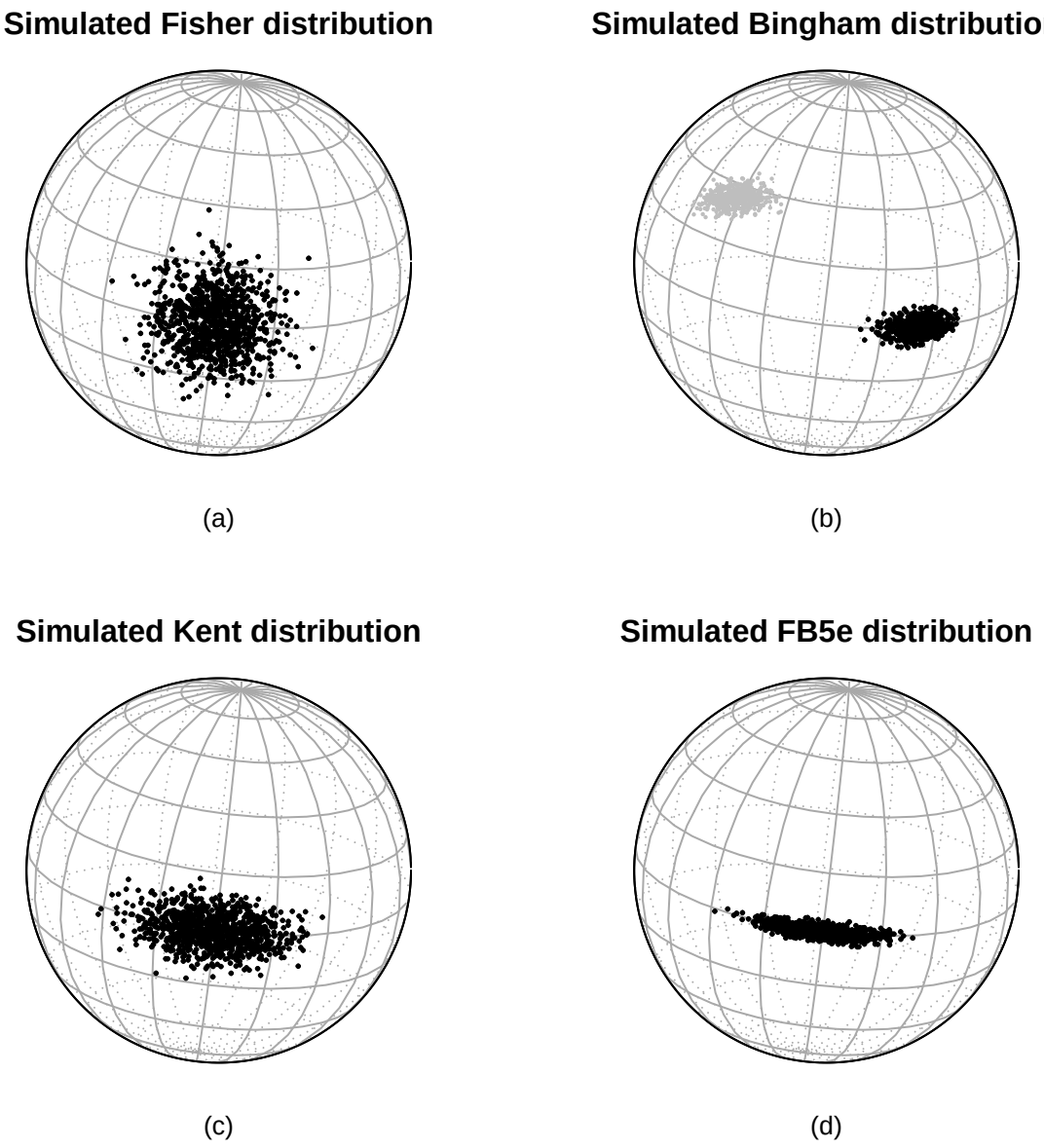

Figure 2: Simulated point clouds for various distributions on S_2. (a) Fisher distribution; (b) Bingham distribution; (c) balanced 5-parameter Fisher-Bingham(FB5b or Kent) distribution; (d) extreme FB5 (FB5e) distribution. Points in gray lie on the opposite side of the sphere.

- If $A = 0$ the model reduces to the von Mises ($q = 2$), the Fisher ($q = 3$), or the von Mises-Fisher (any $q \geq 2$) distribution. This model is the spherical analogue of the *isotropic* $(q-1)$-dimensional normal distribution.

- The case $\lambda_1 = 0,\ \sum_{j=2}^{q} \lambda_j = 0$ is known as the Kent distribution. On S_2, it is also known as the *balanced* 5-parameter Fisher-Bingham (FB5b) distribution (Kent 1982). If $\max |\lambda_j| < \kappa/2$, the distribution is nonsingular unimodal and forms a spherical analogue of the *general* $(q-1)$-dimensional normal distribution. The adjective *balanced* has been added recently to distinguish this model from the following choice.

- On S_2 the case $\lambda_1 = \lambda_2 = 0,\ \lambda_3 = \delta \geq 0$ is known as the *extreme* FB5 (FB5e) distribution (Kent et al. 2016) and is always nonsingular unimodal. It is also a spherical analogue of the *general* bivariate normal distribution, but is better than the balanced model at describing unimodal behaviour closely concentrated near a great circle. For an application see Section 9.2.

The proposed simulation method is defined for any model in the full Fisher-Bingham family. To describe the method, start with the von Mises-Fisher density (4.1) with $A = 0$. The elementary inequality $(1-y)^2 \geq 0$, with $y = x^T\mu_0$, can be re-arranged to give

$$\begin{aligned} f_{\mathrm{F}}^*(x) &\leq \exp\left[(\kappa/2)\left\{\left(x^T\mu_0\right)^2 + 1\right\}\right] \\ &= \exp\left\{\kappa - (\kappa/2)\, x^T A x\right\} \\ &= e^{\kappa} f_{\mathrm{Bing}}^*(x), \end{aligned} \tag{4.2}$$

where $A = I_q - \mu_0\mu_0^T$. Hence an acceptance rejection simulation method for the von Mises-Fisher distribution can be constructed using a Bingham envelope.

The two sides of (4.2) match when $x = \mu_0$ so that it is not possible to get a tighter bound. In relative terms, the two starred densities are maximally different when $x = -\mu_0$. This difference matters most when κ is large, when the efficiency of acceptance-rejection with a Bingham envelope drops to 50%; the efficiency rises to 100% as $\kappa \to 0$. Empirically the efficiency lies between these two extremes for intermediate values of κ.

The inequality (4.2) can be combined with Section 3 to provide a method to simulate the von Mises-Fisher distribution with an ACG envelope. Of course there is no need for a new method for the von Mises-Fisher distribution. Good methods are already available; see the Section 8 for a discussion. However, the bounds of this section can be combined with the previous section to simulate the Fisher-Bingham distribution with an ACG envelope.

The Fisher-Bingham density in (4.1) can be bounded by a Bingham density

$$f^*_{\mathrm{FB}}(x) \le \exp(\kappa - x^T A^{(1)} x),$$

where $A^{(1)} = A + (\kappa/2)(I - \mu_0\mu_0^T)$. Then Section 3 can be used to bound this Bingham density by an ACG density. We shall call the resulting acceptance-rejection algorithm the FBACG algorithm. Some comments on efficiency are given in Section 8.

5 Special cases of the Bingham distribution

There are a number of "accidental isomorphisms" in differential geometry in which a quotient manifold becomes identified with another familiar manifold through a quadratic mapping. These isomorphisms are called "accidental" because there does not seem to be any systematic pattern. In each case the uniform distribution on the first manifold maps to the uniform distribution on the new manifold, and the Bingham distribution maps to a distribution related to the von Mises-Fisher distribution on the new manifold. The implications for simulation are laid out in the next subsections.

5.1 $\mathbb{R}P_1 = S_1$

Real projective space is defined as the quotient space $\mathbb{R}P_{q-1} = S_{q-1}/\{1, -1\}$ in which two antipodal points or "directions" $\pm x$ are identified with one another to represent the same "axis". Since the Bingham and ACG densities have the property of antipodal symmetry, $f(x) = f(-x)$, they can also be viewed as densities on $\mathbb{R}P_{q-1}$.

Next specialize to the circle S_1. A point on the circle can be represented by an angle $\theta \in [0, 2\pi)$ or in Euclidean coordinates $x = (x_1, x_2)^T$ where $x_1 = \cos\theta, x_2 = \sin\theta$. Consider a two-to-one map from S_1 to a new circle defined by $\phi = 2\theta$, with Euclidean coordinates $y = (y_1, y_2)^T$ where $y_1 = \cos\phi = x_1^2 - x_2^2,\ y_2 = \sin\phi = 2x_1x_2$. Note that the antipodal directions $\theta,\ \theta + \pi$ map to the same value of ϕ, so that the map is in fact a one-to-one map between $\mathbb{R}P_1$ and S_1. A quadratic form in x can be rewritten as

$$x^T A x = \frac{1}{2}(a_{11} - a_{22})y_1 + a_{12}y_2 + \frac{1}{2}(a_{11} + a_{22}),$$

which is a linear function of y. Hence a Bingham distribution, whose density is quadratic in x on $\mathbb{R}P_1$ can be identified with a von Mises distribution, whose density is linear in y, on S_1. Similarly, in the ACG density the quadratic form $x^T\Omega x$ becomes a linear function of y, so the density in y reduces to the wrapped Cauchy density (Mardia & Jupp 2000, p. 52).

Suppose A is diagonal, $A = \text{diag}(0, \lambda)$, $\lambda \geq 0$. In this case the dominant axis of the Bingham distribution is the x_1-axis. The corresponding von Mises density takes the form

$$f_{\text{VM}}(y) \propto \exp(\kappa y_1), \quad \kappa = \lambda/2,$$

so that the corresponding von Mises density has its mode in the y_1-direction. The corresponding wrapped Cauchy density, with $\Omega = I + 2A/b$, takes the form,

$$f_{\text{WC}}(y) = \frac{(1-\rho^2)}{1+\rho^2-2\rho y_1}$$

where $\rho = (\beta - 1)/(\beta + 1)$ (Mardia & Jupp 2000, p. 52).

Hence the simulation method for the Bingham distribution with an ACG envelope can be recast as a simulation for the von Mises distribution with a wrapped Cauchy envelope. It turns out that this latter method is identical to the proposal of Best & Fisher (1979), even up to the choice of the optimal tuning constant b.

5.2 $\mathbb{C}P_1 = S_2$

Another quotient space of the sphere is complex projective space, $\mathbb{C}P_{q-1} = S_{2q-1}/S_1$. To understand this space, suppose a unit vector $x \in \mathbb{R}^{2q}$, is partitioned as $x^T = (x_1^T, x_2^T)$ where x_1 and x_2 are q-dimensional. The information in x can also be represented by a q-dimensional complex vector $z = x_1 + ix_2$. Then $\mathbb{C}P_{q-1}$ is obtained from S_{2q-1} by identifying the scalar multiples $e^{i\theta}z$ with one another for all $\theta \in [0, 2\pi)$.

If the $2q \times 2q$ symmetric concentration matrix A for a Bing_{2q-1} distribution can be partitioned in the form

$$A = \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix},$$

where A_1 is symmetric and A_2 is skew symmetric, then then the quadratic form $-x^TAx$ in the exponent of the Bingham density can be expressed in complex notation as $-z^*A_Cz$

where $A_C = A_1 + iA_2$. In terms of z, the density possesses complex symmetry, $f(z) = f(e^{i\theta}z)$ for all $\theta \in [0, 2\pi)$. When expressed in complex notation this distribution is known as the complex Bingham distribution $\mathrm{CB}_{q-1}(A_C)$; it can also be viewed as a distribution on $\mathbb{C}P_{q-1}$ (Kent 1994).

The complex projective space $\mathbb{C}P_{q-1}$ arises in the study of shape for configurations of $q+1$ landmarks in the plane, and the identification with S_1 when $q+1=3$ was used to visualize the shape space for triangles of landmarks (Kendall 1984). Kent (1994) showed that the complex Bingham distribution on $\mathbb{C}P_1$ can be identified with the Fisher distribution on S_2.

5.3 $\mathbb{R}P_3 = SO(3)$

The special orthogonal group $\mathrm{SO}(r)$ is the space of $r \times r$ rotation matrices, $SO(r) = \{X \in \mathbb{R}^{r\times r} : X^TX = I_r,\ |X| = 1\}$. A natural parametric distribution is given by the matrix Fisher distribution $MF_r(F)$, with $r \times r$ parameter matrix F. The density is given by

$$f^*(X) = \exp\{\mathrm{tr}(F^TX)\}. \tag{5.1}$$

To describe the concentration properties of this distribution, it is helpful to give F a *signed singular value decomposition*

$$F = U\Delta V^T. \tag{5.2}$$

The adjective "signed" means that U and V are constrained to be $r \times r$ rotation matrices and the elements of the diagonal matrix Δ satisfy $\delta_1 \geq \cdots \geq \delta_{r-1} \geq |\delta_r|$, where the final element δ_r is negative if and only if $\det(F) < 0$. If $X \sim MF_r(F)$, and if $VFU^T = \exp(S)$ is written in terms of the matrix exponential of a skew symmetric $r \times r$ matrix S, then under high concentration the linearly independent elements of S have asymptotic normal distributions, $s_{ij} \sim N(0, (\delta_i + \delta_j)^{-1}),\ 1 \leq i < j \leq r$. See also Section 5.3 below for further discussion of the case $r = 3$.

There is a quadratic mapping taking an unsigned 4-dimensional unit vector $\pm x$ to a

3×3 rotation matrix $X = M(x) = M(-x)$, say. More specifically

$$M(x) = \begin{bmatrix} x_1^2 + x_2^2 - x_3^3 - x_4^2 & -2(x_1x_4 - x_2x_3) & 2(x_1x_3 + x_2x_4) \\ 2(x_1x_4 + x_2x_3) & x_1^2 + x_3^2 - x_2^2 - x_4^2 & -2(x_1x_2 - x_3x_4) \\ -2(x_1x_3 - x_2x_4) & 2(x_1x_2 + x_3x_4) & x_1^2 + x_4^2 - x_2^2 - x_3^2 \end{bmatrix} \tag{5.3}$$

(Mardia & Jupp 2000, p. 285). Further a random axis $\pm x$ on $\mathbb{R}P_3$ follows a Bingham distribution if and only if the corresponding random matrix $M(x)$ follows a matrix Fisher distribution. In particular, if $A = \Lambda$ is diagonal, then $F = \Delta$ in (5.2) is also diagonal with the parameters related by

$$\lambda_1 = 0, \quad \lambda_2 = 2(\delta_2 + \delta_3), \quad \lambda_3 = 2(\delta_1 + \delta_3), \quad \lambda_4 = 2(\delta_1 + \delta_2). \tag{5.4}$$

Kent et al. (2012) gives some further details.

A simple way to simulate a rotation matrix from the matrix Fisher distribution $MF_3(F)$ for a general parameter matrix F with signed singular value decomposition (5.2) is given as follows. Using the BACG method, simulate x from $\mathrm{Bing}_3(\Lambda)$ with Λ given by (5.4), and let $M(x)$ denote the corresponding rotation matrix using (5.3). Then $UM(x)V^T$ follows the matrix Fisher distribution $MF_3(F)$. From Table 1, the efficiency will always be at least 45%.

6 The matrix Bingham distribution on the Grassmann manifold $\mathcal{G}_{r,q}$

Let $1 \leq r < q$. The Grassmann manifold $\mathcal{G}_{r,q}$ is defined to be the set of all r-dimensional subspaces of $\mathbb{R}^q$. It can be described as a quotient space of a Stiefel manifold $\mathcal{G}_{r,q} = V_{r,q}/O(r)$, where the Stiefel manifold,

$$V_{r,q} = \{X \in \mathbb{R}^{q \times r} : X^T X = I_r\},$$

denotes the space of $q \times r$ column orthonormal matrices X, say. It is convenient to represent an element of $\mathcal{G}_{r,q}$ by a matrix X, where X is identified with XR for all $r \times r$ orthogonal matrices R. It should be noted that the notation for this manifold is not standardized; e.g., some authors write $\mathcal{G}_{r,q-r}$ instead of $\mathcal{G}_{r,q}$.

The matrix Bingham distribution on $V_{r,q}$ is defined by the density

$$f_{\text{MB}}(X) \propto \exp\{\text{tr}(-X^T AX)\}. \tag{6.1}$$

Since $\text{tr}(X^T AX) = \text{tr}(R^T X^T AXR)$ for all $r \times r$ orthogonal matrices, it can also be viewed as a distribution on the Grassmann manifold $\mathcal{G}_{r,q}$. The $q \times q$ concentration matrix A has the same form as for the Bingham distribution in Section 3.

For every r-dimensional subspace in $\mathbb{R}^q$, there is a unique complementary $(q-r)$-dimensional subspace orthogonal to it. If X and $X^\perp$ are column orthonormal matrices, whose columns are bases of these subspaces, then $[X\ X^\perp]$ is a $q \times q$ orthogonal matrix. Further X follows a matrix Bingham distribution on $\mathcal{G}_{r,q}$ with parameter matrix A if and only if $X^\perp$ follows a matrix Bingham distribution on $\mathcal{G}_{q-r,q}$ with parameter matrix $-A$ (but note that the eigenvalues of $-A$ will not have the standardized form given in (3.2)). Hence for simulation purposes, we may without loss of generality suppose that $r \leq q/2$.

The matrix ACG distribution, denoted $MACG_{r,q}(\Omega)$, where Ω is a positive definite symmetric $q \times q$ matrix, is also lies on $\mathcal{V}_{r,q}$. It is also invariant under rotation on the right and hence can also be viewed as a distribution on $\mathcal{G}_{r,q}$. The density takes the form

$$g^*_{MACG}(X) = |X^T \Omega X|^{-q/2}, \quad c_g = |\Omega|^{r/2}$$

(e.g. Chikuse 2003, p. 40). Simulations from this distribution can be constructed as follows. Let Y be a $q \times r$ matrix whose columns are independently normally distributed, $N_q(0, \Omega^{-1})$. Set $X = Y(Y^T Y)^{-1/2}$ using the symmetric square root of a positive definite matrix. Then $X \sim MACG_{r,q}(\Omega)$.

If Ω is related to A by $\Omega = \Omega(b) = I_q + 2A/b$ as in Section 3, then the matrix Bingham density can be bounded by the matrix ACG density by using the inequality in (2.4) r times. Namely, let the eigenvalues of $X^T AX$ be denoted $0 \leq u_1 \leq \cdots \leq u_r$. Since $f^*_{\text{MB}}(X) = \exp(\sum u_i)$ and $f^*_{\text{MACG}}(X) = \{\prod(1 + 2u_i/b)\}^{-q/2}$, applying (2.4) r times yields the envelope bound

$$M_1(b) = c_{\text{MB-bal}}\{M_1^*(b)|\Omega(b)|^{-1/2}\}^r \text{ where } M_1^*(b) = \{e^{-(q-b)/2}(q/b)^{q/2}\}^r. \tag{6.2}$$

Optimizing (6.2) over b yields the same equation (3.6) as before with the same value for the optimal value b_0.

Unfortunately, this bound decreases quickly with r, $1 \le r < q$. However, under high concentration, it is possible to tighten the bound substantially using the following two results.

(a) Let $X^\perp(q \times (q-r))$ be a complementary matrix to X satisfying $X^{\perp T}X^\perp = I_{q-r}$ and $X^T X^\perp = 0$. Then $Z = [X\ X^\perp]$, is a $q \times q$ orthogonal matrix, $Z^T Z = I_q$. Hence the eigenvalues of $Z^T AZ$ are the same as those of A, namely, $\lambda_1, \ldots, \lambda_q$, ordered as in (3.2). Further, $X^T AX$ is a principal submatrix of $Z^T AZ$. Hence by the Cauchy-Poincaré separation or interlacing theorem (e.g. Magnus & Neudecker 1999, pp. 209–211), the eigenvalues of $X^T AX$ satisfy

$$u_1 \ge \lambda_1 = 0, \quad \ldots, \quad u_r \ge \lambda_r.$$

(b) With $b = b_0$, the function $\phi(u)$ in (2.3) is monotone decreasing for $u \ge (q - b_0)/2$. In particular, $\phi(u_j) \le \phi(\lambda_j)$ provided $\lambda_j \ge (q - b_0)/2$. Hence the bound $M_1^*(b_0)$ can be tightened to

$$M_2^*(b_0) = \prod_{j=1}^{r} C_j,$$

where

$$C_j = \begin{cases} e^{-(q-b_0)/2}(q/b_0)^{q/2}, & \lambda_j \le (q-b_0)/2 \\ e^{-\lambda_j}(1 + 2\lambda_j/b_0)^{q/2}, & \lambda_j > (q-b_0)/2. \end{cases}$$

The efficiency is expected to decline as r increases. However, as noted before, we may restrict attention to the case $r \le q/2$. More numerical investigation is needed of the efficiency in this setting.

7 von Mises sine model on the torus

Finally consider the setting of product manifolds, where multivariate versions of directional models can be defined. There are a few special cases where acceptance rejection methods are available (e.g. Mardia et al. (2007, supplementary material) for the sine and cosine versions of the bivariate von Mises distribution on the torus). However, to get good efficiency, it

is often necessary to divide the parameter domain into different regions, each of which requires separate treatment.

To illustrate the potential for the results in this paper, we consider the bivariate von Mises sine model, with density proportional to

$$f^*(\theta, \phi) = \exp\{\kappa_1 \cos\theta + \kappa_2 \sin\phi + \alpha \sin\theta \sin\phi\}, \quad 0 \le \theta, \phi \le 2\pi. \tag{7.1}$$

For example, this distribution is useful in the study of protein structure to model pairs of angles describing the relative orientation of bonds between atoms in amino acids (Mardia et al. 2007). For the discussion here it is supposed that $\kappa_1 > 0,\ \kappa_2 > 0$ and $\alpha^2 < \kappa_1\kappa_2$ so that the distribution mimics a bivariate normal distribution under high concentration. The proposed simulation method proceeds in two steps: (a) first simulate θ from its marginal distribution (7.2) as discussed below, and (b) simulate ϕ given θ from its conditional von Mises distribution; see Section 5.1. We focus on step (a) here.

Integrating (7.1) over ϕ yields the *Bessel* marginal density for θ, proportional to

$$f_1^*(\theta) = \exp\{\kappa_1 \cos\theta\} I_0\left(\sqrt{\kappa_2^2 + \alpha^2 \sin^2\theta}\right), \tag{7.2}$$

where $I_0(\cdot)$ denotes the modified Bessel function of the first order. The derivative of the logarithm of the Bessel function takes the form

$$d/dx\{\log I_0(x)\} = \frac{I_1(x)}{I_0(x)} = A(x), \text{ say,}$$

where it is known that $A(x)$ increases monotonically from 0 to 1 as x ranges from 0 to ∞. Let $A_{\max} = A(\sqrt{\kappa_2^2 + \alpha^2})$. Also note that by a Taylor expansion,

$$\left(\kappa_2^2 + \alpha^2 \sin^2\theta\right)^{1/2} = \kappa_2 \left\{1 + (\alpha^2/\kappa_2^2) \sin^2\theta\right\}^{1/2} \le \kappa_2 + \frac{\alpha^2}{2\kappa_2} \sin^2\theta.$$

Hence by another Taylor expansion,

$$\log\left\{I_0\left(\sqrt{\kappa_2^2 + \alpha^2 \sin^2\theta}\right) / I_0(\kappa_2)\right\} \le \frac{A_{\max}\alpha^2}{2\kappa_2} \sin^2\theta. \tag{7.3}$$

Combining (7.3) with the Bingham bound for the von Mises density in (4.2) and the ACG bound for the Bingham density in (3.4) and (3.6) yields

$$\begin{aligned} f_1^*(\theta) &\le I_0(\kappa_2) \exp\{\frac{\kappa_1}{2}(1 + \cos^2\theta) + \frac{A_{\max}\alpha^2}{2\kappa_2} \sin^2\theta\} \\ &\le I_0(\kappa_2) \exp\{\kappa_1 - \beta \sin^2\theta\} \\ &\le M^*(1 + 2\beta \sin^2\theta/b_0)^{-1}, \end{aligned}$$

where $\beta = \frac{1}{2}(\kappa_1 - A_{\max}\alpha^2/\kappa_2)$, and

$$M^* = (2/b_0)I_0(\kappa_2)e^{\kappa_1 - 1 + b_0/2}.$$

That is, the Bessel density can be bounded by the ACG density. Numerical simulations indicate the efficiency is generally at least 30% under the assumptions given here. The exception is that the efficiency deteriorates under high concentration as $\alpha^2/(\kappa_1\kappa_2)$ increases towards 1. Note that if $\alpha^2/(\kappa_1\kappa_2) = 1$, the bivariate density behaves as a singular normal distribution under high concentration.

8 Review and commentary on different simulation methods

Since the simulation literature for directional distributions is widely scattered, it is useful to summarize the best simulation methods for various distributions of interest. Table 3 lists several common distributions on different spaces, together with the recommended method of simulation. Of course, a uniform envelope can always be used in an acceptance/rejection algorithm. This is a reasonable course of action under low or medium concentration, but the efficiency will decline as the concentration increases; hence a uniform envelope is not recommended when there are better algorithms.

Recently, some MCMC simulation methods on manifolds have been proposed by Kume & Walker (2009) (Fisher-Bingham on S_{q-1}), Green & Mardia (2006) and Habeck (2009) (matrix Fisher on SO(3)), Hoff (2009) (matrix Fisher-Bingham distributions on Stiefel and Grassmann manifolds) and Byrne & Girolami (2013) (more general distributions). However, there is still an ongoing investigation into the efficiency of different MCMC methods, so that the table entry will just state "uniform/MCMC" when there is not a simpler more specific recommendation. Further details are given in the following subsections.

Two principles have guided the recommendations for the acceptance-rejection algorithms given here:

(a) good efficiency across the range of possible concentration parameters, in particular under high concentration and near uniformity, and

Table 3: Recommended simulation methods various distributions on different directional spaces

Distribution	Space	Simulation method
von Mises-Fisher	S_{q-1}	Wood (1987)
Bingham	S_{q-1} or $\mathbb{R}P_{q-1}$	BACG
complex Bingham	S_{2q-1} or $\mathbb{C}P_{q-1}$	Kent et al. (2004)
complex Bingham quartic	S_{2q-1} or $\mathbb{C}P_{q-1}$	Kent et al. (2006)
Kent	S_2	Kent & Hamelryck (2005)/FBACG
aligned Fisher-Bingham	S_{q-1}	FBACG
general Fisher-Bingham	S_{q-1}	uniform/MCMC
matrix Fisher	$SO(3)$	BACG
matrix Fisher	$SO(r),\ r > 3$	uniform/MCMC
matrix Bingham	$\mathcal{V}_{r,q}$ or $\mathcal{G}_{r,q}$	BACG-based
bivariate von Mises sine	$S_1 \otimes S_1$	BACG-based

(b) simplicity (when the efficiencies for different algorithms are similar).

8.1 Uniform distribution on the sphere and Stiefel manifold

The simplest general method to simulate a uniform distribution on the unit sphere $S_{q-1}, q \geq 2$, is to set $x = u/||u||$ where $u \sim N_q(0, I_q)$. In low dimensions there are sometimes simpler methods using polar coordinates. E.g. on the circle S_1, let $\theta \sim \text{Unif}(0, 2\pi)$. On the sphere S_2 with colatitude θ and longitude ϕ, let $\cos\theta \sim \text{Unif}(-1, 1)$ independently of $\phi \sim \text{Unif}(0, 2\pi)$.

On the Stiefel manifold $\mathcal{V}_{r,q},\ 1 \leq r \leq q,\ q \geq 2$, the easiest approach is to simulate $U(q \times r)$ with independent $N(0, 1)$ entries, and set $X = U(U^TU)^{-1/2}$ using e.g. the symmetric square root of a positive definite matrix. Then X is uniformly distributed on $\mathcal{V}_{r,q}$. The same approach works on the Grassmann manifold $\mathcal{G}_{r,q},\ 1 \leq r < q,\ q \geq 2$.

8.2 von Mises-Fisher distribution $\mathbf{F}_{q-1}(\kappa, \mu_0)$ on S_{q-1}

For general $q \geq 2$, the recommended method of simulation is an acceptance/rejection method due to Ulrich (1984), as modified by Wood (1994). This method uses a fractional linear transformation of a beta variate to provide an envelope for $u = x^T \mu_0$. It gives good efficiency across the whole range of values for κ. In particular, for large κ the distribution of $2(1-u)$ is approximately the squared radial part of a multivariate normal distribution under the von Mises-Fisher model and of a multivariate Cauchy distribution under the envelope model, mimicking the efficiency calculations in (2.7).

Once the distribution of $u \in [0, 1]$ has been simulated, it is straightforward to simulate the whole von Mises-Fisher distribution by incorporating a uniformly distributed random direction y, say, on S_{q-2} (so y is a $(q-1)$-vector). More specifically, if $R = [R_1 \ \mu_0]$ is any $q \times q$ rotation matrix whose last column equals μ_0, let $x = u\mu_0 + (1-u^2)^{1/2} R_1^T y$.

For $q = 2$ the Ulrich-Wood method is essentially identical to the Best & Fisher (1979) method. One small exception to the recommendation to use the Ulrich-Wood method is the case $q = 3$ dimensions when u follows a truncated exponential distribution and can be simulated more simply by the inverse method without any need for rejection (Fisher et al. 1987, p. 59).

8.3 Bingham distribution $\mathbf{Bing}_{q-1}(A)$ on S_{q-1} or $\mathbb{R}P_{q-1}$

The BACG method developed in this paper is the first general-purpose acceptance/rejection simulation method for the Bingham distribution. However, earlier methods have been discussed in the literature for some special cases. In particular if $q = 2$, the BACG method reduces to the Best & Fisher (1979) method for the von Mises distribution as noted in Section 5.1.

If $q = 3$ and either $0 = \lambda_1 < \lambda_2 = \lambda_3$ (bipolar case) or $0 = \lambda_1 = \lambda_2 < \lambda_3$ (girdle case), the simulation problem can be reduced to a one-dimensional problem. Best & Fisher (1986) developed effective envelopes in these cases, with efficiencies broadly comparable to the BACG method.

If the eigenvalues appear in pairs then the methods for the complex Bingham can be used. Kent et al. (2004) developed several simulation methods and the best of these is

generally better than BACG. In particular, the efficiency of their "Method 1" approaches 100% under high concentration in contrast to the limiting efficiencies in Table 1 for BACG.

The BACG method here supersedes the MCMC method of Kume & Walker (2006).

Motivated by the accidental isomorphism in Section 5.2, Kent et al. (2006) developed a complex Bingham quartic (CBQ) distribution on $\mathbb{C}P_{q-1},\ q \geq 2$. When $q = 2$, this distribution reduces to the FB5b distribution. Ganeiber (2012) developed an effective and reasonably efficient simulation method for the CBQ distribution for $q > 2$. However, since the technique is not based on an angular central Gaussian envelope, details will not be given here.

8.4 Fisher-Bingham distribution FB(κ, μ_0, A) on S_{q-1}

For the Fisher-Bingham distribution on S_{q-1} and the corresponding FBACG algorithm of Section 4, it is convenient to split the assessment into special cases. When considering efficiency under high concentration, good efficiencies are obtained for the nonsingular unimodal aligned distributions. However, the efficiency can deteriorate to 0 in the singular or nonaligned cases.

- FB5b on S_2, with parameters $\kappa > 0,\ 0 \leq \beta \leq \kappa/2$. An efficient simulation algorithm (KH) for FB5b was developed by Kent & Hamelryck (2005); see also Kent (2012). For small κ, approximately for $\kappa < 10$, FBACG is a bit better than KH, where the FBACG efficiency is at least 34%; for large κ KH is a bit better, with an efficiency of at least 26%. The singular case where κ is large and $2\beta/\kappa$ is very close or equal to 1 needs special consideration; KH maintains its efficiency whereas the efficiency of BACG deteriorates to 0 as $\kappa \to \infty$.

- FB5e on S_2, with parameters $\kappa > 0, \delta \geq 0$. The efficiency is always at least 26% (half the entry in Table 1 for $q = 3$) and increases both as κ decreases and as δ increases.

- More general aligned case on S_{q-1}. Most of the time, if the Fisher-Bingham density (4.1) is unimodal aligned, then the Bingham envelope usually has an efficiency of at least 50% (and so the ACG envelope has an efficiency of at least half the entry in Table 1). The efficiency falls below this level only under high concentration in

the singular or near-singular setting, when the density is excessively flat at its mode. Under high concentration this situation corresponds to the case where the limiting normal distribution would have a singular covariance matrix.

In particular, with the exception of the recommendation to use KH for FB5b on S_2 for large κ, the FBACG method supersedes earlier acceptance-rejection methods developed by Wood (1987) for various unimodal aligned Fisher-Bingham distributions on S_2. It also supersedes the acceptance rejection method of Scealy & Welsh (2011, Appendix A4) for a higher-dimensional version of the Kent distribution, for which the efficiency drops to 0 under high concentration when $q > 3$. In addition it supersedes the MCMC method of Kume & Walker (2009) in the aligned case.

There is also an non-unimodal aligned Fisher-Bingham distribution on S_2 whose mode is a small circle. This case is covered by Wood (1987); the efficiency of the FBACG algorithm drops to 0 under high concentration.

- Non-aligned case. For non-aligned Fisher-Bingham distributions, it is difficult to make any firm theoretical statements about the behaviour of the FBACG algorithm. However, under moderate concentration it is still likely to be preferable to the MCMC methods of Kume & Walker (2009).

8.5 Matrix Fisher distribution MF(F) on $SO(r)$

When $r = 2$, $SO(2)$ is the same as S_1 and the matrix Fisher on SO(2) is identical to the von Mises distribution on S_1, so no new methodology is needed.

When $r = 3$ the accidental isomorphism in Section 5.3 reduces this case to the Bingham distribution on S_3, which can be simulated by the BACG method. Hence the efficiency is always at least 45% for all values of the parameters.

Earlier methods to simulate the matrix Fisher distribution on SO(3), now superseded by BACG, were based on MCMC algorithms. These include Green & Mardia (2006) and Habeck (2009).

The cases $r > 3$ are at least partly covered by the next subsection.

8.6 Matrix Fisher-Bingham distribution on $\mathcal{V}_{r,q}$

A general matrix Fisher-Bingham distribution can be defined on on $\mathcal{V}_{r,q}$; see, e.g. Mardia & Jupp (2000, p. 292). There is not yet a convenient and efficient A/R algorithm other than for the matrix Bingham case, where a solution was given in Section 6. However, the recent MCMC algorithms of Hoff (2009) and Byrne & Girolami (2013) can deal with this case.

9 Applications

Simulation often forms part of the machinery in a larger statistical algorithm. In this section we sketch two ways in which the methodology can be used.

9.1 Markov chain Monte Carlo updating for a rotation matrix

Motivated by problems in protein bioinformatics, Green & Mardia (2006) developed a Bayesian alignment model to match two configurations of "landmarks" $\{x_j, j = 1, 2, \ldots, m\}$ and $\{y_k, k = 1, 2, \ldots, n\}$, in r-dimensional space $\mathbb{R}^r$. The main parameters are: a matching matrix $M(m \times n)$, a shift parameter μ and a rotation matrix R taking one configuration to the other. The matching matrix M consists of zeros and ones, with at most a single one in each row and column. Thus M picks out a subset of pairs of landmarks from the two configurations which can be matched together. Given M, μ and σ^2, the basic model states that for the paired landmarks, one configuration is a rigid body transformation of the other, up to noise. That is, if two landmarks x_j and y_k are paired, i.e. if $M_{jk} = 1$, then the "error" term $y_k - R(x_j - \mu) \sim N(0, \sigma^2)$ is normally distributed, with the errors independent for different pairs.

Green & Mardia (2006) specified prior distributions for each parameter and developed an MCMC algorithm to simulate the posterior distribution. In particular, a Fisher matrix prior was assumed for R, with density proportional to $\exp\{\mathrm{tr}(F_0^T R)\}$. They showed that the posterior distribution for R given the rest of the parameters has the same form with

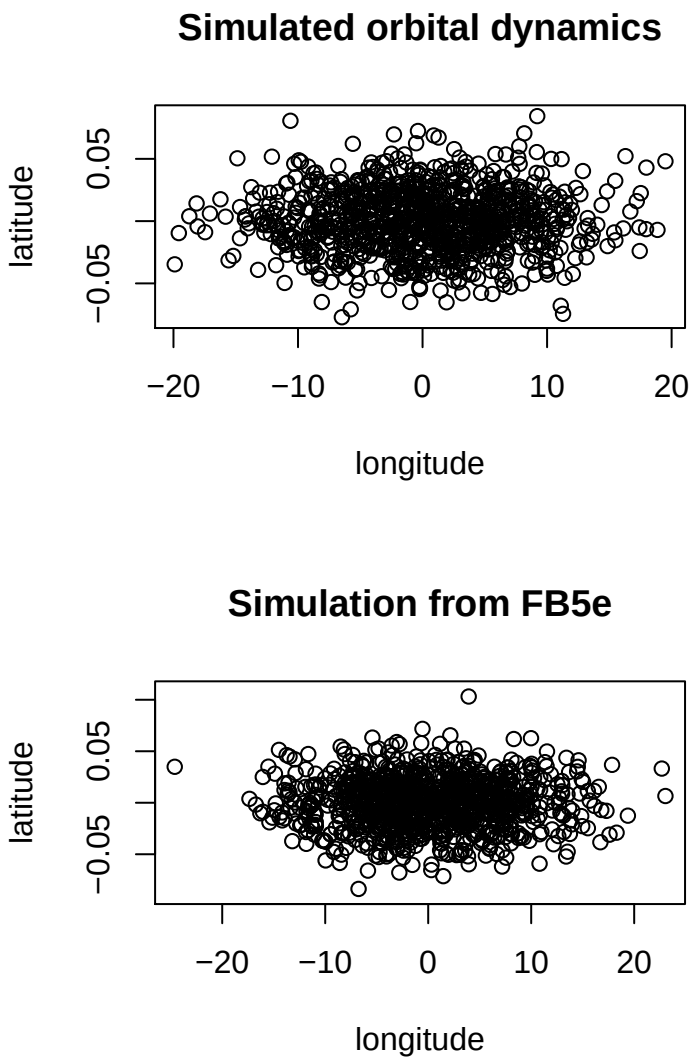

Figure 3: (a) Latitude and longitude for 1000 final orbital positions from noisy initial conditions using orbital dynamics. (b) 1000 simulated values from the FB5e distribution with $\kappa = 68,\ \delta = 2.7e6$. The axes are in degrees and the vertical axis has been blown up by a factor of 100 for clarity.

F_0 replaced by

$$F = F_0 + (1/2\sigma^2) \sum_{j,k:M_{jk}=1} (x_j - \mu) y_k^T,$$

where σ^2 is a variance parameter. In their work in $r = 3$ dimensions, Green & Mardia (2006) used a cumbersome internal MCMC algorithm to simulate R in terms of its Euler

angles; the BACG method developed here is more direct.

9.2 Visualization

One of the key uses of simulation is to visualize distributions that are difficult to understand analytically. Figure 2 illustrates the behaviour of some of the standard directional distributions. Another example arises from recent work on orbital dynamics (Kent et al. 2016). Using the laws of Newtonian motion, it is possible to propagate an initial uncertainty of an object in orbit about the earth to a later time. The predicted path lies near an ellipse but with much greater spread along the path of the ellipse than perpendicular to the path.

Figure 3 shows a simulated point cloud following orbital dynamics and a simulated dataset matched to it from the FB5e distribution (with a sample acceptance ratio of of 27%). For plotting purposes the data have been rotated so that the mode lies at the intersection of the Greenwich meridian and the equator, and the principal axis lies along the equator. In both cases the longitude has a spread of about 60^o, but the latitude is very concentrated with a spread of less than 0.2^o. The FB5e distribution of Section 4 was designed to model this type of pattern. It can be seen that the two plots are very similar.

A common step in particle filtering involves the the approximation of a point cloud by a parametric distribution. The ability to simulate easily from the extreme FB5 distribution facilitates this task.

Appendix

Lemma 1. Consider the function $g(\theta) = \cos\theta + \alpha \sin^2\theta$*, defined on the circle* $\theta \in (-\pi, \pi]$*, where* $\alpha \in \mathbb{R}$ *is a parameter. Then* g *is uniquely maximized over* θ *at* $\theta = 0$ *if and only if* $\alpha \leq 1/2$*.*

Proof. Write $g(\theta)$ as a function of $t = \cos\theta$, to give $h(t) = t + \alpha(1 - t^2)$, with derivative $h'(t) = 1 - 2\alpha t$. If $\alpha > 1/2$, then $h'(t) < 0$ for t near to 1; hence h is not maximized at $t = 1$. If $0 \leq \alpha \leq 1/2$, then $h'(t) > 0$ for $-1 < t < 1$, so $h(t)$ is uniquely maximized at $t = 1$; that is, $g(\theta)$ is uniquely maximized at $\theta = 0$. Similarly, if $\alpha < 0$, then $h(t) < t < 1$

for $-1 < t < 1$, so again $h(t)$ is uniquely maximized at $t = 1$.

Lemma 2. Let $\mu_0 \in S_{q-1}$, $q \geq 2$ be a unit vector and let A^* be a symmetric $q \times q$ matrix such that $A^*\mu_0 = 0$. The function $f(x) = \mu_0^T x - x^T A^* x$ is uniquely maximized over $x \in S_{q-1}$ at $x = \mu_0$ if and only if the matrix $I + 2A^*$ is positive semi-definite.

Proof. It is sufficient to verify the theorem on each great circle passing through μ_0. Let γ denote a unit vector perpendicular to μ_0, $\gamma^T\mu_0 = 0$, and let $g(\theta) = f((\cos\theta)\mu_0 + (\sin\theta)\gamma) = \cos\theta + \alpha\sin^2\theta$, where $\alpha = -\gamma^T A^*\gamma$. By Lemma 1, g is uniquely maximized at $\theta = 0$ if and only if $\alpha \leq 1/2$, i.e. $1 + 2\gamma^T A^*\gamma \geq 0$, which is true for all γ if and only if $I + 2A^*$ is positive semi-definite.

Supplementary material

R-package `simdd`: The package `simdd` (Kent 2016) contains code in R (R Core Team 2017) to implement the methods described in the article for the simulation of directional distributions (GNU zipped tar file).

Acknowledgement

The work of the first author for the FB5e distribution has been supported by the Air Force Office of Scientific Research, Air Force Material Command, USAF under Award No. FA9550-16-1-0099.

References

Best, D. J. & Fisher, N. I. (1979), 'Efficient simulation of the von Mises distribution', *J. Appl. Statist.* **28**, 152–157.

Best, D. J. & Fisher, N. I. (1986), 'Goodness-of-fit and discordancy tests for samples from the Watson distribution on the sphere', *Austr. J. Statist.* **1**, 13–31.

Byrne, S. & Girolami, M. (2013), 'Geodesic Monte Carlo on embedded manifolds', *Scandinavian Journal of Statistics* **40**, 825–845.

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

Fisher, N. I., Lewis, T. & Embleton, B. J. J. (1987), *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge.

Ganeiber, A. M. (2012), Estimation and simulation in directional and statistical shape models, PhD thesis, University of Leeds.

Green, P. J. & Mardia, K. V. (2006), 'Bayesian alignment using hierarchical models, with applications in protein bioinformatics', *Biometrika* **93**, 235–254.

Habeck, M. (2009), 'Generation of three-dimensional random rotations in fitting and matching problems', *Comput. Statist.* **24**, 719–731.

Hoff, P. D. (2009), 'Simulation of the matrix Bingham-von Mises-Fisher distribution with application to multivariate and relational data', *J. of Comput. and Graph. Stat.* **18**, 438–456.

Kendall, D. G. (1984), 'Shape-manifolds, procrusten matrices and complex projective spaces', *Bull. London Math. Soc.* **16**, 81–121.

Kent, J. T. (1982), 'The Fisher-Bingham distribution on the sphere', *Journal of the Royal Statistical Society, Series B* **44**, 71–80.

Kent, J. T. (1994), 'The complex Bingham distribution and shape analysis', *J. Roy. Statist. Soc. B* **56**, 285–299.

Kent, J. T. (2012), Statistical modelling and simulation using the Fisher-Bingham distribution, *in* T. Hamelryck, K. Mardia & J. Ferkinghoff-Borg, eds, 'Bayesian Methods in Structural Bioinformatics', Statistics for Biology and Health, Springer-Verlag, Berlin, pp. 179–188.

Kent, J. T. (2016), *simdd: Simulation of various directional distributions*, Department of Statistics, University of Leeds, Leeds LS2 9JT, UK. R package version 1.1.
URL: *http://www.maths.leeds.ac.uk/ john*

Kent, J. T., Constable, P. D. L. & Er, F. (2004), 'Simulation for the complex Bingham distribution', *Statistics and Computing* **14**, 53–57.

Kent, J. T. & Ganeiber, A. M. (2012), Simulation of random rotation matrices, *in* 'Proceedings of the 46th Scientific Meeting of the Italian Statistical Society'.

Kent, J. T., Ganeiber, A. M. & Mardia (2012), Simulation of the Bingham distribution in directional data analysis, *in* K. V. Mardia, A. Gusnanto, A. D. Riley & J. Voss, eds, 'New Statistics and Modern Natural Sciences', pp. 57–59.

Kent, J. T. & Hamelryck, T. (2005), Using the Fisher-Bingham distribution in stochastic models for protein structure, *in* S. Barber, P. D. Baxter, K. V. Mardia & R. E. Walls, eds, 'Proceedings in Quantitative Biology, Shape Analysis and Wavelets', Leeds University Press, pp. 57–60.

Kent, J. T., Hussein, I. & Jah, M. K. (2016), Directional distributions in tracking of space debris, *in* 'Proceedings of the 19th International Conference on Information Fusion (FUSION), Heidelberg, Germany', IEEE, pp. 2081–2086.

Kent, J. T., Mardia, K. V. & McDonnell, P. (2006), 'The complex Bingham quartic distribution and shape analysis', *J. Roy. Statist. Soc. B* **68**, 747–765.

Kume, A. & Walker, S. G. (2006), 'Sampling from compositional and directional distributions', *Stat. Comput.* **16**, 261–265.

Kume, A. & Walker, S. G. (2009), 'On the Fisher-Bingham distribution', *Stat. Comput.* **19**, 167–172.

Kume, A. & Wood, A. T. A. (2005), 'Saddlepoint approximations for the Bingham and FisherBingham normalising constants', *Biometrika* **92**, 465–476.

Magnus, J. R. & Neudecker, H. (1999), *Matrix Differential Calculus, with Applications in Statistics and Econometrics*, Wiley, Chichester.

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

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979), *Multivariate Analysis*, Academic Press, London.

Mardia, K. V., Taylor, C. C. & Subramaniam, G. K. (2007), 'Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data', *Biometrics* **63**, 505–512.

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Scealy, J. L. & Welsh, A. H. (2011), 'Regression for compositional data by using distributions defined on the hypersphere', *J. Roy. Statist. Soc. B* **73**, 351–375.

Ulrich, G. (1984), 'Computer generation of distributions on the m-sphere', *Appl. Statist.* **33**, 158–163.

Wood, A. T. A. (1987), 'The simulation of spherical distributions in the Fisher-Bingham family', *Communications in Statisics* **B16**, 885–898.

Wood, A. T. A. (1994), 'Simulation of the von Mises-Fisher distribution', *Communications in Statisics* **B23**, 157–164.