**The magic of score matching estimators and approximations for distributions on manifolds and some cutting edge applications to molecular biology**

Kanti V. Mardia
Department of Statistics, University of Leeds, Leeds LS2 9JT, U.K.;
Department of Statistics, University of Oxford, Oxford OX1 3LB, U.K.– k.v.mardia@leeds.ac.uk

**Abstract**

It is well known that one of the major problems for maximum likelihood estimation (MLE) in the well-established directional models is that the normalising constants can be difficult to evaluate. A new general method of "score matching estimation" (SME) will be presented on a compact oriented Riemannian manifold following Mardia et al. (2016). Important applications include von Mises-Fisher, Bingham and joint models on the sphere and related spaces. The estimator is found to be consistent and asymptotically normally distributed under mild regularity conditions. Further, it is easy to compute as a solution of a linear set of equations and requires no knowledge of the normalizing constant. Some examples will be given to demonstrate its good performance. To highlight its properties, in this paper we give another proof of the important result that the SME is the exact MLE for the multivariate normal distribution. Also in this paper, we show how we can approximate in general a given distribution to a member of the exponential family; we call this method "score matching approximation" (SMA). Research for such types of approximations is common, for example for the wrapped normal distribution, and we introduce here the SMA for the multivariate wrapped normal using a multivariate von Mises distribution. Practical examples from structural molecular biology related to protein and RNA will be presented in the talk.

**Keywords**: Exponential family; distributions on torus and sphere; Hyvärinen divergence; von Mises distributions; Riemannian manifold.

# 1 Introduction

We describe here the the Hyvärinen divergence and show how it provides the score matching estimators (SME) and the score matching approximations (SMA) on a Riemannian manifold. Let $f$ and $f^*$ be two probability densities on a Riemannian manifold $M$, defined with respect to the uniform measure $\mu$, where $f$ and $f^*$ are assumed to be everywhere nonzero and twice continuously differentiable. Following Mardia et al. (2016), we define the *Hyvärinen divergence* (Hyvärinen, 2005) between the two densities on $M$ in terms of an integrated gradient inner product for the log ratio,

$$\begin{aligned}\Phi(f;f^*) &= \frac{1}{2}\int_M \langle \log(f/f^*), \log(f/f^*)\rangle\, f^*(x)\mu(dx) \\ &= \frac{1}{2}\int_M \{\langle \log f, \log f\rangle - 2\langle \log f, \log f^*\rangle + \langle \log f^*, \log f^*\rangle\}\; f^*(x)\,\mu(dx). \qquad (1)\end{aligned}$$

Minimizing this divergence measure over $f$ has provided motivation for the SME in general and on manifolds in particular (see, for example, Mardia et al. 2016). Here we consider another problem of approximating a known distribution $f^*$ by another simpler distribution $f$ from an exponential family. We take $f$ to be in an exponential family with parameters $\pi$ and $f^*$ with parameters $\theta$ (which is not necessarily in the exponential family) and the aim is to connect $\pi$ and $\theta$. The simplest example is to approximate the concentration parameter of the wrapped normal by that of the von Mises distribution. In general, suppose $f(x;\pi)$ forms a canonical exponential family on a compact oriented Riemannian manifold $M$ with density

$$f(x) = f(x;\pi) \propto \exp\{\pi^T t(x)\} \tag{2}$$

with respect to the uniform measure $\mu(dx)$, where $\pi$ is an $m$-vector of natural parameters and $t(x) = (t_1(x), \ldots, t_m(x))^T$ is a vector of sufficient statistics. Then it can be shown that the optimal approximation from (1) is given by

$$\pi = W^{-1}d. \tag{3}$$

where $W = (w_{\ell_1\ell_2})$ and $d = (d_\ell)$ have elements (see, equation (7) for the inner product used here)

$$w_{\ell_1\ell_2} = E\{\langle t_{\ell_1}, t_{\ell_2}\rangle(x)\}, \quad d_\ell = -E\{\Delta_M t_\ell(x)\} \quad (\ell, \ell_1, \ell_2 = 1, \ldots, m). \tag{4}$$

This approximation will be called Score Matching Approximation (SMA). Here, we will be dealing with only the circle/torus case so the gradient in $W$ and the Laplace-Beltrami operator in the quantity $d$ are straightforward. The expectation in $W$ and $d$ are with respect to the true distribution $f^*$ and the exponential family $f$ provides the functions $t(x)$, e.g. for the SMA for the wrapped normal case, the expectation is with respect to the wrapped normal distribution, and $f$ has a von Mises density.
For this paper, let the manifold $M$ be the unit sphere $S_p = \{z \in \mathbb{R}^q : \sum z_j^2 = 1\}$, a $p$-dimensional manifold embedded in $\mathbb{R}^q$, $q = p + 1$. The eigenfunctions of $\Delta_M$ on $S_p$, $p \geq 1$, are known as the spherical harmonics. A spherical harmonic of degree $k \geq 0$ has eigenvalue $-\lambda_k$ where $\lambda_k = k(k + p - 1) = k(k + q - 2)$. The action of $\Delta_M$ on the linear and quadratic spherical harmonics, expressed in Euclidean coordinates, can be summarized as follows (for circle $p = 1, q = 2$) with $i \neq j \in \{1, \ldots, q\}$.

$$\Delta_M z_j = -\lambda_1 z_j, \Delta_M(z_i^2 - z_j^2) = -\lambda_2(z_i^2 - z_j^2), \Delta_M(z_i z_j) = -\lambda_2(z_i z_j). \tag{5}$$

Mardia et al. (2016) have given a full coverage of SME for a compact oriented Riemannian manifold and here we give complementary results, e.g. Section 2 gives another proof of the well known important result that the SME are the exact MLE for the multivariate normal distribution. Sections 3 – 6 deal with the proposed SMA with a particular focus on the multivariate wrapped normal distribution via the multivariate von Mises distribution introduced in Mardia et al. (2008). We end the paper with some conclusions.

## 2 The SME for the normal cases revisited

Let us assume that the random vector $x \in \mathbb{R}^p$ is distributed as $N(\mu, \Sigma)$; its density can be recast as

$$f(x) \propto \exp\left\{\sum_{\ell=1}^{m} \pi_\ell t_\ell(x)\right\}, x \in \mathbb{R}^p, \text{where} \tag{6}$$

$$\pi = (\sum_{j=1}^{p} \sigma^{ij}\mu_j, \ \ \sigma^{ii}, i = j;\ a_{ij} = \sigma^{ij}, i < j;\ i, j = 1, \ldots, p)^T,$$
$$t(x) = (x_i, \ \ -x_i^2/2; i = 1, \ldots, p \ \ , -x_i x_j, i < j;\ i, j = 1, \ldots, p)^T,$$

Here the vector $\pi$ denotes the $m = p + p + p(p-1)/2$ parameters in the order listed, and the vector $t = t(x)$ denotes the corresponding functions of $x$. Let

$$\begin{aligned} W_0(x) = (w_{0,\ell_1\ell_2}),\ w_{0\ell_1\ell_2} = \nabla_E t_{\ell_1}(x)^T\ \nabla_E t_{\ell_2}(x); \\ d_0(x) = (d_{0\ell}),\ d_{0\ell} = -\Delta_E t_\ell(x) \quad (\ell, \ell_1, \ell_2 = 1, \ldots, m), \end{aligned} \tag{7}$$

with the Euclidean gradient and the Laplacian given by

$$\nabla_E t_\ell(x) = (\partial t_\ell(x)/\partial x_1, \ldots, \partial t_\ell(x)/\partial x_p)^T, \quad \Delta_E t_\ell(x) = \sum_{i=1}^{p} \partial^2 t_\ell(x)/\partial x_i^2.$$

Let us assume the random sample from the normal distribution is $\{x_h,\ h = 1, \ldots, n\}$ then the score matching estimator $\hat{\pi}_{\text{SME}}$ is

$$\hat{\pi}_{\text{SME}} = \bar{W}_0^{-1}\ \bar{d}_0, \text{where} \tag{8}$$

$$\bar{W}_0 = \sum_{h=1}^{n} W_0(x_h)/n,\ \bar{d}_0 = \sum_{h=1}^{n} d_0(x_h)/n,$$

and $W_0(x)$ and $d_0(x)$ are given by (7).

**The univariate case.** Let us consider the univariate case of $x$ distributed as $N(\mu, \sigma^2)$ so in (6), we have $t_1(x) = x,\ t_2(x) = -x^2/2,\ \pi_1 = \mu/\sigma^2,\ \pi_2 = 1/\sigma^2$. Here, $\nabla t_\ell(x) = dt_\ell(x)/dx$ so that $\nabla t(x) = (1, -x)^T$ and

$$W_0(x) = \begin{bmatrix} 1 & -x \\ -x & x^2 \end{bmatrix}, \bar{W}_0(x) = \begin{bmatrix} 1 & -\bar{x} \\ -\bar{x} & S^2 + \bar{x}^2 \end{bmatrix},$$

where $\bar{x}$ and $S^2$ are respectively the sample mean and variance. Further, from $\Delta t_\ell(x) = d^2 t_\ell(x))/dx^2$, we get $\Delta t_1(x) = 0, \Delta t_2(x) = -1$ so $d_0(x) = (0, -1)^T = \bar{d}_0$ . Hence, we have from (8)

$$\hat{\pi} = \begin{bmatrix} \hat{\mu}/\hat{\sigma}^2 \\ 1/\hat{\sigma}^2 \end{bmatrix} = \bar{W}_0^{-1}\bar{d}_0 = \begin{bmatrix} \bar{x}/S^2 \\ 1/S^2 \end{bmatrix}$$

so that as expected, we obtain

$$\hat{\mu} = \bar{x},\ \hat{\sigma}^2 = S^2.$$

Thus we have shown that the SME is the exact MLE in this case.

**The multivariate case.** The corresponding result holds for the multivariate case but a direct proof is somewhat involved so we limit our attention here to the bivariate normal case. For this case, in (6), we have

$$t_1(x) = x_1, t_2(x) = x_2, t_3(x) = -x_1^2/2, t_4(x) = -x_2^2/2, t_5(x) = -x_1 x_2;$$
$$\pi_1 = \mu_1\sigma^{11} + \mu_2\sigma^{12}, \pi_2 = \mu_1\sigma^{12} + \mu_2\sigma^{22}, \pi_3 = \sigma^{11}, \pi_4 = \sigma^{22}, \pi_5 = \sigma^{12}.$$

We will show that the following result holds:

$$\bar{W}_0\hat{\pi}_{\text{MLE}} = \bar{d}_0 \tag{9}$$

and $\hat{\pi}_{\text{MLE}}$ is obtained from $\pi$ on plugging in $\hat{\mu} = \bar{x},\ \hat{\Sigma} = S$, where $\bar{x}$ and $S$ are respectively the sample mean mean vector and covariance matrix. Using $\nabla t_1(x) = (1, 0)^T,\ \nabla t_2(x) = (0, 1)^T,\ \nabla t_3(x) = (-x_1, 0)^T,\ \nabla t_4(x) = (0, -x_2)^T,\ \nabla t_5(x) = (-x_2. - x_1)^T$, we can construct $W_0(x)$ matrix, which leads to

$$\bar{W}_0(x) = \begin{bmatrix} 1 & 0 & -\bar{x}_1 & 0 & -\bar{x}_2 \\ 0 & 1 & 0 & -\bar{x}_2 & -\bar{x}_1 \\ -\bar{x}_1 & 0 & -S_{11} - \bar{x}_1^2 & 0 & S_{12} + \bar{x}_1\bar{x}_2 \\ 0 & -\bar{x}_2 & 0 & -S_{22} - \bar{x}_2^2 & S_{12} + \bar{x}_1\bar{x}_2 \\ -\bar{x}_2 & -\bar{x}_1 & S_{12} + \bar{x}_1\bar{x}_2 & S_{12} + \bar{x}_1\bar{x}_2 & S_{11} + S_{22} + \bar{x}_1^2 + \bar{x}_2^2 \end{bmatrix}. \tag{10}$$

Let $S^{-1} = (S^{ij})$, then we have

$$\hat{\pi}_{\text{MLE}} = (\bar{x}_1 S^{11} + \bar{x}_2 S^{12}, \bar{x}_1 S^{12} + \bar{x}_2 S^{22}, S^{11}, S^{22}, S^{12})^T. \tag{11}$$

Now, we calculate $\bar{W}_0^{-1}\hat{\pi}_{\text{MLE}}$ from (10)–(11).Then using the following relationships from $SS^{-1} = I$ into the resulting expression,

$$S_{11}S^{11} + S_{12}S^{12} = S_{22}S^{22} + S_{12}S^{12} = 1, S_{11}S^{12} + S_{12}S^{22} = S_{12}S^{11} + S_{22}S^{12} = 0,$$

we find that $\bar{W}_0^{-1}\hat{\pi}_{\text{MLE}} = (0, 0, 1, 1, 0)^T$. Further, $\Delta t_1(x) = 0,\ \Delta t_2(x) = 0,\ \Delta t_3(x) = -1,\ \Delta t_4(x) = -1,\ \Delta t_5(x) = 0$, so $d = (0, 0, 1, 1, 0)^T = \bar{d}_0$. Hence the proof.

It is to be noted that Hyvärinen (2005) has given a proof which avoids using the the canonical parametric representation of the normal distribution.

# 3 The multivariate von Mises distribution

The probability density function of $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \ldots, \theta_p)$ (w.r.t. the uniform measure) with zero mean vector of the multivariate von Mises (MvM) distribution (Mardia et al., 2008) is proportional to

$$\exp\{\boldsymbol{\kappa}^T c(\boldsymbol{\theta}) + \frac{1}{2} s(\boldsymbol{\theta})^T \Lambda\, s(\boldsymbol{\theta})\}, \tag{12}$$

where $-\pi < \theta_i \leq \pi,\ \kappa_i \geq 0,\ -\infty < \lambda_{ij} < \infty,\ \boldsymbol{\kappa}^T = (\kappa_1, \ldots, \kappa_p),\ (\boldsymbol{\Lambda})_{ij} = \lambda_{ij} = \lambda_{ji},\ i \neq j,\ \lambda_{ii} = 0$. Further, $c(\boldsymbol{\theta})^T = (\cos(\theta_1), \ldots, \cos(\theta_p))$, $s(\boldsymbol{\theta})^T = (\sin(\theta_1, \ldots, \sin(\theta_p))$, so the number of parameters are $m = p(p+1)/2$. Write

$$C_i = \cos\theta_i,\ S_i = \sin\theta_i,\ i = 1, \ldots, p.$$

We have $t(\theta)^T = (t_1(\theta)^T, t_2(\theta)^T)$ where $t_1(\theta)^T = (C_1, \ldots, C_p)$ and $t_2(\theta)^T = (S_1S_2, S_1S_3, \ldots, S_{p-1}S_k)$ in lexicographical order. It can be seen that the $E(d)$ and $E(W)$ in (3) for this case depend only on the following moments of the multivariate wrapped normal distribution.

$$E(C_i), E(S_iS_j), E(C_iS_iS_j), E(S_iC_jS_j), E(C_i^2S_j^2) + E(C_j^2S_i^2), E(C_i^2S_jS_k) \tag{13}$$

where $i, j, k = 1, \ldots, p,\ i \neq j \neq k$.

# 4 The SMA for univariate wrapped normal distribution

Using the von Mises distribution with the concentration parameter $\kappa$ as $f$ and the wrapped normal with the concentration parameter $\sigma^2$ as $f^*$ in equation (3), it is found that

$$\kappa = E(\cos\theta)/E(\sin^2\theta), \tag{14}$$

where the expectations are with respect to the wrapped normal distribution. This leads to the following SMA for the wrapped normal case:

$$\kappa = \frac{2c}{(1-c^4)}, \tag{15}$$

where $c = \exp\{-\frac{1}{2}\sigma^2\}$. The standard approximation (see, for example, Mardia and Jupp, 2000) is $\kappa = A^{-1}(c)$, where $A(c) = I_1(c)/I_0(c)$ and $I_p(\kappa)$ is the modified Bessel function of the first kind of $p$th order given by

$$I_p(\kappa) = \sum_{r=0}^{\infty} (\kappa)^{2r+p}/(r!(r+p)!).$$

On plotting the two approximations, we have found that the SMA values are always a bit smaller but there is hardly any difference for small or large $\sigma$. In fact, the maximum relative difference in the two functions is about 0.17 at $c$=0.74, $\kappa$= 2.29 and $\sigma^2$=0.60.

# 5 Moments for the multivariate wrapped normal distribution

Suppose $\theta^T = (\theta_1, \theta_2, \ldots, \theta_p)$ has a multivariate wrapped normal (MWN) distribution; that is,

$$\theta_j = x_j \mod 2\pi,\ j = 1, \ldots, p\ ,$$

where $x \sim N_p(0, \Sigma)$. One reason for choosing this distribution is that its trigonometric moments have explicit expressions. If $\delta$ is a $p \times 1$ vector with integer coefficients, then we can obtain the following from the joint characteristic function of $\theta$ (see, for example, Kent and Mardia, 2009; Kurz, 2015)

$$E\{\cos(\delta^T\theta)\} = \exp\{-\frac{1}{2}\delta^T\Sigma\delta\},\ E\{\sin(\delta^T\theta)\} = 0. \tag{16}$$

In particular, for $j, k = 1, \dots, p$,

$$E\{\cos(\theta_j)\} = \exp\{-\frac{1}{2}\sigma_{jj}\} = c_j, \text{ say, } E\{\sin(\theta_j)\} = 0,$$
$$E\{\cos(\theta_j \pm \theta_k)\} = \exp\{-\frac{1}{2}(\sigma_{jj} \pm 2\sigma_{jk} + \sigma_{kk})\}, \ E\{\sin(\theta_j \pm \theta_k)\} = 0.$$

Combining the two versions of the last two expressions yields

$$E\{\cos(\theta_j)\cos(\theta_k)\} = c_j c_k \cosh(\sigma_{jk}), E\{\sin(\theta_j)\sin(\theta_k)\} = c_j c_k \sinh(\sigma_{jk}), E\{\sin(\theta_j)\cos(\theta_k)\} = 0.$$

**Higher moments.** We will now obtain third order moments for the MWN distribution as it was pointed out that for the multivariate SMA, we only need the additional moment of the form $E(S_1 S_2 C_3^2)$ for the trivariate vector $(\theta_1, \theta_2, \theta_3)$. In fact, we have

$$E(S_1 S_2 C_3^2) = E(S_1 S_2)/2 + E(S_1 S_2 \cos(2\theta))/2 \tag{17}$$

so we need only the moment $E(S_1 S_2 \cos 2\theta)$ to obtain $E(S_1 S_2 \cos 2\theta)$. We have from the characteristic function (16),

$$E\{\cos(\theta_1 \pm \theta_2 + \theta_3)\} = c_1 c_2 c_3 \exp\{\mp\sigma_{12} \mp \sigma_{23} - \sigma_{13}\}.$$

and after some algebra, we get

$$\begin{aligned} E(S_1 S_2 \cos(2\theta)) &= \frac{1}{2}\ [c_1 c_2 c_3^4((\exp\{-\sigma_{23}\}(\sinh(\sigma_{12} + 2\sigma_{13}) \\ &+ \exp\{-\sigma_{13}\}\sinh(\sigma_{12} + 2\sigma_{23}) + \exp\{-\sigma_{12}\}\sinh 2(\sigma_{13} + \sigma_{23}))]. \end{aligned} \tag{18}$$

Hence, we can write down the expression for $E(S_1 S_2 C_3^2)$ noting $E\{\sin(\theta_1)\sin(\theta_2)\} = c_1 c_2 \sinh(\sigma_{12})$.

# 6 The SMA for the MWN case

First, we start with the bivariate case. Let us now write the two angles as $\theta$ and $\phi$. Further, let $c_\theta = \cos\theta, s_\theta = \sin\theta, c_\phi = \cos\phi, s_\phi = \sin\phi$. Then, we have

$$W = \begin{bmatrix} E(s_\theta^2) & 0 & -E(s_\theta c_\theta s_\phi) \\ 0 & E(s_\phi^2) & -E(s_\theta s_\phi c_\phi) \\ -E(s_\theta c_\theta s_\phi) & -E(s_\theta s_\phi c_\phi) & E(c_\theta^2 s_\phi^2) + E(s_\theta^2 c_\phi^2) \end{bmatrix}, \quad d = \begin{bmatrix} E(c_\theta) \\ E(c_\phi) \\ 2E(s_\theta s_\phi) \end{bmatrix}. \tag{19}$$

The parameters for the wrapped normal are $\sigma_{11}, \sigma_{12}, \sigma_{22}$ and will be linked to the parameters of the BvM $\kappa_1, \kappa_2, \lambda_{12}$. We need the the following moments of the wrapped normal distribution for the SMA:

$$E(c_\theta), E(c_\phi), E(s_\theta s_\phi), E(s_\theta^2), E(s_\phi^2), E(s_\theta c_\theta s_\phi), E(s_\theta s_\phi c_\phi), E(c_\theta^2 s_\phi^2) + E(s_\theta^2 c_\phi^2).$$

Let $W = (w_{ij})$ and $d^T = (d_1, d_2, d_3)$. From Section 3 using the joint characteristic function (3.1), we find that

$$d_1 = E(c_\theta) = \exp\{-\frac{1}{2}\sigma_{11}\} = c_1, \text{say}, d_2 = E(c_\phi) = \exp\{-\frac{1}{2}\sigma_{22}\} = c_2, \text{say}, d_3 = E(s_\theta s_\phi) = c_1 c_2 \sinh(\sigma_{12}). \tag{20}$$

Further, for the symmetric matrix $W = (w_{ij})$, the elements are given by

$$\begin{aligned} w_{11} = E(s_\theta^2) = (1 - c_1^4)/2, w_{22} = E(s_\phi^2) = (1 - c_2^4)/2, \\ w_{33} = E(c_\theta^2 s_\phi^2) + E(s_\theta^2 c_\phi^2) = (1 - c_1^4 c_2^4 \cosh(4\sigma_{12}))/2, w_{12} = 0, \\ w_{13} = -E(s_\theta c_\theta s_\phi) = -c_1^4 c_2 \sinh(2\sigma_{12})/2, w_{23} = -E(s_\theta s_\phi c_\phi) = -c_1 c_2^4 \sinh(2\sigma_{12})/2. \end{aligned} \tag{21}$$

Hence we can now substitute these in (3) to get the SMA for this case. Golden et al. (2017) have applied this approximation for a diffusion process for the wrapped normal case. Kurz et al. (2015) have presented other methods for parameter estimation of bivariate circular densities, the bivariate wrapped normal in particular, and it would be interesting to compare these different methods. Further, we can obtain algorithmically the SMA for the MWN as all moments are known in terms of the covariance matrix $\Sigma$ from Section 5.

# 7 Conclusions

The SME solves one of the challenging problems for directional distributions as it bypasses the normalising constants; in contrast, even for the univariate von Mises case (Section 4), we need to compute the Bessel functions for the MLE. For MvM, we simply need to solve a linear equation (namely of the form of equation (8)). The talk will give some cutting edge applications from structural molecular biology (Boomsma, et al. 2008; Mardia 2013) but this work also has potential for streaming data as SME is easy to compute in real time. The paper introduces for the first time a general method for approximating two distributions such as multivariate wrapped normal distribution by a multivariate von Mises distribution; in contrast, even for the univariate case, the standard approximation involves a ratio of two Bessel functions. There are other approaches for approximating distributions but these are not as unified or simplistic in comparison to the SMA method.

**Acknowledgments**
The author wishes to thank Eduardo García-Portugués, John Kent, Gerhard Kurz, and Arnab Laha for their helpful comments.

**References**

Boomsma, W., Mardia, K.V., Taylor, C.C., Ferkinghoff-Borg, J., Krogh, A. and Hamelryck, T. (2008) A generative, probabilistic model of local protein structure. *Proceedings of the National Academy of Sciences* **105**, 8932-8937.

Golden, M., García-Portugués, E., Sorensen, M., Mardia, K.V., Hamelryck, T. and Hein, J. (2017) A generative angular model of protein structure evolution. *Molecular Biology and Evolution.* In press.

Hyvärinen, A. (2005) Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research.* **6**:695-709.

Kent, J.T., and Mardia, K.V. (2009) Principal component analysis for the wrapped normal torus model. *LASR 2009 Proceedings*, A. Gusnanto, K.V. Mardia, and C.J. Fallaize (eds), Leeds University Press, 39-41.

Kurz, G. (2015) *Directional Estimation for Robotic Beating Heart Surgery.* Karlsruhe Institute of Technology, Intelligent Sensor-Actuator-Systems Laboratory.

Kurz, G., Gilitschenski, I., Dolbov, M. and Hanebeck, U.D. (2015) Bivariate angular estimation under consideration of dependencies using directional statistics. *Proceedings of the 53rd IEEE Conference on Decision and Control (CDC 2015)*, Los Angeles, California, USA.

Mardia, K.V., Hughes, G., Taylor, C.C. and Singh, H. (2008) A multivariate von Mises distribution with applications to bioinformatics. *Canadian Journal of Statistics.* **36**:1, 99-109.

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

Mardia, K. V. (2013) Statistical approaches to three key challenges in protein structural bioinformatics, *J. Roy. Statist. Soc. C*, **62**, 487-514.

Mardia, K.V., Kent, J.T and Laha, A.K. (2016) Score matching estimators for directional distributions. $http://arxiv.org/abs/1604.08470$