It thus remains only to calculate the determinant of the Fisher information matrix. We hope it causes no confusion if we adopt the termonological and notational conventions of referring to the determinant of the Fisher information matrix as the Fisher information, F. Since $h/\sqrt{det(F(.))}$ is invariant under parameter transformation and MML is likewise invariant under parameter transformation, we can discuss the issue of calculating F in several alternative ways.

4.5.2 The Fisher information

Recalling Sections 2 and 3, it follows quite clearly from the likelihood function that

$$\frac{\partial L}{\partial \kappa} = \frac{-N}{\kappa} + N \coth \kappa - \text{(terms independent of } \kappa) \tag{6}$$

Hence ([14, Page 245],[22, Theorem 3(c)],[9]),

$$E(\frac{\partial^2 L}{\partial \kappa^2}) = \frac{\partial^2 L}{\partial \kappa^2} = \frac{N}{\kappa^2} - \frac{N}{\sinh^2(\kappa)} \tag{7}$$

Regarding the Fisher information terms in the longitude (β), we firstly note that

$$\frac{\partial L}{\partial \beta} = -\kappa \sin\theta \sin\alpha \sin(\phi - \beta) \tag{8}$$

Hence,

$$e^{\kappa(\sin\theta \sin\alpha \cos(\phi-\beta))} \frac{\partial L}{\partial \beta} = \frac{1}{\kappa}\frac{\partial}{\partial \phi} e^{\kappa(\sin\theta \sin\alpha \cos(\phi-\beta))} \tag{9}$$

Although we are not directly interested in $\frac{\partial L}{\partial \beta}$, integrating over the longitude, ϕ, from 0 to 2π would give that $E(\frac{\partial L}{\partial \beta}) = 0$.

Arguing along these lines will give (c.f. [27, 6]) that

$$E(\frac{\partial^2 L}{\partial \beta \partial \kappa}) = E(\frac{\partial^2 L}{\partial \kappa \partial \beta}) = 0 \quad \text{and} \quad E(\frac{\partial^2 L}{\partial \beta \partial \alpha}) = E(\frac{\partial^2 L}{\partial \alpha \partial \beta}) = 0 \tag{10}$$

Hence, the Fisher information (i.e., the determinant of the Fisher information matrix), F, can be written under this parameterisation as

$$F = E(\frac{\partial^2 L}{\partial \beta^2}) \times (E(\frac{\partial^2 L}{\partial \kappa^2}) E(\frac{\partial^2 L}{\partial \alpha^2}) - (E(\frac{\partial^2 L}{\partial \alpha \partial \kappa})^2)) \tag{11}$$

Regarding one of the Fisher information terms in the co-latitude (α),

$$E(\partial^2 L/\partial \alpha^2) = N\kappa E(\sin\theta \sin\alpha \cos(\phi - \beta) + \cos\theta \cos\alpha) = N\kappa E(\mathbf{x}.\boldsymbol{\mu}) \tag{12}$$

since the second derivative with respect to α re-captures many of the terms in the log-likelihood function.

Defining $A(\kappa)$, or (in alternative notation, [9, Page 87][15, Page 123]) $\rho(\kappa)$, to be $E(\mathbf{x}.\boldsymbol{\mu})$, recalling our transformation in the Appendix with Jacobian $J = 1$ from the surface of a sphere to the surface of a cylinder, we thus have([14, Page 245], [22, 9]) that

$$\begin{aligned} E(\partial^2 L/\partial\alpha^2) &= N\kappa A(\kappa) \;=\; N\kappa E(\mathbf{x}.\boldsymbol{\mu}) \\ &= \frac{N\kappa^2}{2\pi\sinh(\kappa)} \int_0^{2\pi} d\theta\,(\frac{1}{2}\int_{-1}^{1} he^{\kappa h}dh) \\ &= \frac{N\kappa^2}{\sinh(\kappa)} \times \frac{1}{2}\int_{-1}^{1} he^{\kappa h}dh \\ &= \frac{N\kappa}{2\kappa\sinh(\kappa)}\int_{-\kappa}^{\kappa} ue^u du, \text{ letting } u = \kappa h, \\ &= N\kappa\left[\frac{1}{2\kappa\sinh(\kappa)}(u-1)e^u\right]_{-\kappa}^{\kappa} \\ &= N\kappa(\coth(\kappa) - \frac{1}{\kappa}) \end{aligned}$$

At this point, rather than proceeding with calculations for the remaining terms, $E(\partial^2 L/\partial\beta^2)$ and $E(\partial^2 L/\partial\kappa\partial\alpha)$, we cite a result[15, Page 124][1, Section 5.3.2] in this parameterisation and then present an alternative parameterisation. As we know from a general theorem[31, 12], $h/\sqrt{F}$ is invariant and so will be identical in both parameterisations.

It is known [1, Section 5.3.2] that under certain regularity conditions the determinant of the Fisher Information matrix approximates the asymptotic variance of the maximum likelihood estimator.

Mardia [15, Page 124] gives expressions for the asymptotic variance of the maximum likelihood parameter estimates, $\hat{\kappa}$, $\hat{\alpha}$ and $\hat{\beta}$:

$$var(\hat{\kappa}) = \frac{1}{NA'(\kappa)}$$

$$var(\hat{\alpha}) = \frac{1}{N\kappa A(\kappa)}$$

$$var(\hat{\beta}) = \frac{1}{N\kappa A(\kappa)\sin^2\alpha}$$

In addition, Mardia [15, Page 124] states that the maximum likelihood estimates are asymptotically independently Normal. Hence[1, Section 5.3.2],

$$det(F(\alpha,\ \beta,\ \kappa)) = N^3\kappa^2 A(\kappa)^2 A'(\kappa)\sin^2\alpha$$

giving us the Fisher information as desired.

Recalling our choice of $h_{\alpha,\,\beta,\,\kappa}(\alpha,\ \beta,\ \kappa) = \kappa^2\ \sin(\alpha)/(\pi^2\ (1+\kappa^2)^2)$ for this parameterisation gives $h/\sqrt{F} = \kappa/(\pi^2\ (1+\kappa^2)^2\sqrt{N^3 A'(\kappa)A(\kappa)^2})$.

4.5.3 The Fisher information and the prior distribution transformed

If we were to take a different parameterisation, where $h_{\mu}(\mu) = 1/4\pi$ is (locally) uniform on the surface of the sphere, then it can be argued in the general d-dimensional case that we get $F_{\mu,\kappa} = (N\kappa A(\kappa))^{d-1} \times N \frac{\partial^2 L}{\partial \kappa^2}$. (The form of this argument is similar to the one used to obtain MML estimates in multiple factor analysis[24], and is based on perturbing the parameter values near the MML estimate and on symmetry.)

With $d = 3$, this gives $F_{\mu,\kappa} = N^3 \kappa^2 A(\kappa)^2 \frac{\partial^2 L}{\partial \kappa^2} = N^3 \kappa^2 A(\kappa)^2 A'(\kappa)$.

With $h_{\mu,\kappa}(\mu, \kappa) = (1/4\pi) \times 4\,\kappa^2/(\pi\,(1+\kappa^2)^2) = \kappa^2/(\pi^2\,(1+\kappa^2)^2)$, we do indeed get again, as invariance results[31, 26] told us, that

$$h/\sqrt{F} = \kappa/(\pi^2\,(1+\kappa^2)^2 \sqrt{N^3 A'(\kappa) A(\kappa)^2})\,.$$

The expression for the message length, $MessLen$, which MML seeks to minimise, is given immediately before Equation (5).

5 The Kullback-Leibler distance

The Kullback-Leibler distance is a distance between two probability distributions based on information theory. Unlike the mean absolute error and the mean squared error (or mean bias, etc.), it is invariant under parameter transformation. We have used it in simulation tests for the von Mises circular distribution[27], although we have not used it in the experimental results presented in the next section.

Assuming the true distribution to be $f = f_3(\mu, \kappa)$ and an estimating distribution $g = \hat{f}$ to be given by $g = \hat{f} = f_3(\hat{\mu}, \hat{\kappa})$, the Kullback-Leibler distance is given by

$$d_{K-L}(f, g) = \int f \log(f/g) = \log\left(\frac{\kappa}{\hat{\kappa}} \frac{\sinh(\hat{\kappa})}{\sinh(\kappa)}\right) + (\kappa - \hat{\kappa})A(\kappa) + \hat{\kappa}(1 - \hat{\mu}.\mu)$$

for the spherical Fisher distribution.

6 Results

We tested the estimation techniques by running the following simulations: we generated N directions from a spherical von Mises-Fisher distribution with concentration parameter κ, and (without loss of generality) with co-latitude (α) and longitude (β) both set to 0. (κ varies as in the tables.) We then applied the estimation methods previously discussed, namely

1. the Maximum Likelihood estimator (MaxLik),
2. the Marginalised Maximum Likelihood estimator (Schou) [22, 11],
3. the MAP estimators in Cartesian Co-ordinates (defined in Expression (4) and Section 4.3.1) and Spherical Co-ordinates (defined in Expression (3)),
4. the MML estimator.

$\kappa = 0.00$	Max Lik	Schou	MAP (sph)	MAP (xyz)	MML
2 points					
MAE	22.43	10.64	1.97	0.32	0.36
MSE	3.4e4	8.5e3	6.31	0.12	0.15
5 points					
MAE	1.61	0.69	1.32	0.47	0.55
MSE	3.47	1.44	1.99	0.29	0.49
10 points					
MAE	0.96	0.38	1.00	0.47	0.51
MSE	1.14	0.45	1.08	0.29	0.35
20 points					
MAE	0.66	0.27	0.78	0.43	0.45
MSE	0.52	0.21	0.65	0.23	0.26
50 points					
MAE	0.40	0.15	0.53	0.32	0.33
MSE	0.19	0.07	0.30	0.13	0.13
100 points					
MAE	0.28	0.11	0.40	0.25	0.25
MSE	0.09	0.04	0.16	0.07	0.08

$\kappa = 0.50$	Max Lik	Schou	MAP (sph)	MAP (xyz)	MML
2 points					
MAE	63.07	31.24	1.58	0.18	0.16
MSE	1.8e6	4.4e5	6.65	0.05	0.04
5 points					
MAE	1.20	0.82	0.86	0.22	0.30
MSE	2.71	1.38	1.14	0.10	0.36
10 points					
MAE	0.62	0.54	0.58	0.22	0.26
MSE	0.62	0.41	0.45	0.10	0.14
20 points					
MAE	0.37	0.40	0.39	0.20	0.21
MSE	0.22	0.22	0.21	0.07	0.08
50 points					
MAE	0.21	0.25	0.22	0.16	0.17
MSE	0.07	0.10	0.07	0.04	0.04
100 points					
MAE	0.14	0.16	0.14	0.12	0.12
MSE	0.03	0.04	0.03	0.02	0.02

Table 1. Results for $\kappa = 0.00$ and $\kappa = 0.50$

Tables 1, 2 and 3 give mean absolute error (MAE) and mean squared error (MSE) of $\hat{\kappa}$ for each of the above estimators averaged over 1000 simulations.

7 Conclusions and Discussion

Maximum Likelihood is renowned for under-estimating parameters of scale, hence its tendency to under-estimate σ for a Normal distribution[17, 8] and its tendency[2, 22, 27] to over-estimate κ for the von Mises circular distribution (especially[27] for small N). The functional form of the $\text{MAP}_{x,y,z}$ prior, $h_{x,y,z}$, is to decrease monotonically in κ. It is little wonder that this estimator, whose objective function (the posterior) is the (normalised) product of its monotonically decreasing prior and the likelihood function, reliably out-performs κ_{MaxLik}. Related comments apply concerning MAP_{sphere}, whose prior decreases for $\kappa > 1$.

For the case of large κ (not simulated here), the Normal approximation comes into vogue for the invariant estimators (Maximum Likelihood (ML), marginalised ML and MML). An earlier theorem[27][22, Theorem 2(b)] carries over to the spherical case, and gives that, for $\kappa_{MML} > 1$, $\kappa_{MaxLik} > \kappa_{Schou} > \kappa_{MML} > 1$.

We note that the apparent inferiority of the Maximum Likelihood estimator based on the results presented for this problem is stark. Maximum Likelihood is typically the worst of all estimators considered.

The marginalised ML estimator is out-performed by the MML estimator and typically also by the two MAP estimators, except for sufficiently large N when

$\kappa = 1.00$	Max Lik	Schou	MAP (sph)	MAP (xyz)	MML
2 points					
MAE	23.12	11.39	1.15	0.66	0.61
MSE	3.0*e*4	7.5*e*3	3.81	0.45	0.40
5 points					
MAE	1.13	1.02	0.59	0.48	0.54
MSE	2.63	1.63	0.79	0.29	0.49
10 points					
MAE	0.60	0.64	0.40	0.40	0.41
MSE	0.64	0.62	0.32	0.22	0.26
20 points					
MAE	0.40	0.43	0.31	0.31	0.32
MSE	0.27	0.29	0.18	0.15	0.16
50 points					
MAE	0.22	0.23	0.19	0.21	0.21
MSE	0.08	0.08	0.06	0.06	0.07
100 points					
MAE	0.15	0.15	0.14	0.14	0.14
MSE	0.04	0.03	0.03	0.03	0.03

$\kappa = 2.00$	Max Lik	Schou	MAP (sph)	MAP (xyz)	MML
2 points					
MAE	27.14	13.41	0.95	1.62	1.57
MSE	3.4e4	8.7e3	2.92	2.63	2.48
5 points					
MAE	1.71	1.45	0.89	1.11	1.22
MSE	7.49	4.56	1.98	1.43	2.36
10 points					
MAE	0.83	0.76	0.60	0.74	0.75
MSE	1.42	1.10	0.71	0.76	0.88
20 points					
MAE	0.50	0.48	0.43	0.48	0.48
MSE	0.45	0.39	0.31	0.35	0.36
50 points					
MAE	0.30	0.29	0.28	0.30	0.30
MSE	0.14	0.13	0.12	0.13	0.13
100 points					
MAE	0.19	0.19	0.19	0.19	0.19
MSE	0.06	0.06	0.06	0.06	0.06

Table 2. Results for $\kappa = 1.00$ and $\kappa = 2.00$

$\kappa = 0$. This is because the marginalised ML estimator has an in-built preference for the value of $\kappa = 0$. If there is prior belief that $\kappa = 0$, then the prior used by the MML estimator could be correspondingly modified[27] to account for this, putting a point mass of the prior at $\kappa = 0$.

From a study of the results given in Tables 1, 2 and 3, we conclude that

- Firstly, the Bayesian methods for point estimation out-performed the classical point estimators (Maximum Likelihood and marginalised Maximum Likelihood), and very convincingly so for small N.
- Secondly, the MML estimator was competitive with the Bayesian MAP estimators. We found that typically the MML results were in between the results of MAP_{sph} and MAP_{xyz}; rarely was MML the worst of the three, and sometimes it was the best of the three.
- Thirdly, unlike the MAP estimator (and some other estimators), the MML scheme is invariant and avoids the issue of choice of parameterisation.
- Fourthly, the results using the MAP estimate in Cartesian co-ordinates were superior to the results using spherical co-ordinates (the "obvious" parameterisation) for small κ, and vice versa for large κ.

These encouraging results for MML in the spherical von Mises-Fisher case follow upon similar success for MML for the von Mises circular distribution[27, 28, 6].

The authors therefore advocate MML as the best of the methods considered, but note again that the Bayesian estimators outperformed the Classical methods.

$\kappa = 5.00$	Max Lik	Schou	MAP (sph)	MAP (xyz)	MML
2 points					
MAE	59.12	28.50	2.24	4.55	4.48
MSE	1.1e5	2.7e4	7.07	20.66	20.06
5 points					
MAE	3.99	2.86	2.08	2.84	2.36
MSE	45.05	24.29	11.40	9.00	12.29
10 points					
MAE	1.83	1.54	1.35	1.56	1.42
MSE	6.74	4.62	3.26	3.25	3.42
20 points					
MAE	1.11	1.00	0.94	0.99	0.96
MSE	2.25	1.80	1.49	1.41	1.51
50 points					
MAE	0.59	0.57	0.55	0.57	0.55
MSE	0.58	0.53	0.50	0.51	0.50
100 points					
MAE	0.42	0.42	0.41	0.42	0.41
MSE	0.28	0.27	0.26	0.27	0.26

$\kappa = 10.00$	Max Lik	Schou	MAP (sph)	MAP (xyz)	MML
2 points					
MAE	176.89	86.05	5.92	9.51	9.43
MSE	1.0e6	2.5e5	39.25	90.47	89.01
5 points					
MAE	7.33	5.35	4.16	6.21	4.31
MSE	129.71	67.59	32.51	41.24	33.34
10 points					
MAE	3.60	3.02	2.73	3.20	2.77
MSE	28.28	19.43	13.89	13.76	13.95
20 points					
MAE	2.22	2.03	1.92	2.04	1.93
MSE	9.12	7.39	6.23	6.11	6.24
50 points					
MAE	1.25	1.22	1.20	1.25	1.20
MSE	2.60	2.38	2.23	2.27	2.23
100 points					
MAE	0.84	0.82	0.81	0.83	0.81
MSE	1.12	1.07	1.03	1.05	1.03

Table 3. Results for $\kappa = 5.00$ and $\kappa = 10.0$

The mixture modelling of von Mises circular distributions [14, pp128–130] has also been addressed [29, 30, 4] by MML. The authors intend to extend the current work to the mixture modelling of spherical von Mises-Fisher distributions[20], with an eye to applications in for example, proteins, exploratory geological, stellar and micro-wave background radiation data. In particular, we wish to explore the question of whether the available data suggests that the universe has a preferred direction.

Acknowledgments

This work was supported by Australian Research Council (ARC) Grant A49330656 and ARC Postdoctoral Research Fellowship F39340111. The authors thank some anonymous referees for quite rightly challenging an earlier version of this paper. The authors also acknowledge amusing encouragement from a relevant rhyme by Rudyard Kipling[13].

Appendix - the constant in the likelihood function

It follows from our earlier expression [22, Page 369] for the likelihood function, $f_d(\kappa, \boldsymbol{\mu})$, that the function $I_{\frac{1}{2}}(\kappa)$ is given by

$$I_{\frac{1}{2}}(\kappa) = \frac{\kappa^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}} \int e^{\kappa\ \mathbf{x}.\boldsymbol{\mu}}\ d\mathbf{x}$$

where **x** ranges uniformly over the surface of the sphere. Transforming onto the surface of a cylinder (also with area $2 \times 2\pi = 4\pi$) with Jacobian $J = 1$, this then gives us

$$
\begin{aligned}
I_{\frac{1}{2}}(\kappa) &= \frac{\kappa^{\frac{1}{2}}}{\sqrt{2}\pi^{3/2}} \int_0^{2\pi} d\theta \, (\frac{1}{2} \int_{-1}^{1} e^{\kappa h} dh) \\
= \sqrt{\frac{2\kappa}{\pi}} \times \left[\frac{1}{2\kappa} e^{\kappa h}\right]_{h=-1}^{h=1} &= \sqrt{\frac{2}{\pi\kappa}} \times \frac{1}{2}(e^{\kappa} - e^{-\kappa}) \\
= \sqrt{\frac{2}{\pi\kappa}} \sinh(\kappa)
\end{aligned}
$$

References

1. J.M. Bernardo and A.F.M. Smith. *Bayesian Theory*. Wiley, New York, 1994.
2. D.J. Best and N.I. Fisher. The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. *Communications in Statistics (B) – Simulation and Computation*, B 10(5):493–502, 1981.
3. P. Cheeseman and J. Stutz. Bayesian classification (AUTOCLASS): Theory and results. In U.M. Fayyad, G. Piatetsky-Shapiro, P Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*. The AAAI Press, Menlo Park, 1995.
4. D.L. Dowe, L. Allison, T.I. Dix, L. Hunter, C.S. Wallace, and T. Edgoose. Circular clustering of protein dihedral angles by Minimum Message Length. In *Proceedings of the 1st Pacific Symposium on Biocomputing (PSB-1)*, pages 242–255, Hawaii, U.S.A., January 1996.
5. D.L. Dowe, R.A. Baxter, J.J. Oliver, and C.S. Wallace. Point Estimation using the Kullback-Leibler Loss Function and MML. In *Proceedings of the Kullback Memorial Research Conference*, Washington, D.C., U.S.A., May 1996.
6. D.L. Dowe, J.J. Oliver, R.A. Baxter, and C.S. Wallace. Bayesian estimation of the von Mises concentration parameter. In *Proc. 15th Maximum Entropy Conference, Santa Fe, New Mexico*, August 1995. Available on the WWW from `http://www.cs.monash.edu.au/~jono`.
7. D.L. Dowe, J.J. Oliver, and C.S. Wallace. MML estimation of the parameters of the spherical von Mises-Fisher distribution. Technical report TR no. 96/272, Dept. of Computer Science, Monash University, Clayton, Victoria 3168, Australia, 1996. Available on the WWW from `http://www.cs.monash.edu.au/~dld`.
8. D.L. Dowe and C.S. Wallace. Resolving the Neyman-Scott problem by Minimum Message Length. In *Proc. Sydney International Statistical Congess (SISC-96)*, pages 197–198, Sydney, Australia, 1996.
9. N.I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, 1993.
10. N.I. Fisher, T. Lewis, and B.J.J. Embleton. *Statistical Analysis of Spherical Data*. Cambridge University Press, Cambridge, 1987.
11. R.A. Fisher. Dispersion on a sphere. *Journal of the Royal Statistical Society (Series A)*, 217:295–305, 1953.
12. H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proc. of the Royal Soc. of London A*, 186:453–454, 1946.

13. R. Kipling. *Just so stories : for little children.* Macmillan, 1950.
14. K.V. Mardia. *Statistics of Directional Data.* Academic Press, 1972.
15. K.V. Mardia. Distribution theory for the von Mises-Fisher distribution and its application. In S. Kotz G. P. Patil and J. K. Ord, editors, *Statistical Distributions for Scientific Work*, pages 113–130. Reidel, Boston, 1975.
16. R.M. Neal. *Bayesian Learning for Neural Networks.* PhD thesis, Graduate Dept. of Computer Science, Univ. of Toronto, March 1995.
17. J. Neyman and E.L. Scott. Consistent estimates based on partially consistent observations. *Econometrika*, 16:1–32, 1948.
18. J. Neyman and E.L. Scott. A theory of the spatial distribution of the galaxies. *Astrophysics Journal*, 116:144–163, 1952.
19. J.J. Oliver and R.A. Baxter. MML and Bayesianism: Similarities and differences. Technical report TR 206, Dept. of Computer Science, Monash University, Clayton, Victoria 3168, Australia, 1995. Available on the WWW from `http://www.cs.monash.edu.au/~jono`.
20. J.J. Oliver and D.L. Dowe. Minimum Message Length Mixture Modelling of Spherical von Mises-Fisher distributions. In *Proc. Sydney International Statistical Congess (SISC-96)*, page 198, Sydney, Australia, 1996.
21. J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific, Singapore, 1989.
22. G. Schou. Estimation of the concentration parameter in von Mises-Fisher distributions. *Biometrika*, 65:369–377, 1978.
23. M.R. Spiegel. *Schaum's Outline Series: Theory and Problems of Vector Analysis.* McGraw-Hill Book Company, London, 1974.
24. C.S. Wallace. Multiple Factor Analysis by MML Estimation. Technical Report 95/218, Dept. of Computer Science, Monash University, Clayton, Victoria 3168, Australia, 1995.
25. C.S. Wallace and D.M. Boulton. An information measure for classification. *Computer Journal*, 11:185–194, 1968.
26. C.S. Wallace and D.M. Boulton. An invariant Bayes method for point estimation. *Classification Society Bulletin*, 3(3):11–34, 1975.
27. C.S. Wallace and D.L. Dowe. MML estimation of the von Mises concentration parameter. Technical report TR 93/193, Dept. of Computer Science, Monash University, Clayton, Victoria 3168, Australia, 1993. prov. accepted, Aust. J. Stat.
28. C.S. Wallace and D.L. Dowe. Estimation of the von Mises concentration parameter using Minimum Message Length. In *Proc. 12th Australian Statistical Soc. Conf.*, Monash University, Australia, 1994.
29. C.S. Wallace and D.L. Dowe. Intrinsic classification by MML – the Snob program. In C. Zhang, J. Debenham, and D Lukose, editors, *Proc. 7th Australian Joint Conf. on Artificial Intelligence*, pages 37–44. World Scientific, Singapore, 1994. See ftp://ftp.cs.monash.edu.au/pub/snob/Snob.README.
30. C.S. Wallace and D.L. Dowe. MML mixture modelling of Multi-state, Poisson, von Mises circular and Gaussian distributions. In *Proc. Sydney International Statistical Congess (SISC-96)*, page 197, Sydney, Australia, 1996.
31. C.S. Wallace and P.R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society (Series B)*, 49:240–252, 1987.
32. G.S. Watson and E.J. Williams. On the construction of significance tests on the circle and the sphere. *Biometrika*, 43:344–352, 1956.
33. S. Wolfram. *Mathematica: A System for Doing Mathematics by Computer.* Addison-Wesley, Reading, Massachusetts, 1988.