BAYESIAN ESTIMATION OF THE VON MISES CONCENTRATION PARAMETER

David L. Dowe, Jonathan J. Oliver, Rohan A. Baxter and Chris S. Wallace,
Department of Computer Science,
Monash University, Clayton, Vic. 3168, Australia
Email: {dld,jono,rohan,csw}@cs.monash.edu.au

ABSTRACT. The von Mises distribution is a maximum entropy distribution. It corresponds to the distribution of an angle of a compass needle in a uniform magnetic field of direction, μ, with concentration parameter, κ. The concentration parameter, κ, is the ratio of the field strength to the temperature of thermal fluctuations.

Previously, we obtained a Bayesian estimator for the von Mises distribution parameters using the information-theoretic Minimum Message Length (MML) principle. Here, we examine a variety of Bayesian estimation techniques by examining the posterior distribution in both polar and Cartesian co-ordinates. We compare the MML estimator with these fellow Bayesian techniques, and a range of Classical estimators. We find that the Bayesian estimators outperform the Classical estimators.

1. Introduction

The von Mises distribution, $M_2(\mu, \kappa)$, is a maximum entropy distribution. It corresponds to the distribution of an angle of a compass needle in a uniform magnetic field of direction, μ, with concentration parameter, κ. The concentration parameter, κ, is the ratio of the field strength to the temperature of thermal fluctuations. This distribution is a circular analogue of the Gaussian distribution, to which it converges for large κ and small σ.

Circular distributions and the von Mises distribution in particular are of interest in a wide range of fields, such as biology, geography, geology, geophysics, medicine, meteorology and oceanography[3], and protein dihedral angles[10].

We consider a range of estimators for the parameters of the von Mises distribution, both Classical and Bayesian, including a Bayesian estimator [9] obtained using the information-theoretic Minimum Message Length (MML) principle [11], and a variety of Bayesian estimators obtained by examining the posterior distribution in both polar and Cartesian co-ordinates. This work raises questions about the effect of parameterisations on Bayesian inference (e.g., [1, 4]). We find that the Bayesian estimators outperform the Classical estimators on a series of simulations.

2. The Likelihood Function

The von Mises distribution has probability density function $f(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(\theta-\mu)}$

where θ is in a range of 2π and $I_p(\kappa)$ is the modified Bessel function:

$$I_p(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \cos(p\theta) e^{\kappa\cos\theta} d\theta \quad \text{and so} \quad I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} e^{\kappa\cos\theta} d\theta \qquad (1)$$

For data $D = \{\theta_1,\ \theta_2,\ \ldots\ \theta_N\}$, the likelihood function and negative log-likelihood are:

$$p(D|\kappa,\mu) = \prod_{i=1}^{N} \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(\theta_i-\mu)} \quad \text{and} \quad L = N\log(2\pi I_0(\kappa)) - \kappa\sum_{i=1}^{N}\cos(\theta_i-\mu) \qquad (2)$$

3. Estimation Methods

We consider and compare several Bayesian and Classical approaches to estimating the concentration parameter κ of the von Mises distribution. We consider the Bayesian MAP estimate (the posterior mode) and the MML estimate. We consider the following Classical estimates: Maximum Likelihood, Marginalised Maximum Likelihood [7], and an estimator proposed by N.I. Fisher [3].

Whereas we were able to obtain an analytical closed form estimator of μ, all of our estimators for κ will have to be obtained by numerical methods.

3.1. Estimation of μ

To find the maximum likelihood estimator for μ, with $x = \sum_{i=1}^{N}\cos(\theta_i),\ y = \sum_{i=1}^{N}\sin(\theta_i)$, and $R = \sqrt{x^2+y^2}$ we have that $(\cos\hat{\mu},\ \sin\hat{\mu}) = (\frac{x}{R},\ \frac{y}{R})$. Because of the uniform nature of our prior on μ (see Section 4.1.), the MAP estimates for μ and the MML estimate for μ agree with the maximum likelihood estimator for μ.

4. Bayesian Estimators

We note that the Bayesian estimators require a prior distribution. We consider two prior distributions for κ to see if the Bayesian methods are sensitive to the choice of prior. We also note that the Bayesian MAP estimate requires a choice of parameterisation, an issue we discuss in Section 4.2..

4.1. Prior Distributions

We assume a prior $h_\mu(\mu) = \frac{1}{2\pi}$ on μ independent of κ. We consider two priors on κ [9]:

$$h_2(\kappa) = \frac{2}{\pi(1+\kappa^2)} \quad \text{and} \quad h_3(\kappa) = \frac{\kappa}{(1+\kappa^2)^{3/2}}$$

4.2. The MAP Estimate

It is generally known that the mode of the posterior (and hence the MAP estimate) is not invariant under non-linear parameter transformations [2]. Therefore, we consider these estimates in both polar co-ordinates (κ,μ), and Cartesian co-ordinates $(X,Y) = (\kappa.\cos(\mu),\kappa.\sin(\mu))$, since in some sense, both these representations can be considered "natural". Oliver and Baxter [6] illustrated the manner in which the mode moves given the h_3 prior and the data:

$$D = \{\theta_1,\ \theta_2,\ \ldots,\ \theta_{10}\} = \{279^o,\ 143^o,\ 307^o,\ 153^o,\ 35^o,\ 203^o,\ 325^o,\ 45^o,\ 20^o,\ 74^o\}$$

The MAP estimate in polar co-ordinates is $(\kappa = 0.53,\ \mu = 27.8)$, while the MAP estimate in Cartesian co-ordinates is $(x = 0.22,\ y = 0.12)$, which is equivalent to $(\kappa = 0.25,\ \mu = 27.8)$ as shown in Figure 1.

Since we are considering two priors and two parameterisations, we find that there are four MAP estimates for κ. These estimates are the values of κ which maximise the following expressions:

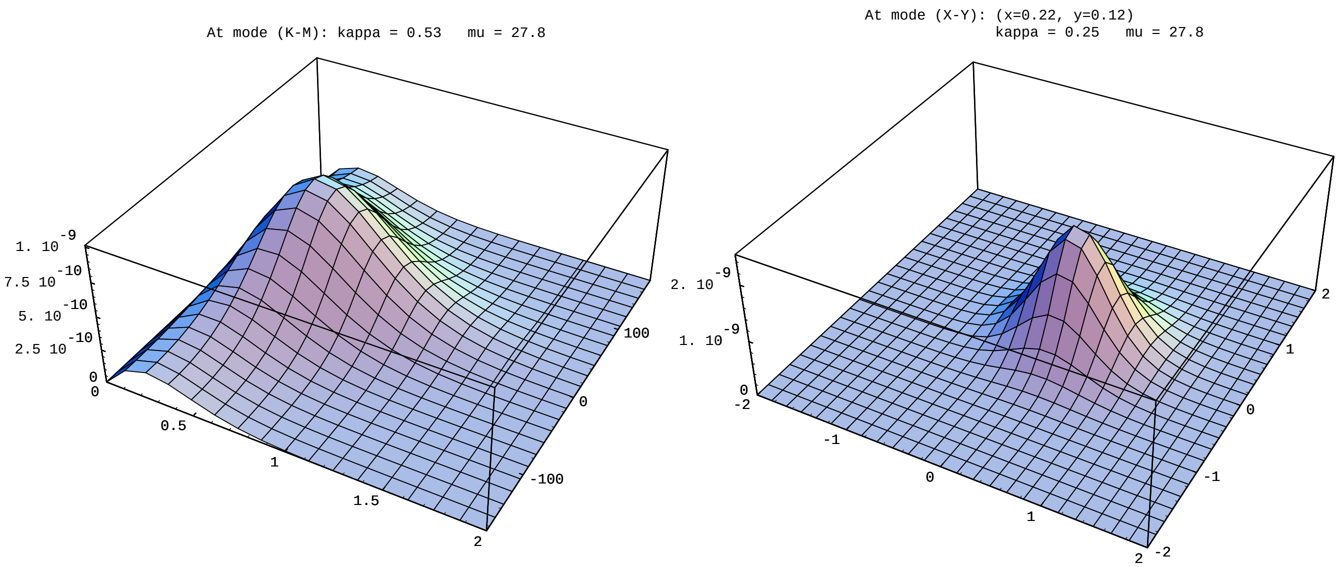

Figure 1: The Posterior Density in Polar and x-y Co-ordinates

$$\mathrm{MAP}_{h2}^{\kappa,\mu} = \max_{\kappa} \log(h_2(\kappa)) - N\log(2\pi I_0(\kappa)) + \kappa\sum_{i=1}^{N}\cos(\theta_i - \hat{\mu})$$

$$\mathrm{MAP}_{h3}^{\kappa,\mu} = \max_{\kappa} \log(h_3(\kappa)) - N\log(2\pi I_0(\kappa)) + \kappa\sum_{i=1}^{N}\cos(\theta_i - \hat{\mu})$$

$$\mathrm{MAP}_{h2}^{x,y} = \max_{\kappa} \log(\frac{h_2(\kappa)}{\kappa}) - N\log(2\pi I_0(\kappa)) + \kappa\sum_{i=1}^{N}\cos(\theta_i - \hat{\mu})$$

$$\mathrm{MAP}_{h3}^{x,y} = \max_{\kappa} \log(\frac{h_3(\kappa)}{\kappa}) - N\log(2\pi I_0(\kappa)) + \kappa\sum_{i=1}^{N}\cos(\theta_i - \hat{\mu})$$

We note that $\mathrm{MAP}_{h2}^{x,y}$ is not a sensible estimator of κ since $\frac{h_2(\kappa)}{\kappa} = \frac{2}{\kappa\pi(1+\kappa^2)}$ diverges as $\kappa \to 0$. Hence, $\mathrm{MAP}_{h2}^{x,y}$ always gives $\kappa = 0$ independent of the data.

5. The MML Estimate

MML is a Bayesian point estimation method proposed by Wallace et al. [11]. For the MML estimates, we use the prior distributions from Section 4.1.. Unlike the MAP estimate, we do not need to consider the parameterisation for the MML estimate, since the MML estimate is invariant under 1-1 differentiable transformations (see [8], [11, p245] or [6, Section 5.4]).

The MML estimate is the value of $(\kappa,\ \mu)$ which minimises the expression [11]:

$$MessLen(\mu\ \&\ \kappa\ \&\ D) \ = \ -log_2 \frac{h(\kappa,\ \mu)\ p(D|\kappa,\ \mu)}{\sqrt{det(F(\kappa,\ \mu))}} \ + \ \text{Constants},$$

where we interpret the term $\frac{1}{det(F(\kappa,\ \mu))}$ as a volume of uncertainty. Minimising the Message Length is then equivalent to maximising:

$$\text{Exp} \ = \ \frac{h(\kappa,\ \mu)\ p(D|\kappa,\ \mu)}{\sqrt{det(F(\kappa,\ \mu))}} \tag{3}$$

In Appendix 1, we derive the Fisher matrix for the von Mises distribution [9]:

$$F(\mu,\ \kappa) \ = \ \begin{bmatrix} E(\frac{\partial^2 L}{\partial \mu^2}) & E(\frac{\partial^2 L}{\partial \mu \partial \kappa}) \\ E(\frac{\partial^2 L}{\partial \kappa \partial \mu}) & E(\frac{\partial^2 L}{\partial \kappa^2}) \end{bmatrix} \ = \ \begin{bmatrix} \kappa N A(\kappa) & 0 \\ 0 & N A'(\kappa) \end{bmatrix}$$

where $A(\kappa)$ is defined to be (see Mardia [5]):

$$A(\kappa) \ = \ \frac{I_1(\kappa)}{I_0(\kappa)} \ = \ \frac{I_0'(\kappa)}{I_0(\kappa)} \tag{4}$$

The determinant of the Fisher matrix is therefore: $det(F(\mu, \kappa)) \ = \ \kappa N^2 A(\kappa) A'(\kappa)$

and thus the expression to be maximised is:

$$\text{Exp} \ = \ \frac{h(\kappa,\ \mu)\ p(D|\kappa,\ \mu)}{N\sqrt{\kappa A(\kappa) A'(\kappa)}} \tag{5}$$

We maximise this numerically. We note there are potential problems in doing this as $\kappa \to 0$, since $det(F(\mu, \kappa)) \to 0$ as $\kappa \to 0$. This problem is discussed in Appendix 2.

6. Results

We tested the estimation techniques by running the following simulations; we generated N angles from a von Mises distribution with concentration parameter κ, and with $\mu \ = \ 0$. We then applied the estimation methods previously discussed, namely (A) the Maximum Likelihood estimator (MaxLik), (B) the Marginalised Maximum Likelihood estimator (Schou) [7], (C) an estimator proposed by N.I. Fisher (NF) [3], (D) the MML estimator [9] with the h_2 and h_3 priors (MML_{h2} and MML_{h3}), and (E) the three sensible MAP estimators discussed ($\text{MAP}_{h2}^{\kappa,\mu}$, $\text{MAP}_{h3}^{\kappa,\mu}$ and $\text{MAP}_{h3}^{x,y}$).

Tables 1 and 2 give the mean absolute error (MAE), the mean squared error (MSE), and the mean Kullback-Leibler distance (MKL) for each of the above methods averaged over 1000 simulations. We did not present the NF estimator in Table 2, since it is defined to be the Maximum Likelihood estimator for $N > 15$.

7. Conclusions and Discussion

From the results given in Section 6, we draw the following conclusions. Firstly, the Bayesian methods for point estimation outperformed the classical point estimators, very convincingly for small N. An extensive comparison of the MML estimator and the Classical estimators was performed in [9].

Secondly, the Bayesian methods were not overly sensitive to the choice of prior. Since $h_3(\kappa) = 0$ at $\kappa = 0$, and $h_2(\kappa) > 0$ for $\kappa = 0$, the Bayesian results using the h_2 prior were superior to the results using the h_3 prior for small κ, and vice versa for large κ.

$N=2$				
κ		MAE	MSE	MKL
0.0	MaxLik	243.3	9.0×10^6	241.9
	Schou	121.6	2.2×10^6	120.7
	NF	24.43	9.0×10^4	23.86
	MML_{h2}	0.2762	0.1038	0.0256
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	0.5889	0.486	0.1152
	MML_{h3}	0.5068	0.3644	0.0873
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	1.0622	1.2614	0.2866
	$\mathrm{MAP}_{h3}^{x,y}$	0.3724	0.1827	0.0448
0.5	MaxLik	758.2	1.4×10^8	675.4
	Schou	379.2	3.6×10^7	337.4
	NF	75.83	1.4×10^6	67.16
	MML_{h2}	0.2097	0.0700	0.0599
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	0.3465	0.1605	0.1269
	MML_{h3}	0.2948	0.1146	0.1049
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	0.6046	0.506	0.2713
	$\mathrm{MAP}_{h3}^{x,y}$	0.1888	0.0547	0.0714
1.0	MaxLik	206.3	4.0×10^6	145.7
	Schou	103.3	1.0×10^6	72.55
	NF	20.70	40203	14.25
	MML_{h2}	0.6774	0.4846	0.1526
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	0.3727	0.2279	0.1615
	MML_{h3}	0.4125	0.2659	0.1534
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	0.3478	0.1627	0.2413
	$\mathrm{MAP}_{h3}^{x,y}$	0.5700	0.3642	0.1464
10.0	MaxLik	1263	4.5×10^7	96.03
	Schou	629.3	1.1×10^7	47.65
	NF	126.7	4.5×10^5	9.4915
	MML_{h2}	9.5083	90.41	1.1481
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	8.9035	79.28	0.8111
	MML_{h3}	9.0301	81.55	0.8706
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	8.4152	70.83	0.6289
	$\mathrm{MAP}_{h3}^{x,y}$	9.3715	87.83	1.0584

$N=5$				
κ		MAE	MSE	MKL
0.0	MaxLik	1.0712	2.2207	0.3599
	Schou	0.5426	1.2513	0.2002
	NF	0.5626	0.7443	0.1509
	MML_{h2}	0.3421	0.4092	0.0745
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	0.5947	0.5618	0.1239
	MML_{h3}	0.5784	0.6569	0.1303
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	0.8892	0.9488	0.212
	$\mathrm{MAP}_{h3}^{x,y}$	0.4599	0.3078	0.0727
0.5	MaxLik	0.8330	2.8412	0.3787
	Schou	0.7709	1.8765	0.2537
	NF	0.5696	0.7528	0.1674
	MML_{h2}	0.4247	0.5881	0.1173
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	0.3866	0.393	0.1387
	MML_{h3}	0.4034	0.6369	0.1510
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	0.4893	0.5215	0.2007
	$\mathrm{MAP}_{h3}^{x,y}$	0.2736	0.1416	0.086
1.0	MaxLik	1.1863	5.7859	0.4593
	Schou	1.1153	3.8451	0.3647
	NF	0.6635	1.2661	0.2047
	MML_{h2}	0.8235	1.3712	0.2101
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	0.5405	0.7788	0.1589
	MML_{h3}	0.6815	1.2755	0.2065
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	0.4445	0.744	0.1772
	$\mathrm{MAP}_{h3}^{x,y}$	0.4639	0.2875	0.1149
10.0	MaxLik	14.79	1126	0.6755
	Schou	11.28	683.7	0.5045
	NF	7.3595	1565	0.3161
	MML_{h2}	6.2488	149.7	0.2589
	$\mathrm{MAP}_{h2}^{\kappa,\mu}$	5.522	47.63	0.2204
	MML_{h3}	6.1960	149.2	0.2557
	$\mathrm{MAP}_{h3}^{\kappa,\mu}$	5.975	131.2	0.2108
	$\mathrm{MAP}_{h3}^{x,y}$	7.604	58.00	0.417

Table 1: Results for $N=2$ and $N=5$

Thirdly, the MML estimator was competitive with the Bayesian MAP estimators. For example, when using the h_3 prior for κ, we find that typically the results using MML_{h3} were in between the results of $\mathrm{MAP}_{h3}^{\kappa,\mu}$ and $\mathrm{MAP}_{h3}^{x,y}$; rarely was MML_{h3} the worst of the three, and sometimes it was the best of the three. In addition, the MML scheme avoids the issue of choice of parameterisation. The results using the MAP estimate in Cartesian coordinates were superior to the results using polar coordinates (the obvious parameterisation) for small κ, and vice versa for large κ.

The authors therefore advocate the MML method, but note again that the Bayesian estimators outperformed the Classical methods.

8. Acknowledgments

This work was supported by Australian Research Council (ARC) Postdoctoral Research Fellowship

$N=25$				
κ		MAE	MSE	MKL
0.0	MaxLik	0.3625	0.1705	0.0417
	Schou	0.1509	0.0867	0.0197
	MML$_{h2}$	0.1554	0.0503	0.0124
	MAP$^{\kappa,\mu}_{h2}$	0.3266	0.1412	0.0347
	MML$_{h3}$	0.3076	0.1242	0.0305
	MAP$^{\kappa,\mu}_{h3}$	0.4880	0.2571	0.0629
	MAP$^{x,y}_{h3}$	0.3056	0.1246	0.0306
1.0	MaxLik	0.3008	0.1523	0.0447
	Schou	0.3116	0.1613	0.0460
	MML$_{h2}$	0.3263	0.1638	0.0457
	MAP$^{\kappa,\mu}_{h2}$	0.2753	0.1316	0.0394
	MML$_{h3}$	0.2841	0.1272	0.0398
	MAP$^{\kappa,\mu}_{h3}$	0.2582	0.1242	0.0387
	MAP$^{x,y}_{h3}$	0.2664	0.1194	0.0372
10.0	MaxLik	2.5336	13.81	0.0479
	Schou	2.3778	11.93	0.0448
	MML$_{h2}$	2.2158	9.3569	0.0409
	MAP$^{\kappa,\mu}_{h2}$	2.3792	9.6396	0.0428
	MML$_{h3}$	2.2129	9.3462	0.0409
	MAP$^{\kappa,\mu}_{h3}$	2.3769	9.6302	0.0427
	MAP$^{x,y}_{h3}$	2.4508	9.0782	0.0443

$N=100$				
κ		MAE	MSE	MKL
0.0	MaxLik	0.1766	0.0390	0.00971
	Schou	0.0732	0.0174	0.00435
	MML$_{h2}$	0.0810	0.0127	0.00318
	MAP$^{\kappa,\mu}_{h2}$	0.1743	0.0381	0.0095
	MML$_{h3}$	0.1685	0.0356	0.00886
	MAP$^{\kappa,\mu}_{h3}$	0.2512	0.0667	0.0166
	MAP$^{x,y}_{h3}$	0.1664	0.0346	0.0086
1.0	MaxLik	0.1359	0.0293	0.0101
	Schou	0.1367	0.0295	0.0101
	MML$_{h2}$	0.1402	0.0308	0.0102
	MAP$^{\kappa,\mu}_{h2}$	0.1353	0.0296	0.0100
	MML$_{h3}$	0.1352	0.0287	0.0099
	MAP$^{\kappa,\mu}_{h3}$	0.1312	0.0285	0.0099
	MAP$^{x,y}_{h3}$	0.1353	0.0295	0.0100
10.0	MaxLik	1.1340	2.1041	0.0103
	Schou	1.1179	2.0167	0.0101
	MML$_{h2}$	1.1007	1.8976	0.0098
	MAP$^{\kappa,\mu}_{h2}$	1.1463	1.9549	0.0100
	MML$_{h3}$	1.1004	1.8970	0.0098
	MAP$^{\kappa,\mu}_{h3}$	1.1457	1.9521	0.0100
	MAP$^{x,y}_{h3}$	1.1404	1.9049	0.0100

Table 2: Results for $N = 25$ and $N = 100$

F39340111 and ARC Grant A49330656.

References

[1] J.A. Achcar and A.F.M. Smith. Aspects of reparameterization in approximate Bayesian inference. In *Bayesian and Likelihood Methods in Statistics and Econometrics*, pages 439–452. Elsevier Science, Amsterdam, 1993.

[2] J.O. Berger. *Statistical Decision Theory and Bayesian analysis.* Springer-Verlag, 1993.

[3] N.I. Fisher. *Statistical Analysis of Circular Data.* Cambridge University Press, 1993.

[4] J.A. Hills and A.F.M. Smith. Parameterization issues in Bayesian inference. In *Bayesian Statistics 4*, pages 227–246. Oxford University Press, London, 1992.

[5] K.V. Mardia. *Statistics of Directional Data.* Academic Press, 1972.

[6] 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, 1994. Available on the WWW from `http://www.cs.monash.edu.au/~jono`.

[7] G. Schou. Estimation of the concentration parameter in von Mises-Fisher distributions. *Biometrika*, 65:369–377, 1978.

[8] C.S. Wallace and D.M. Boulton. An invariant Bayes method for point estimation. *Classification Society Bulletin*, 3(3):11–34, 1975.

[9] C.S. Wallace and D.L. Dowe. MML estimation of the von Mises concentration parameter. Technical report TR 193, Dept. of Computer Science, Monash University, Clayton, Victoria 3168, Australia, 1993; submitted to Australian J. Statistics.

[10] C.S. Wallace and D.L. Dowe. Intrinsic classification by MML – the Snob program. In *Proc. of the 7th Australian Joint Conf. on Artificial Intelligence*, pages 37–44. World Scientific, 1994.

[11] 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.

9. Appendix 1 — Determining the Fisher Information Matrix

9.1. Simplifying the Negative Log-Likelihood

We can simplify the negative log-likelihood (Equation 2) by letting

$$x = \sum_{i=1}^{N} \cos(\theta_i) \qquad \text{and} \qquad y = \sum_{i=1}^{N} \sin(\theta_i) \qquad \text{giving}$$

$$x\cos\mu + y\sin\mu = \cos\mu \sum_{i=1}^{N} \cos\theta_i + \sin\mu \sum_{i=1}^{N} \sin\theta_i = \sum_{i=1}^{N} \cos(\theta_i - \mu) \qquad \text{and}$$

$$-x\sin\mu + y\cos\mu = -\sin\mu \sum_{i=1}^{N} \cos\theta_i + \cos\mu \sum_{i=1}^{N} \sin\theta_i = \sum_{i=1}^{N} \sin(\theta_i - \mu)$$

Thus Equation (2) becomes: $L = N\log(2\pi I_0(\kappa)) - \kappa\,(x\cos\mu + y\sin\mu)$

9.2. The Fisher Information Matrix

For the von Mises distribution, the Fisher matrix has the form: $F(\mu,\kappa) = \begin{bmatrix} E(\frac{\partial^2 L}{\partial \mu^2}) & E(\frac{\partial^2 L}{\partial \mu \partial \kappa}) \\ E(\frac{\partial^2 L}{\partial \kappa \partial \mu}) & E(\frac{\partial^2 L}{\partial \kappa^2}) \end{bmatrix}$

The first partial derivatives of the negative log-likelihood are:

$$\frac{\partial L}{\partial \mu} = -\kappa\,(-x\sin\mu + y\cos\mu) \qquad \text{and} \qquad \frac{\partial L}{\partial \kappa} = NA(\kappa) - (x\cos\mu + y\sin\mu)$$

where $A(\kappa) = \frac{I_0'(\kappa)}{I_0(\kappa)}$ was defined in Equation (4). We derive each entry of the Fisher in turn.

- For the first entry, we find the partial derivative:

$$\frac{\partial^2 L}{\partial \mu^2} = \kappa\,(x\cos\mu + y\sin\mu) = \kappa \sum_{i=1}^{N} \cos(\theta_i - \mu), \qquad \text{and writing } \phi_i = \theta_i - \mu \text{ gives}$$

$$E(\frac{\partial^2 L}{\partial \mu^2}) = \kappa \sum_{i=1}^{N} E(\cos\phi_i) = \kappa N E(\cos\phi_i) \tag{6}$$

We find $E(\cos\phi_i)$ by taking the expectation over every possible data value $0 \le \phi_i \le 2\pi$:

$$E(\cos\phi_i) = \frac{1}{2\pi I_0(\kappa)} \int_0^{2\pi} \cos\phi_i e^{\kappa\cos\phi_i} d\phi_i$$

Noting the definitions of the modified Bessel function (Equation 1) and $A(\kappa)$ (Equation 4):

$$E(\cos\phi_i) = \frac{1}{I_0(\kappa)} \left(\frac{1}{2\pi} \int_0^{2\pi} \cos\phi_i e^{\kappa\cos\phi_i} d\phi_i \right) = \frac{I_1(\kappa)}{I_0(\kappa)} = A(\kappa) \tag{7}$$

Substituting Equation (7) into (6) gives: $E(\frac{\partial^2 L}{\partial \mu^2}) = \kappa N A(\kappa)$

- For the second and third entries, we find the partial derivative

$$\frac{\partial^2 L}{\partial\mu\partial\kappa} = -(-x\sin\mu + y\cos\mu) = -\sum_{i=1}^{N}\sin(\theta_i - \mu) \qquad \text{and hence}$$

$$E(\frac{\partial^2 L}{\partial\mu\partial\kappa}) = -E(\sum_{i=1}^{N}\sin(\theta_i - \mu)) = -NE(\sin(\theta - \mu))$$

We find the expectation, $E(\sin(\theta-\mu))$, integrating over every possible data value $0 \leq \theta \leq 2\pi$:

$$E(\sin(\theta-\mu)) = \frac{1}{2\pi I_0(\kappa)}\int_0^{2\pi}\sin(\theta-\mu)e^{\kappa\cos(\theta-\mu)}d\theta = \frac{1}{2\pi I_0(\kappa)}\left[\frac{e^{\kappa\cos(\theta-\mu)}}{\kappa}\right]_0^{2\pi} = 0$$

Hence, $E(\frac{\partial^2 L}{\partial\mu\partial\kappa}) = 0$. Since $\frac{\partial^2 L}{\partial\kappa\partial\mu} = \frac{\partial^2 L}{\partial\mu\partial\kappa}$, $E(\frac{\partial^2 L}{\partial\kappa\partial\mu}) = 0$.

- For the fourth entry we find the partial derivative

$$\frac{\partial^2 L}{\partial\kappa^2} = NA'(\kappa) \qquad \text{and since } \frac{\partial^2 L}{\partial\kappa^2} \text{ does not depend on the data} \qquad E(\frac{\partial^2 L}{\partial\kappa^2}) = NA'(\kappa)$$

Thus the Fisher matrix is [9]: $F(\mu,\ \kappa) = \begin{bmatrix} \kappa NA(\kappa) & 0 \\ 0 & NA'(\kappa) \end{bmatrix}$

10. Appendix 2 — Evaluating the Message Length as $\kappa \to 0$

We overcame this problem using two techniques [9].

10.1. The h_3 Prior

We wish to evaluate the limit of Expression (3) as $\kappa \to 0$ for the h_3 prior. To evaluate this limit, Mardia[5] gave the following expansion for $A(\kappa)$ for small κ:

$$A(\kappa) = \frac{\kappa}{2}\left[1 - \frac{\kappa^2}{8} + \frac{\kappa^4}{48} + \ldots\right] \text{giving } \lim_{\kappa\to 0} A(\kappa) = 0, \quad \lim_{\kappa\to 0} A'(\kappa) = \frac{1}{2}, \quad \lim_{\kappa\to 0}\frac{A(\kappa)}{\kappa} = \frac{1}{2}$$

and thus

$$\lim_{\kappa\to 0}\text{Exp} = \lim_{\kappa\to 0} p(D|\kappa,\ \mu)\frac{\kappa}{(1+\kappa^2)^{3/2}N\sqrt{\kappa A(\kappa)A'(\kappa)}}$$

$$= p(D|\kappa=0,\ \mu)\frac{1}{N\sqrt{\frac{1}{2}\times\frac{1}{2}}} = p(D|\kappa=0,\ \mu)\frac{2}{N}$$

10.2. The h_2 Prior

For the h_2 prior, the limit of Expression (3) as $\kappa \to 0$ is infinite. However, Wallace and Dowe [9] offered a solution for evaluating the message length as $\kappa \to 0$. They noted that while the Fisher tells us that the uncertainty in μ tends to infinity, we would never actually encode μ to less precision than 2π. They therefore modify the expression to be maximised to:

$$\text{Exp2} = \frac{h_2(\kappa,\ \mu)\ p(D|\kappa,\ \mu)}{\sqrt{(\kappa A(\kappa) + \frac{3}{N\pi^2})A'(\kappa)}} \tag{8}$$