MONASH

UNIVERSITY

AUSTRALIA

MML Estimation of the von Mises Concentration Parameter

C.S. Wallace[1]
D.L. Dowe[2]

TECHNICAL REPORT NO. 93/193

December 1993

DEPARTMENT OF COMPUTER SCIENCE

MONASH UNIVERSITY,
CLAYTON, VICTORIA.
AUSTRALIA, 3168.

MML Estimation of the von Mises Concentration Parameter

C.S. Wallace[1]
D.L. Dowe[2]

TECHNICAL REPORT NO. 93/193

December 1993

1. Department of Computer Science, Monash University, Clayton, Victoria 3168 AUSTRALIA csw@cs.monash.edu.au
2. Department of Computer Science, Monash University, Clayton, Victoria 3168 AUSTRALIA dld@cs.monash.edu.au

MML Estimation of the von Mises Concentration Parameter

C.S. Wallace and D.L. Dowe

Department of Computer Science, Monash University,
Wellington Road, Clayton, Victoria, 3168, Australia.

Abstract

We apply the information-theoretic Minimum Message Length (MML) principle to the problem of estimating the concentration parameter, κ, of the von Mises distribution. We restrict ourselves here to the circular case, $d = 2$. Using various measures of estimator performance (such as mean squared error, etc.), we find that both the MML estimator derived here and the marginal maximum likelihood estimator of Schou (1978) reliably outperform Maximum Likelihood (ML) in estimating the parameter values used to generate simulated data. We further find that the MML estimator compares favourably to Schou's estimator, particularly for small sample sizes.

1. Introduction

Elementary information-theoretic coding results tell us that an event, E, of probability P(E) > 0 can be encoded by a (binary) message of length MessLen(E) = $-\log_2$ P(E) .

For a practical communication problem, this expression may have to be rounded up or down to the nearest integer but will remain non-increasing in probability. Such rounding effects are negligible for all relevant intents and will not be further discussed here.

In general problems of model selection and parameter estimation, given observed data, D, we seek an hypothesis, H, which optimally explains D. Assuming prior probabilities P(H) on hypothesis H, this can be regarded as a problem of maximising
P(D).P(H|D) = P(H & D) = P(H).P(D|H),
where the above equality follows by Bayes's theorem.

Alternatively, given the data D, we seek to minimise the message length
MessLen(H & D) = MessLen(H).MessLen(D|H).
This gives us a criterion for comparing two hypotheses H and H′ (one of which may indeed be the null theory asserting that there is no structure to the data). The method holds valid for continuous-valued data as well as discrete data since any finitely recorded measurements on continuous data can only be to a finite degree of precision.

Correspondingly, we can only state our estimates to a finite precision. In applying MML to problems of parameter estimation, the message length is minimised by stating our estimate to a degree of precision such that the marginal (information) cost of stating the estimate more accurately is equal to the (expected) marginal saving that would be obtained in stating the data.

For a 1-dimensional problem of estimating a single parameter, z, with prior density h, likelihood function p, and negative of log-likelihood function, $L = -\log p$, stating our estimates to precision δ gives an expected message length [WaFr87, p244]

$$\text{MessLen(H\&D)} = \text{MessLen(H)} + \text{MessLen(D|H)} = -\log(\delta h(z)) + \frac{1}{\delta}\int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} L(z+\Delta)\, d\Delta$$

$$= -\log(\delta h(z)) + \frac{1}{\delta}\int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} L(z) + \Delta \frac{\partial L}{\partial z} + \frac{\Delta^2}{2!}\frac{\partial^2 L}{\partial z^2} + \cdots \, d\Delta$$

$$\approx -\log h(z) - \log \delta + L(z) + \frac{\delta^2}{24} F(z)\,,$$

where $F(z) = E\left(\frac{\partial^2 L}{\partial z^2}\right)$ is the Fisher information.

More generally, with $k_1 = \frac{1}{12}$, $k_2 = \frac{5}{36\sqrt{3}} \approx 0.080188$ and the D-dimensional lattice constants, k_D, as given in [CoSl88, pp59-61], generalising the 1-dimensional case above to the D-dimensional problem of estimating $\underset{\sim}{z}$ gives [WaFr87]

$$\text{MessLen(H\&D)} = -\log h(\underset{\sim}{z}) - \log \delta + L(\underset{\sim}{z}) + \frac{D k_D \delta^{2/D}}{2} F(\underset{\sim}{z})^{1/D} \quad \text{to order } \delta^{2/D}.$$

Minimising the message length thus gives [WaFr87]

$0 = \frac{\partial(\text{MessLen})}{\partial \delta} = -\frac{1}{\delta} + k_D \delta^{(2/D)-1} F(\underset{\sim}{z})^{1/D}$, and so [WaFr87] $\delta = (k_D^D F(\underset{\sim}{z}))^{-\frac{1}{2}}$; and thus

$$\text{MessLen} = -\log h(\underset{\sim}{z}) + L(\underset{\sim}{z}) + \tfrac{1}{2}\log F(\underset{\sim}{z}) + \frac{D}{2}\log k_D + \frac{D}{2}$$

$$= -\log\left(\frac{h(\underset{\sim}{z})p(\underset{\sim}{z})}{\sqrt{F(\underset{\sim}{z})}}\right) + \frac{D}{2}(1 + \log k_D)\,.$$

(In this paper, in which we estimate the von Mises parameters μ and κ, we shall only require the case $D = 2$.) Stating the estimate to accuracy $\delta = 1/\sqrt{(k_D^D F(\underset{\sim}{z}))}$, the MML estimate thus seeks to maximise $\frac{h(\underset{\sim}{z})p(\underset{\sim}{z})}{\sqrt{k_D^D F(\underset{\sim}{z})}}$ or, equivalently, $\frac{h(\underset{\sim}{z})p(\underset{\sim}{z})}{\sqrt{F(\underset{\sim}{z})}}$.

$h(\underset{\sim}{z})$ is the prior density of the point estimate $\underset{\sim}{z}$, and $\frac{h(\underset{\sim}{z})}{\sqrt{k_D^D F(\underset{\sim}{z})}}$ approximates the prior probability of the estimate, $\underset{\sim}{z}$, being within an uncertainty region of volume $\frac{1}{\sqrt{k_D^D F(\underset{\sim}{z})}}$.

MML differs from Maximum Likelihood firstly in its use of a (Bayesian) prior distribution, $h(\underset{\sim}{z})$.

Unlike Bayesian methods such as the mean or the mode of the posterior, MML estimates are invariant under 1-1 differentiable transformations. (In fact, MML is an approximation to a more general, SMML method, which is invariant under countably measurable 1-1 transformations [WaFr87, Wall89b].)

A second difference between MML and Maximum Likelihood is the accuracy to which estimates are stated. The MML estimator acknowledges the square root of the Fisher information (multiplied by a power of the appropriate lattice constant) as a measure of uncertainty. Whereas Maximum Likelihood estimators maximise the likelihood function at a point, MML estimators acknowledge a region of uncertainty (of size proportional to $\frac{1}{\sqrt{F}}$) and maximise the average of the likelihood function over this region [WaFr87, p245].

Acknowledging the uncertainty in the statement of the estimates gives us a probability function (*not* a density function) for *discrete* theories. Stating the estimates to the appropriate accuracy, the MML estimates are those from the theory with the highest posterior probability.

The von Mises distribution in 2 dimensions on the circle, variously known as $M_2(\mu, \kappa)$ or $VM(\kappa, \mu)$, has probability density function

$$f(\theta) = \frac{1}{2\pi I_o(\kappa)} \cdot e^{\kappa . \cos(\theta - \mu)}, \quad \theta \varepsilon [-\pi, \pi) \quad (\text{or } \theta \varepsilon [0, 2\pi)\ ),$$ where the modified Bessel function

$$I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) . e^{\kappa . \cos\theta}\, d\theta = \sum_{r=0}^{\infty} \frac{(\frac{\kappa}{2})^{2r+p}}{\Gamma(p+r+1) . \Gamma(r+1)}, \text{ and so } I_0(\kappa) = \sum_{r=0}^{\infty} \frac{(\frac{\kappa}{2})^{2r}}{(r!)^2} .$$

Given data a set of angles $\{\theta_i : i = 1, 2, ..., N\}$ independently and identically distributed from $M_2(\mu, \kappa)$, the negative of the log-likelihood function is given by

$$L = -\log\left(\prod_{i=1}^{N} \frac{1}{2\pi I_0(\kappa)} \cdot e^{\kappa . \cos(\theta_i - \mu)} \right) = N . \log(2\pi I_0(\kappa)) - \kappa \sum_{i=1}^{N} \cos(\theta_i - \mu) \quad .$$

Letting $x = \sum_{i=1}^{N} \cos\theta_i$ and $y = \sum_{i=1}^{N} \sin\theta_i$, $\quad L = N . \log(2\pi I_0(\kappa)) - \kappa . (x . \cos\mu + y . \sin\mu)$.

Hence, x and y are jointly sufficient for μ and κ.

$\frac{\partial L}{\partial \mu} = -\kappa(-x . \sin\mu + y . \cos\mu)$ and so, equating $\frac{\partial L}{\partial \mu}$ to zero gives that the Maximum Likelihood estimate, $\hat{\mu}$ (or μ_{ML}) of μ, is given by

$$(\cos\hat{\mu}, \sin\hat{\mu}) = \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right), \quad x^2 + y^2 \neq 0 \quad .$$

For $x \neq 0$, we can write this as $\tan\hat{\mu} = \frac{y}{x}$.

Assuming the uniform prior distribution $h(\mu) = \frac{1}{2\pi}$ on μ, μ_{MML} can be shown fairly easily to equal $\hat{\mu}$. Estimation of μ is not conceptually difficult. It consists of taking unit vectors pointing in the sample directions $\{\theta_i : i = 1, 2, \ldots, N\}$, joining them head-to-tail, and then looking at the resultant direction. (In the case that $x^2 + y^2 = 0$ and the resultant sum of the vectors is $\underset{\sim}{0}$, we have no preferred choice of $\hat{\mu}$). Hence, unless we are concerned with issues of robustness and the discounting of outliers (e.g. [KoCh]), there is no disagreement on how best to estimate μ. Henceforth, we take $\hat{\mu} = \hat{\mu}_{ML}$.

2. Estimating the von Mises Concentration Parameter, κ

We now address the issue of estimating κ using MML. Various prior distributions on κ are discussed in Section 2.2 .

Let $\phi_i = \theta_i - \hat{\mu}$ for $i = 1, 2, \ldots, N$, and let $R = \sqrt{x^2 + y^2}$.

Note that $I_0'(\kappa) = \frac{d}{d\kappa}\left(\sum_{r=0}^{\infty} \frac{(\frac{\kappa}{2})^{2r}}{(r!)^2} \right) = \sum_{r=1}^{\infty} \frac{2r}{2} \cdot \frac{(\frac{\kappa}{2})^{2r-1}}{(r!)^2} = \sum_{r=0}^{\infty} \frac{(\frac{\kappa}{2})^{2r+1}}{(r+1)!r!} = I_1(\kappa)$

and also that $\left| \sum_{i=1}^{N} \cos \phi_i \right| = \sum_{i=1}^{N} \cos \phi_i = x \cos \hat{\mu} + y \sin \hat{\mu} = \frac{x^2}{R} + \frac{y^2}{R} = R$.

(Just as x and y are jointly sufficient for μ and κ, it follows from the fact that $L = N . \log(2\pi I_0(\kappa)) - \kappa . \sum_{i=1}^{N} \cos(\theta_i - \hat{\mu} + \hat{\mu} - \mu) = N . \log(2\pi I_0(\kappa)) - \kappa . R . \cos(\hat{\mu} - \mu)$ that R is sufficient for κ).

To obtain κ_{ML}, the Maximum Likelihood estimator of κ,

$$0 = \frac{\partial L}{\partial \kappa} = N . \frac{I_0'(\kappa)}{I_0(\kappa)} - (x \cos \mu + y \sin \mu) \Big|_{\mu = \hat{\mu}} = N . \frac{I_1(\kappa)}{I_0(\kappa)} - R \ .$$

Defining $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$ and $\bar{R} = \frac{R}{N}$,

$A(\kappa_{ML}) = \bar{R}$ and so $\kappa_{ML} = A^{-1}(\bar{R})$, where A^{-1} is the inverse function of A.

It follows from its definition that $A(\kappa)$ is the expected value of $\cos\theta$ under the $M_2(0, \kappa)$ distribution.

We also have [Mard, p288] that for large κ, $A(\kappa) = 1 - \frac{1}{2\kappa} - \frac{1}{8\kappa^2} + 0(\kappa^{-3})$

and for small κ, $A(\kappa) = \frac{\kappa}{2}\left\{ 1 - \frac{\kappa^2}{8} + \frac{\kappa^4}{48} + 0(\kappa^6) \right\}$.

Schou's estimator, κ_S, is defined [Scho, p370] as that value of κ which maximises the marginal density of R, and is given by [Scho,p371; UpFi,p230]

$$\kappa_S = 0 \qquad R^2 \le N$$

$$RA(R\kappa_S) = NA(\kappa_S) \qquad R^2 > N \ .$$

The uniqueness and existence of the solution for κ_S to the functional equation $RA(R\kappa_S) = NA(\kappa_S)$ is shown by Schou [Scho, p371, Theorem 1]. We note in passing that this functional equation has an interesting physical interpretation. Consider N independent pendula of unit length and unit mass free to swing in a vertical plane which are subject to random thermal fluctuations in a uniform (gravitational) field of strength κ. The expected total potential of the N bodies will be N times the expected value of $\cos\theta$ under the $M_2(0,\kappa)$ distribution - namely, $N.A(\kappa)$. An alternative view is that the centre of mass of the system has length $\frac{R}{N}$ and mass N, and is under a (gravitational) force of strength $R\kappa$. With the expected value of $\cos\theta$ under the $M_2(0, R\kappa)$ distribution being $A(R\kappa)$, the centre of mass has expected potential $N.\frac{R}{N}.A(R\kappa) = R.A(R\kappa)$. So, given that the resultant length is observed to be R, κ_S is the value of κ which balances the equation of expected potential energies.

Theorem 1: If the sample size $N = 2$ and the true von Mises concentration parameter $\kappa = 0$ (i.e., the true distribution is uniform about the circle), then $E(\kappa_{ML})$, the expected value of the Maximum Likelihood estimate, will be infinite.

Proof: For two observations θ_1 and θ_2 (s.t. $\theta_1 - \theta_2 \ne \pi \bmod 2\pi$),

$\hat{\mu} = \mu_{ML}$ is given as before and let $\phi_i = \theta_i - \hat{\mu}$. With $|\theta_1 - \theta_2| \le \pi$,

$$R = \sum_{i=1}^{2} \cos\phi_i = \cos\left(\theta_1 - \frac{\theta_1+\theta_2}{2}\right) + \cos\left(\theta_2 - \frac{\theta_1+\theta_2}{2}\right) = 2\cos\left(\frac{|\theta_1-\theta_2|}{2}\right).$$

From [Mard; p288 and pp297-8, Appendices 2.2-2.3],

$$A(\kappa) = 1 - \frac{1}{2\kappa} - \frac{1}{8\kappa^2} + 0(\kappa^{-3}) \qquad \text{for large } \kappa \text{, and}$$

$$A(\kappa) \le 1 - \frac{1}{2\kappa} \qquad \text{for } \kappa \ge 1\ ;$$

and so $A^{-1}(A(\kappa)) = \kappa \ge \dfrac{1}{2(1-A(\kappa))}$ for $\kappa \ge 1$ and $A(\kappa) \ge 0.45$.

For our problem,

$$\kappa_{ML} = A^{-1}(\bar{R}) = A^{-1}\left(\cos\left(\frac{|\theta_1-\theta_2|}{2}\right)\right) \ge A^{-1}\left(1 - \tfrac{1}{2}\left(\frac{\theta_1-\theta_2}{2}\right)^2\right) \ge \frac{1}{2.\frac{1}{2}}\left(\frac{\theta_1-\theta_2}{2}\right)^2 = \frac{4}{(\theta_1-\theta_2)^2} \ .$$

So, $$\mathrm{E}(\kappa_{ML} \mid \kappa = 0, N = 2) \ge \frac{1}{\pi^2}\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}}\int_{\frac{-\pi}{2}}^{\frac{\pi}{2}} \frac{4}{(\theta_1-\theta_2)^2}\, d\theta_1\, d\theta_2$$

$$= \frac{4}{\pi^2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left[\frac{1}{\theta_2-\theta_1}\right]_{\theta_1=\theta_2}^{\theta_1=\frac{\pi}{2}} + \left[\frac{1}{\theta_1-\theta_2}\right]_{\theta_1=\theta_2}^{\theta_1=\frac{-\pi}{2}} d\theta_2\ ,$$ which is infinite. Q.E.D.

Theorem 2:[Scho, p372, Theorem 2]:

Let κ_{ML} and κ_S denote respectively the maximum likelihood estimate and (Schou's) marginal maximum likelihood estimate of κ, based on a sample of $N > 1$ independent observations from the $M_d(\mu, \kappa)$ distribution. Let R be the resultant length. Then

a. $\kappa_{ML}(R)$ and $\kappa_S(R)$ are increasing functions of R which map $[0, N)$ onto $[0, \infty)$. Further $\kappa_{ML}(R)$ is differentiable in the whole interval, $\kappa_S(R)$ except for $R = \sqrt{N}$;

b. $\kappa_S(R) < \kappa_{ML}(R)$ if $0 < R < N$;

c. $\kappa_{ML}(R) = \frac{1}{2}(d-1)\left(\frac{N}{N-R}\right) - \frac{1}{4}(d-3) + O(N-R)$ as $R \to N$;

d. $\kappa_S(R) = \frac{1}{2}(d-1)\left(\frac{N-1}{N-R}\right) - \frac{1}{4}(d-3)\left(\frac{N-1/R}{N-1}\right) + O(N-R)$ as $R \to N$.

Combining Theorems 2(c) and (d) gives that $\frac{\kappa_S(R)}{\kappa_{ML}(R)} \to \frac{N-1}{N}$ as $R \to N$.

Hence, since the singularity in Theorem 1 only occurs when $\theta_1 \approx \theta_2$ and $R = 2\bar{R} \approx 2$, $\mathrm{E}(\kappa_S \mid \kappa = 0,\ N = 2) \approx \frac{1}{2}\mathrm{E}(\kappa_{ML} \mid \kappa = 0,\ N = 2) = \infty$.

N.I. Fisher's [Fish] estimator, κ_{NF}, given by

$$\kappa_{NF} = \begin{cases} \max(\kappa_{ML} - \dfrac{2}{N\kappa_{ML}},\ 0) & N \le 15,\ \kappa_{ML} < 2 \\ \dfrac{(N-1)^3 \kappa_{ML}}{N^3 + N} & N \le 15,\ \kappa_{ML} \ge 2 \\ \kappa_{ML} & N \ge 16 \end{cases}$$

clearly suffers this same divergence, with $\mathrm{E}(\kappa_{NF} \mid \kappa = 0,\ N = 2) \approx \frac{1}{10}\infty = \infty$.

Empirical tests show that with an appropriate choice of prior, such as $h_2(\kappa)$ or $h_3(\kappa)$ from Section 2.2, MML estimators do not suffer from this divergence.

As we recall from earlier comments and calculations, as well as using the (log-)likelihood function, the MML estimator also uses the Fisher information (and relates it to precision [WaFr87, p245]) and Bayesian priors. We return later to discuss the priors.

2.1 Fisher information

The Fisher information is the determinant of the Fisher information matrix, given by

$$F = \begin{vmatrix} \mathrm{E}\left(\dfrac{\partial^2 L}{\partial \mu^2}\right) & \mathrm{E}\left(\dfrac{\partial^2 L}{\partial \mu \partial \kappa}\right) \\ \mathrm{E}\left(\dfrac{\partial^2 L}{\partial \kappa \partial \mu}\right) & \mathrm{E}\left(\dfrac{\partial^2 L}{\partial \kappa^2}\right) \end{vmatrix}$$

From before, $L = N.\log(2\pi I_0(\kappa)) - \kappa(x\cos\mu + y\sin\mu)$ and $\frac{\partial L}{\partial \mu} = -\kappa(-x\sin\mu + y\cos\mu)$

So, $E\left(\frac{\partial^2 L}{\partial\kappa\partial\mu}\right) = E\left(\frac{\partial^2 L}{\partial\mu\partial\kappa}\right) = (-x\sin\mu + y\cos\mu)|_{\mu=\hat{\mu}} = -x\sin\hat{\mu} + y\cos\hat{\mu} = 0$, by symmetry.

$$\frac{\partial^2 L}{\partial\mu^2} = \kappa(x\cos\mu + y\sin\mu) = -(L - N.\log(2\pi I_0(\kappa))) = \kappa(x\cos\mu + y\sin\mu)$$
$$= \kappa\sum_{i=1}^{N}\cos(\theta_i - \mu)$$

So, $E\left(\frac{\partial^2 L}{\partial\mu^2}\right) = \kappa\sum_{i=1}^{N}\cos(\theta_i - \hat{\mu}) = \kappa R = \kappa\sum_{i=1}^{N}\cos\phi_i = \kappa N E(\cos\phi_i) = \kappa N\frac{I_1(\kappa)}{I_0(\kappa)} = \kappa N A(\kappa)$.

$\frac{\partial L}{\partial\kappa} = N.\frac{I_0'(\kappa)}{I_0(\kappa)} - (x\cos\mu + y\sin\mu) = N.A(\kappa) - (x\cos\mu + y\sin\mu)$, from before.

$\frac{\partial^2 L}{\partial\kappa^2} = N.A'(\kappa) = N.\left(1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)\right)$ [Mard, p288]

So, $F = E\left(\frac{\partial^2 L}{\partial\mu^2}\right).E\left(\frac{\partial^2 L}{\partial\kappa^2}\right) = \kappa N^2 A(\kappa).A'(\kappa) = \kappa N^2 A(\kappa).\left(1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)\right)$

We recall from Section 1 that the message length to be minimised is given (with $d = 2$) by

$\text{MessLen} = -\log\left(\frac{h(z)p(z)}{\sqrt{F(z)}}\right) + (1 + \log k_2)$, where $p(z) = e^{-L(z)}$ is the likelihood function and $F(z)$ is the Fisher information just discussed.

We assume a uniform prior distribution $h_\mu(\mu) = \frac{1}{2\pi}$ independent of the distribution of κ. We look at three different prior distributions h_κ on κ. As it should not cause any confusion, we notationally replace h_κ by h.

2.2 Prior distributions

We consider the MML estimators obtained from each of the three priors: $h_1(\kappa) = \frac{1}{\kappa}$, the Cauchy prior $h_2(\kappa) = \frac{2}{\pi(1+\kappa^2)}$, and $h_3(\kappa) = \frac{\kappa}{(1+\kappa^2)^{3/2}}$. We then compare these results with Maximum Likelihood, κ_{ML}, and Schou's estimator, κ_S. For large κ, $e^{\kappa.\cos\theta} \approx e^\kappa . e^{-\frac{\kappa\theta^2}{2}}$ for $-\pi \le \theta < \pi$ and so $M_2(0,\kappa) \approx N(0, \frac{1}{\kappa})$. With κ corresponding to $\frac{1}{\sigma^2}$, the uniform prior on $\ln\sigma$ corresponds to a $\frac{1}{\kappa}$ prior on κ. While the approximation is valid for large κ, for small κ the $h_1(\kappa) = \frac{1}{\kappa}$ prior puts a large (and unnormalisable) amount of weight on very similar distributions with $\kappa \approx 0$.

Where κ and μ respectively correspond to the strength and direction of a (magnetic or other) physical field, the Cartesian co-ordinates $(X,Y) = (\kappa \cos \mu,\ \kappa \sin \mu)$ correspond respectively to the $x-$ and $y-$components of the field strength.

The transformation $(\kappa, \mu) \rightarrow (X,Y) = (\kappa \cos \mu,\ \kappa \sin \mu)$ has Jacobian $J = \kappa$, and hence both the unnormalised prior $h_1(\kappa) = \frac{1}{\kappa}$ and the Cauchy prior $h_2(\kappa) = \frac{2}{\pi(1+\kappa^2)}$ give singularities at the origin in $(X,Y)-$space. The singularity of the $h_2(\kappa)$ prior in $(X,Y)-$space can be thought of as saying that if it is X and Y rather than κ and μ that one wants to measure, then the accuracy to which κ is measured can afford to decrease for very small κ. If it is X and Y that one wants to measure, then the prior $h_3(\kappa) = \frac{\kappa}{(1+\kappa^2)^{3/2}}$ is bounded and normalised with similar asymptotic behaviour to h_2 but does not diverge in $(X,Y)-$space. For the prior $h_3(\kappa)$, the MML estimate is (indeed) obtained by minimising $-\log h + L + \frac{1}{2}\log F$, and hence by equating $\frac{\partial L}{\partial \kappa} - \frac{\partial}{\partial \kappa}(\log h) + \frac{1}{2}\frac{\partial}{\partial \kappa}(\log F)$ to 0.

Substituting for F in the above, within a constant, the message length is
$L - \log h + \frac{1}{2}\log(\kappa A) + \frac{1}{2}\log(1 - \frac{A}{\kappa} - A^2)$, where $\frac{\partial L}{\partial \kappa} = NA(\kappa) - (x \cos \mu + y \sin \mu)$.

Since $\lim_{\kappa \to 0} \frac{A}{\kappa} = \frac{1}{2}$, $\lim_{\kappa \to 0} \frac{1}{2}\log\left(\frac{\kappa A}{h^2}\right)$ is finite for the $h_3(\kappa)$ prior but is infinite for the priors $h_1(\kappa)$ and $h_2(\kappa)$.

We recall that the Fisher information gives a measure of our certainty in the estimate. Since $\mathrm{E}\left(\frac{\partial^2 L}{\partial \kappa \partial \mu}\right) = 0$, MML gives that $k_2 \mathrm{E}\left(\frac{\partial^2 L}{\partial \mu^2}\right) = k_2 N \kappa A$ is our certainty in μ and that we should state our estimate of μ to precision $\frac{1}{\sqrt{k_2 \mathrm{E}\left(\frac{\partial^2 L}{\partial \mu^2}\right)}} = \frac{1}{\sqrt{k_2 N \kappa A}} \approx \sqrt{\frac{12}{N \kappa A}}$. We recall [WaFr87, p245] that the MML estimate remains invariant under transformation from (κ, μ) to (X,Y), and note both that the MML approximation assumes smooth priors and that the $h_1(\kappa)$ and $h_2(\kappa)$ priors diverge at the origin in (X,Y)-space.

For estimates near the origin, the divergence of h in Cartesian co-ordinates could cause $\frac{h}{k_2\sqrt{F}}$ to be an inaccurate approximation of the probability mass within the uncertainty region. By invariance, $\frac{h}{k_2\sqrt{F}}$ suffers the same divergence in polar co-ordinates - the problem for $h_1(\kappa)$ and $h_2(\kappa)$ being the divergence of $\frac{12}{\sqrt{N\kappa A}}$ as $\kappa \rightarrow 0$. Supposing that we will always wish to state μ to no less precision than 2π, for the h_1 and h_2 priors we modify the message length expression to be minimised to be (within a constant)

$$L - \log h + \tfrac{1}{2}\log(\kappa A + \frac{3}{\pi^2 N}) + \tfrac{1}{2}\log(1 - \frac{A}{\kappa} - A^2)$$

This gives $\lim_{\bar{R}\to 0} \kappa_{MML:h_1} = \lim_{\bar{R}\to 0} \kappa_{MML:h_2} = 0$ and, for $N = 2$, $\kappa_{MML:h_2} \approx 0.338\bar{R}$ for small $\bar{R}$.

For the h_3 prior, for the top (unadjusted) form of the message length, $\lim_{\bar{R}\to 0} \kappa_{MML:h_3} = 0$ and for $N = 2$, $\kappa_{MML:h_3} \approx 0.57\bar{R}$ for small $\bar{R}$.

Since N and $\bar{R}$ are sufficient to characterise a distribution, we consider the behaviour of the estimators for various $\bar{R}$.

As we compare the behaviour of the estimators in the limits as $\bar{R} \to 0$ and $\bar{R} \to 1$, we recall from Theorem 2 that $\kappa_S \approx \frac{(N-1)\kappa_{ML}}{N}$ as $\bar{R} \to 1$. We also find that the estimators fall into groups under these limits.

As $\bar{R} \to 0$, κ_{ML}, κ_{NF}, κ_S , $\kappa_{MML:h_1}$ $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ all tend to 0.

As $\bar{R} \to 1$ the asymptotic similarity of h_2 and h_3 discussed earlier will see that $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ asymptotically concur (as can be seen for $N = 16$, $\bar{R} \geq 0.75$). (As discussed earlier, this grouping will be valid for all N.) Another grouping, larger in value and slightly curious is, $\kappa_{MML:h_1}$ and κ_S (as can be seen for $N = 16$, $\bar{R} \geq 0.75$). The value κ_{ML} returned by Maximum Likelihood is larger still (as suggested by Theorem 2), and falls into a group of its own. Notice how closely κ_S and $\kappa_{MML:h_1}$ follow one another. (See Table 1 and Figure 1 in Section 2.3 .)

Also note from the factor of N in the log-likelihood term in the expression for the Message Length that if our prior is continuous and normalisable and the prior is zero on a set of at most measure zero (within a neighbourhood of the true value of κ), such as with $h_2(\kappa)$ or $h_3(\kappa)$, then as the amount of data tends to infinity, the MML estimates will all converge to the same value.

The prior $h_1(\kappa) = 1/\kappa$, which is uniform in log κ, is continuous. It can be normalised by restricting it to be 0 outside some range [L, U] (and by then scaling it). If the true κ is in [L,U], then, with sufficient data, the MML estimate will converge to κ. Empirical tests show that if we contravene the spirit of MML and permit h_1 to have a range of $[0, \infty)$, then this estimator still performs fairly well but it is overly prone to estimating $\kappa = 0$ for small values of N. e.g. When $N = 2$, κ is estimated to be 0 until $\bar{R} \approx$ 0.99990421. When $N = 3$, the threshold is at $\bar{R} \approx$ 0.915; and, for $N = 4$, at $\bar{R} \approx$ 0.807. (Thereafter, as $\bar{R} \to 1$, empirical tests show that $\kappa_{MML:h_1}$ approaches κ_S .)

2.3 Empirical Estimates of κ from $\bar{R}$

We have acknowledged earlier in Section 2 that $\bar{R}$ is sufficient for κ. We present here the various estimates of κ as a function of $\bar{R}$ for N=16. (See also Figure 1.)

$\bar{R}$	κ_{ML}	$\kappa_{MML:h1}$	$\kappa_{MML:h2}$	κ_S	$\kappa_{MML:h3}$
0.01	0.020001	0.000000	0.004453	0.000000	0.015240
0.05	0.100125	0.000000	0.022476	0.000000	0.076355
0.10	0.201008	0.000000	0.046355	0.000000	0.153706
0.15	0.303440	0.000000	0.073583	0.000000	0.233085
0.20	0.408277	0.000000	0.107747	0.000000	0.315603
0.25	0.516490	0.000000	0.157488	0.000000	0.402457
0.30	0.629215	0.000000	0.246409	0.439033	0.495050
0.35	0.747833	0.000000	0.389409	0.613547	0.594938
0.40	0.874080	0.000000	0.542040	0.763158	0.704001
0.45	1.010221	0.000000	0.694544	0.911444	0.824526
0.50	1.159320	0.000000	0.853530	1.067327	0.959410
0.55	1.325697	1.032584	1.025562	1.237005	1.112558
0.60	1.515739	1.265206	1.217641	1.427431	1.289562
0.65	1.739446	1.520119	1.439324	1.648462	1.499105
0.70	2.013628	1.819396	1.705464	1.916027	1.755430
0.75	2.369301	2.195912	2.042538	2.258977	2.084461
0.80	2.871287	2.712417	2.502989	2.737067	2.538307
0.85	3.680408	3.513565	3.209755	3.498384	3.239122
0.90	5.304689	5.045979	4.538331	5.015780	4.561753
0.91	5.852232	5.554469	4.974557	5.527357	4.996577
0.92	6.539389	6.192734	5.520734	6.169883	5.541288
0.93	7.425719	7.017353	6.226402	6.999305	6.245397
0.94	8.610342	8.122404	7.174046	8.108592	7.190892
0.95	10.271689	9.675733	8.510162	9.664998	8.524691
0.96	12.766781	12.011434	10.524670	12.003180	10.536538
0.97	16.928871	15.910475	13.894881	15.904270	13.903772
0.98	25.257906	23.716978	20.650692	23.711947	20.656354
0.99	50.253847	47.147710	40.948489	47.144911	40.950983

Table 1. Various estimates of κ as a function of $\bar{R}$ for N = 16

Estimates of kappa

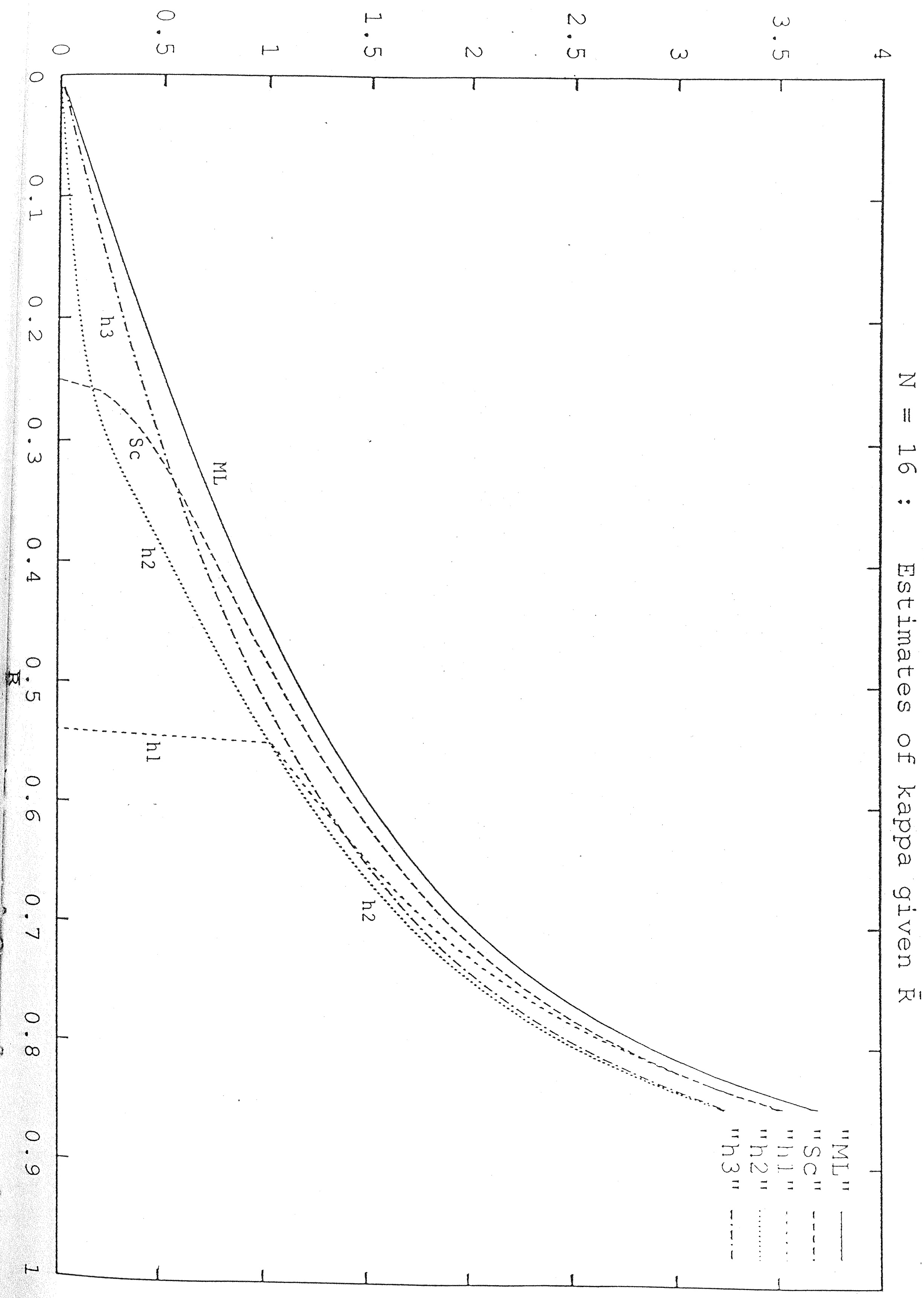

Figure 1

3. Discussion of Results and Conclusion

3.1 Discussion of Results

The reader is referred to the tabulated results in the Appendix. When N=2, we have empircial constant upper bounds on $\kappa_{MML:h_2}(\bar{R})$ and $\kappa_{MML:h_3}(\bar{R})$ as $\bar{R} \to 1$ contrasted with the divergence results following Theorem 1. The preferability of $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ over the other estimators under all four error criteria is thus not surprising in this case. When N=5, we still find that for all values of κ, $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ out-perform κ_{ML} under all criteria. (Indeed, apart from the occasional exception of the estimator $\kappa_{MML:h_1}$ for κ near 1.0 and 2.0, κ_{ML} is the least reliable estimator and, like $\kappa_{MML:h_1}$, need not be considered further here.) For N=5, $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ both out-perform κ_S and κ_{NF} in mean squared error for all test κ. Of the other criteria, we see a slight preference for κ_{NF} over one of $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ when κ=1.0 or κ=2.0, and likewise for κ_{NF} and κ_S over one of the κ_{MML} in the mean absolute error (m.a.e.) case when κ=0.0 or in the mean bias case when κ=0.0 or κ=0.25. On other tests when N=5, $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ are to be preferred. When N=10, $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ both out-perform κ_{ML}, κ_S and κ_{NF} in m.s.e. and mean Kullback-Leibler distance for all tested κ. Although the trends understandably weaken for large N, when N=10, 25 or 100, for each of m.a.e., m.s.e. and mean Kullback-Leibler distance, both $\kappa_{MML:h_2}$ and $\kappa_{MML:h_3}$ simultaneously out-perform both κ_{NF} and κ_S for at least four of our seven test values of κ. (We recall that $\kappa_{NF} = \kappa_{ML}$ for $N \geq 16$.)

The comparatively large standard deviations in our simulated data are due to the distributions of the estimated values of κ being correspondingly skewed. An alternative way of highlighting our results to the discussion immediately above would be to gather empirical statistics of differences between estimators (e.g. $\kappa_{MML:h_2} - \kappa_S$) or of differences between their Kullback-Leibler distances. Doing this takes advantage of the monotonicity in $\bar{R}$ of all estimators save κ_{NF}, and improves the statistical significance of the results discussed above.

3.2 Conclusion

The information-theoretic Minimum Message Length (MML) principle [WaBo68, WaFr87] (and, e.g., [BoWa, WaBo75]) has been successfully applied to several problems of inductive and statistical inference (e.g., [WaGe, Wall90, WaFr92, WaPa, OlWa]). MML is a tractable approximation to the efficient coding mechanism of Strict MML [WaFr87, Wall89b]. Applying MML to parameter estimation of the von Mises distribution has offered encouraging results in comparison with other published estimators using a variety of error criteria. The results are particularly encouraging when the sample size, N, is small.

Appendix: Simulation Results

We compare the various estimators by four error criteria from Monte Carlo simulations: mean bias (mb), mean absolute error (mae), mean squared error (mse) and mean Kullback-Leibler distance (KL). Values of N from 2,5,10,25,100 and 500 were chosen to generate the tables. (Thus the names of the tables.) The rows of the tables refer to the estimation techniques used (where h_1, h_2 and h_3 refer to MML priors discussed in Section 2.2). The columns of the table are headed by the value of κ used for the simulation, with the number of simulation runs in brackets above. Finally, the entry in a given row (estimator) and column (true κ) of a table is the mean of the relevant error criterion accompanied by its sample standard deviation in brackets.

These results were obtained using a pseudo-random number generator [Wall89a] of period $2^{32}(2^{32}-1)$.

$N=2$	(10000) $\kappa=0.0$	(1024) $\kappa=0.25$	(10000) $\kappa=0.50$	(10000) $\kappa=1.0$	(1024) $\kappa=2.0$	(10000) $\kappa=5.0$	(1024) $\kappa=10.0$
ML	12357 (549331)	273.4 (3346)	17091 (915407)	12003 (852831)	15139 (367697)	43578 (3.2e+06)	5234 (53712)
Sc	6178 (274665)	136.5 (1673)	8545 (457703)	6001 (426415)	7568 (183848)	21786 (1.6e+06)	2612 (26856)
h_1	2980 (188527)	118.6 (1671)	2599 (181300)	578.4 (25268)	592.6 (6504)	3958 (151778)	1108 (10931)
h_2	.280 (.164)	.0278 (.160)	-.206 (.163)	-.674 (.158)	-1.607 (.127)	-4.534 (.0653)	-9.506 (.0355)
h_3	.515 (.327)	.258 (.319)	.0425 (.327)	-.394 (.322)	-1.257 (.272)	-4.097 (.164)	-9.023 (.103)
NF	1235 (54933)	27.24 (334.6)	1708 (91540)	1199 (85283)	1512 (36769)	4353 (317053)	514.4 (5371)

N=2.Table2.mb

$N=5$	(1000000) $\kappa=0.0$	(1000000) $\kappa=0.25$	(1000000) $\kappa=0.50$	(1000000) $\kappa=1.0$	(102400) $\kappa=2.0$	(102400) $\kappa=5.0$	(102400) $\kappa=10.0$
ML	1.058 (1.809)	.870 (2.177)	.812 (3.215)	1.010 (3.760)	2.241 (9.744)	6.877 (22.80)	14.58 (53.97)
Sc	.523 (1.539)	.334 (1.828)	.269 (2.649)	.435 (3.109)	1.427 (7.827)	4.627 (18.23)	9.776 (43.18)
h_1	.186 (1.481)	-.0253 (1.785)	-.146 (2.638)	-.109 (3.195)	.939 (7.986)	4.648 (18.25)	9.835 (43.16)
h_2	.335 (.806)	.116 (.955)	-.0366 (1.378)	-.169 (1.686)	.0311 (4.050)	.589 (9.113)	.715 (21.53)
h_3	.568 (.823)	.353 (.968)	.211 (1.381)	.102 (1.661)	.279 (4.002)	.708 (9.083)	.783 (21.51)
NF	.548 (.985)	.340 (1.152)	.213 (1.635)	.124 (1.869)	.219 (4.769)	.852 (11.22)	2.100 (26.57)

N=5.Table2.mb

$N=10$	(10000) $\kappa=0.0$	(102400) $\kappa=0.25$	(102400) $\kappa=0.50$	(102400) $\kappa=1.0$	(102400) $\kappa=2.0$	(10000) $\kappa=5.0$	(10000) $\kappa=10.0$
ML	.619 (.374)	.419 (.404)	.318 (.493)	.299 (.740)	.616 (1.468)	1.977 (4.209)	4.153 (9.040)
Sc	.281 (.435)	.0887 (.477)	.0118 (.581)	.0658 (.791)	.393 (1.348)	1.345 (3.782)	2.795 (8.133)
h_1	.0348 (.237)	-.197 (.289)	-.376 (.450)	-.482 (.913)	.135 (1.587)	1.374 (3.782)	2.816 (8.126)
h_2	.235 (.267)	.0208 (.303)	-.117 (.404)	-.209 (.660)	-.0337 (1.167)	.234 (2.922)	.203 (6.291)
h_3	.436 (.287)	.225 (.314)	.0916 (.392)	-.0198 (.605)	.0659 (1.128)	.274 (2.912)	.227 (6.283)
NF	.336 (.401)	.141 (.435)	.0508 (.513)	.0269 (.631)	-.0020 (.0020)	.0367 (3.037)	.216 (6.525)

N=10.Table2.mb

$N=25$	(102400) $\kappa=0.0$	(102400) $\kappa=0.25$	(102400) $\kappa=0.50$	(102400) $\kappa=1.0$	(102400) $\kappa=2.0$	(102400) $\kappa=5.0$	(10000) $\kappa=10.0$
ML	.366 (.199)	.187 (.233)	.117 (.284)	.100 (.365)	.187 (.598)	.631 (1.680)	1.327 (3.481)
Sc	.160 (.243)	-.0040 (.296)	-.0288 (.360)	.0299 (.386)	.118 (.582)	.433 (1.610)	.897 (3.341)
h1	.0013 (.0381)	-.245 (.0800)	-.460 (.215)	-.516 (.658)	.0559 (.641)	.447 (1.609)	.903 (3.339)
h2	.159 (.163)	-.0298 (.213)	-.107 (.292)	-.0990 (.380)	-.0180 (.568)	.0862 (1.467)	.0652 (3.054)
h3	.311 (.174)	.123 (.205)	.0329 (.256)	-.0215 (.344)	.0111 (.560)	.100 (1.464)	.0734 (3.052)

N=25.Table2.mb

	(10000)	(1024)	(10000)	(10000)	(10000)	(10000)	(10000)
$N = 100$	$\kappa = 0.0$	$\kappa = 0.25$	$\kappa = 0.50$	$\kappa = 1.0$	$\kappa = 2.0$	$\kappa = 5.0$	$\kappa = 10.0$
ML	.179 (.0938)	.0504 (.132)	.0301 (.149)	.0233 (.172)	.0427 (.258)	.141 (.705)	.293 (1.449)
Sc	.0773 (.117)	-.018 (.171)	.0034 (.161)	.0088 (.173)	.0270 (.257)	.0964 (.698)	.195 (1.434)
h1	.0000 (.0000)	-.250 (.000)	-.500 (.0212)	-.521 (.569)	.0147 (.260)	.100 (.698)	.197 (1.434)
h2	.0847 (.0854)	-.039 (.147)	-.0320 (.162)	-.0224 (.173)	-.0043 (.256)	.0185 (.682)	.0064 (1.404)
h3	.171 (.0898)	.0373 (.126)	.0092 (.145)	-.0071 (.170)	.0021 (.255)	.0219 (.682)	.0083 (1.404)

N=100.Table2.mb

	(10000)	(10000)	(10000)	(1024)	(1024)	(1024)	(1024)
$N = 500$	$\kappa = 0.0$	$\kappa = 0.25$	$\kappa = 0.50$	$\kappa = 1.0$	$\kappa = 2.0$	$\kappa = 5.0$	$\kappa = 10.0$
ML	.0800 (.0413)	.0084 (.0636)	.0058 (.0656)	.0080 (.0752)	.0105 (.109)	.0398 (.302)	.0664 (.636)
Sc	.0349 (.0517)	-.0010 (.0673)	.0013 (.0661)	.0052 (.0753)	.0075 (.109)	.0311 (.301)	.0474 (.635)
h1	.0000 (.0000)	-.149 (.144)	-.0114 (.0679)	-.0010 (.0757)	.0050 (.109)	.0319 (.301)	.0476 (.635)
h2	.0390 (.0391)	-.0113 (.0691)	-.0056 (.0664)	-.0009 (.0753)	.0013 (.109)	.0159 (.300)	.0104 (.632)
h3	.0792 (.0409)	.0060 (.0631)	.0016 (.0652)	.0019 (.0750)	.0026 (.109)	.0165 (.300)	.0108 (.632)

N=500.Table2.mb

	(10000)	(1024)	(10000)	(10000)	(1024)	(10000)	(1024)
$N = 2$	$\kappa = 0.0$	$\kappa = 0.25$	$\kappa = 0.50$	$\kappa = 1.0$	$\kappa = 2.0$	$\kappa = 5.0$	$\kappa = 10.0$
ML	12357 (549331)	273.5 (3346)	17091 (915407)	12004 (852831)	15140 (367697)	43578 (3.2e+06)	5235 (53712)
Sc	6178 (274665)	136.7 (1673)	8545 (457703)	6002 (426415)	7569 (183848)	21788 (1.6e+06)	2614 (26855)
h_1	2980 (188527)	119.1 (1671)	2600 (181300)	580.4 (25268)	596.5 (6503)	3968 (151778)	1127 (10929)
h_2	.280 (.164)	.140 (.0820)	.210 (.158)	.674 (.158)	1.607 (.127)	4.534 (.0653)	9.506 (.0355)
h_3	.515 (.327)	.326 (.249)	.284 (.167)	.407 (.305)	1.257 (.272)	4.097 (.164)	9.023 (.103)
NF	1235 (54933)	27.39 (334.6)	1709 (91540)	1200 (85283)	1514 (36769)	4358 (317053)	523.5 (5370)

N=2.Table2.mae

	(1000000)	(1000000)	(1000000)	(1000000)	(102400)	(102400)	(102400)
$N = 5$	$\kappa = 0.0$	$\kappa = 0.25$	$\kappa = 0.50$	$\kappa = 1.0$	$\kappa = 2.0$	$\kappa = 5.0$	$\kappa = 10.0$
ML	1.058 (1.809)	.880 (2.173)	.875 (3.199)	1.230 (3.694)	2.633 (9.646)	7.656 (22.55)	16.18 (53.52)
Sc	.523 (1.539)	.626 (1.749)	.794 (2.542)	1.169 (2.913)	2.139 (7.663)	5.853 (17.88)	12.36 (42.51)
h_1	.186 (1.481)	.440 (1.730)	.751 (2.533)	1.429 (2.860)	2.620 (7.602)	5.918 (17.87)	12.35 (42.51)
h_2	.335 (.806)	.262 (.926)	.434 (1.308)	.843 (1.470)	1.513 (3.756)	3.177 (8.561)	6.753 (20.45)
h_3	.568 (.823)	.392 (.953)	.412 (1.335)	.704 (1.507)	1.359 (3.775)	3.099 (8.568)	6.700 (20.46)
NF	.548 (.985)	.527 (1.079)	.578 (1.545)	.687 (1.742)	1.273 (4.602)	3.833 (10.58)	7.987 (25.43)

N=5.Table2.mae

$N = 10$	(10000) $\kappa = 0.0$	(102400) $\kappa = 0.25$	(102400) $\kappa = 0.50$	(102400) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.619 (.374)	.440 (.382)	.416 (.414)	.529 (.597)	.964 (1.267)	2.749 (3.751)	5.750 (8.119)
Sc	.281 (.435)	.378 (.304)	.481 (.327)	.575 (.547)	.874 (1.099)	2.378 (3.234)	4.975 (7.015)
h_1	.0348 (.237)	.285 (.202)	.544 (.220)	.929 (.451)	1.097 (1.155)	2.386 (3.240)	4.969 (7.020)
h_2	.235 (.267)	.201 (.228)	.341 (.247)	.549 (.422)	.831 (.821)	1.865 (2.261)	3.973 (4.882)
h_3	.436 (.287)	.268 (.278)	.281 (.289)	.434 (.422)	.777 (.820)	1.849 (2.266)	3.959 (4.884)
NF	.336 (.401)	.344 (.302)	.416 (.304)	.490 (.399)	.630 (.777)	1.993 (2.292)	4.133 (5.054)

N=10.Table2.mae

$N = 25$	(102400) $\kappa = 0.0$	(102400) $\kappa = 0.25$	(102400) $\kappa = 0.50$	(102400) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(102400) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.366 (.199)	.227 (.194)	.238 (.195)	.289 (.243)	.457 (.428)	1.267 (1.271)	2.662 (2.607)
Sc	.160 (.243)	.258 (.144)	.298 (.204)	.297 (.248)	.440 (.399)	1.193 (1.165)	2.505 (2.386)
h1	.0013 (.0381)	.253 (.0460)	.506 (.0463)	.752 (.367)	.475 (.435)	1.193 (1.168)	2.503 (2.387)
h2	.159 (.163)	.177 (.121)	.264 (.164)	.312 (.238)	.437 (.362)	1.093 (.982)	2.309 (1.999)
h3	.311 (.174)	.178 (.159)	.202 (.162)	.271 (.213)	.428 (.362)	1.089 (.984)	2.307 (2.000)

N=25.Table2.mae

$N = 100$	(10000) $\kappa = 0.0$	(1024) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(10000) $\kappa = 1.0$	(10000) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.179 (.0938)	.1119 (.086)	.120 (.0925)	.137 (.106)	.205 (.163)	.558 (.454)	1.144 (.936)
Sc	.0773 (.117)	.1430 (.096)	.126 (.100)	.137 (.105)	.203 (.160)	.550 (.441)	1.126 (.910)
h1	.0000 (.0000)	.250 (.000)	.500 (.0014)	.641 (.428)	.206 (.160)	.550 (.441)	1.126 (.910)
h2	.0847 (.0854)	.1293 (.080)	.130 (.101)	.140 (.105)	.204 (.155)	.538 (.420)	1.104 (.868)
h3	.171 (.0898)	.1050 (.080)	.115 (.0879)	.135 (.103)	.202 (.156)	.538 (.420)	1.104 (.868)

N=100.Table2.mae

$N = 500$	(10000) $\kappa = 0.0$	(10000) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(1024) $\kappa = 1.0$	(1024) $\kappa = 2.0$	(1024) $\kappa = 5.0$	(1024) $\kappa = 10.0$
ML	.0800 (.0413)	.0512 (.0388)	.0527 (.0395)	.0605 (.0453)	.0855 (.0688)	.243 (.184)	.508 (.388)
Sc	.0349 (.0517)	.0532 (.0414)	.0529 (.0396)	.0604 (.0452)	.0853 (.0684)	.242 (.182)	.507 (.385)
h1	.0000 (.0000)	.183 (.0968)	.0551 (.0414)	.0608 (.0451)	.0855 (.0684)	.242 (.182)	.507 (.385)
h2	.0390 (.0391)	.0553 (.0429)	.0533 (.0400)	.0605 (.0449)	.0853 (.0680)	.241 (.179)	.504 (.381)
h3	.0792 (.0409)	.0505 (.0383)	.0522 (.0391)	.0602 (.0448)	.0852 (.0680)	.241 (.179)	.504 (.381)

N=500.Table2.mae

$N=2$	(10000) $\kappa = 0.0$	(1024) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(10000) $\kappa = 1.0$	(1024) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(1024) $\kappa = 10.0$
ML	3.0e+11 (1.8e+13)	1.1e+07 (1.9e+08)	8.4e+11 (7.0e+13)	7.3e+11 (6.8e+13)	1.4e+11 (4.1e+12)	1.0e+13 (1.0e+15)	2.9e+09 (4.8e+10)
Sc	7.5e+10 (4.6e+12)	2.8e+06 (4.6e+07)	2.1e+11 (1.8e+13)	1.8e+11 (1.7e+13)	3.4e+10 (1.0e+12)	2.5e+12 (2.5e+14)	7.3e+08 (1.2e+10)
h_1	3.6e+10 (2.8e+12)	2.8e+06 (4.6e+07)	3.3e+10 (3.2e+12)	6.4e+08 (5.3e+10)	4.3e+07 (8.5e+08)	2.3e+10 (1.5e+12)	1.2e+08 (2.9e+09)
h_2	.106 (.0931)	.0265 (.0234)	.0689 (.0744)	.479 (.225)	2.598 (.427)	20.57 (.599)	90.37 (.677)
h_3	.373 (.368)	.168 (.206)	.109 (.0994)	.259 (.288)	1.654 (.742)	16.82 (1.376)	81.43 (1.878)
NF	3.0e+09 (1.8e+11)	112695 (1.9e+06)	8.4e+09 (7.0e+11)	7.3e+09 (6.8e+11)	1.4e+09 (4.1e+10)	1.0e+11 (1.0e+13)	2.9e+07 (4.8e+08)

N=2.Table2.mse

$N=5$	(1000000) $\kappa = 0.0$	(1000000) $\kappa = 0.25$	(1000000) $\kappa = 0.50$	(1000000) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(102400) $\kappa = 5.0$	(102400) $\kappa = 10.0$
ML	4.393 (965.8)	5.497 (1570)	11.00 (4732)	15.16 (1135)	99.98 (8683)	567.0 (21437)	3125 (243336)
Sc	2.641 (618.3)	3.452 (1005)	7.091 (3028)	9.852 (726.3)	63.29 (5553)	353.8 (13698)	1959 (155626)
h_1	2.229 (617.9)	3.186 (1004)	6.978 (3027)	10.22 (726.0)	64.65 (5550)	354.5 (13692)	1959 (155616)
h_2	.762 (155.1)	.925 (251.7)	1.900 (756.3)	2.870 (181.6)	16.40 (1385)	83.39 (3396)	464.0 (38757)
h_3	1.000 (155.1)	1.061 (251.7)	1.951 (756.3)	2.768 (181.6)	16.09 (1385)	83.01 (3396)	463.4 (38756)
NF	1.270 (234.1)	1.442 (380.4)	2.720 (1146)	3.507 (274.0)	22.79 (2097)	126.6 (5158)	710.5 (58797)

N=5.Table2.mse

$N=10$	(10000) $\kappa = 0.0$	(102400) $\kappa = 0.25$	(102400) $\kappa = 0.50$	(102400) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.523 (.785)	.339 (.676)	.344 (1.187)	.636 (5.994)	2.535 (13.02)	21.63 (129.8)	98.97 (612.8)
Sc	.268 (.676)	.235 (.548)	.338 (.950)	.630 (4.831)	1.972 (10.28)	16.12 (102.8)	73.96 (485.6)
h_1	.0576 (.533)	.122 (.452)	.344 (.907)	1.066 (4.813)	2.537 (10.32)	16.19 (102.9)	73.97 (485.7)
h_2	.126 (.426)	.0921 (.345)	.177 (.608)	.479 (2.930)	1.364 (5.945)	8.593 (58.76)	39.61 (276.3)
h_3	.273 (.479)	.149 (.393)	.162 (.642)	.367 (2.944)	1.277 (5.976)	8.552 (58.78)	39.53 (276.3)
NF	.274 (.550)	.209 (.428)	.266 (.606)	.399 (3.005)	5.976 (6.026)	9.227 (62.20)	42.62 (294.6)

N=10.Table2.mse

$N=25$	(102400) $\kappa = 0.0$	(102400) $\kappa = 0.25$	(102400) $\kappa = 0.50$	(102400) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(102400) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.174 (.187)	.0893 (.150)	.0945 (.165)	.143 (.271)	.392 (.941)	3.220 (8.631)	13.88 (35.23)
Sc	.0849 (.169)	.0875 (.119)	.130 (.156)	.150 (.267)	.353 (.830)	2.781 (7.550)	11.97 (30.67)
h1	.0015 (.0430)	.0662 (.0568)	.258 (.0742)	.700 (.445)	.414 (.930)	2.789 (7.563)	11.96 (30.68)
h2	.0520 (.112)	.0461 (.0842)	.0968 (.113)	.154 (.232)	.323 (.669)	2.159 (5.693)	9.329 (22.74)
h3	.127 (.142)	.0573 (.107)	.0667 (.118)	.119 (.206)	.314 (.675)	2.154 (5.706)	9.318 (22.75)

N=25.Table2.mse

$N = 100$	(10000) $\kappa = 0.0$	(1024) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(10000) $\kappa = 1.0$	(10000) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.0409 (.0413)	.0199 (.030)	.0231 (.0337)	.0301 (.0457)	.0686 (.110)	.517 (.923)	2.185 (3.825)
Sc	.0196 (.0378)	.0297 (.032)	.0260 (.0408)	.0300 (.0447)	.0668 (.105)	.496 (.872)	2.095 (3.610)
h1	.0000 (.0000)	.0625 (.000)	.250 (.0016)	.595 (.478)	.0680 (.105)	.497 (.874)	2.095 (3.611)
h2	.0145 (.0289)	.0232 (.024)	.0273 (.0394)	.0306 (.0439)	.0656 (.0990)	.466 (.786)	1.972 (3.244)
h3	.0372 (.0378)	.0174 (.026)	.0210 (.0303)	.0289 (.0422)	.0652 (.0994)	.466 (.787)	1.971 (3.245)

N=100.Table2.mse

$N = 500$	(10000) $\kappa = 0.0$	(10000) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(1024) $\kappa = 1.0$	(1024) $\kappa = 2.0$	(1024) $\kappa = 5.0$	(1024) $\kappa = 10.0$
ML	.0081 (.0080)	.0041 (.0059)	.0043 (.0061)	.0057 (.0080)	.0120 (.0189)	.0929 (.133)	.409 (.588)
Sc	.0039 (.0073)	.0045 (.0070)	.0044 (.0061)	.0057 (.0079)	.0120 (.0187)	.0919 (.131)	.405 (.579)
h1	.0000 (.0000)	.0428 (.0279)	.0047 (.0067)	.0057 (.0079)	.0120 (.0186)	.0919 (.131)	.405 (.579)
h2	.0030 (.0059)	.0049 (.0072)	.0044 (.0062)	.0057 (.0078)	.0119 (.0184)	.0903 (.127)	.400 (.565)
h3	.0079 (.0078)	.0040 (.0057)	.0043 (.0060)	.0056 (.0078)	.0119 (.0184)	.0903 (.127)	.400 (.565)

N=500.Table2.mse

$N = 2$	(10000) $\kappa = 0.0$	(1024) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(10000) $\kappa = 1.0$	(1024) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(1024) $\kappa = 10.0$
ML	12356 (549331)	265.4 (3297)	15333 (768438)	7498 (491127)	5768 (125355)	11785 (997301)	392.2 (4520)
Sc	6177 (274665)	132.4 (1648)	7666 (384218)	3748 (245563)	2883 (62677)	5892 (498650)	195.7 (2260)
h_1	2980 (188527)	116.0 (1647)	2219 (147793)	428.0 (20207)	269.0 (3648)	618.1 (25522)	84.16 (989.9)
h_2	.0261 (.0229)	.0341 (.0315)	.0605 (.0508)	.151 (.0785)	.387 (.0910)	.824 (.0545)	1.147 (.0298)
h_3	.0893 (.0866)	.0892 (.0886)	.105 (.109)	.151 (.137)	.291 (.156)	.600 (.102)	.868 (.0623)
NF	1235 (54933)	26.23 (329.4)	1532 (76843)	749.5 (49112)	576.5 (12535)	1178 (99730)	39.04 (451.8)

N=2.Table2.KL

$N = 5$	(1000000) $\kappa = 0.0$	(1000000) $\kappa = 0.25$	(1000000) $\kappa = 0.50$	(1000000) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(102400) $\kappa = 5.0$	(102400) $\kappa = 10.0$
ML	.360 (1.656)	.369 (1.838)	.403 (2.574)	.495 (2.034)	.655 (2.996)	.716 (2.592)	.743 (3.103)
Sc	.202 (1.320)	.219 (1.464)	.271 (2.048)	.394 (1.597)	.522 (2.364)	.538 (2.034)	.557 (2.447)
h_1	.108 (1.316)	.134 (1.462)	.212 (2.049)	.426 (1.592)	.673 (2.349)	.558 (2.035)	.557 (2.447)
h_2	.0764 (.659)	.0886 (.730)	.125 (1.018)	.224 (.772)	.334 (1.133)	.287 (.945)	.283 (1.154)
h_3	.130 (.668)	.137 (.737)	.160 (1.024)	.220 (.781)	.295 (1.136)	.275 (.945)	.280 (1.154)
NF	.153 (.784)	.159 (.873)	.179 (1.232)	.222 (.928)	.286 (1.400)	.342 (1.180)	.347 (1.444)

N=5.Table2.KL

$N = 10$	(10000) $\kappa = 0.0$	(102400) $\kappa = 0.25$	(102400) $\kappa = 0.50$	(102400) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.121 (.151)	.121 (.147)	.124 (.161)	.134 (.212)	.160 (.272)	.173 (.305)	.174 (.325)
Sc	.0614 (.135)	.0717 (.126)	.0970 (.132)	.133 (.182)	.143 (.233)	.148 (.260)	.149 (.278)
h_1	.0120 (.0958)	.0295 (.0914)	.0789 (.108)	.203 (.162)	.215 (.277)	.149 (.260)	.149 (.278)
h_2	.0295 (.0847)	.0380 (.0798)	.0608 (.0902)	.110 (.133)	.132 (.172)	.115 (.179)	.115 (.190)
h_3	.0647 (.0973)	.0667 (.0941)	.0743 (.105)	.0958 (.139)	.121 (.168)	.114 (.179)	.115 (.190)
NF	.0639 (.119)	.0712 (.111)	.0883 (.107)	.106 (.130)	.101 (.162)	.122 (.182)	.121 (.198)

N=10.Table2.KL

$N = 25$	(102400) $\kappa = 0.0$	(102400) $\kappa = 0.25$	(102400) $\kappa = 0.50$	(102400) $\kappa = 1.0$	(102400) $\kappa = 2.0$	(102400) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.0425 (.0446)	.0426 (.0448)	.0427 (.0453)	.0440 (.0486)	.0470 (.0567)	.0490 (.0627)	.0494 (.0626)
Sc	.0207 (.0406)	.0302 (.0367)	.0433 (.0382)	.0448 (.0495)	.0449 (.0527)	.0459 (.0574)	.0462 (.0572)
h1	.0003 (.0100)	.0164 (.0141)	.0618 (.0164)	.151 (.0852)	.0526 (.0790)	.0460 (.0575)	.0462 (.0573)
h2	.0128 (.0271)	.0205 (.0263)	.0344 (.0296)	.0447 (.0451)	.0441 (.0480)	.0417 (.0486)	.0421 (.0481)
h3	.0312 (.0342)	.0320 (.0348)	.0339 (.0362)	.0390 (.0410)	.0431 (.0476)	.0416 (.0486)	.0421 (.0481)

N=25.Table2.KL

$N = 100$	(10000) $\kappa = 0.25$	(1024) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(10000) $\kappa = 1.0$	(10000) $\kappa = 2.0$	(10000) $\kappa = 5.0$	(10000) $\kappa = 10.0$
ML	.0102 (.0102)	.0101 (.010)	.0103 (.0103)	.0104 (.0107)	.0104 (.0109)	.0107 (.0114)	.0104 (.0112)
Sc	.0049 (.0094)	.0105 (.009)	.0107 (.0112)	.0103 (.0106)	.0103 (.0107)	.0106 (.0111)	.0103 (.0109)
h1	.0000 (.0000)	.0154 (.000)	.0597 (.0002)	.127 (.0989)	.0104 (.0107)	.0106 (.0111)	.0102 (.0109)
h2	.0036 (.0072)	.0088 (.008)	.0107 (.0107)	.0104 (.0105)	.0103 (.0104)	.0103 (.0106)	.0100 (.0104)
h3	.0093 (.0094)	.0093 (.010)	.0097 (.0097)	.0101 (.0103)	.0102 (.0104)	.0103 (.0106)	.0100 (.0104)

N=100.Table2.KL

$N = 500$	(10000) $\kappa = 0.0$	(10000) $\kappa = 0.25$	(10000) $\kappa = 0.50$	(1024) $\kappa = 1.0$	(1024) $\kappa = 2.0$	(1024) $\kappa = 5.0$	(1024) $\kappa = 10.0$
ML	.0020 (.0020)	.0020 (.0020)	.0020 (.0020)	.0020 (.0021)	.0019 (.0020)	.0020 (.0019)	.0020 (.0020)
Sc	.0010 (.0018)	.0021 (.0022)	.0020 (.0020)	.0020 (.0020)	.0019 (.0020)	.0020 (.0019)	.0020 (.0020)
h1	.0000 (.0000)	.0109 (.0065)	.0020 (.0020)	.0020 (.0020)	.0019 (.0020)	.0020 (.0019)	.0020 (.0020)
h2	.0008 (.0015)	.0021 (.0022)	.0020 (.0020)	.0020 (.0020)	.0019 (.0020)	.0020 (.0019)	.0020 (.0019)
h3	.0020 (.0020)	.0020 (.0020)	.0020 (.0019)	.0020 (.0020)	.0019 (.0020)	.0020 (.0019)	.0020 (.0019)

N=500.Table2.KL

Acknowledgements

This work was supported by Australian Research Council grant A49030439 and Australian Research Council 1992 small grant 9103169.

References:

[BoWa] Boulton, D.M. and C.S. Wallace, 'A program for numerical classification', *Computer Journal*, 13, pp63-69, 1970.

[CoSl88]Conway, J.H. and N.J.A. Sloane, 'Sphere Packings, Lattices and Groups', Springer-Verlag, London 1988.

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

[KoCh]Ko, D. and T. Chang, 'Robust M-Estimators on Spheres', J. Multivariate Analysis Vol. 45 No.1 (April 1993), pp104-136.

[Mard]Mardia, K.V., 'Statistics of directional data', Academic Press, 1972.

[OlWa]Oliver, J. and C.S. Wallace, 'Inferring Decision Graphs', Workshop 8, 12th International Joint Conference on Artificial Intelligence (IJCAI), 1991, Sydney.

[Scho] Schou, G., 'Estimation of the concentration parameter in von Mises-Fisher distributions', *Biometrika*, Vol.65, No. 1 (1978), pp369-77.

[UpFi]Upton, G.J.G. and B. Fingleton, 'Spatial Data Analysis by example', Vol.2 - Categorical and Directional Data, John Wiley and sons, 1989.

[Wall89a]Wallace, C.S., 'A Long-Period Pseudo-Random Generator', Technical Report No. 89/123, Dept. of Computer Science, Monash University, Australia, February 1989.

[Wall89b]Wallace, C.S., 'False Oracles and SMML Estimators', Technical Report No. 89/128, Department of Computer Science, Monash University, Australia, June 1989.

[Wall90]Wallace, C.S., 'Classification by Minimum-Message-Length Inference', S.G. Akl et al. (eds.) Advances in Computing and Information - ICCI'90, Niagara Falls, Lecture Notes in Computer Science, No.468, Springer-Verlag, 1990, pp72-81.

[WaBo68]Wallace, C.S. and D.M. Boulton, 'An Information Measure for Classification', *Computer Journal*, Vol.11, No.2, 1968, pp185-194.

[WaBo75] Wallace, C.S. and D.M. Boulton, 'An invariant Bayes method for point estimation', *Classification Soc. Bull.* 3, pp11-34, 1975.

[WaFr87]Wallace, C.S. and P.R. Freeman, 'Estimation and Inference by Compact Coding', *Journal of the Royal Statistical Society Series B*, Vol.49, No.3, pp240-265, 1987.

[WaFr92]Wallace, C.S. and P.R. Freeman, 'Single Factor Analysis by MML Estimation', *Journal of the Royal Statistical Society Series B*, Vol.54, No.1, pp195-209, 1992.

[WaGe]Wallace, C.S. and Georgeff, M.P, 'A general selection criterion for inductive inference', *ECAI '84: Advances in Artificial Intelligence*, Elsevier, Amsterdam, 1984, pp1-18.

[WaPa]Wallace, C.S. and J.D. Patrick, 'Coding Decision Trees', *Machine Learning*, 11, 7-22 (1993).