The Maximum Likelihood Solution to Inclination-only Data

Þórður Arason[1] and Shaul Levi[2]

(1) Veðurstofa Íslands – The Icelandic Meteorological Office, IS 150 Reykjavík, ICELAND

(2) College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, USA

arason@vedur.is, Shaul_Levi@msn.com

1 Abstract

The The arithmetic means of inclination-only data are known to introduce a shallowing bias. Several methods have been proposed to estimate unbiased means of the inclination along with measures of the precision.

Most of the inclination-only methods were designed to maximize the likelihood function of the marginal Fisher distribution. However, the exact analytical form of the maximum likelihood function is fairly complicated, and all these methods require various assumptions and approximations that are inappropriate for many data sets. For some steep and dispersed data sets, the estimates provided by these methods are significantly displaced from the peak of the likelihood function to systematically shallower inclinations. The problem in locating the maximum of the likelihood function is partly due to difficulties in accurately evaluating the function for all values of interest. This is because some elements of the log-likelihood function increase exponentially as precision parameters increase, leading to numerical instabilities.

In this study we succeeded in analytically cancelling exponential elements from the likelihood function, and we are now able to calculate its value for any location in the parameter space and for any inclination-only data set, with full accuracy. Furtermore, we can now calculate the partial derivatives of the likelihood function with desired accuracy. Locating the maximum likelihood without the assumptions required by previous methods is now straight forward.

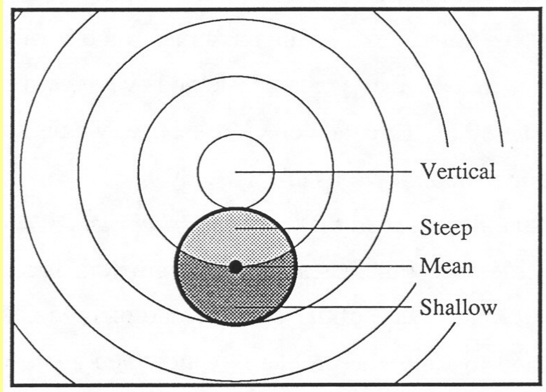
The information to separate the mean inclination from the precision parameter will be lost for very steep and dispersed data sets. It is worth noting that the likelihood function always has a maximum value. However, for some dispersed and steep data sets with few samples, the likelihood function takes its highest value on the boundary of the parameter space, i.e. at inclinations of ±90°, but with relatively well defined dispersion. Our simulations indicate that this occurs quite frequently for certain data sets, and relatively small perturbations in the data will drive the maxima to the boundary. We interpret this to indicate that, for such data sets, the information needed to separate the mean inclination and the precision parameter is permanently lost.

To assess the reliability and accuracy of our method we generated large number of random Fisher-distributed data sets and used seven methods to estimate the mean inclination and precision paramenter. These comparisons are described by Levi and Arason at the 2006 AGU Fall meeting. The results of the various methods is very favourable to our new robust maximum likelihood method, which, on average, is the most reliable, and the mean inclination estimates are the least biased toward shallow values.

Further information on our inclination-only analysis can be obtained from:

http://www.vedur.is/~arason/paleomag

2 Mean Bias of Inclination-only Data



The geometry of the sphere dictates that any circularly symmetric distribution about a true mean will be represented by more shallow inclinations than steep as compared to the mean. Arithmetic average of inclinations will therefore result in a too shallow estimate of the mean. From Arason [1991, Fig. 5.1, p. 207].

The "normal" distribution of three dimensional directions is the Fisher-distribution. By using Fisher-statistics one can obtain unbiased directional mean of a sample drawn from such a distribution (*Fisher*, Proc. R. Soc. London, Ser. A, 217, 295-305, 1953). Sometimes one has only access to inclinations and not declinations. Paleomagnetic directions from borecores usually lack declinations, but inclinations can be reliable. *Briden and Ward* (Pure Appl. Geophys., 63, 133-152, 1966) showed that for such inclination-only data, the arithmetic mean is biased toward shallow inclinations.

In paleomagnetic applications this inclination shallowing bias is usually less than a few degrees. For individual studies such a discrepancy is of a minor importance and usually well within the confidence limits of the study. However, since this is a one sided bias, attempts to combine results of many studies may lead to errors. Therefore, improper procedures for estimating mean inclinations in individual studies can seriously affect combined average estimates.

3 The problem and our Solution

Forty years ago *Briden and Ward* [1966] pointed out that the arithmetic mean of inclination-only data introduces a shallowing bias. Furthermore, they derived the likelihood function assuming the directions follow the Fisher-distribution, and presented a graphical method to estimate the true mean inclination along with the precision parameter (κ).

The likelihood function includes exponential elements, that are very difficult to accurately evaluate.

Several workers have attempted to derive a method to calculate the maximum likelihood estimates of mean inclination and the precision [e.g., *Kono*, 1980; *McFadden and Reid*, 1982]. Those methods make certain assumptions and approximations, which sometimes are inappropriate leading to inaccurate estimates of the maximum likelihood, and a bias toward shallow inclinations.

In this study we present a simple and robust method to calculate simultaneously the maximum likelihood estimates of the mean inclination and precision parameter without the assumptions and approximations of previous workers.

The exact mathematical form of the log-likelihood function and its derivatives is available. The task is to accurately identify the pair of (θ, κ) that maximize the likelihood function.

However, there are two major obstacles in directly identifying the maximum:

1. There are exponential elements in the likelihood function that become impossible to directly evaluate, and attempts in ordinary programming languages will often lead to an overflow or very inaccurate values, even for ordinary paleomagnetic data.

2. The likelihood function and its derivatives include Bessel functions that are difficult to accurately evaluate.

Our direct solution to the maximum likelihood problem includes:

A. We were successful in analytically cancelling all the exponential terms from the log-likelihood function.

B. We use an accurate estimation of the Bessel functions, many orders of magnitude more accurate than previous attempts on the problem.

Once these obstacles are cleared, accurate calculation of the maximum is straight forward.

4 Conclusions

The problem of estimating unbiased means of paleomagnetic inclination-only data was described forty years ago.

Several methods have been proposed to solve the problem. Some of these methods have evaluated the maximum likelihood estimates. However, these methods are based on various approximations and assumptions that turn out to be inappropriate for steep and dispersed data. Unfortunately, these estimates are sometimes inaccurate and on average biased toward shallow inclinations.

Analytical cancellations of exponential elements in the functions of the problem are essential to calculate the estimates accurately.

We present a method with accurate representations of the functions needed to solve the problem.

The method that we present makes it possible for scientists to accurately calculate the maximum likelihood estimates of the inclination-only problem.

Poster presented at the 2006 Fall meeting of the American Geophysical Union, 11-15 December 2006, San Francisco, California, USA

A The Log-likelihood Function and its Derivatives

The joint density function of the Fisher distribution is

$$P(\theta,\varphi)=\left(\frac{\kappa\sin\theta}{4\pi\sinh\kappa}\right)e^{\kappa\cos\theta_0\cos\theta+\kappa\sin\theta_0\sin\theta\cos(\varphi-\varphi_0)}$$

For inclination-only data we only have co-inclinations θ_i: θ_1, ..., θ_N, (θ = 90° − I, where I is inclination in degrees), which must be regarded as a random sample from the marginal distribution of the Fisher-distribution

$$f(\theta)=\left(\frac{\kappa\sin\theta}{2\sinh\kappa}\right)e^{\kappa\cos\theta_0\cos\theta}\,I_0(\kappa\sin\theta_0\sin\theta)$$

Here $I_0(x)$ is the hyperbolic Bessel function of order zero. The likelihood function is

$$H(\theta,\kappa)=\prod_{i=1}^{N}f(\theta_i)=\left(\frac{\kappa}{2\sinh\kappa}\right)^N\prod_{i=1}^{N}\sin\theta_i\,e^{\kappa\cos\theta\cos\theta_0}\,I_0(\kappa\sin\theta\sin\theta_0)$$

and the log-likelihood function is

$$h(\theta,\kappa)=\ln[H(\theta,\kappa)]$$

$$h(\theta,\kappa)=N\ln\left(\frac{\kappa}{2\sinh\kappa}\right)+\sum_{i=1}^{N}\kappa\cos\theta\cos\theta_i+\sum_{i=1}^{N}\ln[I_0(\kappa\sin\theta\sin\theta_i)]+\sum_{i=1}^{N}\sin\theta_i$$

The maximum value of the likelihood function is where the partial derivatives of the log-likelihood function are both zero

$$\frac{\partial h}{\partial\theta}=-\kappa\sin\theta\sum_{i=1}^{N}\cos\theta_i+\kappa\cos\theta\sum_{i=1}^{N}\left\{\sin\theta_i\left[\frac{I_1(\kappa\sin\theta\sin\theta_i)}{I_0(\kappa\sin\theta\sin\theta_i)}\right]\right\}=0$$

$$\frac{\partial h}{\partial\kappa}=\frac{N}{\kappa}-N\coth\kappa+\sum_{i=1}^{N}\cos\theta\cos\theta_i+\sum_{i=1}^{N}\left\{\sin\theta\sin\theta_i\left[\frac{I_1(\kappa\sin\theta\sin\theta_i)}{I_0(\kappa\sin\theta\sin\theta_i)}\right]\right\}=0$$

Here $I_1(x)$ is the hyperbolic Bessel function of first order. By adjusting these equations we were able to set up two equations that can be iteratively solved.

B Evaluation of the Log-likelihood Function

The likelihood and log-likelihood functions include exponential elements that prevent direct evaluation of the functions for some combinations of (θ, κ). However, it is possible to analytically rewrite the log-likelihood function so that these elements need never to be evaluated. In such a way one can evaluate the log-likelihood function for any combination of (θ, κ).

The first term includes the sinus hyperbolicus function (see fig), which increases exponentially. This term can be written as

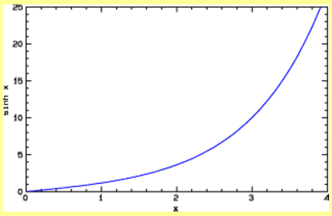
$$N\ln\left(\frac{\kappa}{2\sinh\kappa}\right)=N\left(\ln(\kappa)-\ln(1-e^{-2\kappa})-\kappa\right)$$

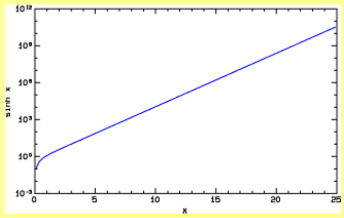
for κ higher than 15 the exponential term can be omitted and for values close to zero this can also be simplified. The second and third terms can be simplified to

$$\sum_{i=1}^{N}\kappa\cos\theta\cos\theta_i+\sum_{i=1}^{N}\ln[I_0(\kappa\sin\theta\sin\theta_i)]$$
$$=\sum_{i=1}^{N}\left(\kappa\cos(\theta-\theta_i)+\ln[B(\kappa\sin\theta\sin\theta_i)]\right)$$

where $B(x)$ = $I_0(x)/e^x$. By cancelling the exponential element, the Bessel function can be accurately calculated for any combination of (θ, κ), see box C.

The only problem with the last term is if one of the observed inclinations is exactly ±90°.





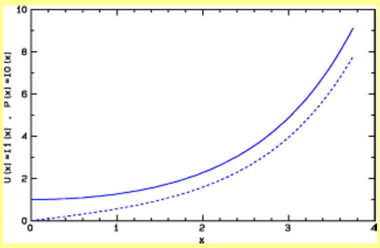
C Evaluation of the Hyperbolic Bessel Functions

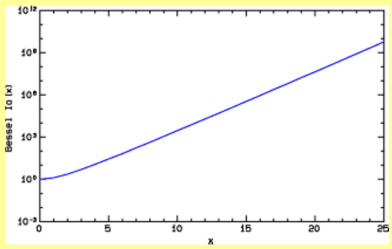
The hyperbolic Bessel functions, can not be expressed as a finite combination of elementary functions.

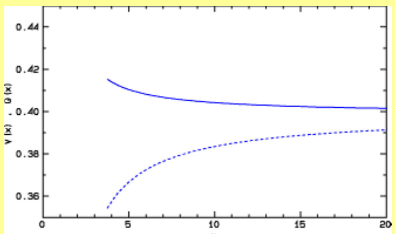
In order to evaluate the Bessel function, $I_0(x)$, and the ratio $I_1(x)/I_0(x)$ we use the approximations of *Press et al.* (Numerical recipes, The art of scientific computing (Fortran version), 1989), which is based on *Olver* (Ch. 9 in *Abramowitz and Stegun*, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, eq. 9.8.1-4, 1972). These approximations are five orders of magnitude more accurate than the approximations used by *Enkin and Watson* (Geophys. J. Int., 126, 495-504, 1996).

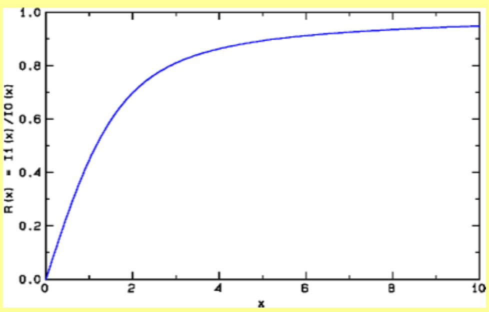
Direct evaluation of the Bessel functions is problematic because they increase exponentially (see figs) and we may need to evaluate the functions for very high values, f.ex. $I_0(1000)\approx 2.5\cdot10^{432}$.

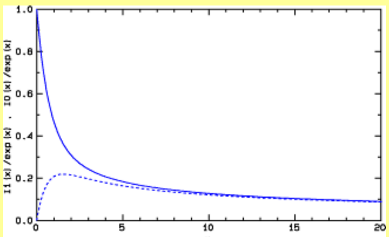
Numerical inaccuracies are avoided by never directly evaluating the Bessel functions. We analytically cancel their exponential elements, prior to the evaluation, and the ratio $I_0(x)/e^x$ is numerically well behaved and takes values between 0.01 and 1 for the range 0 < x < 1000 (see fig). Similarly we evaluate the ratio $I_1(x)/e^x$, without calculating $I_1(x)$. The ratio $I_1(x)/I_0(x)$ is then evaluated by calculating [$I_1(x)/e^x$] / [$I_0(x)/e^x$].

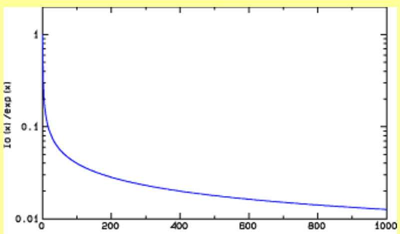












D Our Calculation of the Maximum Likelihood Solution

a. We calculate the solution of the maximum likelihood problem by iteration in a similar fashion as *Enkin and Watson* (1996). As an initial guess of θ and κ, we use the arithmetic mean and the inverse variance of the co-inclinations (θ = 90° - I).

$$\hat{\theta}_0 = \bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta_i \qquad \hat{\kappa}_0 = \left(\frac{1}{N-1}\sum_{i=1}^{N}\left(\theta_i - \bar{\theta}\right)^2\right)^{-1}$$

In calculating κ we need to use θ in radians. Then we iterate alternatively for new estimates of θ and κ through the following two equations:

$$\tan\hat{\theta}_{j+1} = \frac{\sum_{i=1}^{N}\left(\sin\theta_i\left[\frac{I_1\left(\hat{\kappa}_j \sin\hat{\theta}_j \sin\theta_i\right)}{I_0\left(\hat{\kappa}_j \sin\hat{\theta}_j \sin\theta_i\right)}\right]\right)}{\sum_{i=1}^{N}\cos\theta_i}$$

$$\hat{\kappa}_{j+1} = \left(\coth\hat{\kappa}_j - \frac{1}{N}\sum_{i=1}^{N}\left\{\cos\hat{\theta}_{j+1}\cos\theta_i + \sin\hat{\theta}_{j+1}\sin\theta_i\left[\frac{I_1\left(\hat{\kappa}_j \sin\hat{\theta}_{j+1}\sin\theta_i\right)}{I_0\left(\hat{\kappa}_j \sin\hat{\theta}_{j+1}\sin\theta_i\right)}\right]\right\}\right)^{-1}$$

Usually, except for steep and dispersed data sets, the pair (θ, κ) converges in a few iterations (<10). We repeat this process 10 000 times unless the angular difference between successive iterations becomes less than 0.000 001°. For the solution pair we evaluate the log-likelihood function.

b. Then we repeat the iteration process from another intitial guess by halfing the initial values of θ and κ, used in (a). When this second process converges we calculate the value of the log-likelihood function.

c. Now we locate the maximum on the edge, I = ±90°, from the initial value

$$\hat{\kappa}_0 = \left(1 - \cos\theta \cdot \frac{1}{N}\sum_{i=1}^{N}\cos\theta_i\right)^{-1}$$

iterating through

$$\hat{\kappa}_{j+1} = \left(\coth\hat{\kappa}_j - \cos\theta \cdot \frac{1}{N}\sum_{i=1}^{N}\cos\theta_i\right)^{-1}$$

Estimation of κ on the edge usually converges very fast. At the maximum on the edge we evaluate the value of the log-likelihood function.

d. Finally, we compare the solutions in a, b, and c and select the pair (θ, κ), which gives the highest value of the log-likelihood function.

F Fishers Numerical Example

In this example we use the paleomagnetic data used in a numerical example by *Fisher* (1953). His nine inclinations were: 66.1, 68.7, 70.1, 82.1, 79.5, 73.0, 69.3, 58.8, and 51.4.

The following list shows the evolution of the solution through our iteration process. We start with the initial guess of the arithmetic mean I = 68.78° (θ = 21.22°), κ = 36.42. For this pair the value of the log-likelihood function is h = 3.83.

Iter	Theta	Kappa	Log-likelihood
0	21.22	36.42	3.83
1	18.93	35.59	4.12
2	18.57	34.75	4.129
3	18.44	34.13	4.132
4	18.36	33.68	4.1336
5	18.31	33.36	4.1344
10	18.189	32.66	4.13533
20	18.153	32.467	4.1353855
30	18.15129	32.4554	4.1353857157
40	18.151171	32.45473	4.135385716284
44	18.151166	32.45471	4.1353857162853

After 44 iterations the angular difference between successive steps was less than our limit (0.000 001°). The solution indicates that I = 71.85°.

Now we try to start the iteration from another initial values and we select half the initial values of the first process.

Iter	Theta	Kappa	Log-likelihood
0	10.61	18.21	3.79
1	11.96	17.99	3.87
2	12.73	18.45	3.89
3	13.35	19.16	3.92
4	13.93	20.00	3.94
5	14.49	20.96	3.97
10	16.77	26.58	4.09
20	18.06	31.94	4.1351
30	18.146	32.424	4.1353846
40	18.1508	32.4529	4.135385712
50	18.151145	32.45459	4.13538571627
57	18.151161	32.45468	4.1353857162848

After 57 iterations the angular difference between steps was less than our limit. The solution indicates again that I = 71.85°. We note that the value of the log-likelihood function is practically the same, although slightly lower.

Now we find the maximum value on the edge, I = +90°.

Iter	Theta	Kappa	Log-likelihood
0	0.00	12.651665098	3.76
1	0.00	12.651665094914	3.76
2	0.00	12.651665094914	3.76

However, the value of the log-likelihood function on the edge is considerably lower than for our solutions in the interior.

The highest value of the log-likelihood function is from the first iteration process and our maximum likelihood solution is I = 71.85°, κ = 32.45.

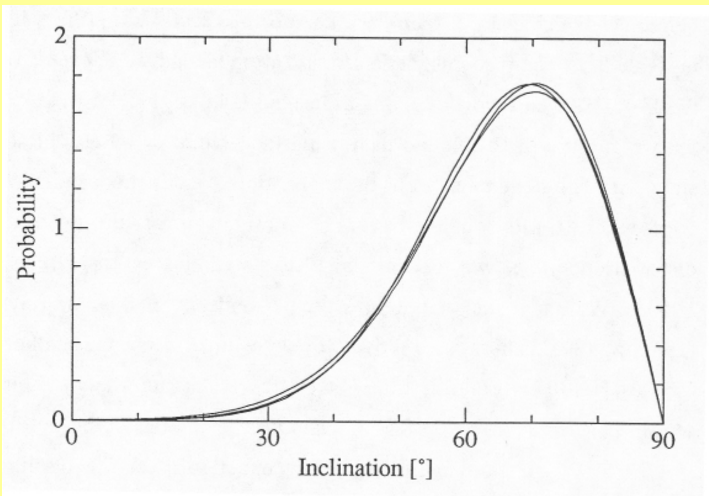
E The Physical Meaning of the Edge Solution

For dispersed and steep data the likelihood function sometimes has its maximum value on the edge (I = ±90°).

The probability distribution for these cases is such that similar distributions may originate from a range of pairs of true inclinations and precision parameters. The fig. on the right shows an example of this effect, where three distributions of different pairs of (θ, κ) are almost identical. For such steep and dispersed true inclinations it becomes impossible to extract unique information on both true inclination and precision parameter from a finite set of observed inclinations, and any attempt to do so will depend critically on the assumptions of the calculation method.

In our maximum likelihood method we make use of this "discrepancy". A mean inclination on the edge, i.e. I = ±90°, indicates that a unique solution does not exist and the information to separate inclination and precision parameter is permanently lost.

However, such a solution indicates that the mean inclination is probably "steep" and the maximum likelihood estimate of precision parameter sets an upper limit to the true value.



The inclination-only distribution for three combinations of the true values (I, κ). The values are (I = 75°, κ = 8.6), (I = 80°, κ = 10), and (I = 85°, κ = 12). It is not important in this context to identify the curves. From Arason (1991, Fig. 5.12, p. 266).