

Interactive comment on “Modelling complex geological angular data with the Projected Normal distribution and mixtures of von Mises distributions” by R. M. Lark et al.

R. M. Lark et al.

mlark@nerc.ac.uk

Received and published: 7 May 2014

We are grateful to this reviewer for his/her thorough reading of our paper and for the suggestions that are made. Our responses and proposals for revision are as follows.

1. Use of the likelihood ratio to choose the number of components in the MVM distribution

We use likelihood ratio tests to compare a model with $g + 1$ components against an

C1186

alternative with g in preference to an information criterion. That is because the information criterion is an informal basis for model comparison, whereas the likelihood ratio allows us to perform a formal hypothesis test, albeit against a bootstrapped distribution under the null hypothesis. This gives us the familiar measure of confidence for an inference in the model comparison (a p -value) which the comparison on an information criterion does not provide. This was also the motivation of the studies we cited (Fu et al., 2008; Aitken et al., 1981) for using the likelihood ratio test. We could add a sentence to explain this in section 2.1

2. MVM distribution with covariates, and categorical covariates.

In principle it might be possible to include covariates in the MVM distribution. For example, one might specify a common set of components for the mixture, and let the weights α_i , $i = 1, 2, \dots, g$ depend on the covariate. However, this would be a much more complex model than the extension of the PN proposed by Wang and Gelfand (2013). For a start, one would need to ensure that the α_i were all non-negative and sum to 1, and it is not apparent that the model would be identifiable. We might add the following sentence to section 2.3:

‘In principle a similar extension of the MVM model in which the mixture weights α_i depend on a set of covariates might be developed. However, such a model would be complex and might not be identifiable. The extension of the PN model is much more straightforward.’

The categorical covariates considered in this paper are in the Bangladesh example (section 3.2.3) where we combine the orientation observations into a single data set with the categorical covariate ‘observation type’ which is either ‘anticlinal axial plane’ or ‘Landsat-derived lineament’. To clarify this we could add a reference to section 3.2.3 in section 2.3, and add a sentence in 3.2.3 to make it clear that ‘observation type’ is the categorical variable.

C1187

3. AIC vs BIC.

The AIC does not favour models with numerous parameters, it has a penalty for model complexity, albeit a smaller penalty than the BIC if there are more than about 16 data points. In our case studies the AIC and BIC lead to the same choice between MVM and PN distributions for all data sets. We might edit section 2.2 to read:

‘For $g \geq 2$ the MVM distribution has more parameters than the PN. Having fitted both to a data set, the question remains whether the greater complexity of the MVM is justified by its goodness of fit. Two commonly-used criteria to select between models of differing complexity, where the models are not nested and so cannot be formally compared on the likelihood ratio, is to use information criteria which combine the maximized log-likelihood for each fitted model with a term than penalized models for the number of parameters that must be estimated. One such criterion is the AIC due to Akaike (1973):

$$A = -2\ell + 2N_p, \quad (1)$$

where ℓ is the maximized (full) log likelihood for a model with N_p parameters. The model with the smallest A is selected, so effectively the selection is based on the likelihood with a penalty for the model complexity as measured by N_p . Another information criterion is the Bayes information criterion (Kass and Raftery, 1995) (BIC):

$$B = -2\ell + N_p\{\log(2\pi n)\}, \quad (2)$$

where n is the number of observations. In both cases one selects the model for which the information criterion is smallest. The BIC penalizes extra parameters more heavily than does AIC unless n is small. More fundamentally, the AIC selects a model which appears to be closest to the underlying but unknown model which generates the data, whereas the BIC selects a model with a maximized posterior probability (Wit et al, 2013). The AIC is a basis for a pragmatic choice of model which seems to explain the data and offer a sound basis for prediction, whereas the BIC aims to identify the ‘true’

C1188

model (Spiegelhalter et al. 2014). The two criteria are therefore not directly commensurate, and the question of which is ‘best’ depends on the principles and purposes of model selection, and Wit et al. (2013) found that there is no consensus. A detailed discussion of the two criteria is outwith the scope of the present paper, so for our present purposes we present both criteria.’

And we then edit section 3.1.2 (second paragraph) to read:

‘In both cases both the AIC and the BIC was smallest for the more complex MVM distribution.’

and we edit section 3.2.2. to read (second sentence)

‘As shown in Table 2, the AIC and BIC support the same conclusion: the MVM model is preferred for the anticline planes, but the PN model is preferred for the the Landsat-derived lineaments; in this latter case the likelihood for the PN distribution was larger than for the more complex MVM distribution.’

4. Table of parameter estimates

We could add a table of the MVM parameter estimates. However, as Wang and Gelfand (2013) note, the parameters of the PN model are not directly very informative, and the density plots on Figures 2 and 3 are much more useful.

We already discuss the contrasts between the distributions in terms of how effectively they capture the geological variation. In particular see page 2190, lines 17–23 of the original version of the paper with respect to the Sherwood Sandstone group, and line 30 on page 2190 to line 14 on page 2191 in respect of the Windermere supergroup. We might add to the end of section 3.1.2. the sentence:

‘The MVM model, giving three distinct modes as seen in Figure 2b, in addition to the broader distribution of dip directions over the interval from due south to north-west,

C1189

captures this complexity better than the PN which shows two close to antipodal modes'.

There is rather less geological information to compare with the distributions in respect of the Bangladesh data, but we do make observations on the relative performance of the two distributions in section 3.2.2.

5. Orientation data

It is not the case that the doubled orientation data are symmetric, they should not be. Orientation data are commonly presented in raw form in terms of a pair of angles, m and $m + \pi$, and therefore typically appear multimodal with pairs of modes separated by π or 180 degrees. These pairs of values are doubled to give a single value, $2m$ (since $2m + 2\pi = 2m$ on a circular support). The new variable does not (necessarily) have a symmetrical or bimodal distribution, its support is on the full circle. That procedure was followed in this case. The support of the data is not on the half circle, it is simply that the doubled angles have a rather restricted distribution because both the anticline axial planes and the Landsat-derived lineaments have somewhat strongly aligned orientations.

To clarify this we can replace the second-last sentence in 3.2.1 with the following:

'As described by Davis and Sampson (2002), all orientation data can be expressed by pairs of values: m and $m + 2\pi$ radians representing the bearing for each end of a linear feature such as a fault or plane. Following Krumbein (1939) these values can be doubled to give a single value $2m$ (since $2m + 2\pi = 2m$ for values on the circle). The doubled orientations may be distributed over the whole circle, and can be analysed with methods appropriate for angular data, and that was done here.'

6. MVM model with covariates

C1190

In principle one might compare the fit of a common MVM and two separate MVM distributions to the two subsets of data, but this would require substantial further work to address the regularity of the log-ratio statistic and the identifiability of the compared models. This is beyond the scope of the current paper, but it could be mentioned as a topic for further research in the Conclusions (section 4).

7. Parametric bootstrap

It is possible, in principle, that a parametric bootstrap might be used for the comparison between models where one parameter estimate goes to the boundary, this would require further investigation beyond the scope of the current, in particular to establish that the procedure is consistent, but it could be mentioned as a topic for further research in the Conclusions (section 4) current paper.

New references.

Kass, R.E., Raftery, A.E. 1995. Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Krumbein, W.C. 1939 Preferred orientation of pebbles in sedimentary deposits. *Journal of Geology*, **47**, 673–706.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2014), The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, (In press).

Wit, E., van den Heuvel, E., Romeijn, J-W. 2012. 'All models are wrong ...': an introduction to model uncertainty. *Statistical Neerlandica* **66**, 217–236.

C1191

C1192