Numerical Investigation on Asymptotic Features of Model Selection Criteria

Dan Lu (dl07f@fsu.edu) and Ming Ye (mye@fsu.edu)

Department of Scientific Computing, Florida State University, Tallahassee, FL 32306

1. INTRODUCTION

Using a synthetic example of four geostatistical models (including the true or data-generating model), this study is to investigate:

- Accuracy of *BIC* and *KIC*;
- Asymptotic features of the model selection criteria, *AIC*, *AICc*, *BIC*, and *KIC;*
- Fisher information matrix's role in model selection;
- Sensitivity of posterior to prior model probability.

2. THEORETICAL BACKGROUND

The posterior distribution of the quantity of interest, Δ, given a set of data **D** is:

$$p(\Delta|\mathbf{D}) = \sum_{k=1}^{K} p(\Delta|M_k,\mathbf{D})\, p(M_k|\mathbf{D})$$

where **M** =$(M_1,\ldots,M_K)$ is the set of models considered.

The posterior model probability of model M_k is given by Bayes' rule:

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k)\,p(M_k)}{\sum_{l=1}^{K} p(\mathbf{D}|M_l)\,p(M_l)}$$

where

$$p(\mathbf{D}|M_k) = \int p(\mathbf{D}|\boldsymbol{\theta}_k, M_k)\, p(\boldsymbol{\theta}_k|M_k)\, d\boldsymbol{\theta}_k$$

The likelihood function can be evaluated using either a simple Monte Carlo method via

$$\hat{p}(\mathbf{D}|M_k) = \frac{1}{N}\sum_{i=1}^{N} p(\mathbf{D}|\boldsymbol{\theta}_k^{(i)}, M_k)$$

or the Laplace approximation via model selection criteria

$$p(\mathbf{D}|M_k) = \exp\left(-\frac{1}{2} IC_k\right)$$

where IC represents *AIC*, *AICc*, *BIC*, and *KIC*

$$AIC_k = -2\ln\left[L(\hat{\boldsymbol{\theta}}_k|\mathbf{D})\right] + 2N_k$$

$$AICc_k = -2\ln\left[L(\hat{\boldsymbol{\theta}}_k|\mathbf{D})\right] + 2N_k + \frac{2N_k(N_k+1)}{N-N_k-1}$$

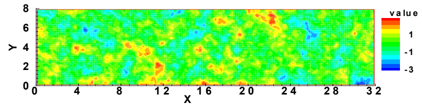
$$BIC_k = -2\ln\left[L(\hat{\boldsymbol{\theta}}_k|\mathbf{D})\right] + N_k \ln N$$

$$KIC_k = -2\ln[L(\hat{\boldsymbol{\theta}}_k|\mathbf{D})] + N_k\ln N - 2\ln p(\hat{\boldsymbol{\theta}}_k) - N_k\ln 2\pi + \ln|\bar{\mathbf{F}}_k|$$

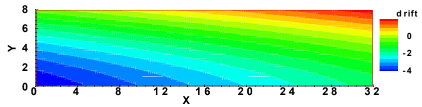
$$\bar{F}_{kij} = \frac{1}{N}F_{kij} = -\frac{1}{N}\left.\frac{\partial^2 \ln\left[L(\hat{\boldsymbol{\theta}}_k|\mathbf{D})\right]}{\partial\theta_{ki}\partial\theta_{kj}}\right|_{\boldsymbol{\theta}_k=\hat{\boldsymbol{\theta}}_k} \quad \text{(Ye et al., 2008)}$$

3. THE SYNTHETIC EXAMPLE

Random data **D** consists of deterministic trend μ and random residual **R**: **D= μ+R**



True data **D**



True drift **μ**

True and Alternative Models

$$M_0:\ \boldsymbol{\mu}(x,y) = a_0 + a_1x + a_2y + a_3y^2$$

Three alternative models, M_1, M_2, and M_3

$$M_1:\ \boldsymbol{\mu}(x,y) = a_0 + a_1x + a_2y + a_3xy + a_4y^2$$

$$M_2:\ \boldsymbol{\mu}(x,y) = a_0 + a_1xy + a_2y + a_3y^2$$

$$M_3:\ \boldsymbol{\mu}(x,y) = a_0 + a_1x + a_2y$$

4. RESULTS AND DISCUSSIONS

4.1 Asymptotic Investigation

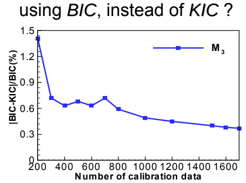
The asymptotic numerical investigation is conducted by gradually increasing the number of calibration data, **D**, from 200 to 1,700. For each experiment, following Ye et al. (2004), the four models are calibrated, and the four criteria and corresponding model probabilities are calculated.

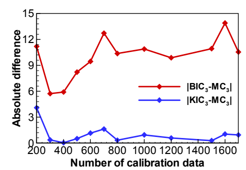
4.2 Accuracy of *BIC* and *KIC*

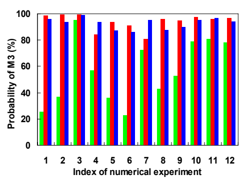
Questions:

When does *KIC* approach to *BIC* ?

How accurate is the posterior model probability calculated using *BIC*, instead of *KIC* ?



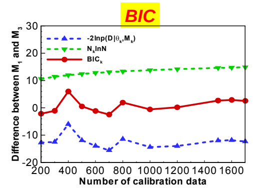


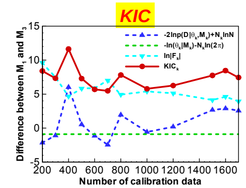


Red (*MC*), Blue (*KIC*), Green (*BIC*)

Taking M_3 as an example, *KIC* asymptotically reduces to *BIC*, But using *MC* results as reference, *KIC* is more accurate than *BIC* for approximating the model likelihood function and for calculating model probabilities.

4.3 Asymptotic Features of *BIC* and *KIC*



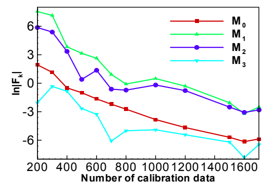


The asymptotic behavior of the difference of *BIC* between M_1 and M_3 is **both** determined by the model fitting term and the complexity penalty term. The $\ln|\bar{\mathbf{F}}_k|$ term of *KIC* imposes **more penalties** to the complex model.

4.4 Fisher Information Matrix

$\ln|\bar{\mathbf{F}}_k|$ is the major reason that *KIC* is more accurate than *BIC* for approximation the model likelihood function.

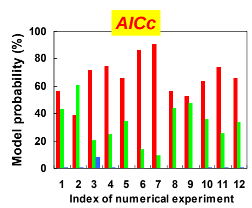
$\ln|\bar{\mathbf{F}}_k|$ values decrease when the number of calibration data increases and its values follow the same order, $M_1 > M_2 > M_0 > M_3$

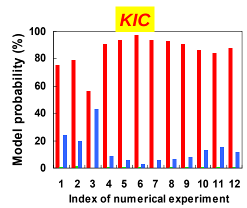


The $\ln|\bar{\mathbf{F}}_k|$ term discriminates alternative models based on not only model complexity but also model structures, a feature is owned only by *KIC*.

4.5 Consistency property of *BIC* and *KIC*

Over all the numerical experiments, M_2 receives negligible probability, due to its most deviation from the true model M_0. When the true model is included, *KIC* selects the true model in all experiments by assigning overwhelmingly large probability to it. While *AICc* selects M_1 for the experiment of *N*=300.

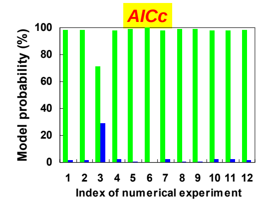


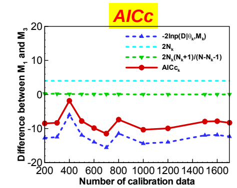


Red (M_0), Green (M_1) and Blue (M_3)

4.6 Asymptotic Features of *AICc*

When the true model is excluded, *AICc* considers M_1 as the best model for all the experiments. The asymptotic behavior of the difference of *AICc* is **solely** determined by the model fitting term.

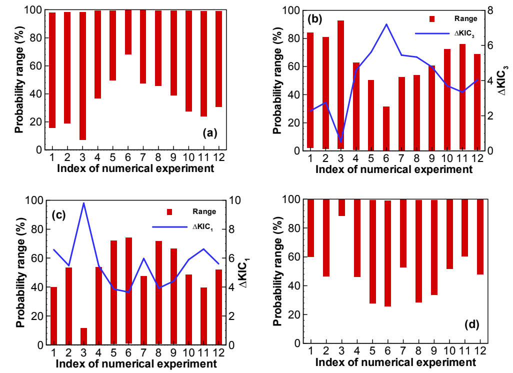




Green (M_1) and Blue (M_3)

4.7 Sensitivity of Posterior Model Probability

- The sensitivity of posterior to prior model probability is measured by the range between the maximum and minimum posterior probability (a smaller range corresponding to a smaller sensitivity).
- For both *AICc*- and *KIC*-based probabilities, sensitivity of posterior to prior model probability is irrelevant to the number of calibration data.
- The sensitivity is inversely related to the Δ*KIC* values, because small Δ*KIC* values give large range of the posterior model probability (Ye et al., 2005).



Sensitivity of posterior probability of (a) M_0 and (b) M_3 (when the true model is included) and (c) M_1 and (d) M_3 (when the true model is excluded) for the twelve numerical experiments.

5. REFERENCES

Lu, D., and M. Ye (2009), Numerical investigation of asymptotic features of model selection criteria: A case study of alternative geostatistical models, *Water Resources Research*, Under Review.

Ye, M., S.P. Neuman, and P.D. Meyer (2004), Maximum Likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resources Research*, 40, W05113, doi:10.1029/2003WR002557.

Ye, M., S.P. Neuman, P.D. Meyer, and K.F. Pohlmann (2005), Sensitivity analysis and assessment of prior model probabilities in MLBMA with application to unsaturated fractured tuff, *Water Resources Research*, 41, W12429, doi:10.1029/2005WR004260.

Ye, M., P.D. Meyer, and S.P. Neuman (2008), On model selection criteria in multimodel analysis, *Water Resources Research*, 44, W03428, doi:10.1029/2008WR006803.