Date 20.07.2012

Implicit complexity

A problem for model selection tools

Presentation for the
VIIth Conference of the SLMFC

Jan-Willem Romeijn
Faculty of Philosophy
University of Groningen

Narrow topic, broad relevance

Scientists regularly consult model selection tools to choose among available statistical models.

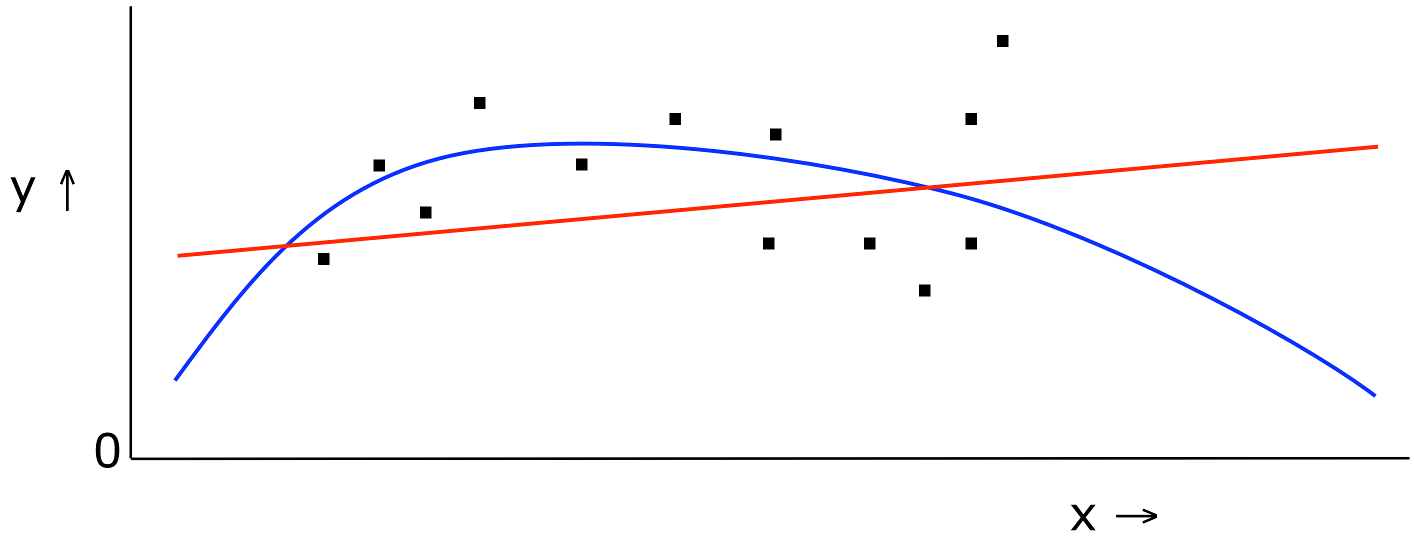
› In cognitive developmental psychology, researchers choose between models that differ in terms of constraints on the parameters expressing task difficulty.

› In psychiatry, the search for subtypes of depression requires a choice on the clustering of symptoms from the DSM as indicators of latent disorders.

My goal here is to investigate and eventually improve model selection tools.

Model selection

In model selection, we choose between families of probability distributions. A simple example of a model selection problem is curve fitting.



Model selection

Several so-called information criteria (AIC, BIC, and the like) provide a score for the overall quality of a model. The ICs present a trade-off:

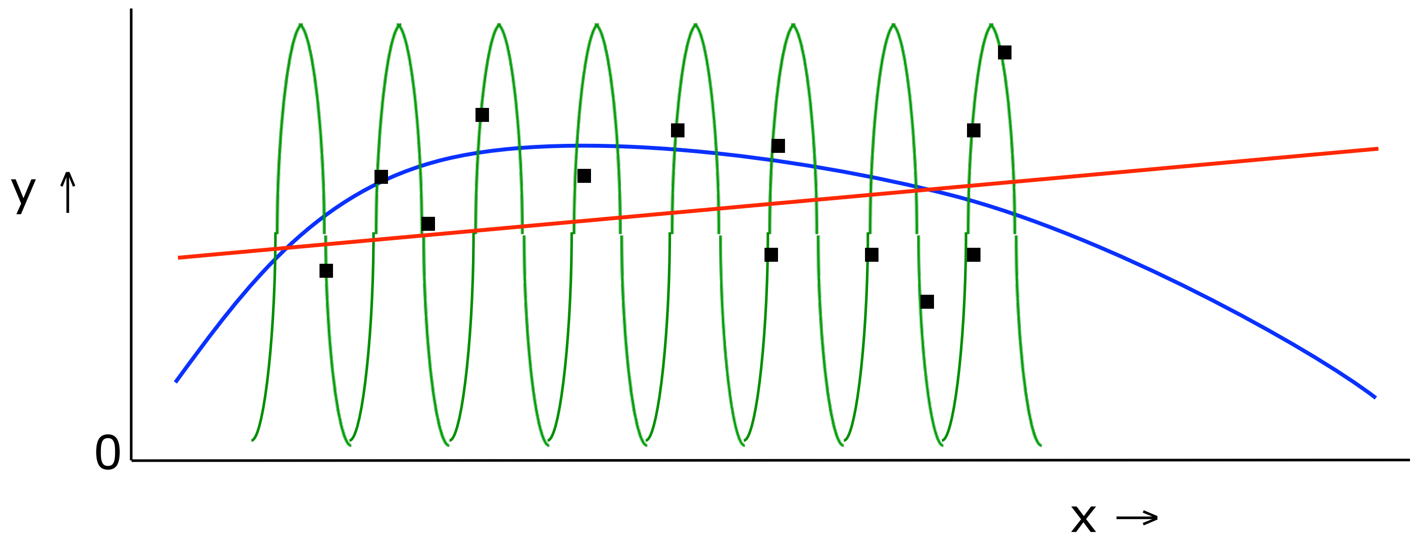
Score(M) = Fit with data – Complexity

= P(Data | Estimate in M) – Dimension(M)

This simple score results from independently motivated criteria for models, and lots of fancy maths.

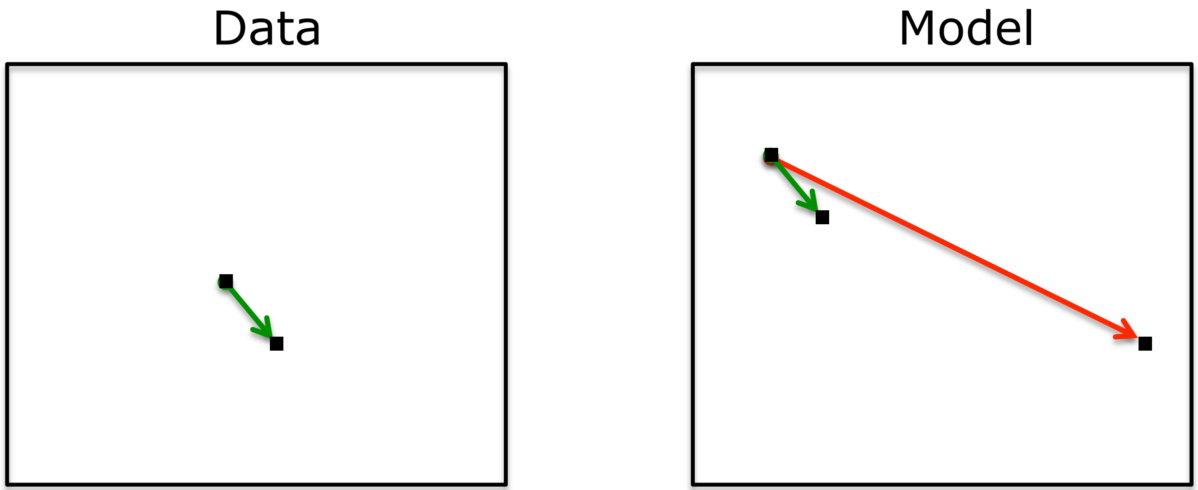
The problem

How to evaluate the model of sine curves below? It has only three dimensions and its fit is perfect. The ICs give the wrong answer.



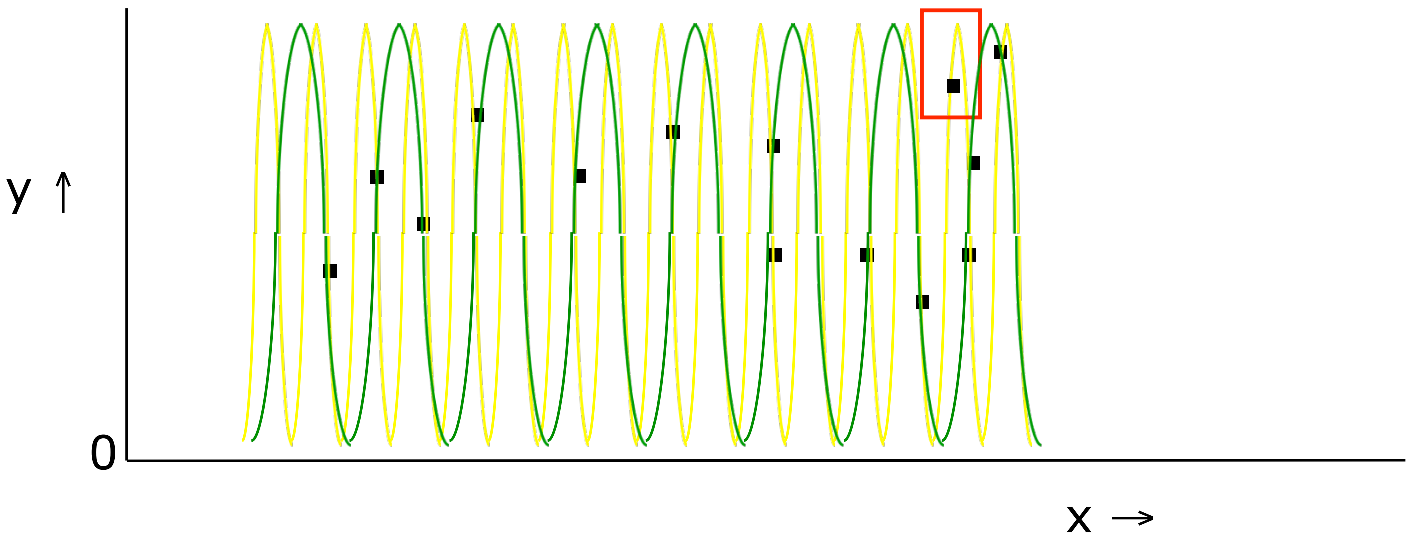
Sensitivity of estimations

Intuitively, the best estimate according to a model must be robust under tiny changes in the data. A good model is not skittish.



Sensitivity of estimations

The sine model suffers from exactly that defect: nudging a single point in data space will cause the best estimate to change radically.



Sensitivity in the BIC

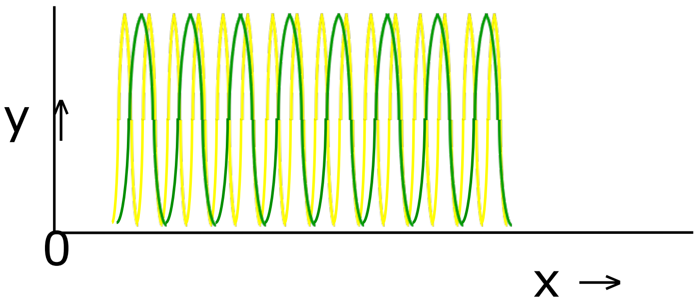
The Bayesian information criterion is an approximation of the marginal likelihood of the model. Interestingly, the sensitivity shows up in this approximation, as the Fisher information:

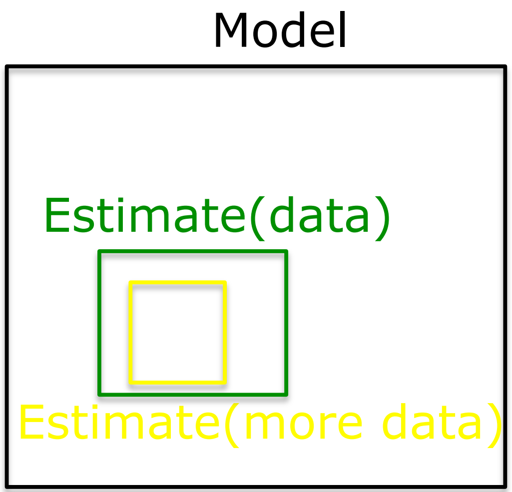
Score(M) = P(Data | Estimate in M) – Dimension(M)
– Fisher Information(Estimate) + …

The problem is that its contribution to the score is too small: the terms for fit and dimension grow with the data size, the sensitivity not.

Priors to the rescue

Actually, the sine model is degenerate: many hypotheses perfectly fit the data. With every data addition, the set of best estimates gets smaller.





Diminishing priors

Effectively, the prior for the best estimate will diminish with the addition of new data points. And this will show up in a properly approximated marginal likelihood.

Score(M) = P(Data | Estimate in M) – Dimension(M)

– Diminishing prior + …

More in general, because highly versatile models like the sine model are all over the place, a counteracting prior term will act as penalty.

The upshot

Bayesian model selection tools give the right answer to model selection problems involving implicitly complex models. But the BIC approximation needs to be adjusted.

› They do so relative to a conceptualization or parameterization of the data. This echoes well-known facts about induction.

› The penalty deriving from the prior may fall away if we happen to choose a prior preselecting the estimated curves. But that would be a freak accident.

I hope the foregoing illustrates how a toy example from confirmation theory can inform real scientific practice.

Thanks

With questions and remarks, please email:

`j.w.romeijn@rug.nl`

Slides will be made available on my website:

`http://www.philos.rug.nl/~romeyn`