

# Assessing Model Preference

H. Lee

Harvard Smithsonian Center for Astrophysics

September 9, 2008

# Motivation

- Vinay called me to get some the 3rd opinion about a.k.a. Vuoung's test.
- Not knowing how their conversation begin, I began to think on my own way. What's wrong with information criteria such as AIC and BIC, we use so often? Why astronomers do not care about in ranking candidate models based on well known model selection criteria? Inference comes after one chooses a model and as a consequence, the results of inference depend on the model choice.
- The shortcoming of IC for astronomers is that IC do not provide the level of significance of choosing one model over the other. They like to know  $\text{Prob}(\text{choosing model A over model B})$  not  $\text{AIC}(\text{model A}) < \text{AIC}(\text{model B}) < \text{AIC}(\text{model C})\dots$

## Vinay and Andrea's example

Power Law vs Blackbody (see Vinay's slog post):

$$\text{PL: } f(x) = A_p \left( \frac{x}{x_{min}} \right)^{-\alpha}$$

$$\text{BB: } f(x) = A_b \frac{x^2}{e^{x/kT} - 1}$$

- Normalization constants,  $A_p$  and  $A_b$  have no commonality except their meanings in physics.
- Both models have two parameters to be estimated. The parameter spaces of these two models are non nested.
- For the convenience' sake, we assume independence within these pairs  $(A_p, \alpha)$  and  $(A_b, T)$ . (In fact, I've kept hearing  $A_p$  and  $\alpha$  correlated).

# Astronomer's Model Selection I. Reduced $\chi^2$

–First, they bin (group) data to increase S/N for  $\chi^2$  fitting. Then, compare reduced  $\chi^2$  for a better fit model (does not produce significance).

–  $(D_i^b - M_i^b) / \sqrt{M_i^b} \sim N(0, 1)$ , where  $D_i^b$  and  $M_i^b$  denote binned data based on observed Poisson counts ( $D_i$ ) and model (predicted) counts ( $M_i$ ).

This approximation is correct if the specified model is a right one. Reduced  $\chi^2$  indicates how good the chosen model fits data. If both models produce reasonable reduced  $\chi^2$  values, we could devise a test similar to F test for testing the ratio of these two reduce  $\chi^2$  values is F distributed under the null hypothesis ( $H_o : red\chi^2(A)/red\chi^2(B) = 1$ ).  
 –instrument effects and pile ups are simplified/ignored/taken care of.

–Note that this  $\chi^2$  fitting process transforms a complex parameter space into a simple Gaussian error space. Degrees of freedom only carry the information in those models. The transformed space cannot tell whether models are nested or not. Residuals do not tell the model properties (PL or BB). It could be viewed as dimension reduction in a degrading fashion because of the loss of information. Furthermore, we must be sure about both models are relatively good-fit prior to comparing two models.

– If  $A_p$  and  $\alpha$  are dependent, reduced  $\chi^2$  with the degree of freedom  $(n-p)$  where  $p=2$  seems not proper.

→ how to get effective degrees of freedom? Will  $\frac{D_i^b - M_i^b}{\sqrt{M_i^b}} \stackrel{iid}{\sim} N(0, 1)$  be valid?

## Astronomer's Model Selection II. LRT

When comparing nested models, with or without binning, astronomers use LRT.

Unfortunately, LRT has its own limits (see papers and slog posts) and is not suitable for comparing PL and BB to produce model preference significance.

The notion of Model Misspecification (if one model is right, the other is misspecified) in general sense is not considered with LRT. The term, model misspecification is not discussed in the astronomical community.

# Model Selection – Statisticians' viewpoint

See Books (there are not many)

- Burnham, K.P. and Anderson, D.R. (2002) Model Selection and Multimodel Inference: A Practical Information-theoretic Approach (2nd Ed.) Springer-Verlag, NY
- Claeskens and Hjort (2008) Model Selection and Model Averaging, Cambridge Univ. Press, UK
- Lahiri, P. (2001) Model Selection, IMS Lecture Notes **38**, IMS
- Linhart, H. and Zucchini, W. (1986). Model Selection, John Wiley & Sons, NY
- McQuarrie, A.D.R. and Tsai, C.-L. (1998) Regression and Time Series Model Selection, World Scientific Publishing Co. INC., NJ

Papers from the slog for testing non nested models.

<http://groundtruth.info/AstroStat/slog/2008/non-nested-hypothesis-tests/>  
More Frequentist and Bayesian Model Selection and Model Averaging papers (heavy on variable selection).

The problem is that the main ideas of these papers cannot be typified into astronomical problems in choosing a right model among candidates. Statistics related to non nested or semi nested model selection heavily rely on regression settings and focused on variable selection.

# P(choosing model A over model B)

We assume the best fit estimators are unbiased.

Devise some likelihood type measure  $L$  so that the goal became assessing  $P(L(A) > L(B))$

$$\begin{aligned}
 P(\text{choosing model A over B}) &= P(L(A) > L(B)) \\
 &= E[P(L(A) > L(B)|D)] \\
 &= \frac{1}{N} \sum P(L(A; D_i) > L(B; D_i)) + o(n^{-\delta}) \quad \delta > 0
 \end{aligned}$$

We know that  $D_i \sim Poi(M_i)$  ( $M_i$  depends on the model choice) and best fits of  $A_p, \alpha, A_b, T$  on  $D$ .