FABADA (Fitting Algorithm for Bayesian Analysis of Data) From parameter determination to model selection

L. C. Pardo[1], G. Sala [2]

[1] Grup de Caracterització de Materials, Universitat Politècnica de Catalunya, Barcelona
[2] Grup d'Astronomia I Caracterització de Materials, Universitat Politècnica de Catalunya, Barcelona

THE METHOD

Bayes Theorem (*applied to fitting*)

Relates the probability that an hypothesis is true given a set of data P(H|Data) (what we want)

with the probability that you get some data, given an hypothesis P(Data|H) (What we measure)

P(H) is the prior knowledge about your hypothesis, P(Data) is a normalization constant

$$P(H|Data) = \frac{P(Data|H)\,P(H)}{P(Data)} \propto P(Data|H)$$

T. Bayes.

...in our case, no prior information will be supposed

Bayesian versus Frequentist

(related to fitting)

Frequentist	Bayesian
We get „values" for parameters ($P_i \pm \varepsilon_i$) and for χ^2	We get Probability Distribution Functions (PDF) for parameters **and for** χ^2
Supposes a single minimum for χ^2 (P_i) and quadratic in all parameters	No suppositions are needed
Model selection only possible if conditions above are fulfilled	Model selection always possible, taking only data into account.
Least Square can get stuck in local minima	Fitting procedure does not get stuck

What is hidden behind the ubiquitous χ^2 ?

If we *suppose* a Gaussian probability distribution of data (D_i) around the values calculated with a fitting function (H_i), with a determined combination of parameters (your hypothesis):

$$P(H|Data) = \prod_{i=1}^{i=Ndata} P(H|Data_i) = \prod_{i=1}^{i=Ndata} \frac{1}{\sigma_i\sqrt{2\pi}}\, e^{-\frac{(H_i-D_i)^2}{2\sigma_i^2}}$$

Let's define now the likelihood as P(H|Data), then its logarithm:

$$\log\, L \propto \sum_{i=1}^{i=Ndata} -\frac{(H_i-D_i)^2}{2\sigma_i^2} = -\chi^2/2$$

Therefore, χ^2 is a measure of the likelihood of our hypothesis. But, in fact the supposition that data (D_i) are distributed as a Gaussian centered in the calculated values (H_i), is only valid for high count rates. ***For low count rates we should use the Poisson distribution, and then***:

$$\log L = \sum_{i=1}^{i=Ndata} \log[P(H|Data_i)] = \sum_{i=1}^{i=Ndata} \log\left[\frac{e^{-H_i} H_i^{D_i}}{D_i!}\right] \propto \sum_{i=1}^{i=Ndata} [-H_i + D_i \cdot \log H_i]$$

Fortunately, spectrometers have usually high enough count rates, and the χ^2 approximation is true. However, understanding what is χ^2 allows us to use different suppositions about how our data are distributed. Therefore, what is hidden behind χ^2 is nothing more than **...PROBABILITY!**

Bayes at work: The Markov Chain Monte Carlo Method

Finding the best fit to data means finding the maximum Likelihood, which in the case of Gaussian distribution is finding the minimum in the $\chi^2(P_i)$ function, where P_i are the values for every parameter of the fitting function.

If we have a set of parameters (P_i^{old}) and we randomly generate a new set of parameters (P_i^{new}), we will accept this new set with a probability:

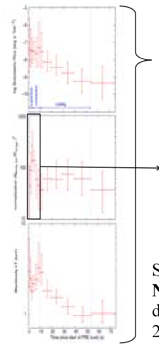
$$\frac{prob(P_i^{new}/Data)}{prob(P_i^{old}/Data)} = e^{-\frac{(\chi^2_{new}-\chi^2_{old})}{2}}$$

Therefore we "explore" the surface $\chi^2(P_i)$ having into account the probability that data are described by the fitting function (or hypothesis), similar to a Monte Carlo simulation where Energy is χ^2 and error is temperature. That has some advantages:

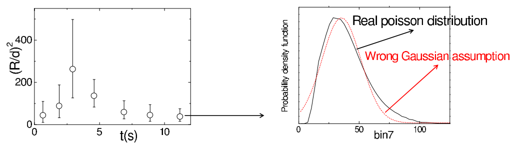
- ✓ All $\chi^2(P_i)$ compatible with data are explored, **without** making any assumption about how the minimum in $\chi^2(P_i)$ depends on parameters (even not supposing that here is a single minimum!)
- ✓ Correlations between parameters are naturally taken into account.
- ✓ We obtain the Probability Distribution Function (PDF) of parameters (containing much more information that simply $P_i \pm \varepsilon_i$)
- ✓ Model selection is easily done **without** making any assumption on $\chi^2(P_i)$, and obtaining (again) a PDF, which can reveal more than one minimum in $\chi^2(P_i)$, i.e. more than one description of data.

THE EXAMPLE

Testing the Photospheric Radius Expansion (PRE) of a type I X-ray burst of the Rapid Burster



Time resolved spectral analysis of Swift/XRT data during 2009 outburst shows a Type I X-ray burst with possible Photospheric Radius Expansion (PRE burst) (Sala et al. 2012, ApJ 752, 178)

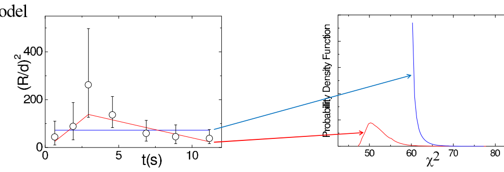


Since **low count rate does not allow to approximate the PDF by a Normal distribution** it is not possible to perform any analysis with χ^2 defined as usually. We redefine it as in Pardo et al (Conf. Ser. 325 012006, 2011) to take into account the asymmetry of the PDF

Model Selection

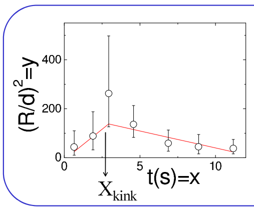
Is the photosphere really expanding?

Test with FABADA the radius evolution: constant radius vs expansion-contraction model



The reduced χ^2 obtained modeling the data by a constant is χ^2=3.3 and by the expansion-contraction model χ^2=1.16. Therefore the expansion is clearly favored. Moreover the analysis performed allows us to calculate the χ^2 associated to *all the combinations of parameters* that are compatible with the data and their errors, i.e. we can obtain the PDF related to χ^2 itself. From the PDF's associated to both models we conclude that no constant calculated compatible with data errors is as good as the worst fit with the expansion-contraction model.

Parameter estimation



The expansion-contraction model:

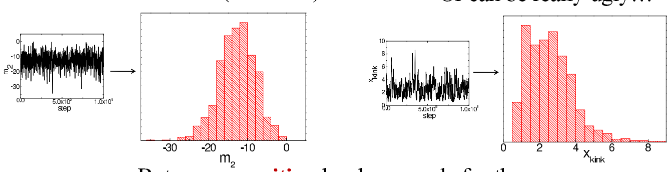
For $x<x_{kink}$ (expansion) $\quad y_1 = a_1 + m_1 x$

For $x>x_{kink}$ (contraction) $\quad y_2 = a_2 + m_2 x$

Continuity: $\quad a_2 = a_1 + (m_1 - m_2)\, x_{kink}$

We obtain **Probability Distribution Functions (PDF)** for each parameter P, instead of $P \pm \delta P$. As an example we show the PDF's associated to two parameters:

The PDF can be "nice" (Gaussian) Or can be really ugly…



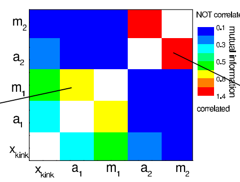
But **no supposition** has been made for them
(in classical fittings it is **imposed** to be a Gaussian)

Parameter correlation

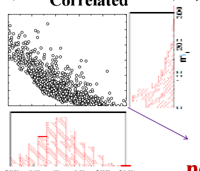
We can quantify correlation between variables using the mutual information: $MI = S(x) + S(y) - S(x,y)$

Where $S(x) = \sum_{\forall x} P(x)\ln P(x)$ and $S(x,y) = \sum_{\forall x,y}^{n} P(x,y)\ln P(x,y)$

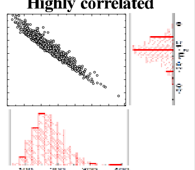
We calculate MI for all combinations of two parameters



Correlated



Highly correlated



Mutual information has also in to account **non linear correlations**!!