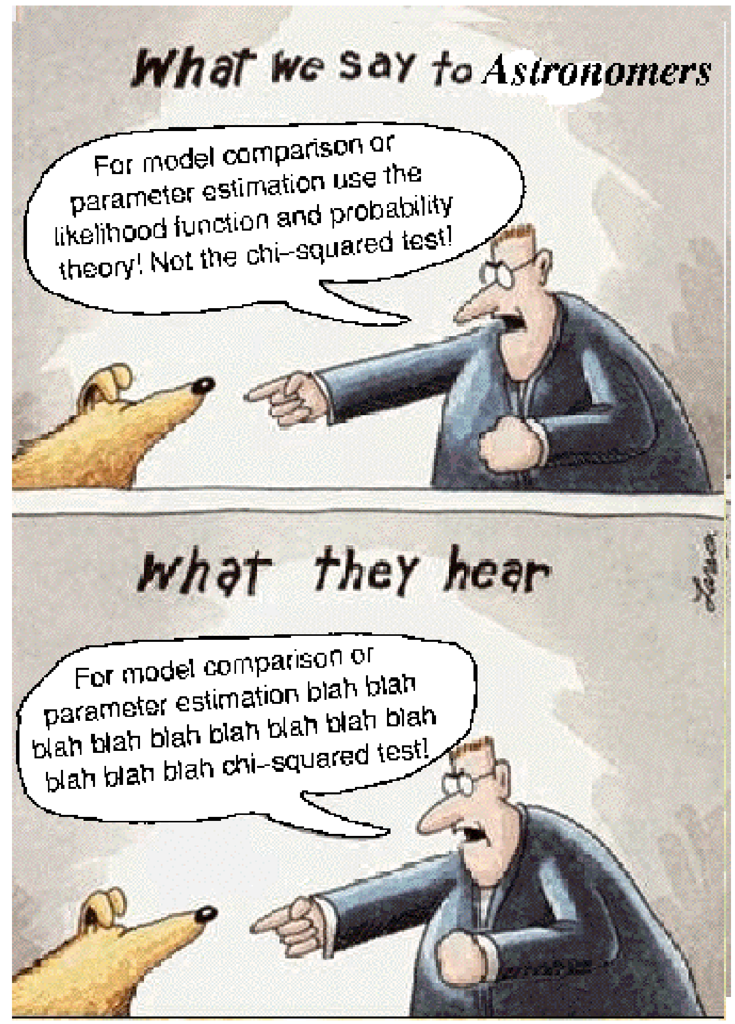
X-ray Astrostatistics Bayesian Methods in Data Analysis

Jeremy Drake, Nov.2005

Aneta Siemiginowska
Vinay Kashyap
and CHASC

X-ray Astrostatistics
Bayesian Methods
in Data Analysis



Jeremy Drake, Nov.2005

Aneta Siemiginowska
Vinay Kashyap
and CHASC

CHASC: California-Harvard Astrostatistics Collaboration

- http://hea-www.harvard.edu/AstroStat/
- History: why this collaboration?
- Regular Seminars: each second Tuesday at the Science Center
- Participate in SAMSI workshop => Spring 2006
- Participants: HU Statistics Dept., Irvine UC, and CfA astronomers
- Topics related mostly to X-ray astronomy, but also sun-spots!
- Papers: MCMC for X-ray data, Fe-line and F-test issues, EMC2, hardness ratio and line detection
- Algorithms are described in the papers => working towards public release

Stat: David van Dyk, Xiao-Li Meng, Taeyoung Park, Yaming Yu, Rima Izem
Astro: Alanna Connors, Peter Freeman, Vinay Kashyap, Aneta Siemiginowska Andreas Zezas, James Chiang, Jeff Scargle

X-ray Data Analysis and Statistics

- Different type analysis: Spectral, image, timing.
- XSPEC and Sherpa provide the main fitting/modeling environments
- X-ray data => counting photons:

  -> normal - Gaussian distribution for high number of counts, but very often we deal with **low counts data**
- Low counts data (< 10)
  - => Poisson data and χ^2 is not appropriate!
- Several modifications to χ^2 have been developed:
  - Weighted $\chi 2$ (.e.g. Gehrels 1996)
- Formulation of Poisson Likelihood (ΔC follows $\Delta\chi 2$ for N>5)
  - Cash statistics: (Cash 1979)
  - C-statistics - goodness-of-fit and background (in XSPEC, Keith Arnaud)

Steps in Data Analysis

- Obtain data – observations!
- Reduce – processing the data, extract image, spectrum etc.
- Analysis – Fit the data
- Conclude – Decide on Model, Hypothesis Testing!
- Reflect

Hypothesis Testing

- How to decide which model is better?

  A simple power law or blackbody?

  A simple power law or continuum with emission lines?

- Statistically decide: how to reject a simple model and accept more complex one?
- Standard (Frequentist!) Model Comparison Tests:
  - Goodness-of-fit
  - Maximum Likelihood Ratio test
  - F-test

Steps in Hypothesis Testing - I

1/ Set up 2 possible exclusive hypotheses:

M0 – null hypothesis – formulated to be rejected

M1 – an alternative hypothesis, research hypothesis

each has associated terminal action

2/ Specify a priori the significance level α

choose a test which:
- approximates the conditions
- finds what is needed to obtain the sampling

distribution and the region of rejection, whose area is a fraction of the total area in the sampling distribution

3/ Run test: reject ***M0*** if the test yields a value of the statistics whose probability of occurance under ***M0*** is $<\alpha$

4/ Carry on terminal action

Steps in Hypothesis Testing - II

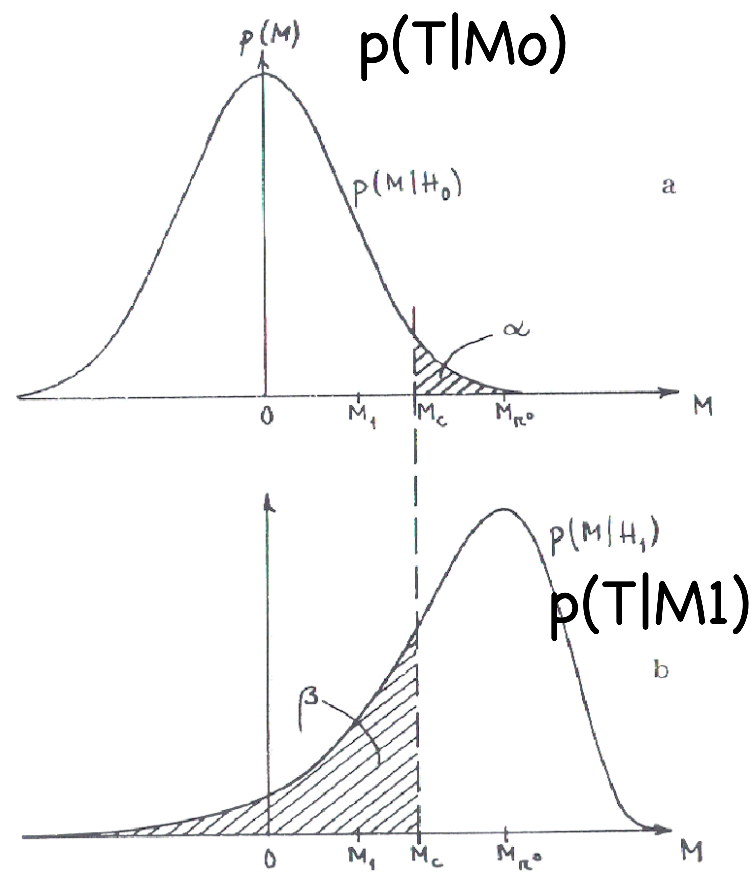
- Two model Mo (simpler) and M1 (more complex) were fit to the data D; Mo => null hypothesis.
- Construct test statistics T from the best fit of two models:

  e.g. $\Delta\chi^2 = \chi^2_0 - \chi^2_1$
- Determine each sampling distribution for T statistics, e.g.

  p(T | Mo) and p(T | M1)
- Determine significance α =>

  Reject Mo when p (T | Mo) < α
- Determine the power of the test =>

  β – probability of selecting Mo when M1 is correct



Conditions for LRT and F-test

- The two models that are being compared have to be **nested:**
  - broken power law is an example of a nested model
  - BUT power law and thermal plasma models are **NOT** nested
- The null values of the additional parameters may **not be on the boundary** of the set of possible parameter values:
  - continuum + emission line

    -> line intensity = 0 on the boundary
- References

Freeman et al 1999, ApJ, 524, 753
Protassov et al 2002, ApJ 571, 545

Simple Steps in Calibrating the Test:

1. Simulate N data sets (e.g. use fakeit in Sherpa or XSPEC):
   => the null model with the best-fit parameters (e.g. power law, thermal)
   => the same background, instrument responses, exposure time as in the initial analysis
2. (A) Fit the null and alternative models to each of the N simulated data sets

   and

   (B) compute the test statistic:

   $$T_{LRT} = -2\log\,[L(\theta_0|sim)/L(\theta_1|sim)]$$

   θ_0 θ_1 – best fit parameters

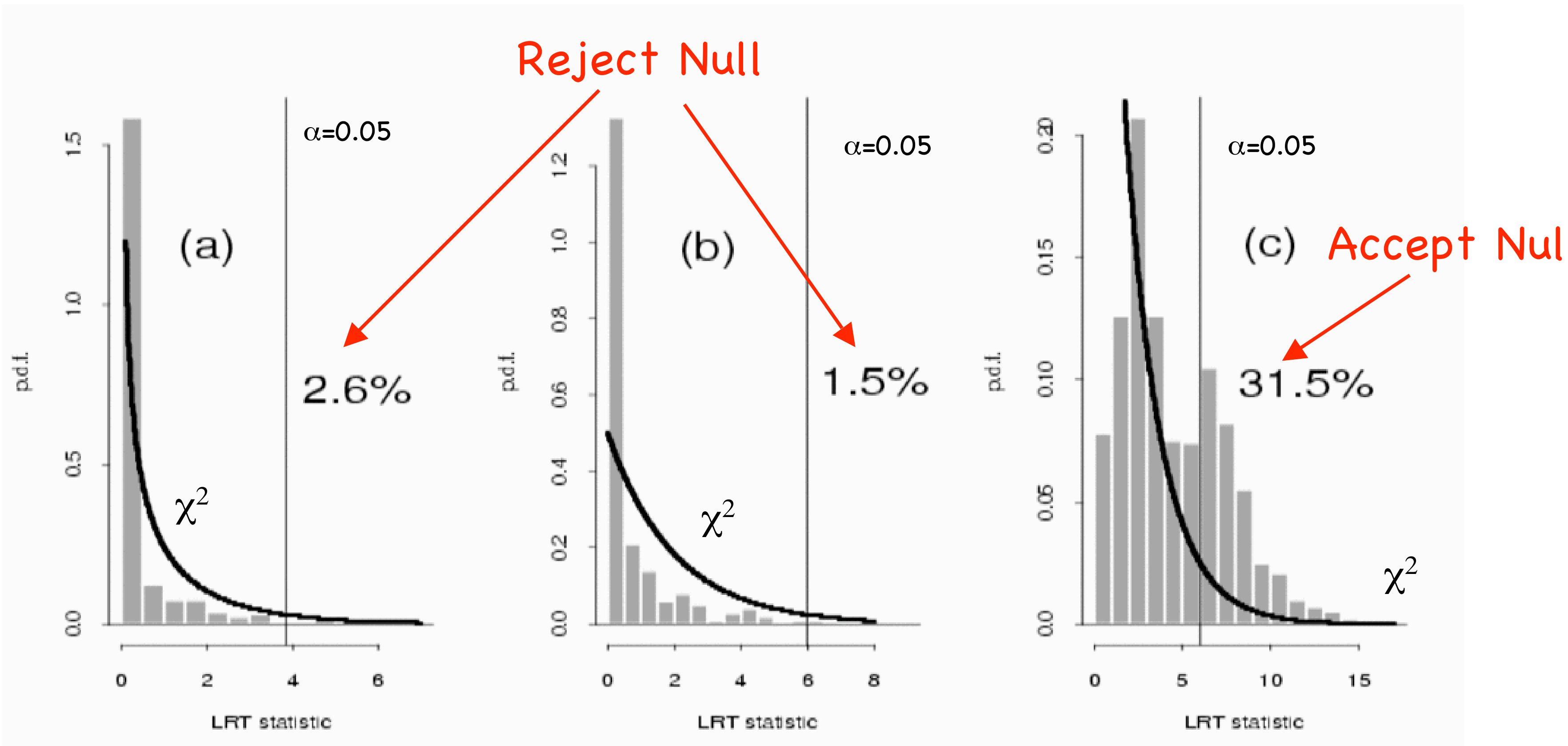
   $$T_F = \Delta\chi^2/\chi^2_\nu$$
3. Compute the p-value - proportion of simulations that results in a value of statistic (T) more extreme than the value computed with the observed data.

   p-value = (1/N) * Number of [T(sim) > T(data)]

Simulation Example

Comparison between p-value
And significance in the χ^2 distribution

M0 - power law
M1 - pl+narrow line
M2 - pl+broad line
M3 - pl+absorption line



M0/M1

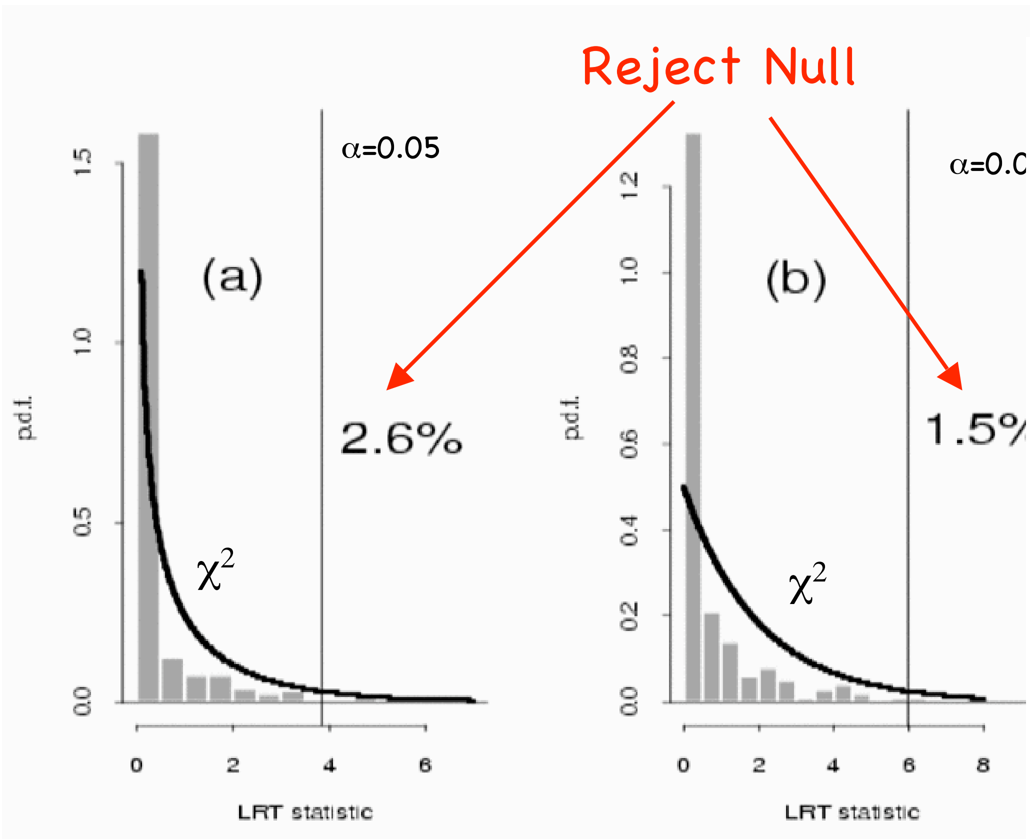
M0/M2

M0/M3

Simulation Example

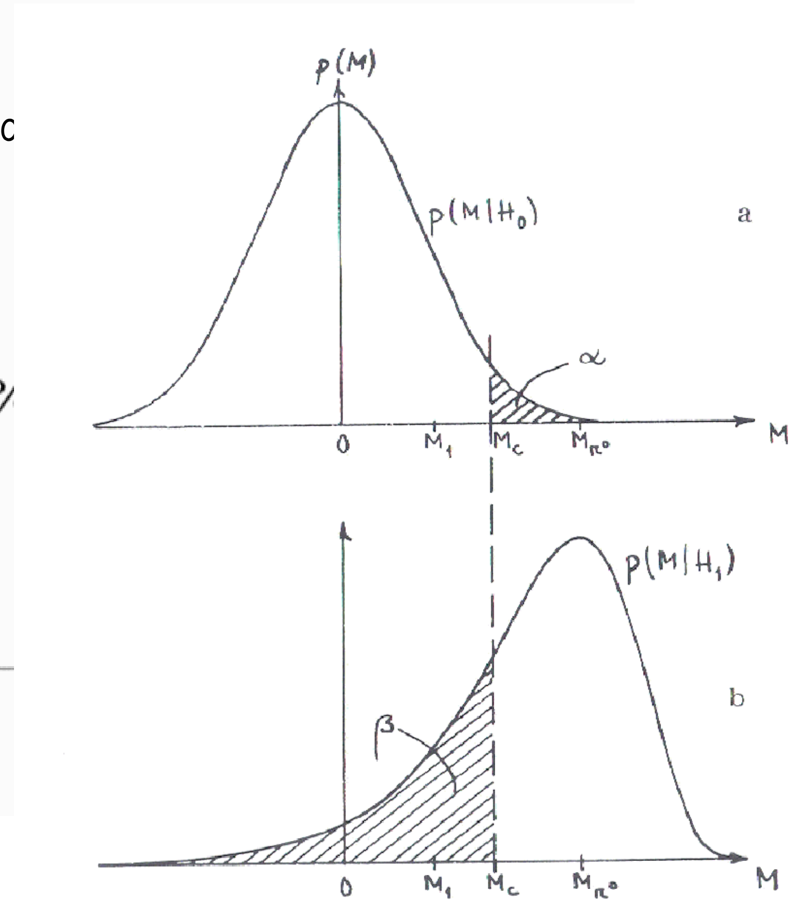
Comparison between p-value
And significance in the χ^2 distribution

M0 - power law
M1 - pl+narrow line
M2 - pl+broad line
M3 - pl+absorption line



M0/M1

M0/M2



Bayesian Methods

- use Bayesian approach - max likelihood, priors, posterior distribution - to fit/find the modes of the posterior (best fit parameters)
- Simulate from the posterior distribution, including uncertainties on the best-fit parameters,
- Calculate posterior predictive p-values
- Bayes factors:

  direct comparison of probabilities P(M1)/P(Mo)

CHASC Projects at SAMSI 2006

- Source and Feature detection Working group
- Issues in Modeling High Counts Data
    - Image reconstructions (e.g. Solar data)
    - Detection and upper limits in high background data (GLAST)
    - Smoothed/unsharp mask images - significance of features
- Issues in Low Counts Data

  Upper limits

  Classification of Sources - point source vs. extended

  Poisson data in the presence of Poisson Background

  Quantification of uncertainty and Confidence

Other Projects in Town:

Calibration uncertainties in X-ray analysis

Emission Measure model for X-ray spectroscopy

(Log N - Log S) model in X-ray surveys

Bayesian Model Comparison

To compare two models, a Bayesian computes the odds, or odd ratio:

$$O_{10} = \frac{p(M_1 \mid D)}{p(M_0 \mid D)}$$

$$= \frac{p(M_1)\,p(D \mid M_1)}{p(M_0)\,p(D \mid M_0)}$$

$$= \frac{p(M_1)}{p(M_0)} B_{10}\ ,$$

where B_{10} is the *Bayes factor*. When there is no *a priori* preference for either model, $B_{10} = 1$ of one indicates that each model is equally likely to be correct, while $B_{10} \geq 10$ may be considered sufficient to accept the alternative model (although that number should be greater if the alternative model is controversial).

Bayesian Model Comparison

we showed how Bayes' theorem is applied in model fits. It can also be applied to model comparison:

$$p(M \mid D) = p(M)\,\frac{p(D \mid M)}{p(D)}.$$

p(***M***) is the prior probability for ***M***;

p(***D***) is an ignorable normalization constant; and

p(***D*** | ***M***) is the average, or global, likelihood:

$$p(D \mid M) = \int d\theta\, p(\theta \mid M)\, p(D \mid M, \theta)$$

$$= \int d\theta\, p(\theta \mid M)\, L(M, \theta).$$

In other words, it is the (normalized) integral of the posterior distribution over all parameter space. Note that this integral may be computed numerically, by brute force, or if the likelihood surface is approximately a multi-dimensional Gaussian ($i.e.$ if $L \propto \exp[-\chi^2/2]$), by the ***Laplace approximation:***

$$p(D \mid M) = p(\hat{\theta} \mid M)(2\pi)^{P/2}\sqrt{\det C}\, L_{\max},$$

where C is the covariance matrix (estimated numerically at the mode).

Model Comparison Tests

- A model comparison test statistic T is created from the best-fit statistics of each fit; it is sampled from a probability distribution p(T). The test significance is defined as the integral of p(T) from the observed value of T to infinity. The significance quantifies the probability that one would select the more complex model when in fact the null hypothesis is correct. A standard threshold for selecting the more complex model is significance < 0.05 (the "95% criterion" of statistics).