AstroStatistics

Aneta Siemiginowska
Harvard-Smithsonian Center for Astrophysics

OUTLINE

- Introduction and general consideration
    - Motivation: why do we need statistics?
    - Probabilities/Distributions
    - Frequentist vs. Bayes
- Statistics in X-ray Analysis:
    - Poisson Likelihood
    - Parameter Estimation
    - Hypothesis Testing
- References and Summary

Why do we need Statistics?

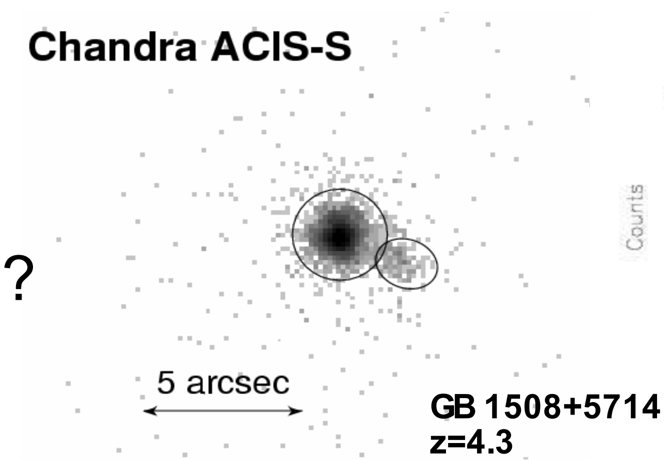
- How do we take decisions in Science?

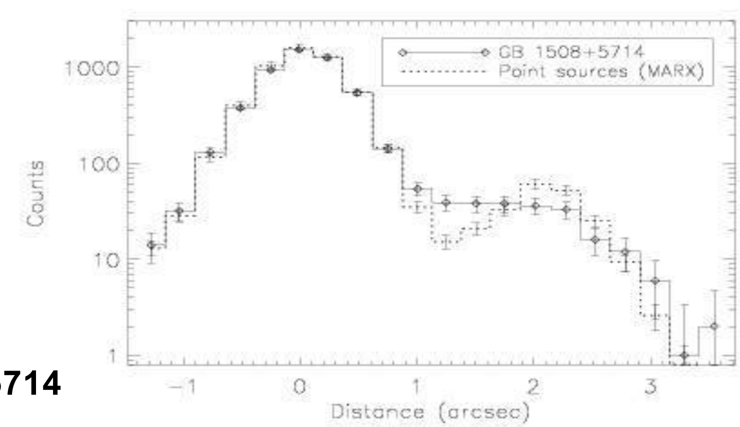
  Tools: instruments, data collections, reduction, classifications – tools and techniques

  Decisions: is this hypothesis correct? Why not? Are these data consistent with other data? Do we get an answer to our question? Do we need more data?

- Comparison to decide:
  - Describe properties of an object or sample:

Example:
Is a faint extension
a jet or a point source?





Siemiginowska et al (2003)

Stages in Astronomy Experiments

Stage	How	Example	Considerations
OBSERVE	Carefully	Experiment design, exposure time (S)	What? Number of objects, Type? (S)
REDUCE	Algorithms	calibration files QE,RMF,ARF,PSF (S)	data quality Signal-to-Noise (S)
ANALYSE	Parameter Estimation, Hypothesis testing (S)	Intensity, positions (S)	Frequentist Bayesian? (S)
CONCLUDE	Hypothesis testing (S)	Distribution tests, Correlations (S)	Belivable, Repeatable, Understandable? (S)
REFLECT	Carefully	Mission achieved? A better way? We need more data! (S)	The next Observations (S)

Wall & Jenkins (2003)

Probability

quantifies randomness and uncertainty

Statistics

uses probability to make scientific inferences based on observations

Bayesian

Probability quantifies degree of belief that an event will occur

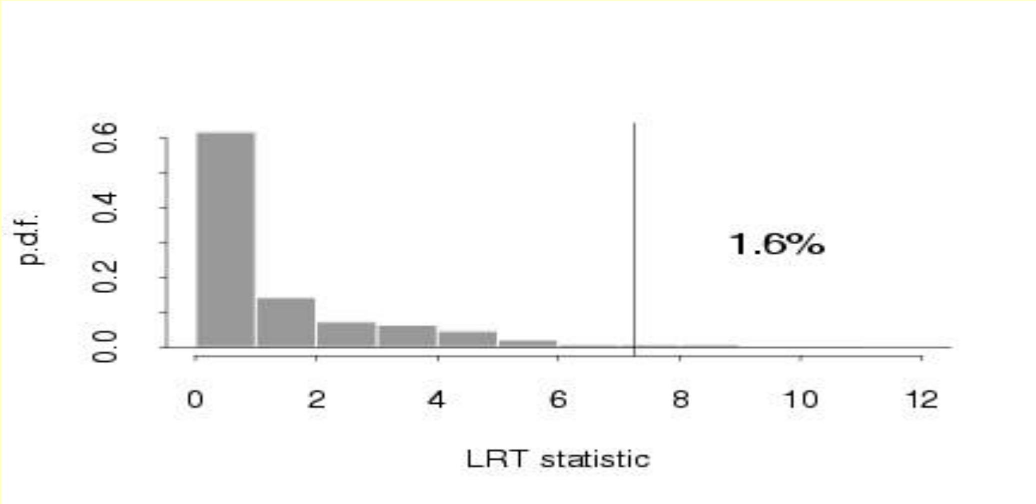
Frequentist

Probability is the relative frequency of an events occuring, in the limit of infinite number of trials.

Probability Distributions

Probability is crucial in decision process:

Example:



Limited data yields only partial idea about the line width in the spectrum. We can only assign the probability to the range of the line width roughly matching this parameter. We decide on the presence of the line by calculating the probability.

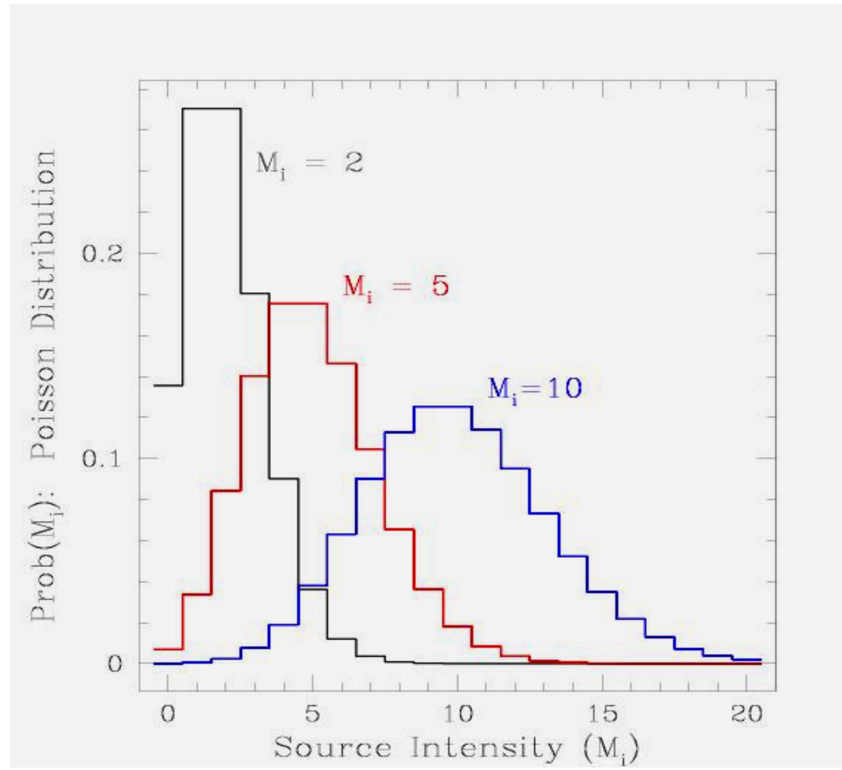
The Poisson Distribution

Collecting X-ray data => Counting individual photons
=> Sampling from Poisson distribution

The discrete Poisson distribution:

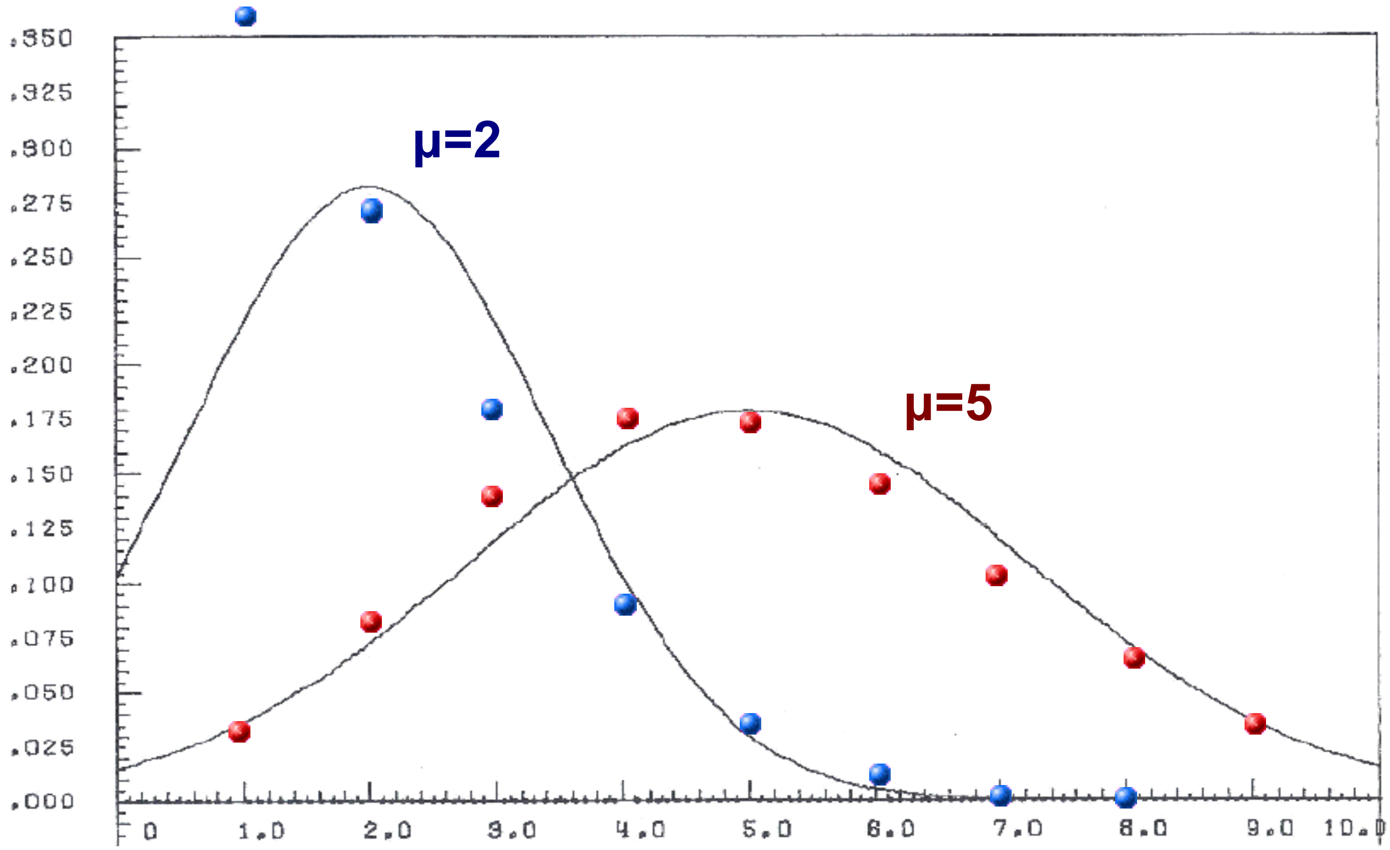
$$p(D_i \mid M_i) = \frac{M_i^{D_i}}{D_i!} e^{-M_i}$$

probability of finding D_i events (*counts*) in bin *i* (*energy rage*) of dataset *D* (*spectrum*) in a given length of time (exposure time), if the events occur independently at a constant rate M_i *(source intensity).*



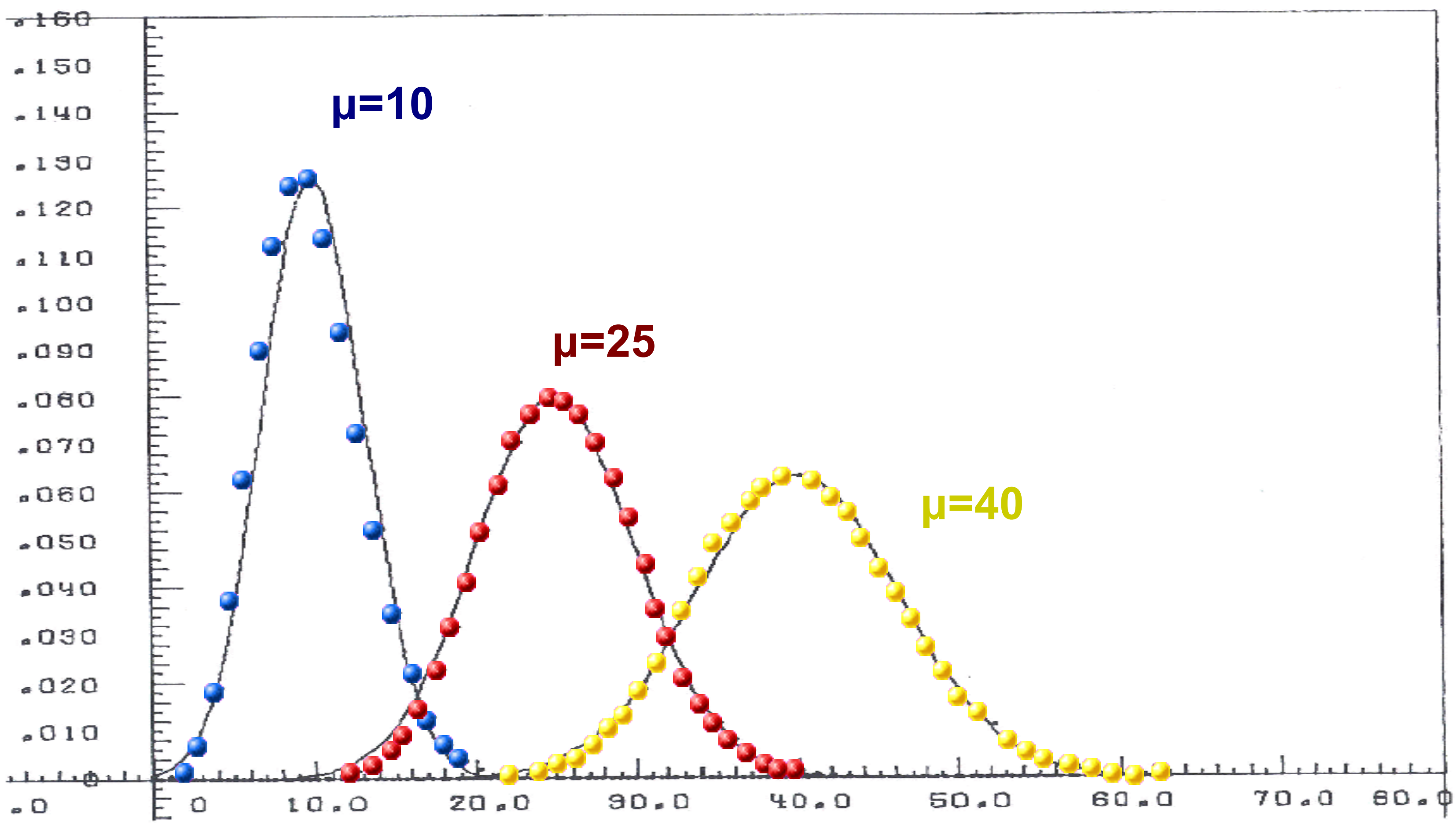
As M_i => ∝ Poisson distribution converges to Gaussian distribution N($\mu = M_i$; $\sigma^2 = M_i$)

Poisson vs. Gaussian Distributions – Low Number of Counts



Comparison of Poisson distributions (dotted) of mean μ = 2 and 5 with normal distributions of the same mean and variance (Eadie *et al.* 1971, p. 50).

Poisson vs. Gaussian Distributions



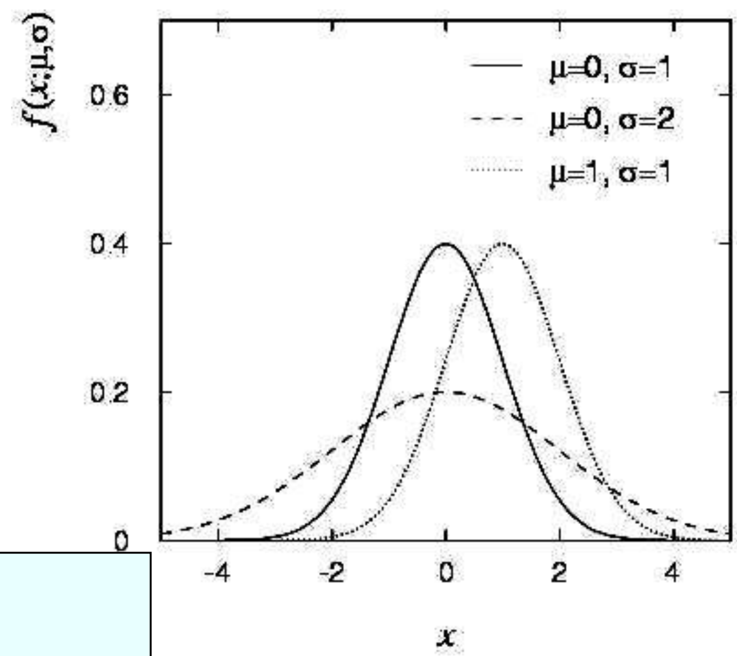
Comparison of Poisson distributions (dotted) of mean $\mu = 10, 25$ and 40 with normal distributions of the same mean and variance (Eadie *et al.* 1971, p. 50).

Gaussian Distribution

For large *counts* Poisson (and the Binomial) distributions converges to Gaussian (normal) distributions.

$$\text{prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\mu)^2/2\sigma^2]$$

Mean - μ
Variance - σ^2



Note: Importance of the Tails!

$\pm 2\sigma$ range covers *95.45%* of the area, so 2σ result has less than *5%* chance of occurring by chance, but because of the error estimates this is not the acceptable result. Usually 3σ or 10σ have to be quoted and the convergence to Gaussian is faster in the center than in the tails!

What do we do in X-rays?

Example:

I've observed my source, reduce the data and finally got my X-ray spectrum – what do I do now? How can I find out what does the spectrum tell me about the physics of my source?

Run **XSPEC** or ***Sherpa*!** But what do those programs really do?

Fit the data => $C(h) = \int R(E,h)\, A(E)\, M(E,\theta)\, dE$

Counts; Response; Effective Area; Model

h- detector channels
E- Energy
θ- model parameters

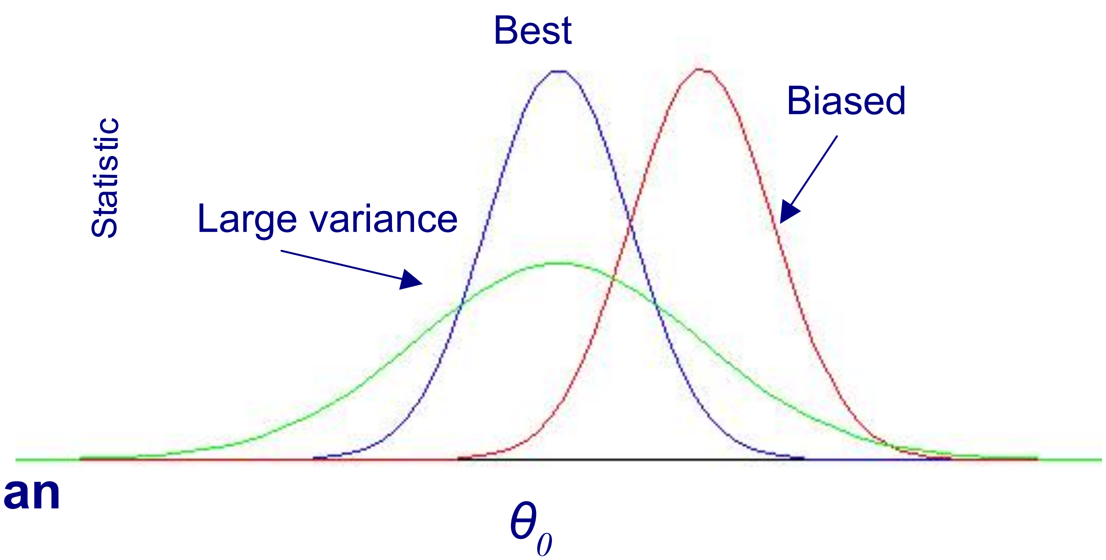
Assume a model and look for the best model parameters which describes the observed spectrum.

Need a Parameter Estimator - Statistics

Parameter Estimators: Statistics

Requirements on Statistics:

- **Unbiased**
  - converge to true value with repeated measurements
- **Robust**
  – less affected by outliers
- **Consistent**
  – true value for a large sample size (Example: rms and Gaussian distribution)
- **Closeness**
  - smallest variations from the truth



Maximum Likelihood: Assessing the Quality of Fit

Use the Poisson distribution to assess the probability of sampling data $\boldsymbol{D_i}$ given a predicted (convolved) model amplitude M_i. Thus to assess the quality of a fit, it is natural to maximize the product of Poisson probabilities in each data bin, *i.e.*, to maximize the Poisson likelihood:

$$\mathcal{L} = \prod_i^N \mathcal{L}_i = \prod_i^N \frac{M_i^{D_i}}{D_i!} \exp(-M_i) = \prod_i^N p(D_i \mid M_i)$$

In practice, what is often maximized is the log-likelihood,

L = log$\mathcal{L}$. A well-known statistic in X-ray astronomy which is related to *L* is the so-called "Cash statistic":

$$C \equiv 2\sum_i^N [M_i - D_i \log M_i] \propto -2L,$$

Likelihood Function

Likelihood

Observed Counts

Probability Distribution

Model parameters

$$L(X_1,X_2,\ldots.X_N) = P(X_1,X_2,\ldots.X_N|\Theta)$$

$$= P(X_1\,|\Theta)\,P(X_2\,|\Theta)\ldots.\,P(X_N\,|\Theta)$$

$$= \prod P(X_i|\Theta)$$

P - Poisson Probability Distribution for X-ray data

$X_1,\ldots.X_N$ - X-ray data - independent

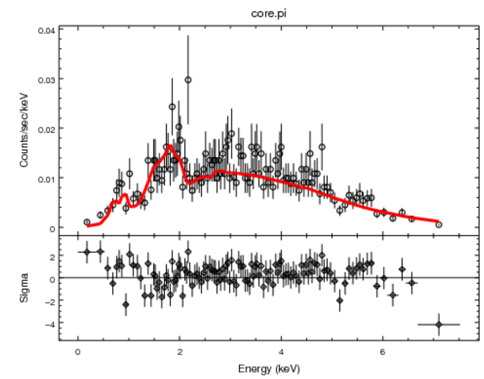
Θ - model parameters

Likelihood Function: X-rays Example

- X-ray spectra modeled by a power law function:

$$f(E)= A * E^{-\gamma}$$

E - energy; A, γ- model parameters: a normalization and a slope



Predicted number of counts:

$$M_i = \int R(E,i) * A(E) \, AE^{-\gamma} dE$$

For A = 0.001 ph/cm^2/sec, Γ=2 then in channels i= (10, 100, 200)

Predicted counts: M_i = (10.7, 508.9, 75.5)

Observed X_i = (15, 520, 74)

Calculate individual probabilities:

Use Incomplete Gamma Function

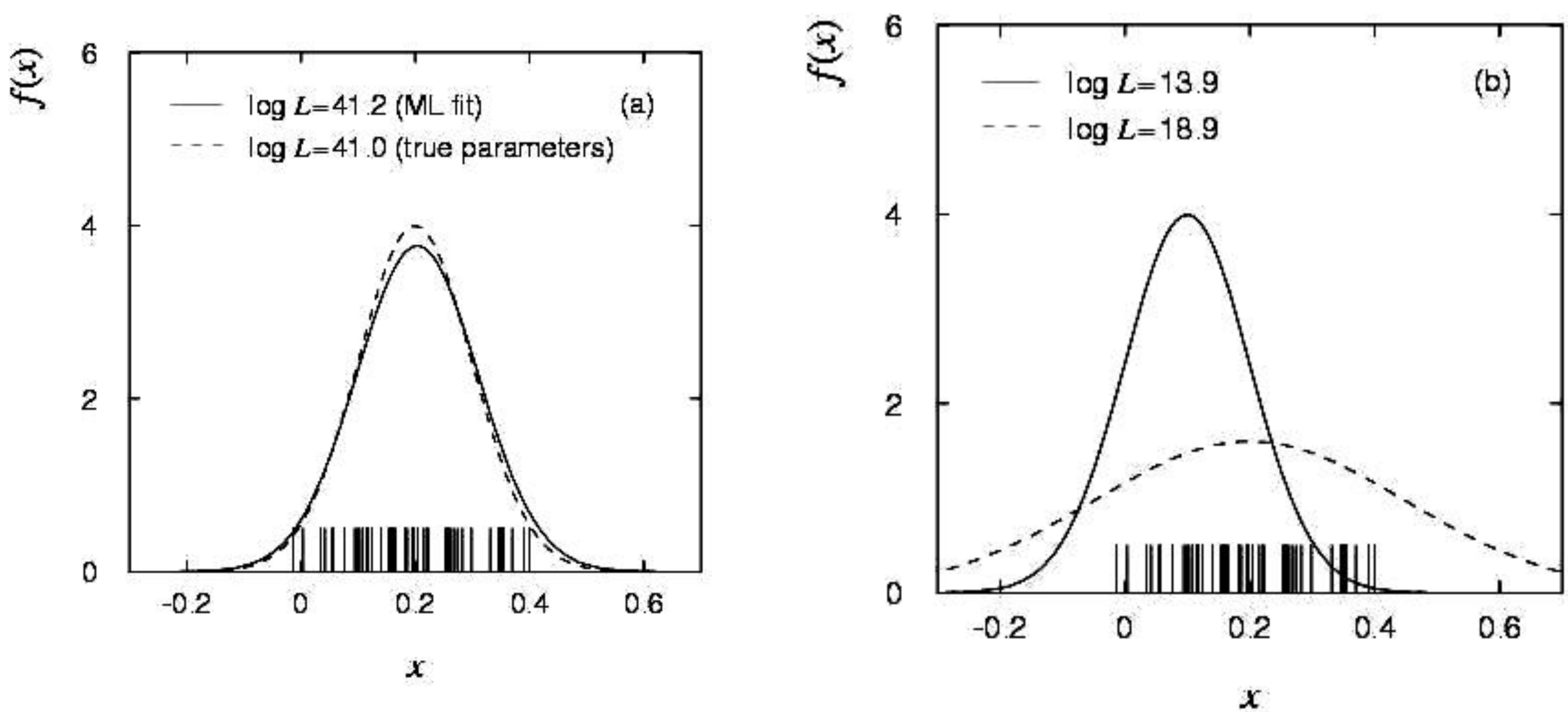
$\Gamma(X_i, M_i)$

$$\mathcal{L}(\{X_i\}) = \prod^{N} \mathcal{P}(X_i; M_i(A, \gamma))$$
$$= \mathcal{P}(15; 10.7)\mathcal{P}(520; 508.9)\mathcal{P}(74; 75.5)$$
$$= 3.37 \times 10^{-5}$$

- Finding the maximum likelihood means finding the set of model parameters that maximize the likelihood function

Maximum Likelihood

If the hypothesized θ is close to the true value, then we expect a high probability to get data like that which we actually found.



(Non-) Use of the Poisson Likelihood

In model fits, the Poisson likelihood is not as commonly used as it should be. Some reasons why include:

- a historical aversion to computing factorials;
- the fact the likelihood cannot be used to fit "background subtracted" spectra;
- the fact that negative amplitudes are not allowed (not a bad thing physics abhors negative fluxes!);
- the fact that there is no "goodness of fit" criterion, i.e. there is no easy way to interpret $\mathcal{L}_{max}$ (however, *cf*. the **CSTAT** statistic); and
- the fact that there is an alternative in the Gaussian limit: the χ^2 statistic.

χ^2–Statistic

$$\chi^2 \equiv \sum_i^N \frac{(D_i - M_i)^2}{\sigma_i^2}$$

The χ^2 statistics is **minimized** in the fitting the data, varying the model parameters until the best-fit model parameters are found for the minimum value of the χ^2 statistic

Degrees-of-freedom = **k-1- N**

N – number of parameters
K – number of spectral bins

“Versions” of the χ^2 Statistic

Generally, the χ^2 statistic is written as:

$$\chi^2 \equiv \sum_i^N \frac{(D_i - M_i)^2}{\sigma_i^2} \quad ,$$

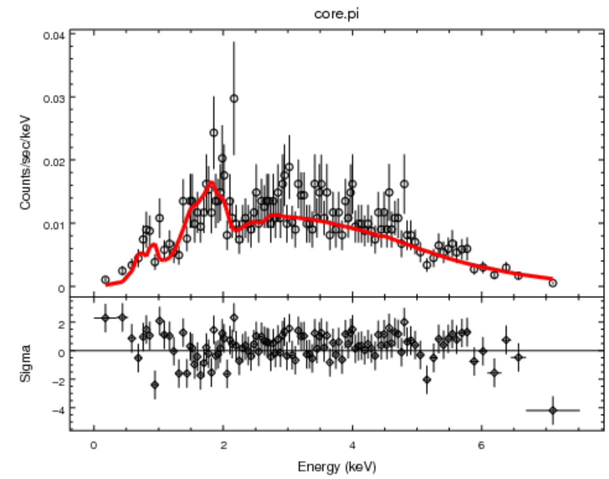
where σ_i^2 represents the (unknown!) variance of the Poisson distribution from which D_i is sampled.

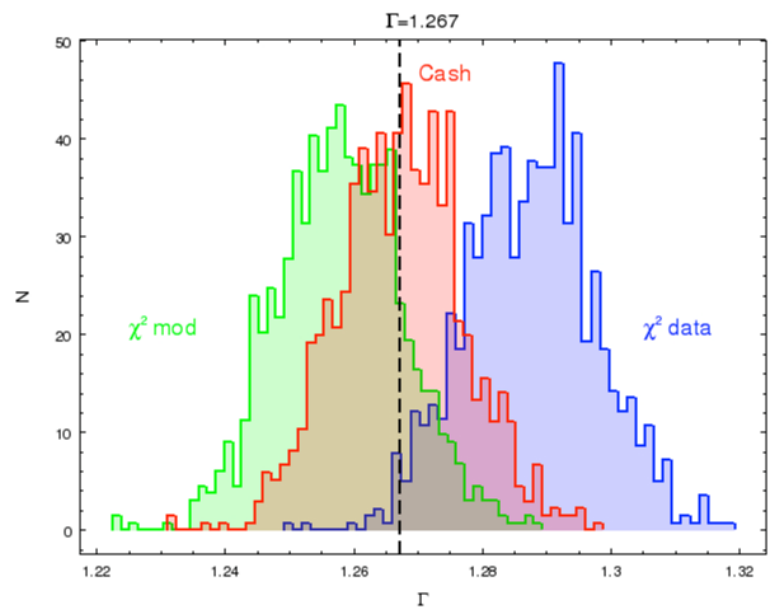
χ^2 **Statistic**	σ_i^2
Data Variance	D_i
Model Variance	M_i
Gehrels	$[1+(D_i+0.75)^{1/2}]^2$
Primini	M_i from previous best-fit
Churazov	based on *smoothed* data D
“Parent”	$\frac{\sum_{i=1}^N D_i}{N}$
Least Squares	1

Note that some X-ray data analysis routines may estimate σ_i for you during data reduction. In PHA files, such estimates are recorded in the **STAT_ERR** column.

Low Counts X-ray Data

- Standard X-ray analysis in XSPEC or Sherpa
- Parameterized Forward Fitting of the data
- Assuming statistics - typically χ^2
- Modified/weight χ^2 to account for low counts
- Bias when the true distributions are not normal.
- Poisson data - need to use the Poisson likelihood (e.g. Cash)
- MCMC methods probe the entire parameter space and do not get stuck in local minima (i.e. it can get out).





Bayesian Model For Low Counts Data

Bayesian Framework

likelihood → prior

Posterior distribution → $$p(\theta|d,I) = \frac{p(d|\theta,I)\,p(\theta|I)}{p(d|I)}$$

$p(d|I)$ constant

θ - model parameters

d - observed data

I - initial information

Poisson Likelihood

$$p(d|\lambda_s,\lambda_b,I) = \frac{\exp^{-(\lambda_s+\lambda_b)}\,(\lambda_s+\lambda_b)^d}{d!}$$

data (d), source (λ_s), background (λ_b)

Bayesian Model For Low Counts Data

Model Predicted X-ray Spectra

Predicted Intensity = Instrument Response (Source Model Intensity x Effective Area) + Background

θ_s parameters

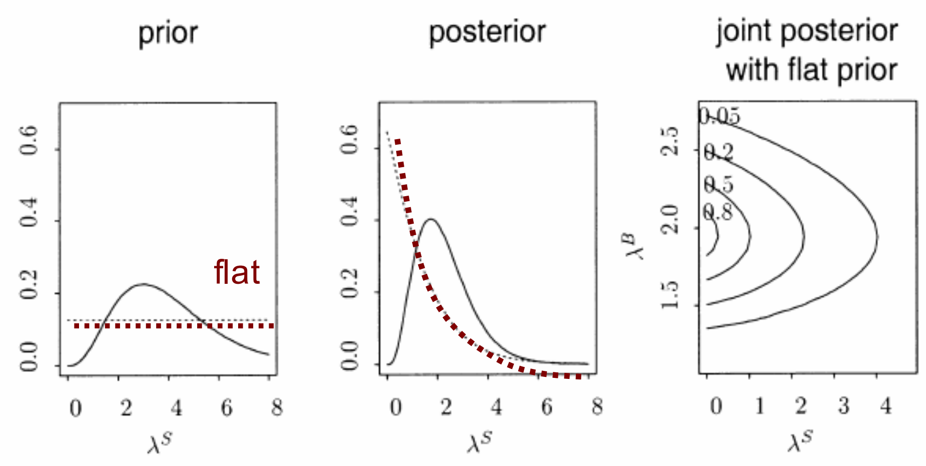
θ_b parameters

Model
$\lambda_s(\theta_s) + \lambda_b(\theta_b)$

Combining information

Prior

- allows us to include a priori knowledge, e.g. range of parameters
- non-informative - e.g. flat within the range
- normal, log-normal, γ - gamma etc.



Simulations from Posterior

- Example:
  - An absorbed power law model => $M_j(a,\Gamma,N_H) = a * E_j^{-\Gamma} * f_j(N_H)$
  - Poisson Likelihood:

$$\prod_{j=1}^{J} \frac{e^{-Mj}\, M_j^{dj}}{d_j!}$$

Log-likelihood $\sum_j -M_j + d_j \log(M_j)$ (similar to Cash)

Gaussian distributions are typical prior distributions for (a,Γ,N_H) and

Log Posterior Distribution is then:

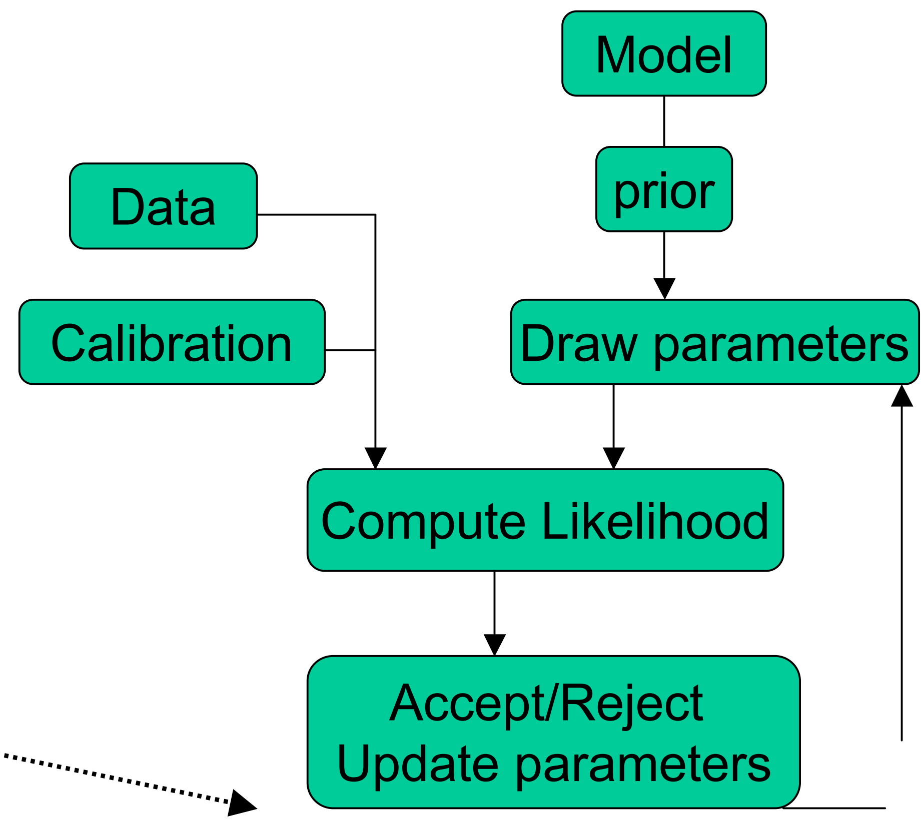
$$\sum_j [-M_j + d_j \log(M_j)\,] + [\,\log G(\log(a),\mu_a,\sigma_a) + \log G(\Gamma,\mu_\Gamma,\sigma_\Gamma) + \log G(N_H,\mu_N,\sigma_N)\,]$$

Simulations from Posterior

$$\sum_j [-M_j + d_j \log(M_j)\ ] + \ [\log\ G(\log(a),\mu_a,\sigma_a) + \log\ G(\Gamma,\mu_\Gamma,\sigma_\Gamma) + \log\ G(N_H\ ,\mu_N,\sigma_N)\ ]$$

Likelihood

prior



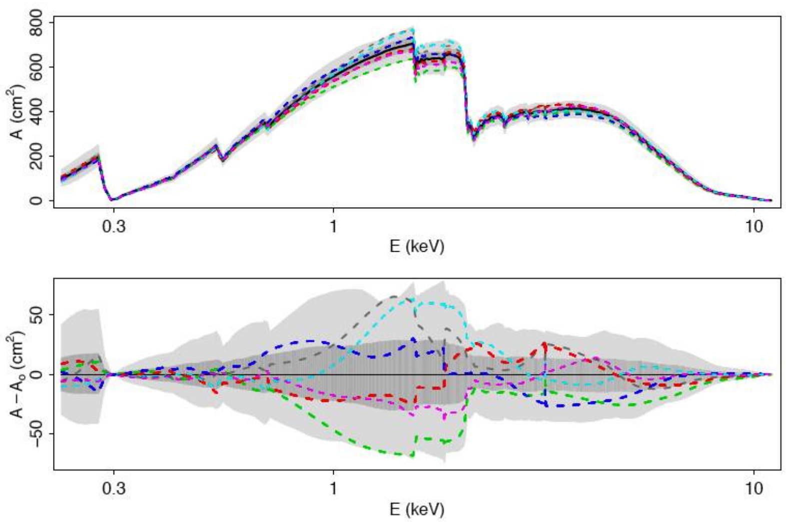
Simulation from the posterior distribution requires careful and efficient algorithms:

Draw parameters from a "proposal distribution", calculate likelihood and posterior probability of the "proposed" parameter value given the observed data, use a Metropolis-Hastings criterion to accept or reject the "proposed" values.

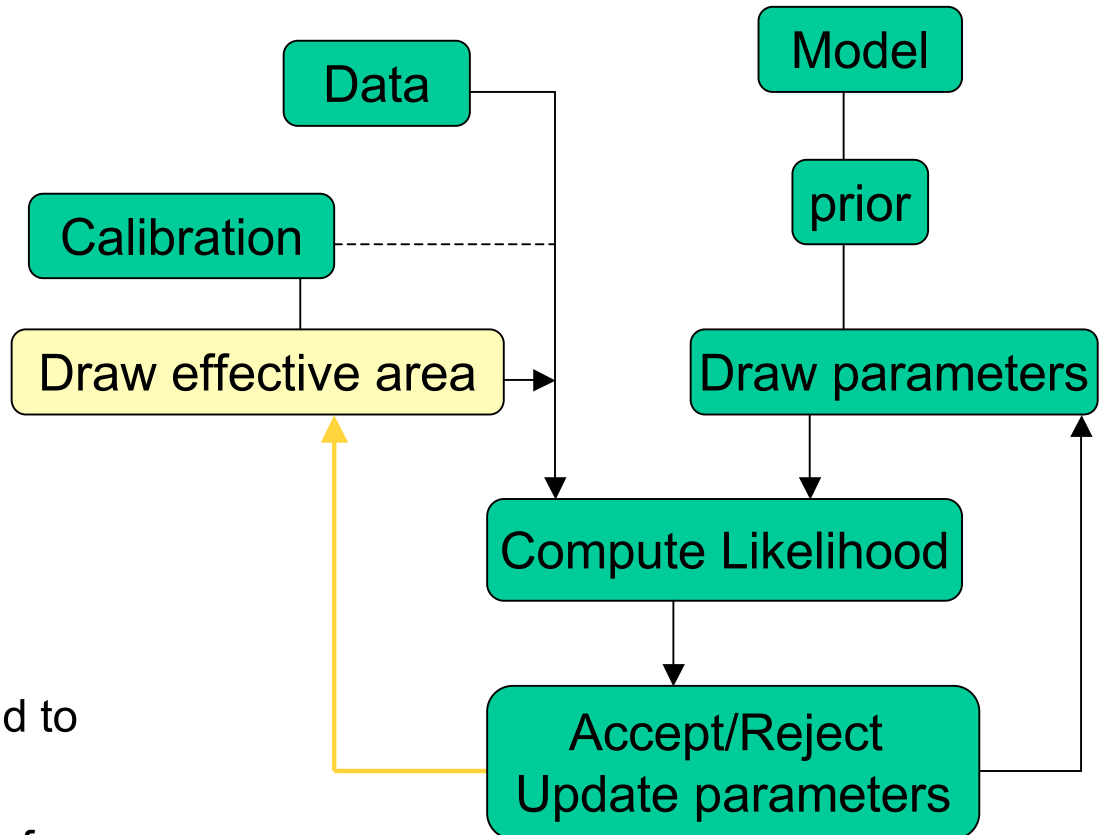
Included in pyBlocxs:
http://hea-www.harvard.edu/AstroStat/pyBLoCXS/

Application: Systematic Errors Calibration Uncertainties

Chandra ACIS-S Effective Area



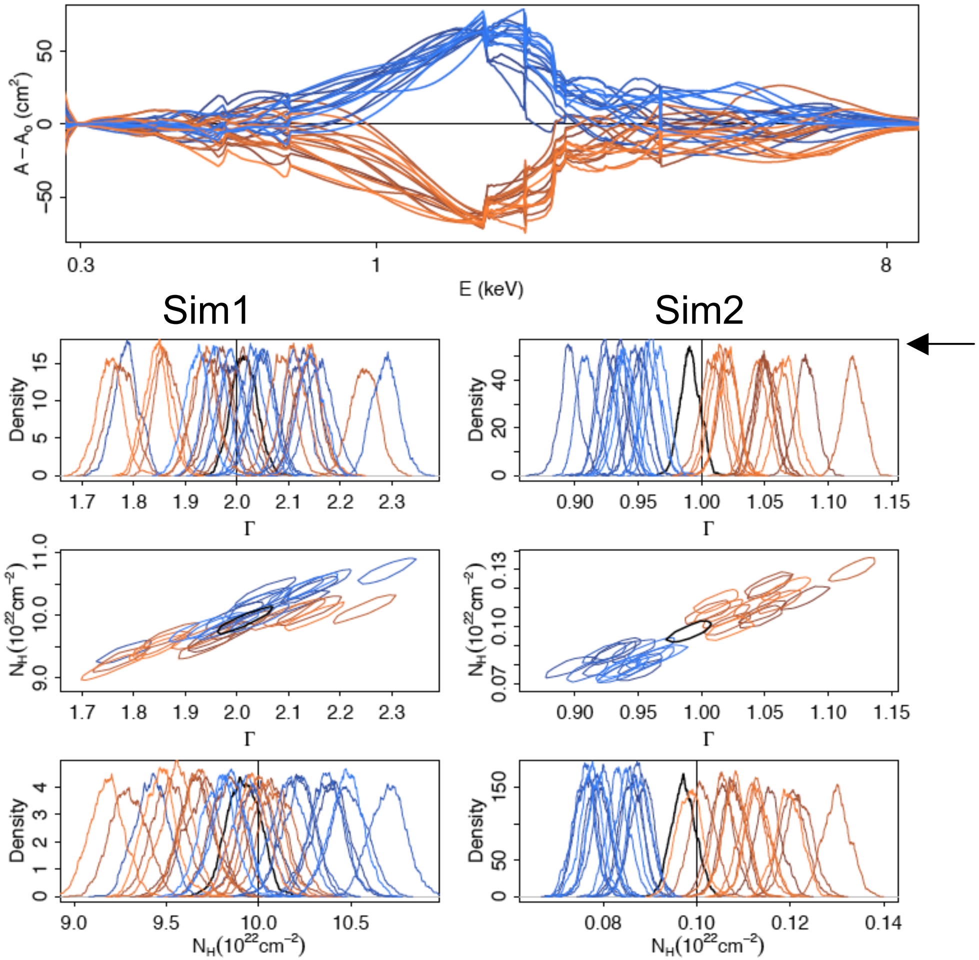
Drake et al. 2006 Proc. SPIE, 6270,49



- Non-linear errors cannot simply add to stats errors.
- Include a draw from an ensemble of effective area curves in the simulations.

Application: Calibration Uncertainties

Deviations from the default ARF (A_0)



Effects of ARF uncertainty on parameters

Simulations of 10^5 counts

Sim1: $\Gamma=2$ N_H=1e23

Sim2: $\Gamma=1$ N_H=1e21

Lee et al 2011, ApJ

Fitting: Optimization Methods

- Optimization - finding a minimum (or maximum) of a function:

  "A general function f(x) may have many isolated local minima, non-isolated minimum hypersurfaces, or even more complicated topologies. No finite minimization routine can guarantee to locate the unique, global, minimum of f(x) without being fed intimate knowledge about the function by the user."

- Therefore:

1. Never accept the result using a single optimization run; always test the minimum using a different method.
2. Check that the result of the minimization does not have parameter values at the edges of the parameter space. If this happens, then the fit must be disregarded since the minimum lies outside the space that has been searched, or the minimization missed the minimum.
3. Get a feel for the range of values of the fit statistic, and the stability of the solution, by starting the minimization from several different parameter values.
4. Always check that the minimum "looks right" using a plotting tool.

Fitting: Optimization Methods

- "Single - shot" routines, e.g, Simplex and Levenberg-Marquardt

  start from a guessed set of parameters, and then search to improve the parameters in a continuous fashion:

  - Very Quick
  - Depend critically on the initial parameter values
  - Investigate a local behaviour of the statistics near the guessed parameters, and then make another guess at the best direction and distance to move to find a better minimum.
  - Continue until all directions result in increase of the statistics or a number of steps has been reached

- "Scatter-shot" routines, e.g. Monte Carlo

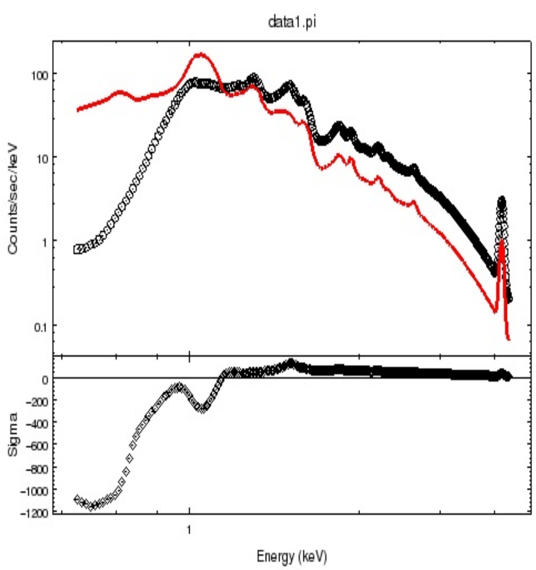
  examines parameters over the entire permitted parameter space to see if there are better minima than near the starting guessed set of parameters.

Optimization Methods: Comparison

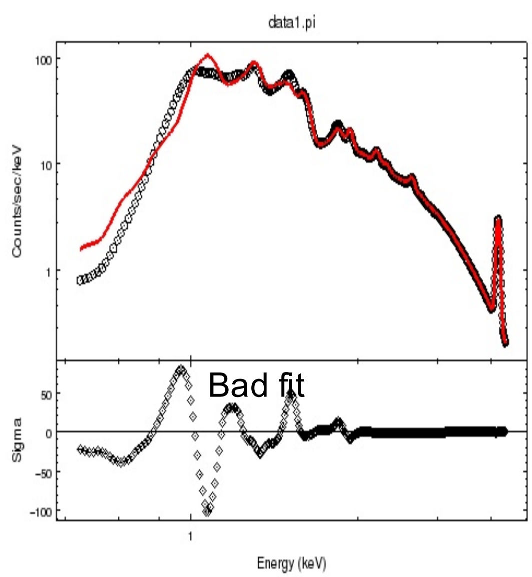
Data: high S/N simulated ACIS-S spectrum of the two temperature plasma
Model: photoelectric absorption plus two MEKAL components (correlated!)

Method	Number of Iterations	Final Statistics
Levmar	31	1.55e5
Simplex	1494	0.0542
Moncar	13045	0.0542

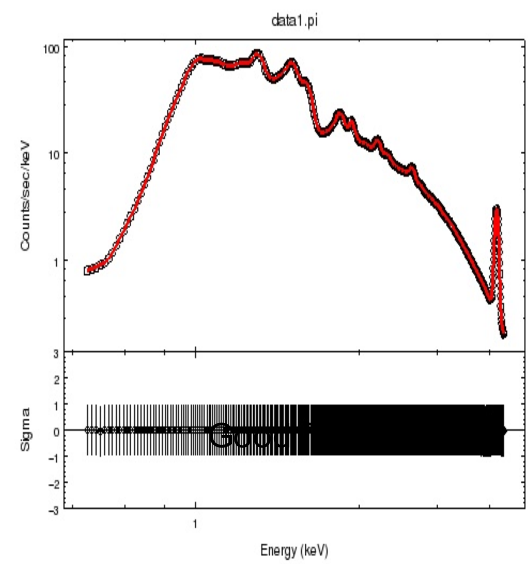
Start fit from the same initial parameters
Figures and Table compares the efficiency and final results



Data and Model with initial parameters

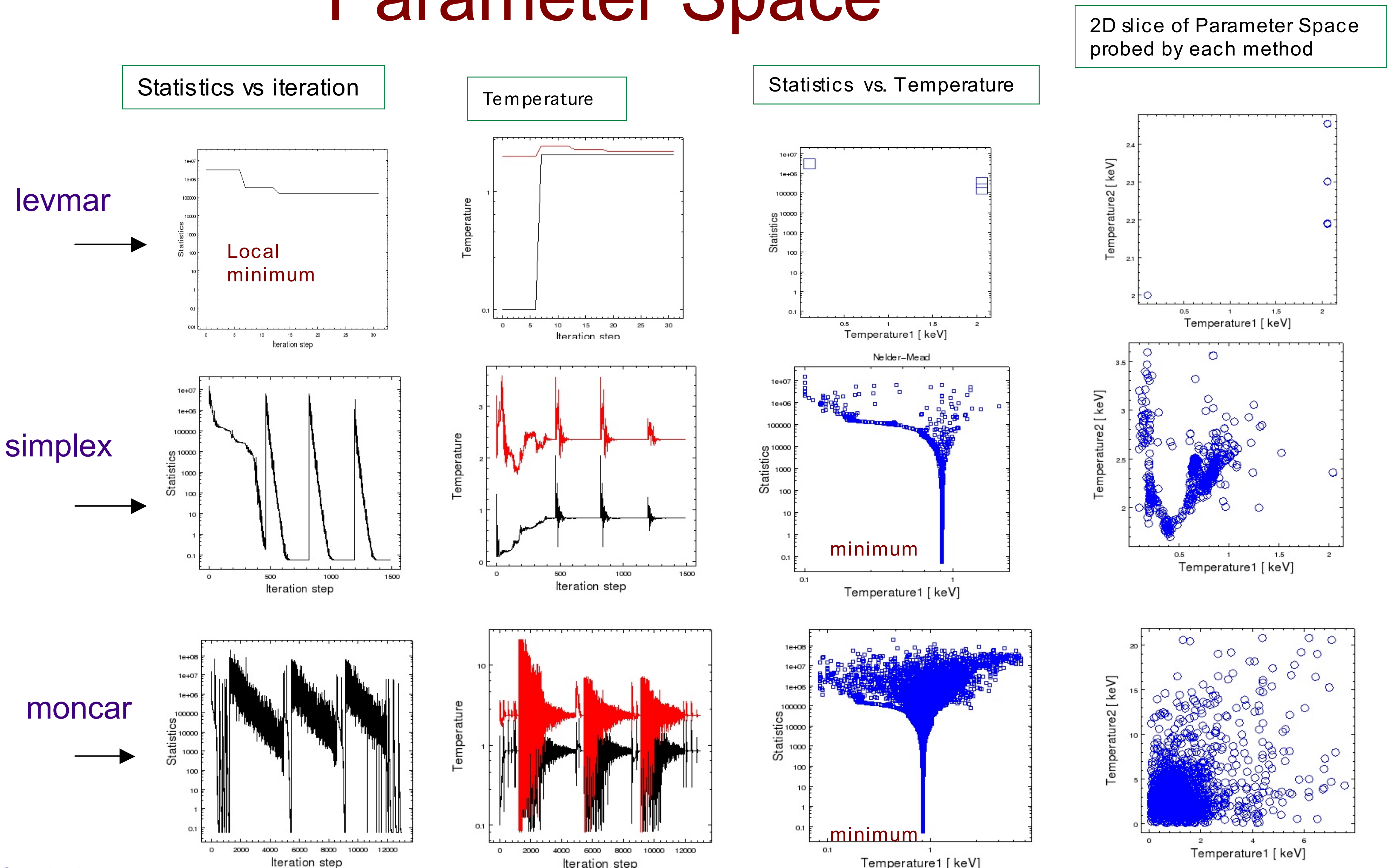


Levmar fit



Simplex and Moncar fit

Optimization Methods: Probing Parameter Space



Optimization Methods: Summary

- "levmar" method is fast, very sensitive to initial parameters, performs well for simple models, e.g. power law, one temperature models, but fails to converge in complex models.
- "simplex" and "moncar" are both very robust and converge to global minimum in complex model case.
- "simplex" is more efficient than "moncar", but "moncar" probes larger part of the parameter space
- "moncar" or "neldermead" should be used in complex models with correlated parameters

Final Analysis Steps

- How well are the model parameters constrained by the data?
- Is this a correct model?
- Is this the only model?
- Do we have definite results?
- What have we learned, discovered?
- How our source compares to the other sources?
- Do we need to obtain a new observation?

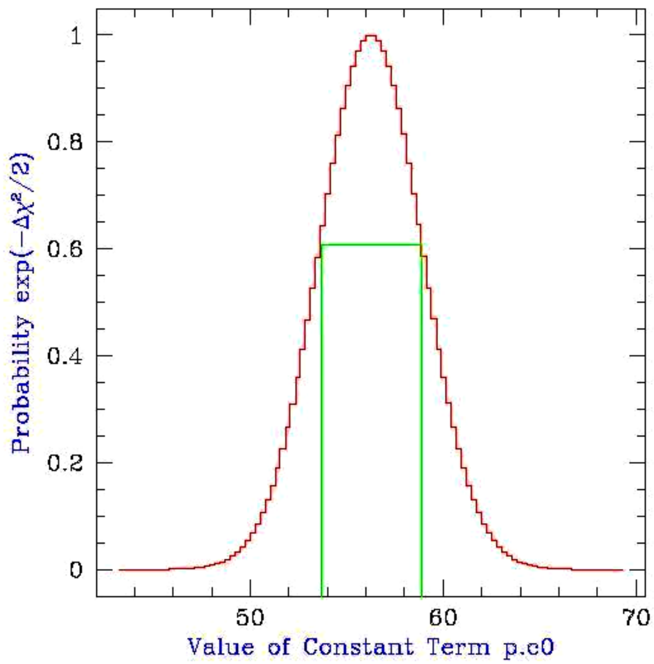
Confidence Limits

Essential issue = after the best-fit parameters are found estimate the confidence limits for them. The region of confidence is given by (Avni 1976):

$$\chi^2_\alpha = \chi^2_{min} + \Delta(\nu,\alpha)$$

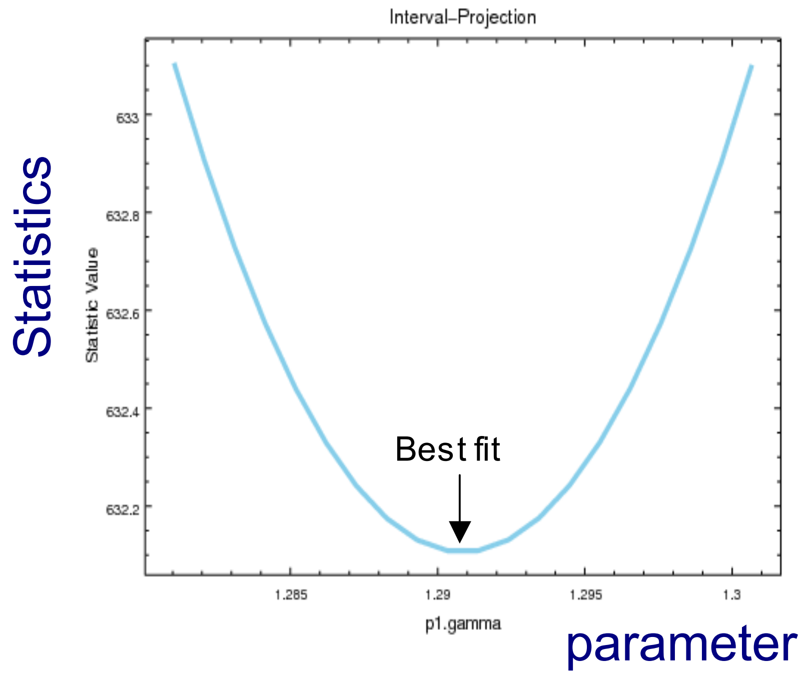
ν - degrees of freedom
α - significance
χ^2_{min} - minimum

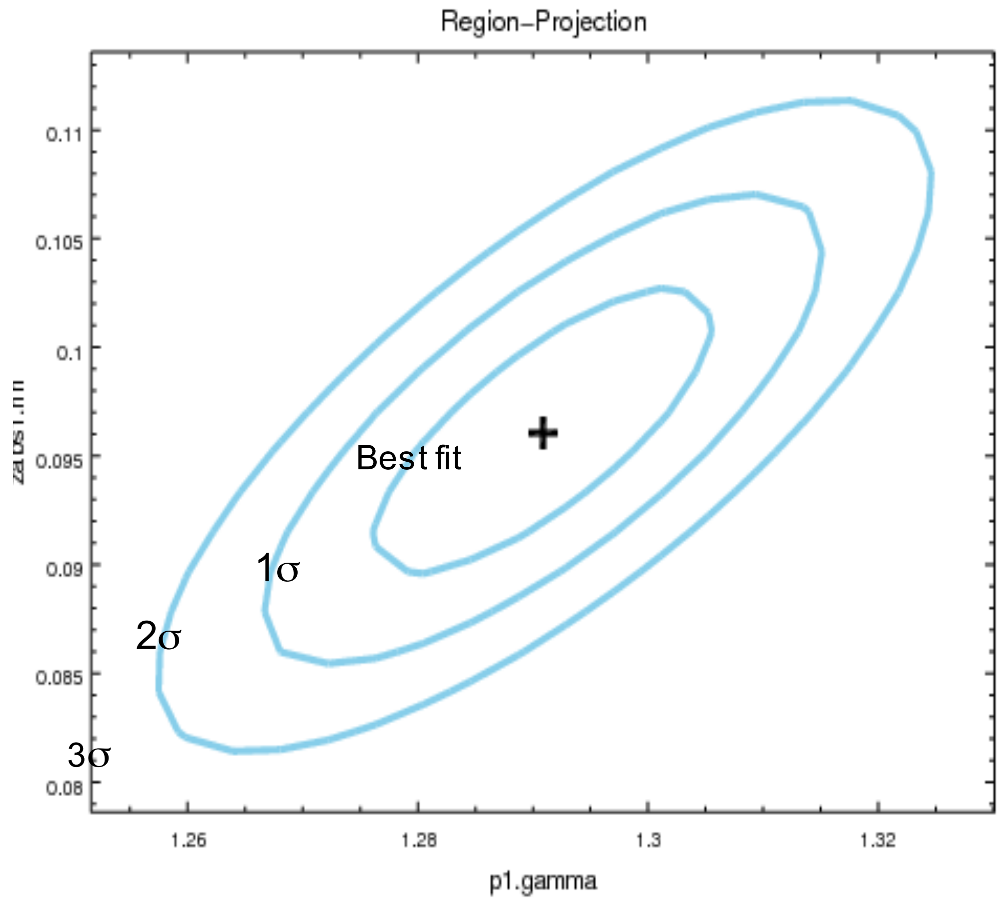
Δ depends only on the number of parameters involved nor on goodness of fit



Significance α	Number of parameters 1	2	3
0.68	1.00	2.30	3.50
0.90	2.71	4.61	6.25
0.99	6.63	9.21	11.30

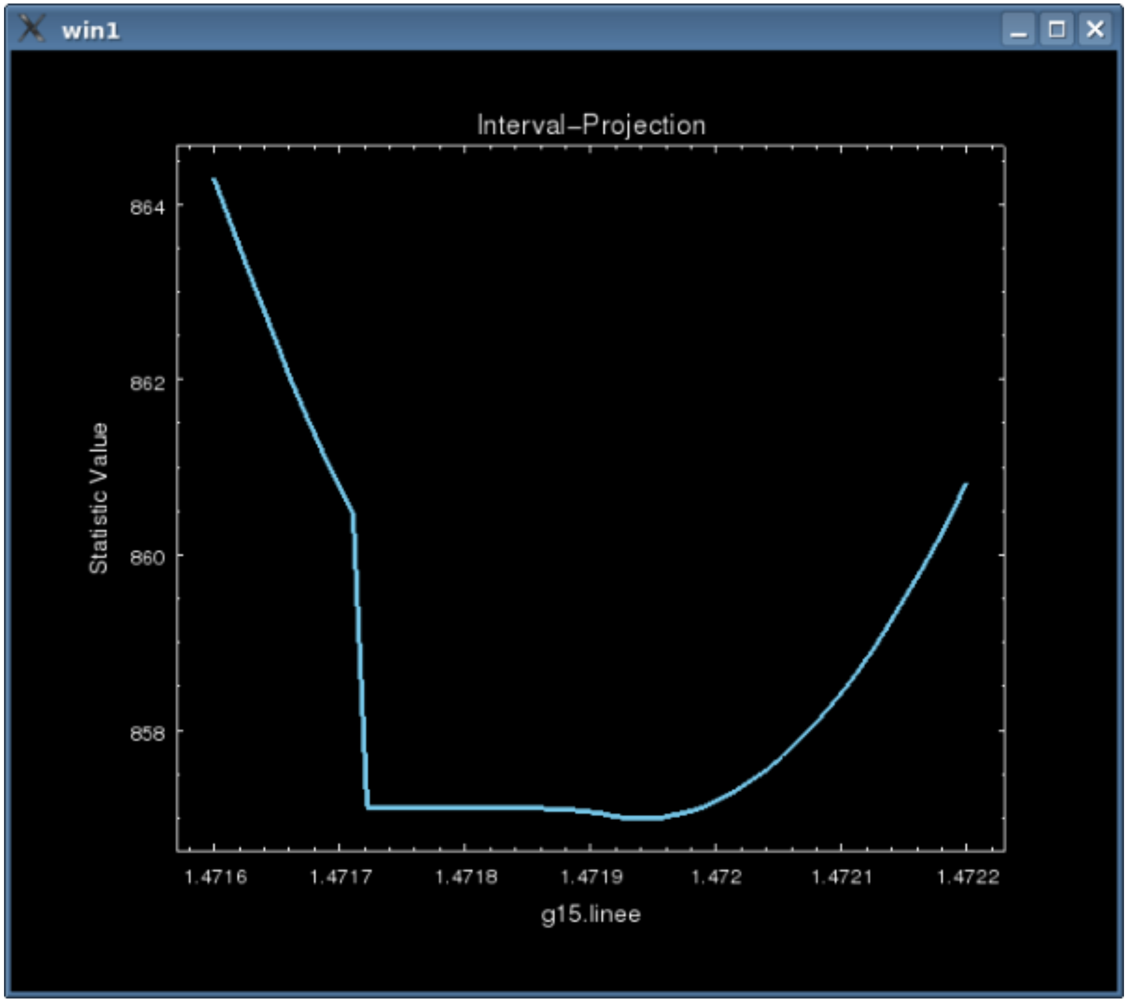
Confidence Regions





Well behaved parameter space

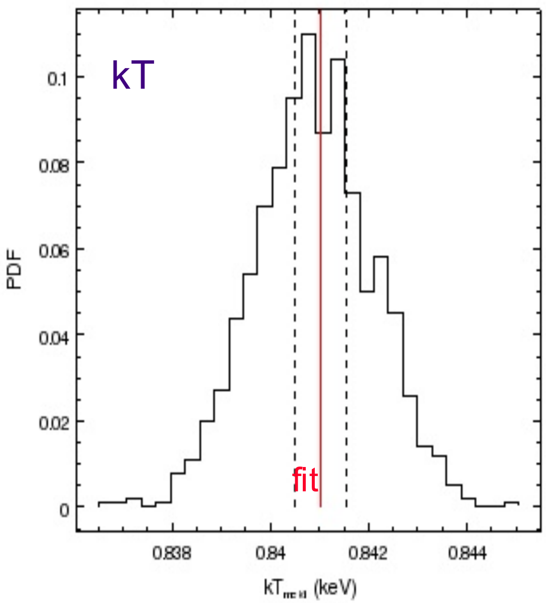
Not well-behaved Surface

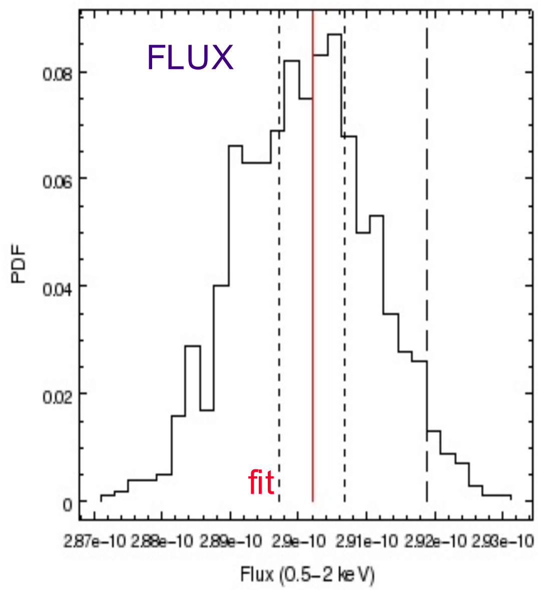


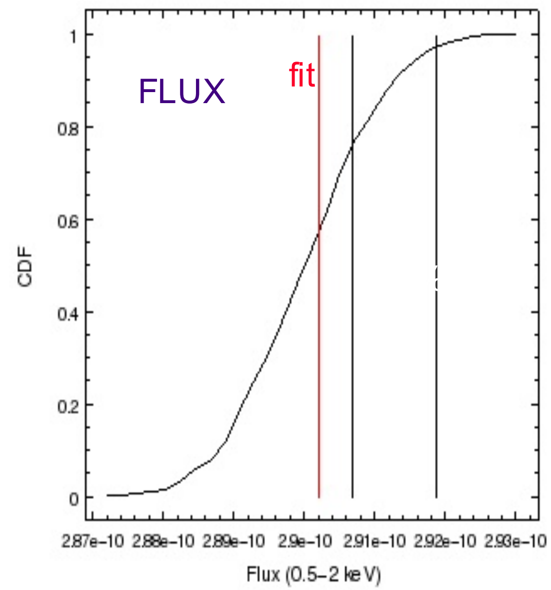
Non-Gaussian Shape

Distributions of Flux and Parameters

Monte Carlo Simulations to characterize parameters and flux and distributions.
Plot the PDF and CDF and calculate Quantiles of 68% and 95%

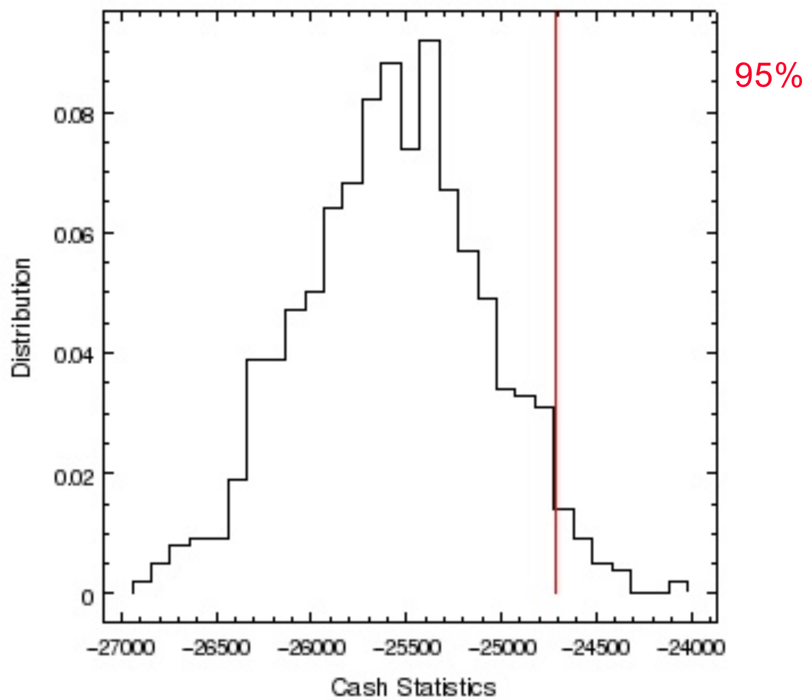


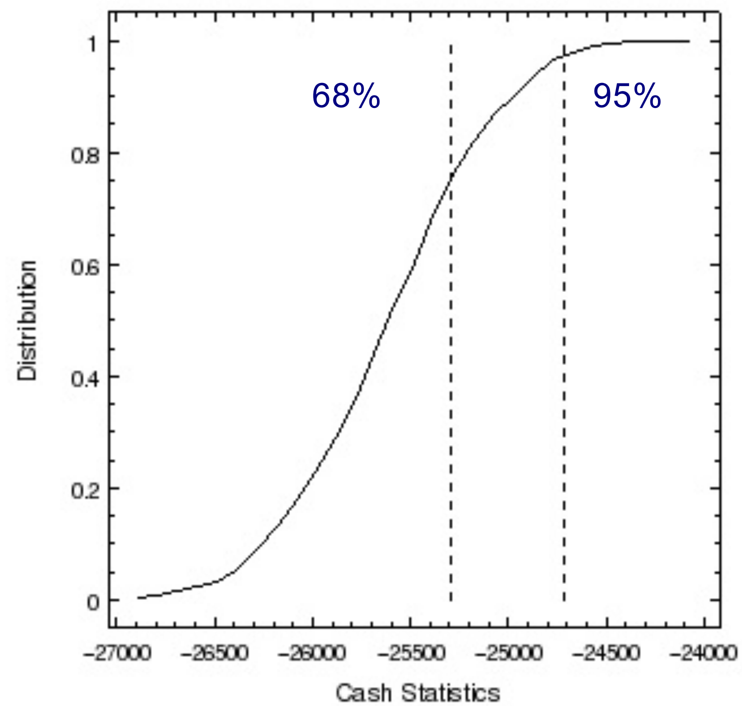




Goodness of Fit

Need simulations for the fit with likelihood statististics (Cash in Sherpa) to obtain the shape of the distribution.





Statistics

Model Selection

How to choose between different models?

Does a more complex model better describe the data?

Steps in Hypothesis Testing

1/ Set up 2 possible exclusive hypotheses - two models:

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 < α

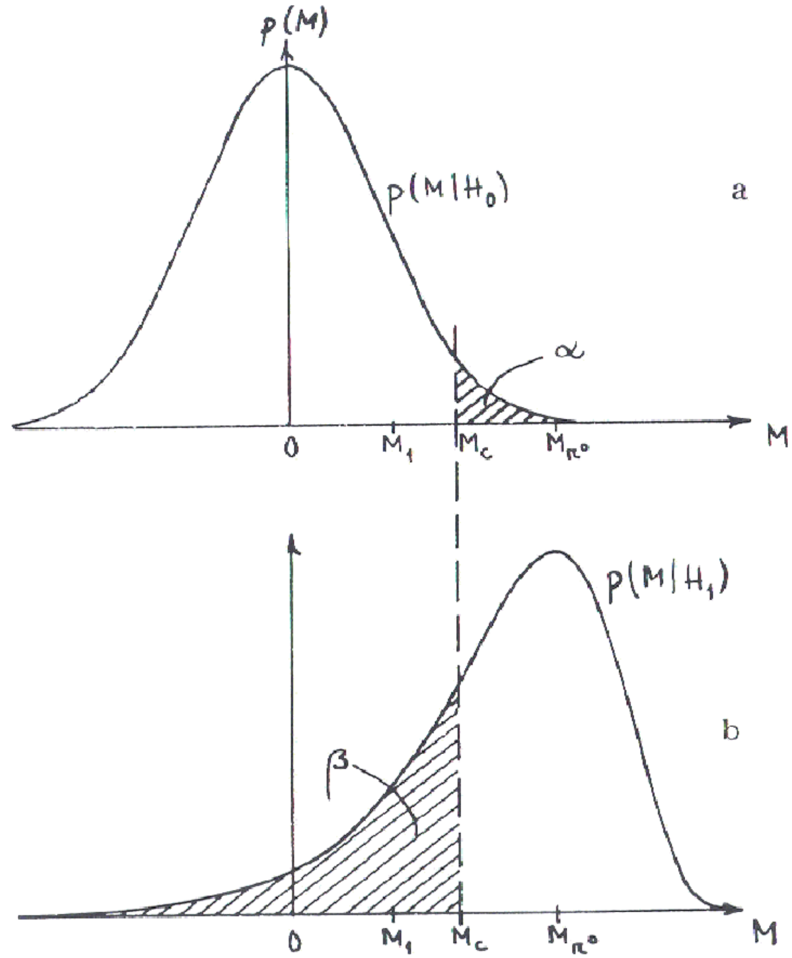
4/ Carry on terminal action

α- significance

=> Probability limit of rejecting the null when it is true.

(1-β) – power of test

=> correctly reject H0 when it is false



Comparison of distributions $p(T \mid M_0)$ (from which one determines the significance α) and $p(T \mid M_1)$ (from which one determines the power of the model comparison test $1 - \beta$) (Eadie *et al.* 1971, p.217)

Test Statistics

- Likelihood Ratio Test

  Ratio of likelihood values:

  $$LRT = 2[\ln p(d|M_1) - \ln p(d|M_0)]$$

- F-test

  For Gaussian data statistic follows F distribution

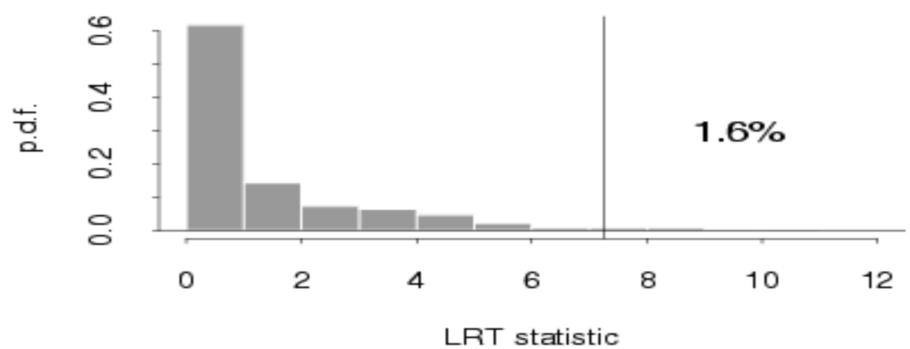
  $$F = \frac{\Delta\chi^2}{\Delta P} / \frac{\chi_1^2}{(N - P_1)} .$$

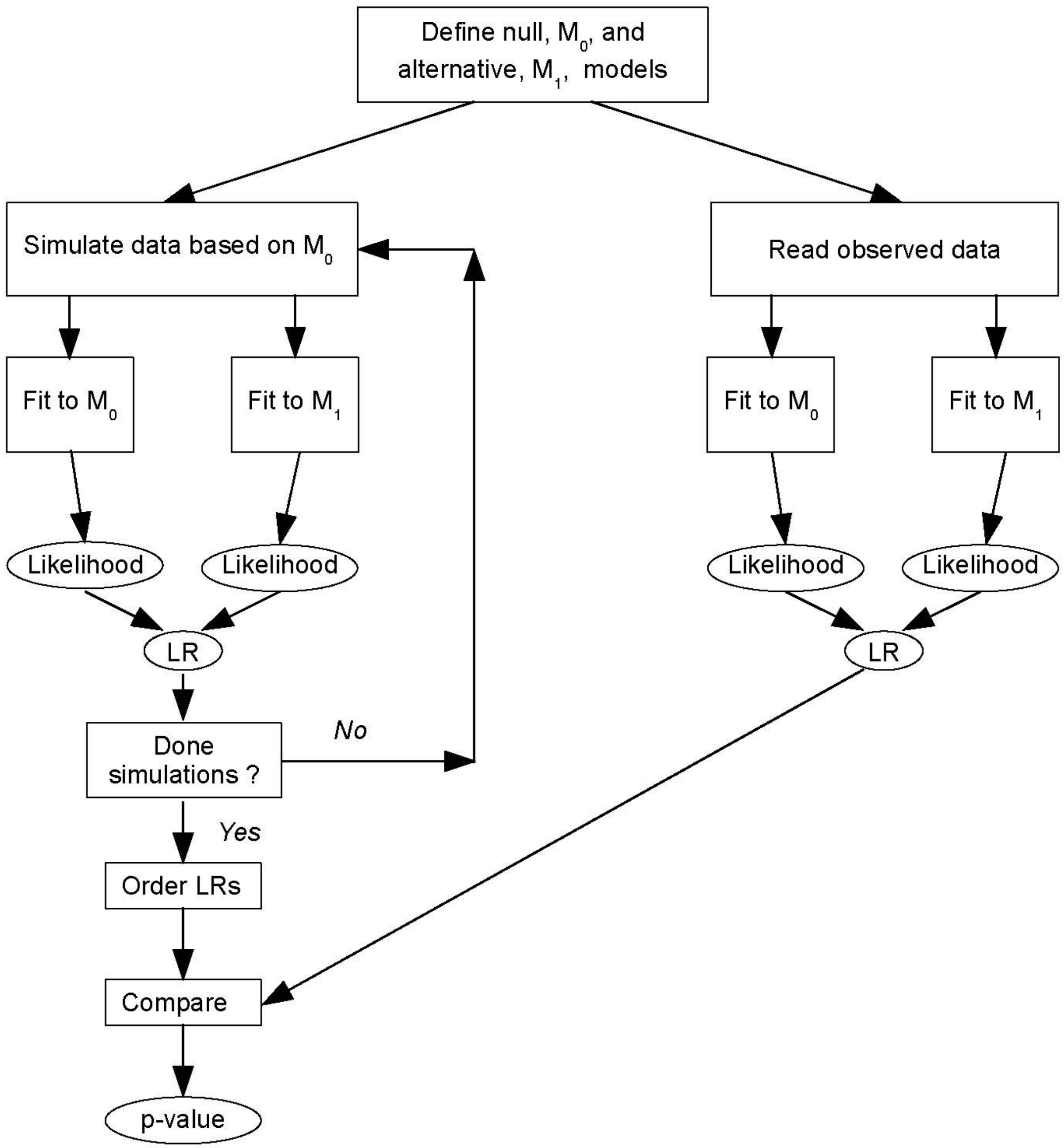
- Tests only valid if
  - The models are nested
  - Not on the boundary of the parameter space
  - Asymptotic limit has been reached

Monte Carlo Simulations

- Simulations to test for more complex models, e.g. addition of an emission line
- Steps:
  - Fit the observed data with both models, M_0, M_1
  - Obtain distributions for parameters
  - Assume a simpler model M_0 for simulations
  - Simulate/Sample data from the assumed simpler model
  - Fit the simulated data with simple and complex model
  - Calculate statistics for each fit
  - Build the probability density for assumed comparison statistics, e.g. LRT and calculate p-value

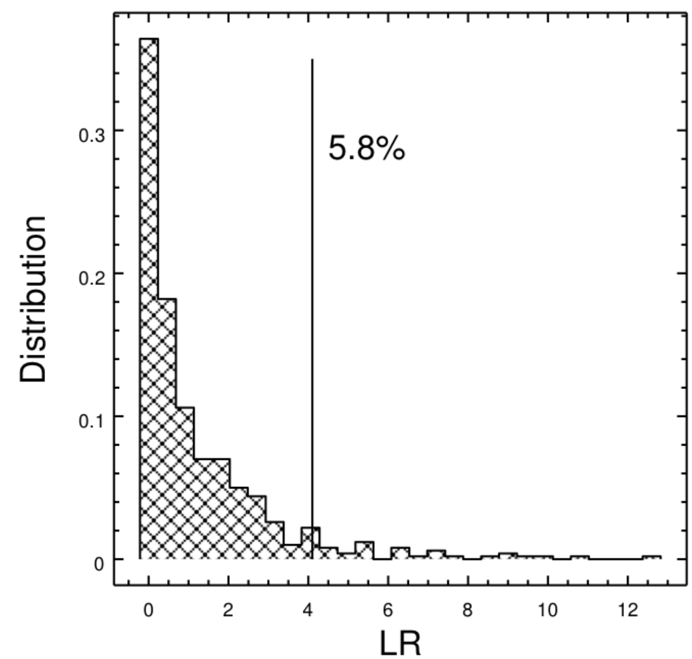
Example:
Visualization, here accept more complex model, p-value 1.6%





From the Statistics chapter in "Handbook of X-ray Astronomy"

Would we accept Model1?



Bayesian Model Comparison

Bayes' theorem 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
- $p(D \mid M)$ is the average, or global, likelihood:

$$\begin{aligned} p(D \mid M) &= \int d\theta\, p(\theta \mid M)\, p(D \mid M, \theta) \\ &= \int d\theta\, p(\theta \mid M)\, \mathcal{L}(M, \theta). \end{aligned}$$

In other words, it is the (normalized) integral of the posterior distribution over all parameter space. Note that this integral may be computed numerically

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). BUT BF ARE NO CALIBRATED IN GENERAL

Summary

- Motivation: why do we need statistics?
- Probabilities/Distributions
- Poisson Likelihood
- Parameter Estimation - Optimization, MC
- Model Selection and Statistical Tests

Conclusions

Statistics is the main tool for any astronomer who need to do data analysis and need to decide about the physics presented in the observations.

Attend Astrostatistics Sessions at the Scientific Meeting.

Astrostatistics Collaboration at CfA:

http://hea-www.harvard.edu/AstroStat/

References:

"Practical Statistics for Astronomers", Wall & Jenkins, 2003
Cambridge University Press

Gelman, Carlin, Stern and Rubin, 2003 2nd edition "Bayesian Data Analysis"
http://www.stat.columbia.edu/~gelman/book/

Eadie et al 1976, "Statistical Methods in Experimental Physics"

Selected Additional References

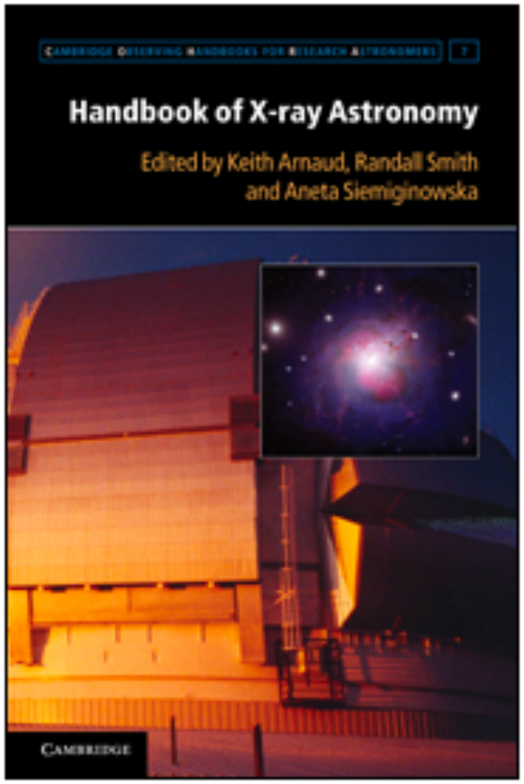
- Astrostatistics:
  - Van Dyk et al 2001, ApJ 548, 224, Analysis of Energy Spectra with Low Photon Counts via Bayesian Posterior Simulation
  - Babu, G. J., Feigelson, E. D. 1996, Astrostatistics (London: Chapman & Hall)
  - Press, W. H., Teukolsky, S. A., Vetterling, W. T.,& Flannery, B. P. 1992, Numerical Recipes (Cambridge: Cambridge Univ. Press)
  - "Statistical Challenges in Modern Astronomy" - ed. E. Feigelson & G. Babu (New York: Springer-Verlag), 275

- Estimators: Modified $\mathcal{L}$ and χ^2 statistics:
  - Cash, W. 1979, ApJ 228, 939
  - Churazov, E., et al. 1996, ApJ 471, 673
  - Gehrels, N. 1986, ApJ 303, 336
  - Avni, Y. 1976, ApJ, 210, 642

- Issues in Fitting and Error Analysis
  - Protassov et al 2002, ApJ 571, 545 Statistics, Handle with Care: Detecting Multiple Model Components with the Likelihood Ratio Test
  - Kelly 2007, ApJ, 665, 1489 Some Aspects of Measurement Error in Linear Regression of Astronomical Data,
  - Lee et al 2011, ApJ, 731, 126, Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting
  - Park et al 2008, ApJ. 688, 807, Searching for Narrow Emission Lines in X-ray Spectra: Computation and Methods
  - Freeman P. et al 1999, ApJ 524, 753 (and references therein)
  - Park et al 2006, ApJ, 652, 610, Bayesian Estimation of Hardness Ratios: Modeling and Computations
  - Kashyap et al 2010, AoJ,719, 900, On Computing Upper Limits to Source Intensities

CAMBRIDGE | Catalogue

Home > Catalogue > Handbook of X-ray Astronomy



Details

60 b/w illus.
Page extent: 224 pages
Size: 228 x 152 mm

Handbook of X-ray Astronomy

Series: Cambridge Observing Handbooks for Research Astronomers (No. 7)

Edited by Keith Arnaud
University of Maryland, College Park

Randall Smith
Harvard-Smithsonian Center for Astrophysics

Aneta Siemiginowska
Harvard-Smithsonian Center for Astrophysics

View list of contributors...

Hardback (ISBN-13: 9780521883733)

Not yet published - available from November 2011

Price is not yet set (inclusive of GST)
Export price AUD$0.00

Modern x-ray data, available through online archives, are important for m