POISSON STATISTICS PROJECT

ALICIA MARTINEZ GONZALEZ
ANDY POLLOCK

alicatex@hotmail.com

OVERVIEW

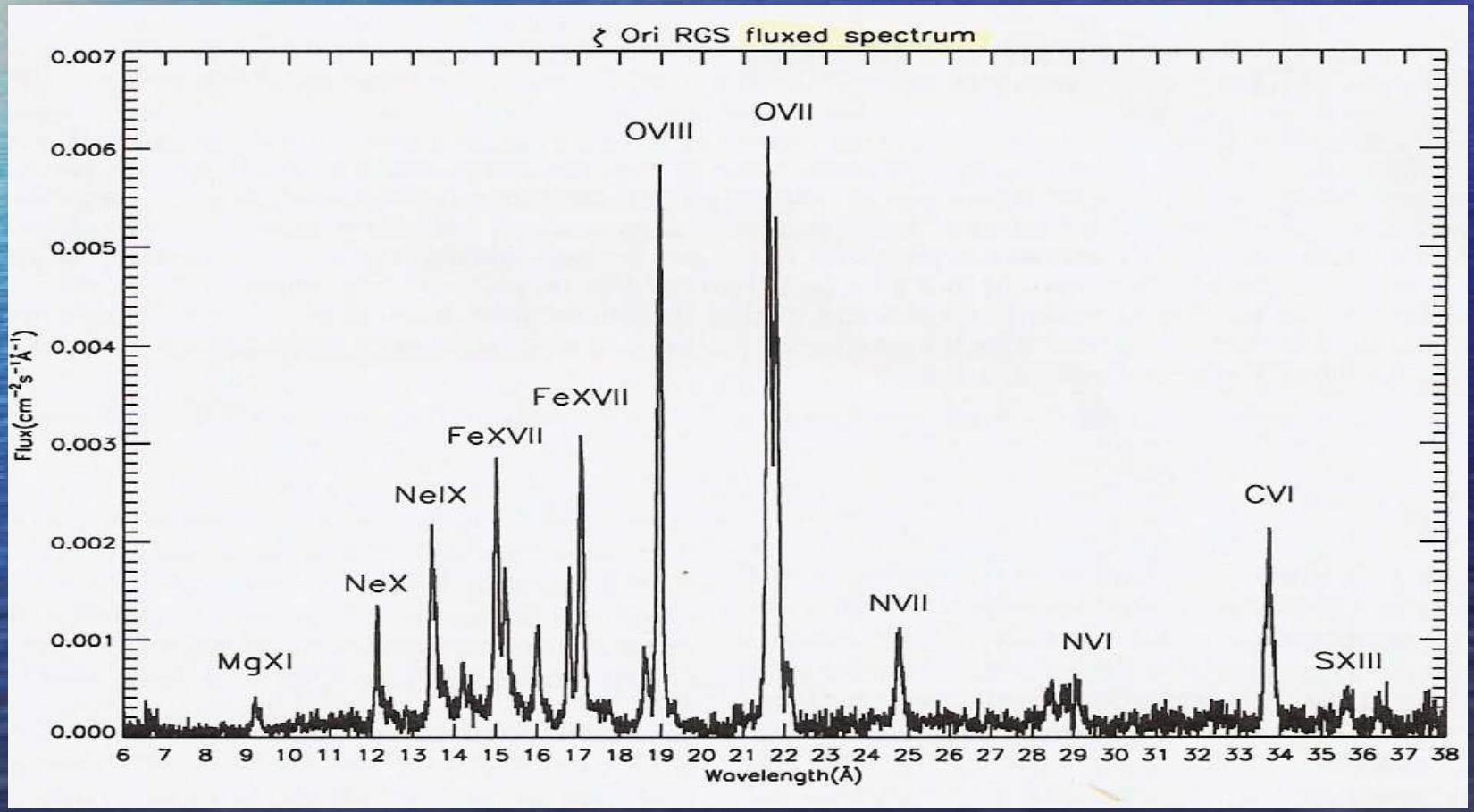
OBJECTIVE:

IMPROVE THE GOODNESS-OF-FIT WHEN THERE ARE FEW OR NO COUNTS PER BIN

- **WHY IS THIS PROJECT NECESSARY?**
  - INTRODUCTION AND POISSON EXAMPLES
  - X-SQUARED AND C-STATISTIC
  - WHAT HAPPENS WHEN WE HAVE A FEW COUNTS PER BIN?
- **WHAT WE NEED? HOW TO GET IT?**
  - NEW: “A”-STATISTIC/GOODNESS-OF-FIT FOR C-STATISTIC
  - XSPEC USES SIMULATIONS, SO WHY DON'T WE?
- **CONCLUSIONS AND FUTURE STEPS**

Starting situation

- Validation of physically model fit to spectra in X-ray astronomy
- *evaluate general techniques to estimate parameters where the correct functional form is known but the parameters are not known*
- *Decide with statistical analysis if the model is true*
- *Use and test with XMM-Newton data*

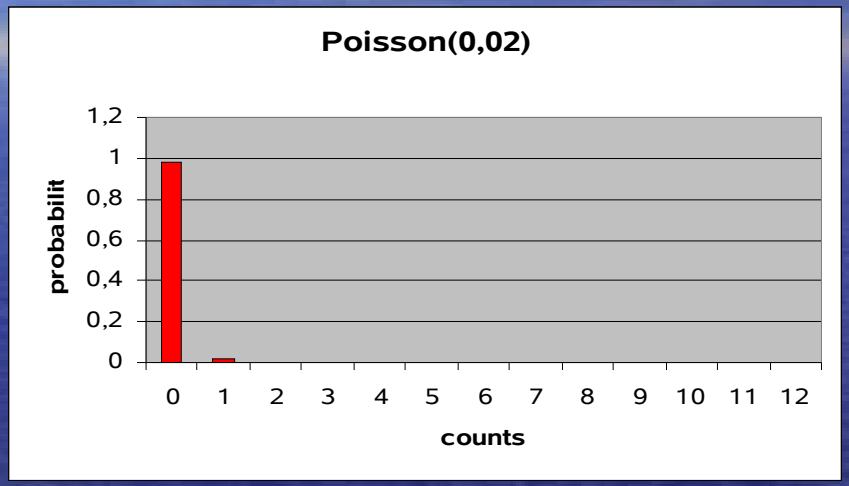


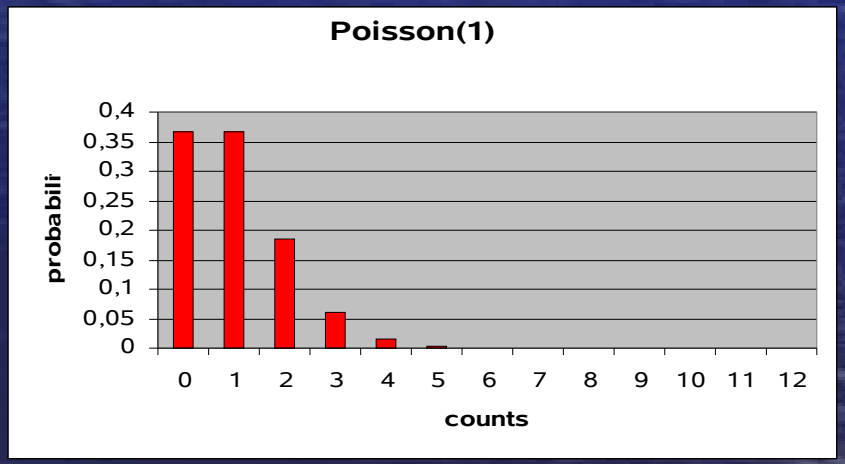
Possion distribution

POISSON EXAMPLES

two examples of the Poisson distribution with different (mean 0.02, 1.0)

$$P_{\mu}(X=x)=\frac{e^{-\mu}\mu^{x}}{x!}$$



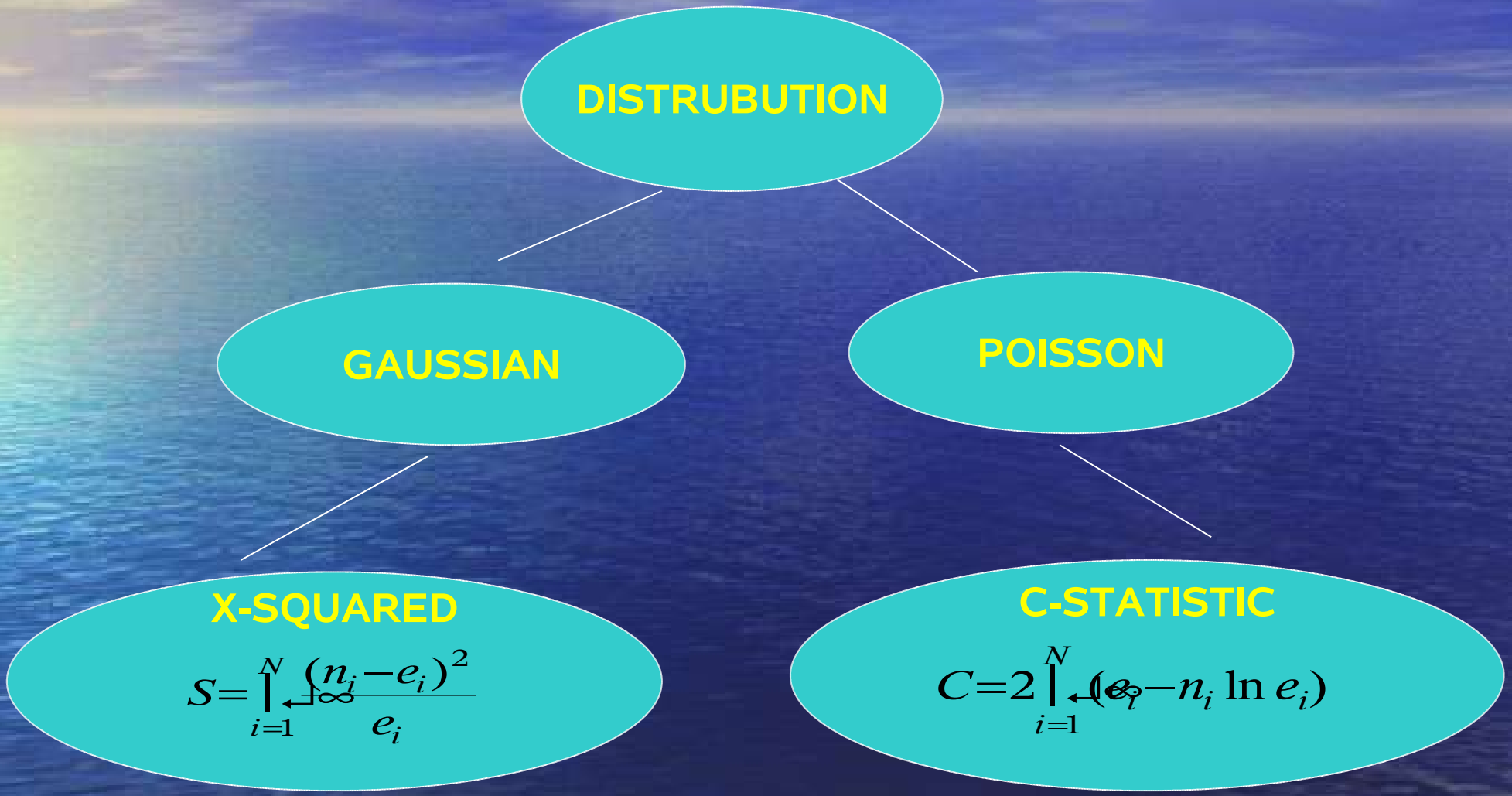


In our data there are a lot of bins with 0 counts

--> mean --> 0

--> perfectly good data value

statistical evaluation



n_i= Observed counts in bin i e_i= expected number

DIFFERENCES BETWEEN X-SQUARED AND C-STATISTIC

- X-squared
- It comes from a Gaussian distribution

Gaussian errors of individual data points

Independence of data points

- Doesn't always work
- how many counts per bin are enough?

Norbert Schartel:
5 for central bins and 4 for first and last bins

- C-statistic
- It comes from a Poisson distribution

Poisson errors of individual data points

- always works
- how many counts per bin are enough???

Norbert Schartel: allways
Webster Cash says 9
Andy Pollock says 1

WHY IS THIS PROJECT NECESSARY ?

WHAT HAPPENS WHEN WE HAVE A FEW OR NO COUNTS PER BIN?

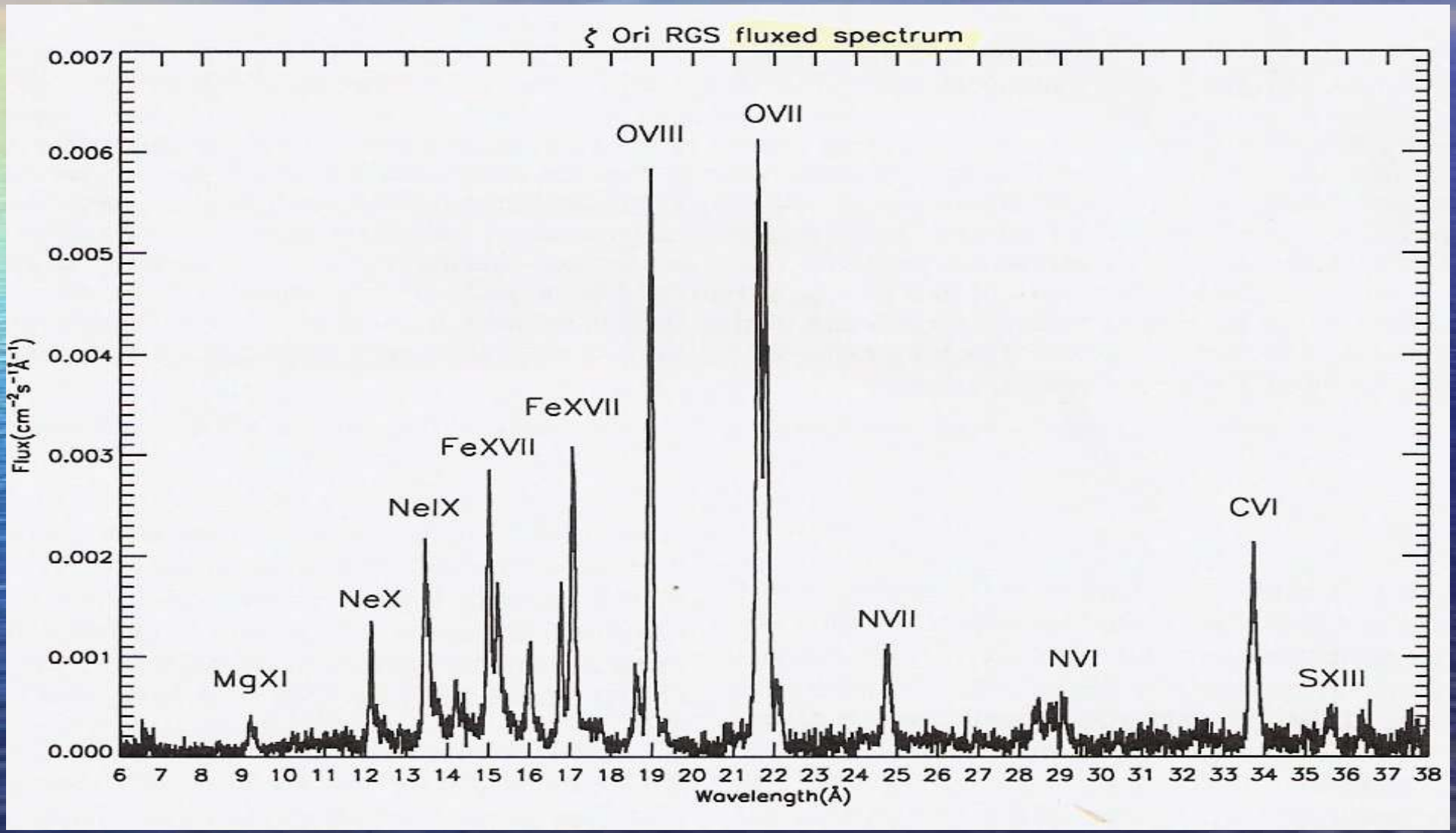
- All statistical proofs and theorems are valid for high numbers of counts only

- X-squared minimization criterion is very bad if any of the observed data bins have few counts
- There are tests in both cases to check if there are enough counts per bin: In out case: not enough counts!!!

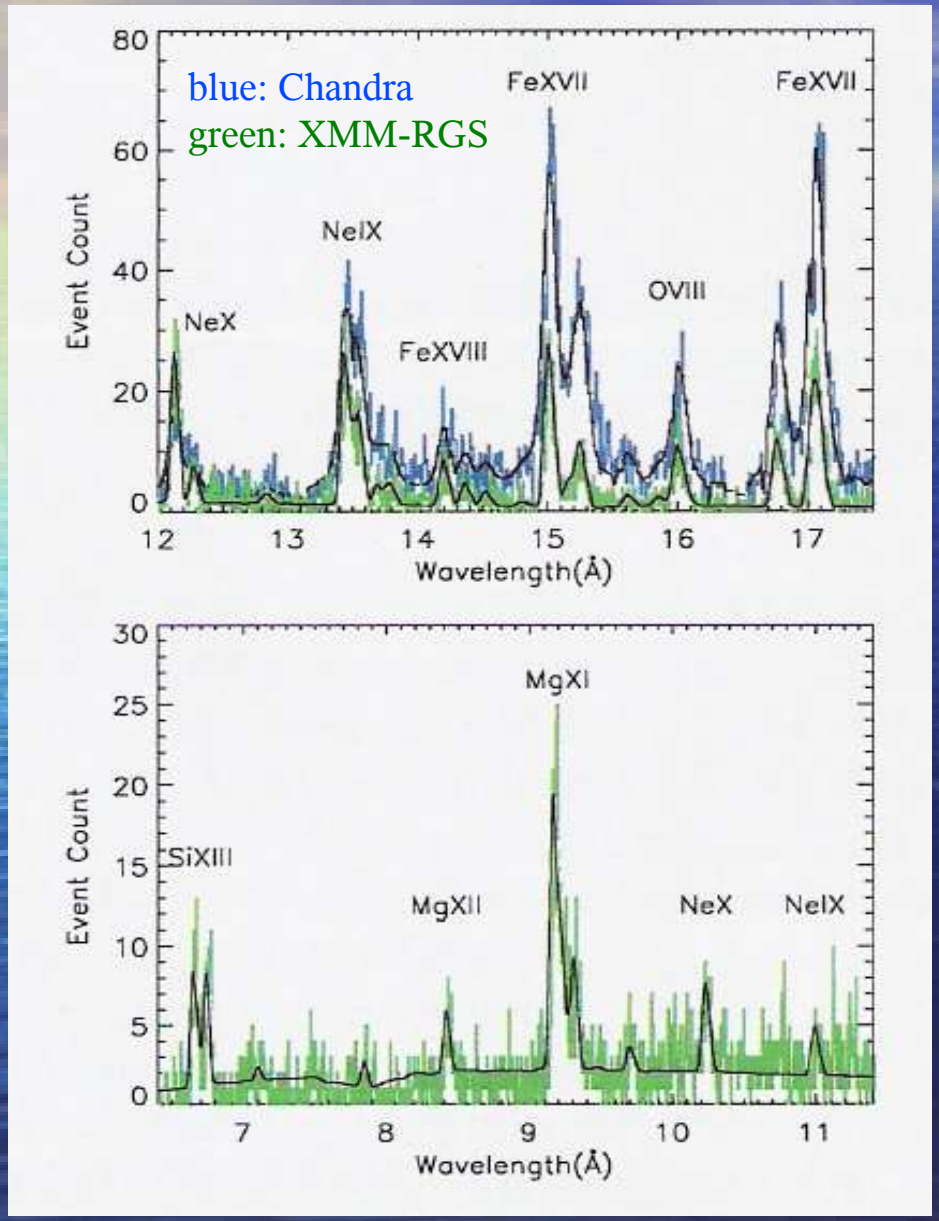
- W. Cash: C is a better criterion to use a likelihood function based on Poisson statistics. It usually yields tighter error boxes than does chi2 but never the opposite!
  But no prove!

Example:

RGS fluxed spectrum of the star zeta ori



Model fits to data



Comparison of high-resolution X-ray data with the best-fit model

--> emission-line velocity parameters

TriLine velocities		
blueV	-1642 ± 22	km s^{-1}
centralV	-302 ± 29	km s^{-1}
redV	$+1646 \pm 26$	km s^{-1}
bremsstrahlung continuum		
N_H	2.5×10^{20}	cm^{-2}
kT	0.494 ± 0.007	keV
normalisation	$5.66 \pm 0.14 \times 10^{-3}$	
C-statistic = 22 862.4 using 26281 PHA bins		

What does the value of C means?

WHAT WE NEED?: Solution 1

NEW "A"-STATISTIC

Requirements:

- The statistics only depends on the data

$$A(x1,x2,\ldots,xn)$$

- should work in our case (few counts per bin)
- independent of errors between data points
- easy to use and to program
- need a new test for this statistics

WHAT WE NEED? Solution 2

--> Derive GOODNESS-OF-FIT FOR C-STATISTIC for few counts per bin

- What does the value of C-statistic means?
- How to decide objectively when there are enough counts per bin?
- Improve xspec
  - 'xspec uses a fitting algorithm to find the best-fit values of model parameters', but usually it finds a local minimum!
  - Understand when and how xspec uses simulation

CONCLUSIONS AND FUTURE STEPS

- Poisson statistics is always applicable
- Xspec sometimes does not find global minimum
- Golden Recipe of the statistical analysis does not exist

- evaluate C-statistic test in detail
- find an algorithm that gives us a global minimum for the best fit parameters
- if no satisfactory test for C-statistic can be found develop “A”-statistic

alicatex@hotmail.com

THANKS AND GOOD BYE!