

Spectral Analysis & Statistics

Presentation for 13th ESAC SAS Workshop

12th June 2013

Norbert Schartel

(XMM-Newton Project Scientist)

XMM-Newton Science Operation Centre, ESAC, ESA,
28691 Villanueva de la Cañada, Spain



Overview

1. **Ideal experiment / X-ray Astrophysics**
2. **Spectral analysis**
3. **Questions to be answered by the spectra**
4. **Statistical tests**
5. **The large view**
6. **Golden rules of statistics**



XMM-Newton

Norbert Schartel

Ideal Experiment

- **expected spectrum: either from theory or short test measurement**
- **resolution of detector \gg intrinsic line-width**
- **effective area: known per resolution element**
- **background: either known or short test measurement**
- **optimization of experiment:**
 - **exposure time per resolution element: such that constant signal/noise ratio is expected**



XMM-Newton

Norbert ScharTEL

X-ray Astrophysics

- spectrum and flux:
- resolution vs. line-width:
- background:
- observation optimization:
- generally not known with sufficient accuracy to plan observation accurate
- pn < MOS << RGS
 - in general:
 - line-width < resolution,
 - **redistribution**
- **not well known in advance and variable**
- only partly possible



XMM-Newton

Norbert Schartel

Ideal Experiment

- **reconstruction of spectrum:**
- **data [counts]:**
 - **Y1, Y2, Y3, Y4 Y2000**
- **effective area:**
 - **a1, a2, a3, a4a2000**
- **exposure time:**
 - **t1, t2, t3 ..t2000**
- **spectra [photons/sec/cm²] :**
 - **f1=Y1/ t1/ a1 + 0 x Y2/ t1/a1**
 - **f2=Y2/ t2/ a2 + 0 x Y1/t1 ..**
 - **f3=Y3/ t3/ a3 + 0 x Y1/....**
- **start with the spectrum:**
- **input spectrum [photons/sec]:**
 - **P1(E1), P2(E2), P3, P4 ...**
- **registered with a certain probability: effective area**
 - **a1x t1, a2 x t2, ...**
- **redistributed with respect of counts channels:**
- **Y1(e1) = P1(E1) x a1(E1) x t1(E1) + P2(E2) X 0**



XMM-Newton

Norbert Schartel

X-ray Astrophysics

- **reconstruction of spectrum:**
- **data [counts]:**
 - **Y1, Y2, Y3, Y4 Y2000**
- **effective area [cm²]:**
 - **a1, a2, a3, a4a2000**
- **exposure time:**
 - **t1, t2, t3 ..t2000**
- **spectra [photons/sec/cm²] :**
 - **f1=Y1/ t1/ a1 + b x Y2/ t1/a1**
 - **f2=Y2/ t2/ a2 + c x Y1/t1 ..**
 - **f3=Y3/ t3/ a3 + c x Y1/....**
- **start with the spectrum:**
- **input spectrum [photons/sec/cm²]:**
 - **f1(E1), f2(E2), f3, f4 ...**
- **registered with a certain probability: effective area**
 - **a1x t1, a2 x t2,**
- **redistributed with respect of counts channels:**
- **Y1(e1) = f1(E1) x a1(E1) x t1(E1) + P2(E2) X b**



XMM-Newton

Norbert Schartel

X-ray Astrophysics

reconstruction of spectrum:

- data [counts]:
 - $Y_1, Y_2, Y_3, Y_4 \dots Y_{2000}$
- effective area [cm²]:
 - $a_1, a_2, a_3, a_4 \dots a_{2000}$
- exposure time:
 - $t_1, t_2, t_3 \dots t_{2000}$
- spectra [counts/sec/cm²] :
 - $f_1 = Y_1 / t_1 / a_1 + b \times Y_2 / t_1 / a_1$
 - $f_2 = Y_2 / t_2 / a_2 + c \times Y_1 / t_1 / a_1 \dots$
 - $f_3 = Y_3 / t_3 / a_3 + c \times Y_1 / t_1 / a_1 \dots$

start with the spectrum:

- input spectrum [photons/sec/cm²]:
 - $P_1(E_1), P_2(E_2), P_3, P_4 \dots$
- registered with a certain probability: effective area
 - $a_1 \times t_1, a_2 \times t_2, a_3, a_3$
- redistributed with respect of counts channels:
- $Y_1(e_1) = P_1(E_1) \times a_1(E_1) \times t_1(E_1) + P_2(E_2) \times b \dots$

Redistribution and Resolution

- the channels in the spectra reflect a purely mathematical sorting
 - a count in a channel, which corresponds to an energy k , can not be identified with a photon of energy k
 - physical sampling, “real” channels, “independent bins”, resolution:
 - simplest case: (Abbe, 1820-1830) what is the minimum distance between two point which can be resolved with a given resolution?
 - today more complex: Shannon information theorem, Nyquist criteria
 - **information content of a given spectra does not increase if the sampling goes below FWHM/3**
 - most of the spectra provided in X-ray missions are over-sample
-



XMM-Newton

Norbert Schartel

Concept Of Spectral Analysis

- 1. physically motivated model spectra which is a function of parameters (x_1, x_2, \dots, x_i) and energy (e_1, e_2, \dots)**
 - **Example:**
 - thermal emission of a blackbody (BB),
 - with parameters: temperature (T) and norm (N)
 - $BB(T, N, e_i)$
- 2. multiplication with effective area (a_1, a_2, \dots): expected flux (f) as function of energy and model**
 - $F(BB(T, N), e_i) = BB(T, N, e_i) * a_{ij}, (i=j \ a_{ij} = k_j, i \text{ not } j \ a_{ij}=0)$
- 3. folding with redistribution matrix (Detector Response Matrix (b_{ij})) and multiplication with exposure time (t): counts as function of channel (k_i) and model:**
 - $counts(c_j) = F(BB(T, N), e_i) * b_{ij}$
- 4. statistical analysis**
 - is model true?
 - hypothesis tests
 - can the description of the spectra be improved?
 - new loop with changed parameter of model

Standard Questions

- given a data set the most common questions are:
 - **distributions**
 - 50 counts are measured. What is there distribution?
 - **hypothesis test:**
 - ASCA showed a power law with $\Gamma=2.1$ and flux =1.5 → does this model describe the pn data?
 - **estimate of parameters:**
 - the data can be described with a power law. What is the index and the normalization?
 - **estimate of confidence level:**
 - what is the error of the power law and what is the error of the normalization?
 - **hypothesis test:**
 - **comparison between two fits:**
 - is a power-law or a power-law + emission line a better description of the data



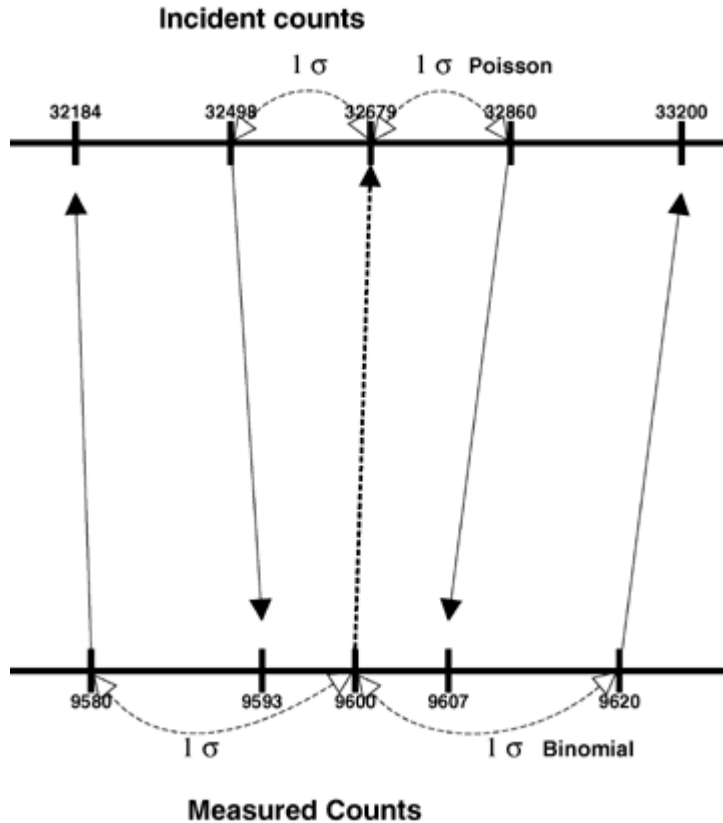
XMM-Newton

Norbert Schartel

Statistical Methods

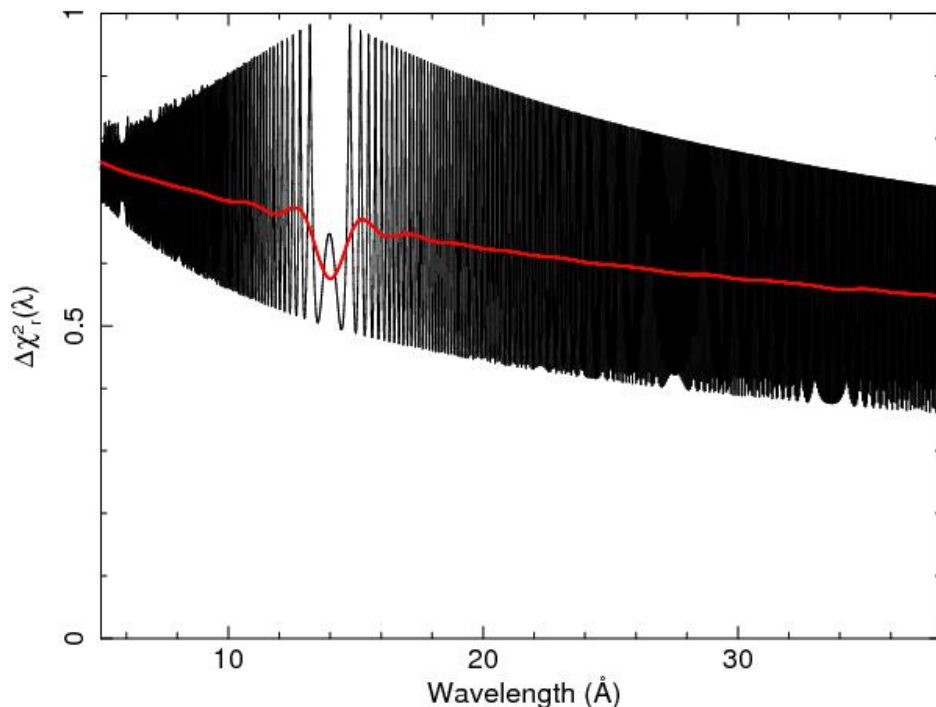
- statistical methods to answer the questions (provided in xspec):
 - **distributions**
 - Poisson distribution versus binomial distribution
 - **hypothesis test:**
 - χ^2 hypothesis test
 - **estimate of parameters**
 - modified minimum χ^2 method
 - minimize C value (C-statistics)
 - **estimate of confidence level**
 - determine parameter for $\chi^2 = \chi^2_{\text{minimum}} + \Delta \chi^2$
 - **comparison between two fits**
 - F-test

Comments to Distributions I: OM



- Kuin & Rosen, 2008, MNRAS 383, 383
- Fordham et al., 2000, MNRAS 312, 83

Comments to Distributions II: RGS



Relative contribution per wavelength bin to the reduced χ^2 . Calculated and shown is the ratio $\text{Var}[Y_i]/E[Y_i]$ (see text). The thick line is a low-resolution spline approximation showing the trends in the average value. (From Kaastra, J. S. et, al., 2011, A&A 534, A37)

- J. S. Kaastra et al. 2011, A&A 534, A37

- Study of 600 ks RGS spectrum of Mrk 509 with excellent quality

- **Fluxed Spectrum:**

 - > 400 counts per bin

 - Expectaion: reduced $\chi^2 \gg 1$

 - But: reduced $\chi^2 \approx 0.6$

 - **Counts are not χ^2 distributed**

 - **use rescaled χ^2**

 - **Small number of counts in fluxed RGS spectra do follow the C-statistic**

Statistical Functions

- χ^2 is defined as: $\chi^2 = \sum (Y_m - Y_e)^2 / (Y_e)$
 - for $n \rightarrow \infty$
 - and requires:
 - Gaussian errors of individual data points
 - independence of data points
- modified χ^2 is defined as: $\chi^2_{\text{mod}} = \sum (Y_m - Y_e)^2 / (\Delta Y_m)^2$
 - assumption for $n \rightarrow \infty$ $\chi^2_{\text{mod}} \rightarrow \chi^2$
 - only valid if: $(\Delta Y_m)^2 \approx Y_e$
 - requirements of χ^2 are fulfilled
- C-statistics is defined as: $C = 2 \sum (Y_e - Y_m \ln Y_e)$
 - W. Cash, 1979, Apj 228, 939
 - assumption for $n \rightarrow \infty$ $C \rightarrow \chi^2 + K$
 - Uses:
 - Poisson errors of individual data points
- F-test is defined as: $F = \chi^2_1 * \text{d.o.f}_2 / \chi^2_2 / \text{d.o.f}_1$



XMM-Newton

Norbert Schartel

Independence Of Data Points

- the measurement of each Y_i is statistically independent
- but the Y_i are correlated (over-sampling in provided spectra, resolution)
- **statistically independent but correlated data points**
- **all (presented) test methods require independent data points**
 - (small sentence at the beginning of each text book)
- **never sample with a bin-size $< FWHM/3$, this provides**
 - complete spectral information
 - avoids over-sampling

Independence Of Data Points

- example: same spectra, same signal/noise for each bin:
1. sampling without considering the resolution:
 - **reduced chi-squared = 0.8191388**
 - for 3364 degrees of freedom
 - **null hypothesis probability = 1.00**
 2. sampling with bin-size $> \text{FWHM}/3$:
 - **reduced chi-squared = 1.094959**
 - for 138 degrees of freedom
 - **null hypothesis probability = 0.210**

Asymptotic Distributed ...

- **all quantities used for the statistical analysis, are itself of statistical nature, i.e. they have an error**
 - **example: $\text{r.m.s.}(\chi^2) = \sqrt{2(\text{d.o.f.})}$**
- **all statistical proofs and theorems are valid for $n \rightarrow \infty$**
- **but we have often not an infinite number of counts**
 - **Example:**
 - **χ^2 tests requires:**
 - **signal to noise ≥ 5 for central Y_i and**
 - **signal to noise ≥ 4 for Y_1 and Y_{last}**
 - **number of $Y \geq 5$**

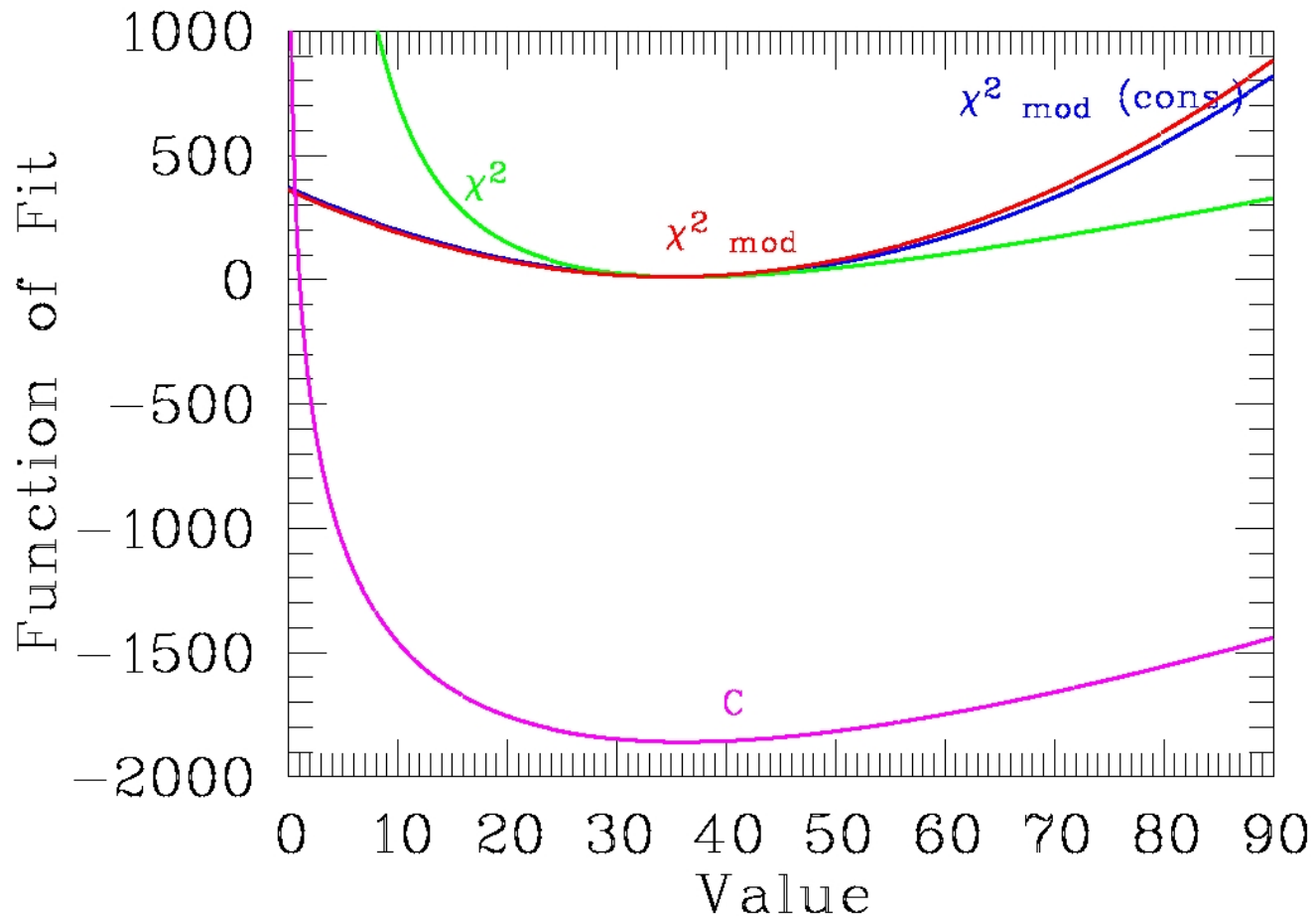
Parameter Estimation

- **intrinsic correlation between error and data!**
- **example:**
 - $\langle Y \rangle = 36$,
 - 10 data points $Y_1, Y_2 \dots$: 30, 42, 24, 48, 33, 39, 34, 38, 36, 36
 - error of $Y_1 = \sqrt{Y_1}$
- mean $Y = 360 / 10 = 36$
- minimum of $\chi^2 = ?$
- minimum of $\chi^2_{\text{mod}} = ?$
- minimum of $C = ?$

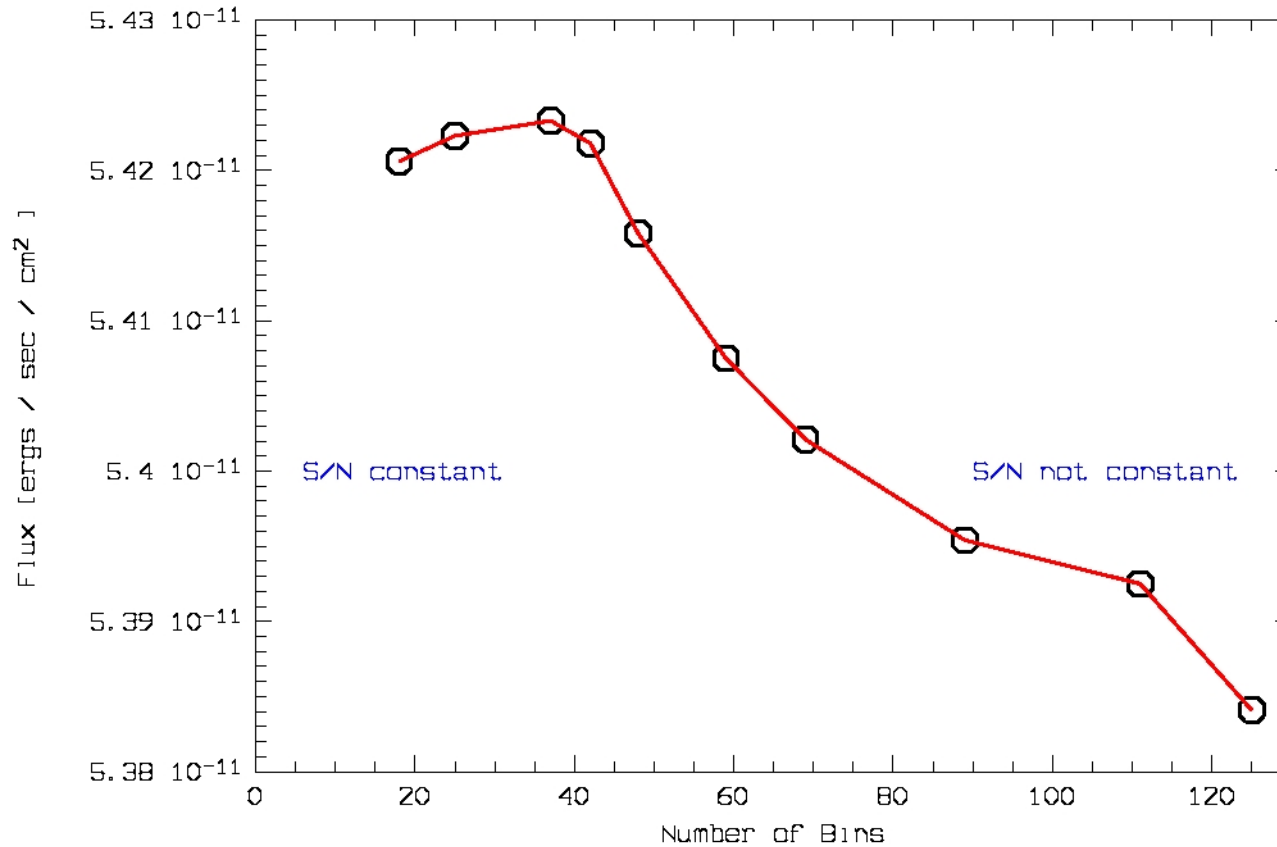
Parameter Estimation

- intrinsic correlation between error and data!
 - example:
 - $\langle Y \rangle = 36$,
 - 10 data points $Y_1, Y_2 \dots$: 30, 42, 24, 48, 33, 39, 34, 38, 36, 36
 - Error of $Y_1 = \sqrt{Y_1}$
- | | | |
|---|---|------|
| – Mean $Y = 360 / 10$ | = | 36.0 |
| – Minimum of χ^2 | = | 36.5 |
| – Minimum of χ^2_{mod} | = | 34.9 |
| – Minimum of C | = | 36.0 |
| – Minimum of χ^2_{mod} (constant error) | = | 36.0 |

Parameter Estimation



Real Example



Parameter Estimation

our data show a correlation between data points and their error: $\Delta Y \approx \sqrt{Y}$

bias for parameter estimation!

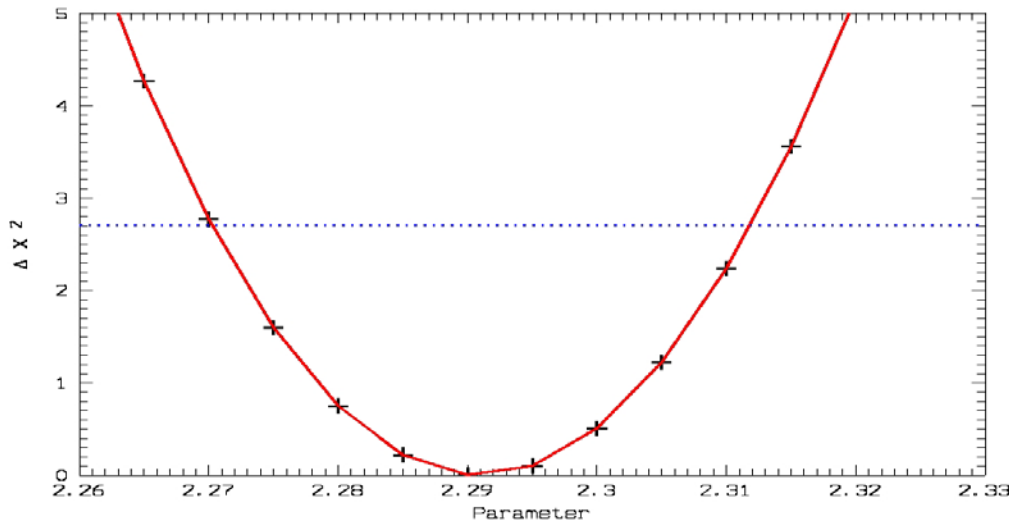
possible solutions:

- **estimator which is independent of error,**
 - e.g. C-statistics
- **ensure, that data points have equal errors,**
 - e.g. binning with constant signal/noise

the higher the numbers of counts the less important are bias effects!

Estimation Of Errors

- **method:** search for the value of the parameter which corresponds to a certain confidence level, i.e. $\Delta \chi^2$
 - **example:**
 - 90 % confidence level of one interesting parameter =>
 - $\Delta \chi^2 = 2.71$ ($\chi^2 = \chi^2_{\text{minimum}} + 2.71$)



Estimation Of Errors

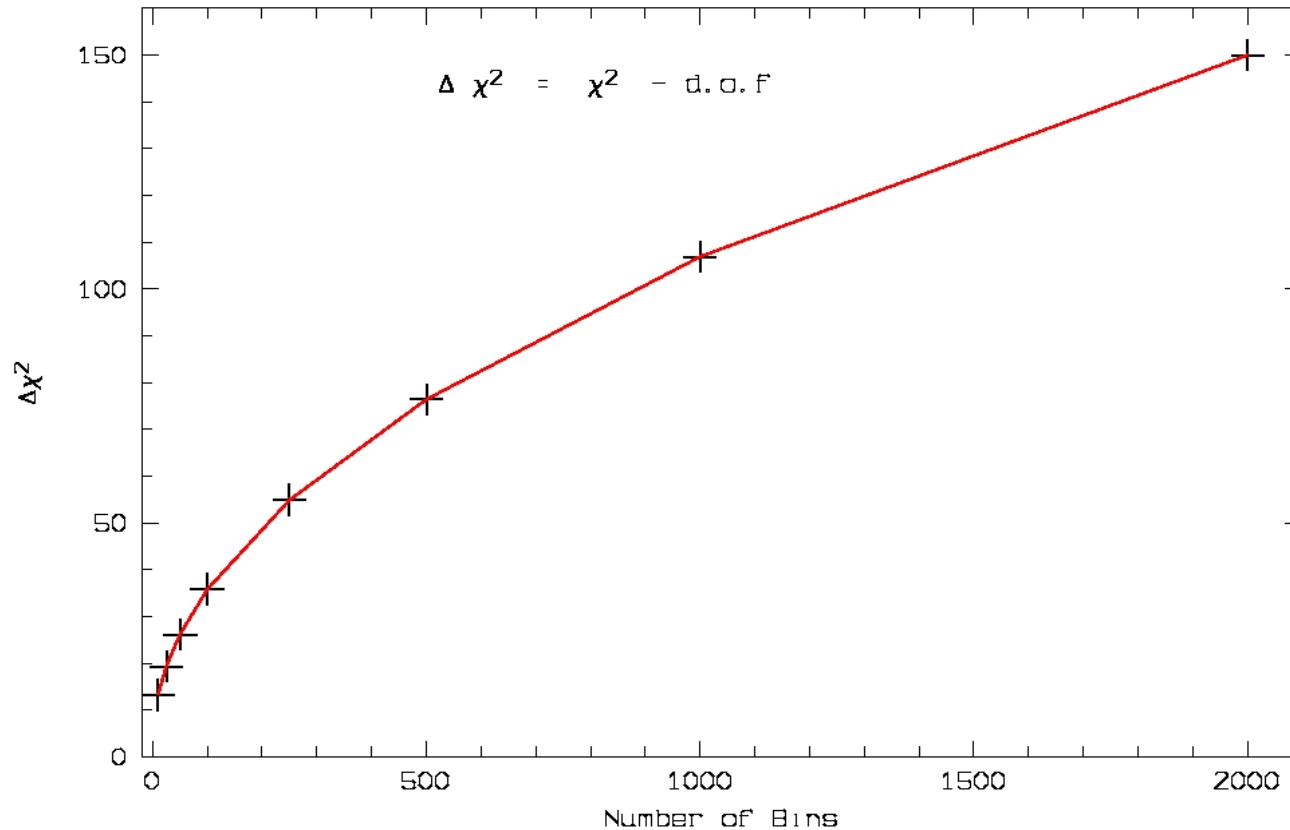
- **be careful:**
- $\Delta \chi^2$ (parameter) is in general not a smooth function with a single minimum
 - plot $\Delta \chi^2$ as function of the parameter
- the interesting parameter is often not independent from other parameters
 - plot two-dimensional $\Delta \chi^2$ contours

Hypothesis Test

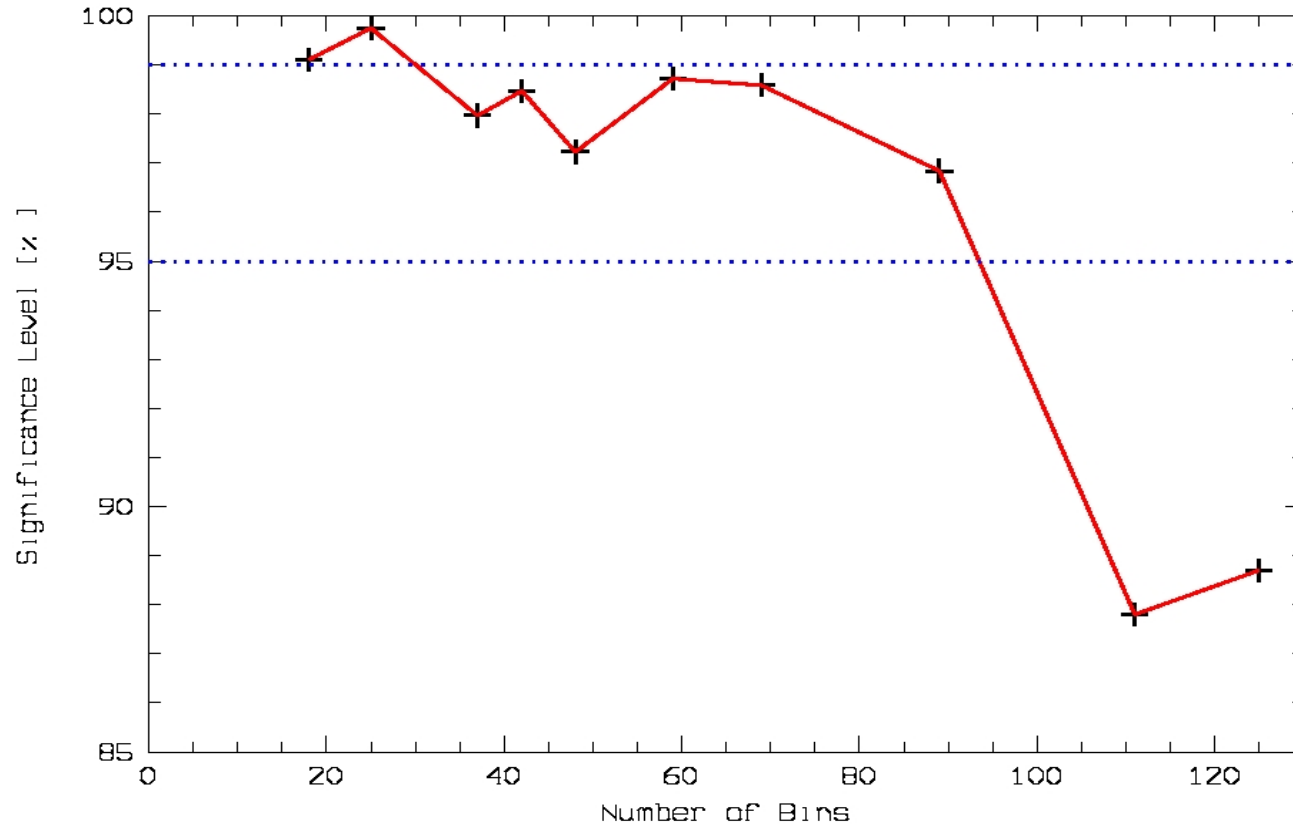
- Literature: J. Kaastra in X-ray spectroscopy in astrophysics, 1999, p 276
- correct model:
 - $\chi^2 = n - m$ (n: number of data points, m:=number of free parameters)
 - $\text{r.m.s.}(\chi^2) = \sqrt{2(n-m)}$
- false model:
 - $\chi^2 = n - m + r$
 - $r \sim \text{number of photons}$ (but independent from n)
- rejection of model
 - $\chi^2 \geq n - m + f \sqrt{2(n-m)} \rightarrow r > \sqrt{2(n-m)}$

Hypothesis Test

99% confidence level for hypothesis rejection



Real Example



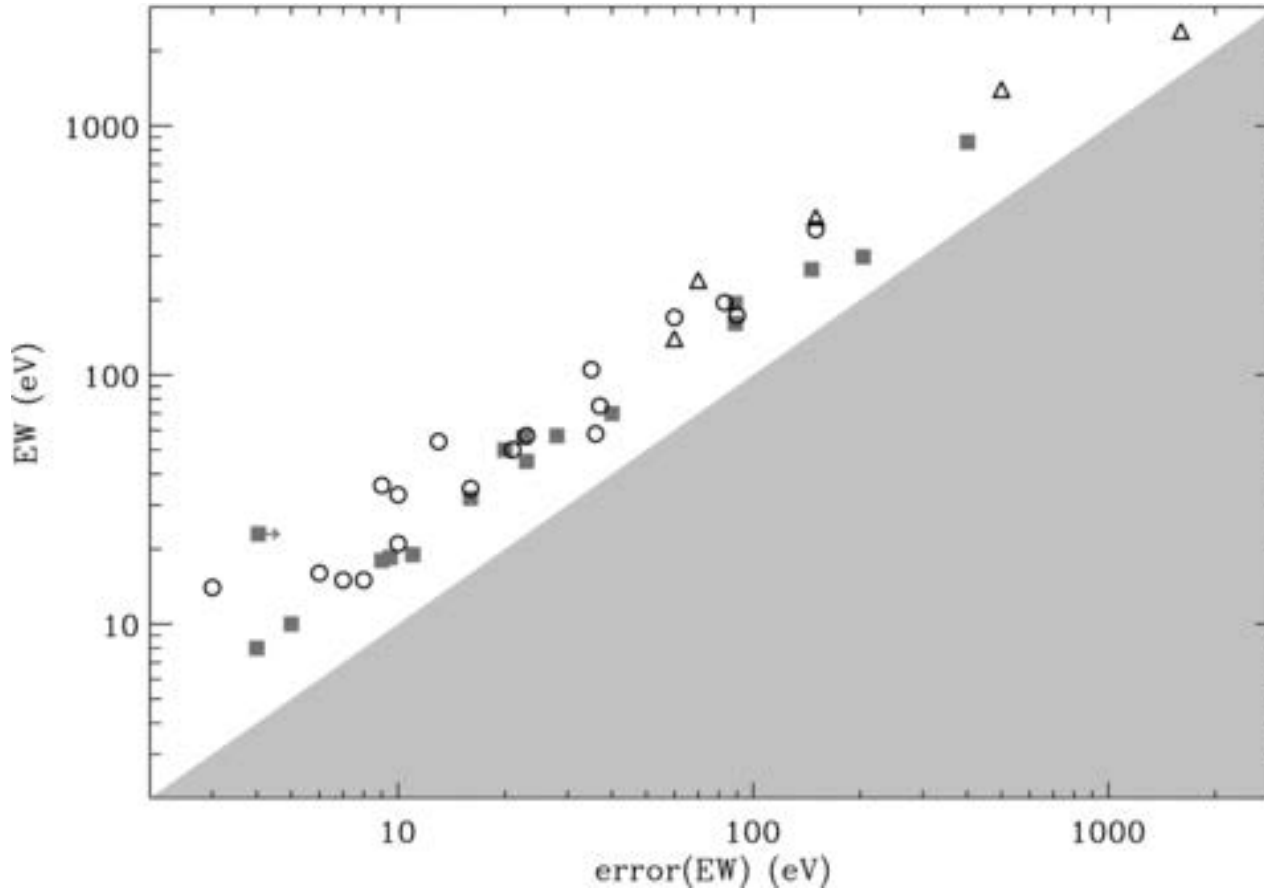
F-Test

- literature: J. Kaastra in X-ray spectroscopy in astrophysics, 1999, p 276
- F is defined as:
 - $F = \chi^2_1 \text{ d.o.f}_2 / \chi^2_2 \text{ d.o.f}_1$
 - $F = \chi^2_1(n-m_2) / \chi^2_2(n-m_1)$
 - where
 - n number of data points
 - m number of fitted parameters
- F is a **statistical quantity!**
 - For large n:
 - **r.m.s (F) = $2 / \sqrt{(n)}$**
 - **$F = 1 + (r_1+r_2)/n$**

F-Test

- **Example:**
 - $r_1 + r_2 = 15$ (independent from n)
 - $m_1=3$,
 - $m_2=6$,
- $n = 40$,
 - d.o.f.₁=37, $\chi^2_1 = 52$, d.o.f.₂=34 $\chi^2_2=34$
 - R.m.s.(F) = 0.3162
 - Probability of improvement:
 - $F \Rightarrow 99.951897\%$
 - $F+R.m.s(F) \Rightarrow 99.997780\%$
 - $F-R.M.S(F) \Rightarrow 98.051537\%$
- $n = 100$,
 - d.o.f.₁=97, $\chi^2_1 = 112$, d.o.f.₂=94 $\chi^2_2=94$
 - R.m.s.(F) = 0.2
 - Probability of improvement:
 - $F \Rightarrow 99.912766\%$
 - $F+R.m.s(F) \Rightarrow 99.999939\%$
 - $F-R.M.S(F) \Rightarrow \text{NO, model 1 is better than model 2}$

The Large View



Vaughan, S.; Uttley, P., 2008, MNRAS 390, 421

Golden Recipe Of The Statistical Analysis

1. golden Recipe of the statistical analysis does not exist.
2. every scientific question is a new statistical challenge and often requires a new concepts
3. In general:
 1. never over-sample a spectra
 2. a bin-width $< 1/3$ FWHM does not add any spectral information, but increases the d.o.f.
 3. if possible bin with constant signal/noise
4. every spectral analysis requires a new optimal solution depending on the analyzed spectra and on the specific scientific question