CIAO Workshop, NCRA Pune, 2017 Oct 23-27

Statistics
for High-Energy Astrophysics

Vinay Kashyap
CXC/CfA

What is AstroStatistics for?

Obtain *estimates* and *uncertainties* on quantities useful for astrophysical inference,

while taking into account instrument sensitivities, statistical fluctuations, and circumstances of observation,

and avoid the pitfalls of making incorrect inferences.

Outline

- Properties of X-ray data
- Making peace with jargon
- Statistical concepts
    1. Error Propagation
    2. Bootstrap
    3. Distributions
    4. *p*-values and Hypothesis Tests
    5. Bayesian analysis
    6. MCMC
    7. Model Fitting
    8. Things to be afraid of
- Tools at our disposal

X-ray data is not like optical data

- A list of events {x,y,t,E}→ marked Poisson process
- Calibration
- Poisson likelihood:

  Prob(k counts when intensity is θ) = $\theta^k e^{-\theta}/\Gamma(k+1)$
- Gaussian approximation is widely used: $\mu = \sigma^2 = k$

Jargon

- Probability, p(·) — *frequency of occurrence* or *degree of belief*
- Likelihood, $\mathcal{L} \equiv p(D \mid \theta)$ — probability of obtaining observed data assuming a particular model
- Fitting
  - χ^2 — measure of closeness, also goodness of fit ≡ –2 ln(Gaussian likelihood)
  - cstat/cash ≡ –2 ln(Poisson Likelihood)
- *p*-values/Null Hypothesis Significance Testing
- Tests of dissimilarity: Kolmogorov-Smirnoff, F-test

1. Error Propagation

- How to propagate the uncertainty from one stage to another
- Simple case: assume everything is distributed as a Gaussian, and has well-defined means and standard deviations
- $g=g(a_i)$

  $\Rightarrow \sigma^2(g) = \sum_i (\partial g/\partial a_i)^2\, \sigma^2(a_i)$

1. square adding

$g = g(a_i)$

$\sigma^2(g) = \sum_i (\partial g/\partial a_i)^2\ \sigma^2_i$

$g = C \cdot a$

$\rightarrow \sigma_g = C \cdot \sigma_a$

$g = 1/a$

$\rightarrow \sigma_g/g = \sigma_a/a$

$g = \ln(a)$

$\rightarrow \sigma_g = \sigma_a/a$

$g = a + b$

$\rightarrow \sigma^2_g = \sigma^2_a + \sigma^2_b$

2. Bootstrap

- How to estimate the uncertainty within almost any set of measurements
- Steps:
  - 1: construct summary statistic
  - 2: extract random sample of same size from original dataset and recompute summary statistic from Step 1
  - 3: repeat Step 2 a large number of times and compute mean and variance of summary statistic
- Quick and easy
- Accurate, if sample in hand is a good representation of population (e.g., don't try this with power-laws)

3. Distributions

- **Binomial** — one or the other, with probability ρ

  k of one out of a total of N, $p(k|N,\rho) = {}^{N}C_{k}\, \rho^{k}\, (1-\rho)^{N-k}$

- **Poisson** — events occur randomly

  $p(k|\,\theta) = (1/k!)\, \theta^{k}\, e^{-\theta}$

- **Gaussian** (aka **Normal**)— all summary statistics that have a sufficiently large sample

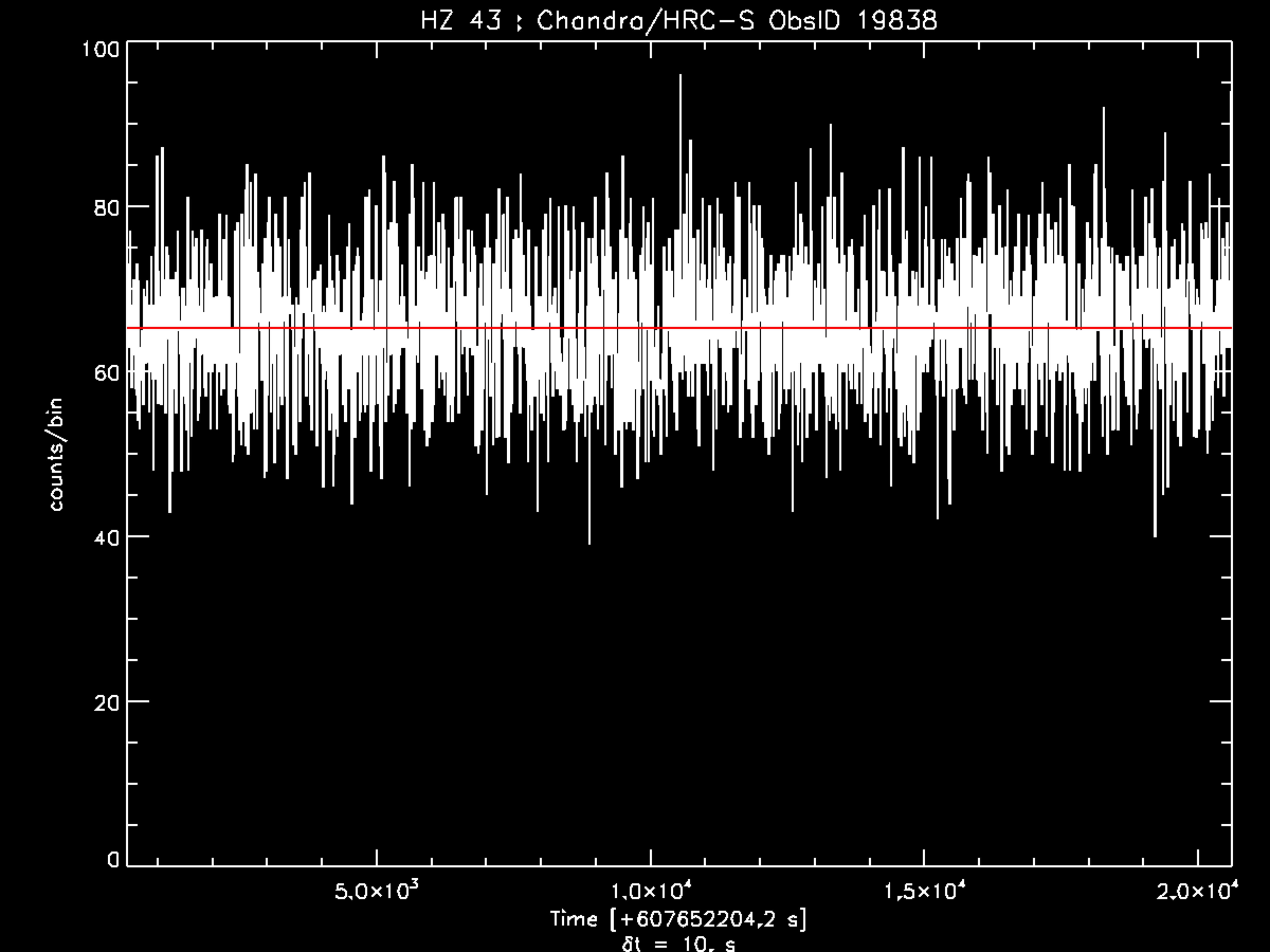
  $f(x;\mu,\sigma^2) = (1/\sqrt{2\pi\sigma^2})\, \exp[-(x-\mu)^2/(2\sigma^2)]$

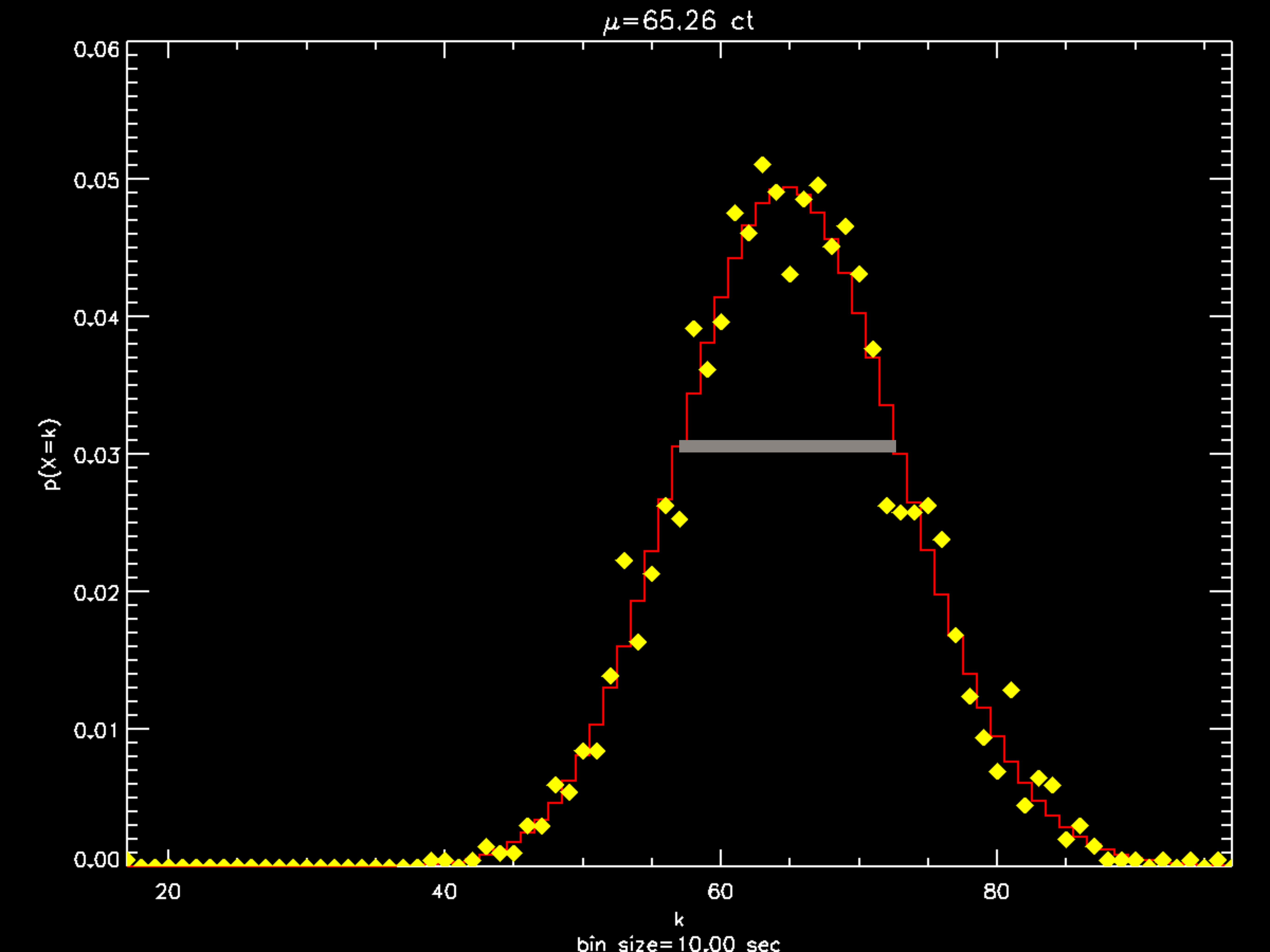
- **Gamma** — continuous variable conjugate to Poisson

  $p(x;\alpha, \beta) = \beta^{\alpha}/\, \Gamma(\alpha) \cdot x^{\alpha-1}\, e^{-\beta x}$, $x\geq 0$, $\alpha\geq 0$, $\beta\geq 0$; Poisson for $\beta=1$ and $\alpha=k+1$

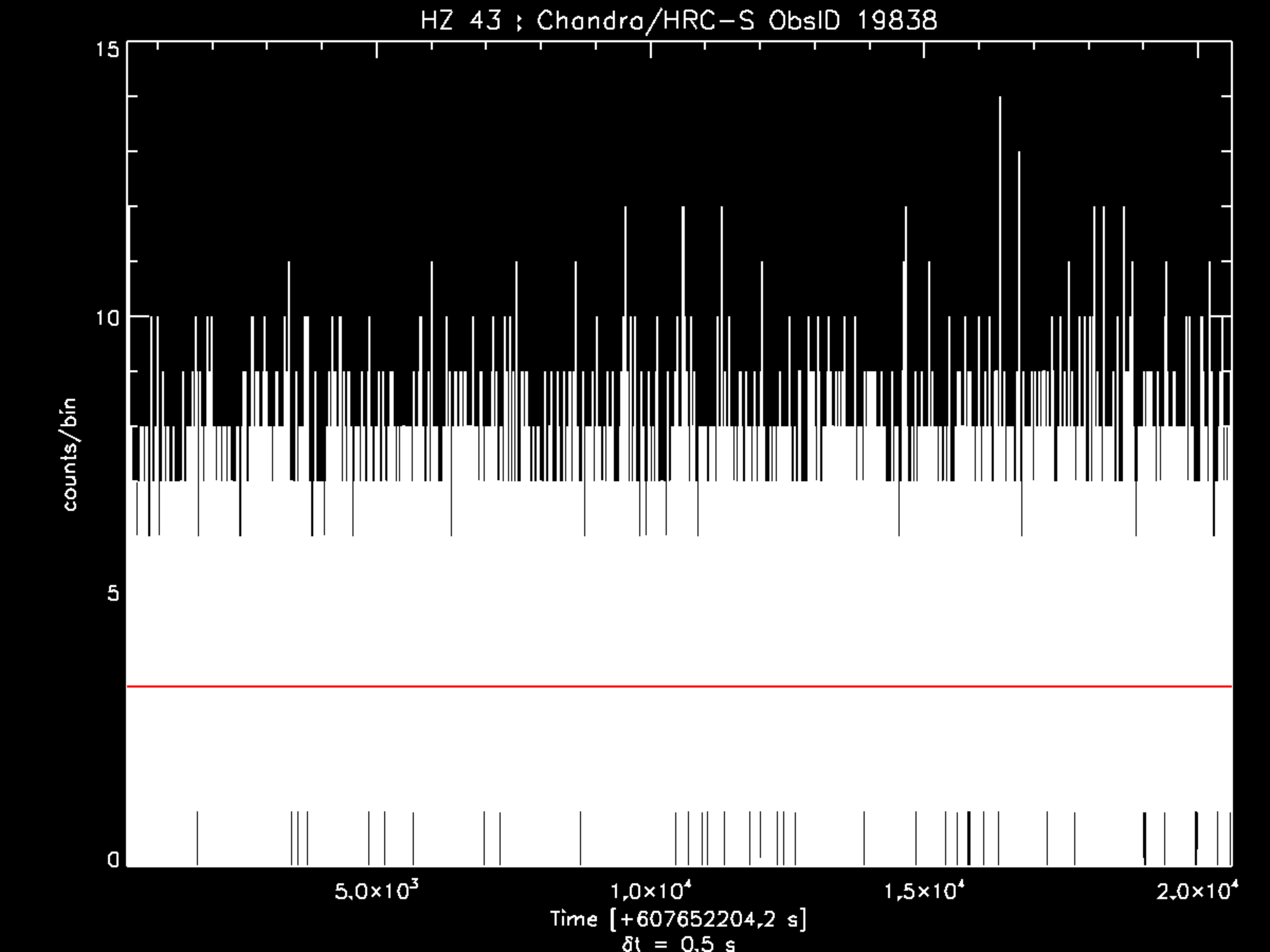
- χ^2 — measure of similarity and distance between samples

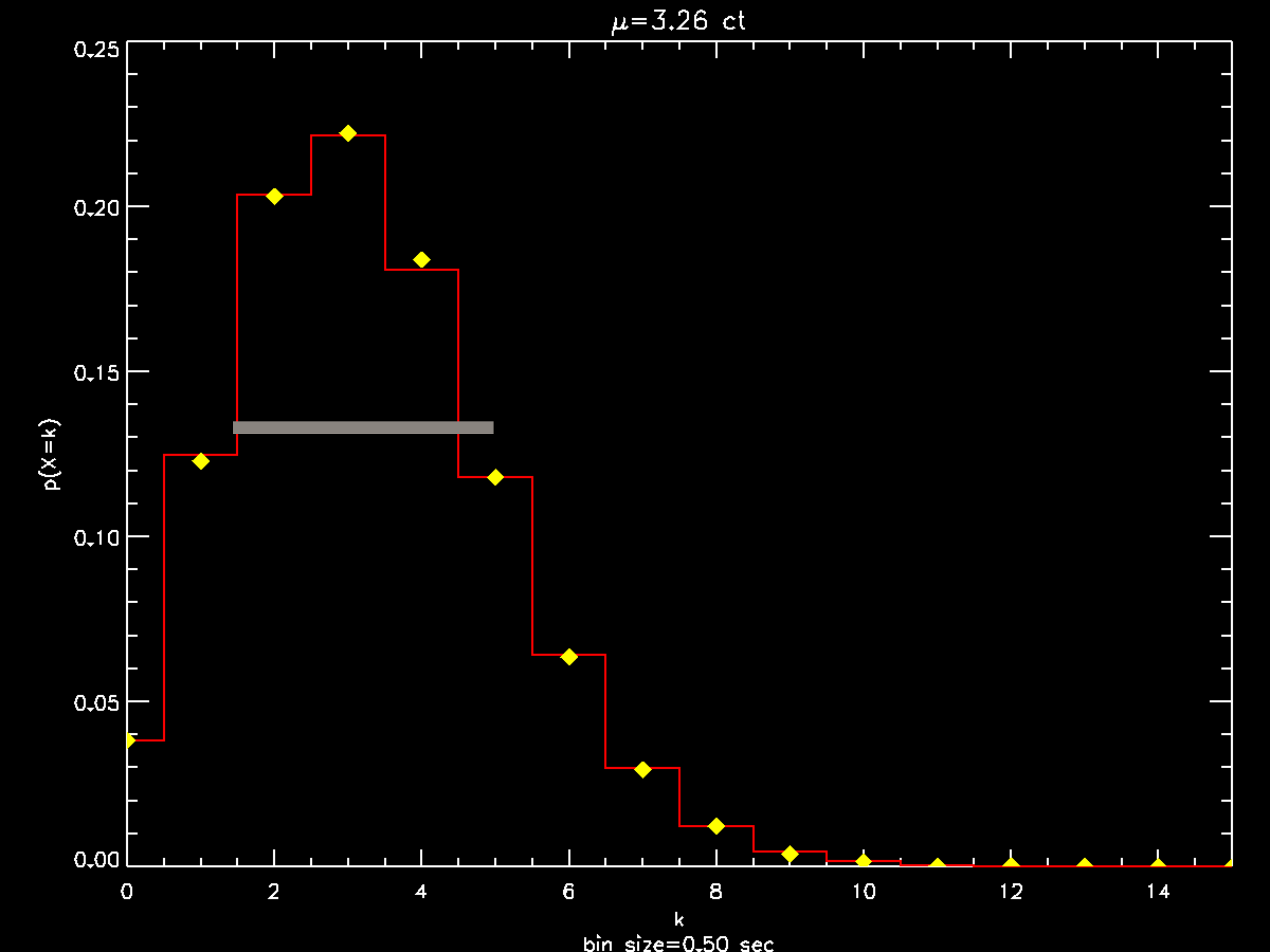
  $p(\chi^2|n) = (2^{-n/2}/(n/2-1)!)\, (\chi^2)^{(n-2)/2}\, \exp[-\chi^2/2] \propto (\chi^2)^{(n/2-1)}\, \exp[-\chi^2/2] \equiv \mathrm{Gamma}(\chi^2;n/2,-1/2)$


HZ 43 ; Chandra/HRC-S ObsID 19838
counts/bin
0
20
40
60
80
100
5.0×10^3
1.0×10^4
1.5×10^4
2.0×10^4
Time [+607652204.2 s]
δt = 10. s



μ=65.26 ct
0.06
0.05
0.04
0.03
0.02
0.01
0.00
p(X=k)
20
40
60
80
k
bin size=10.00 sec


HZ 43 ; Chandra/HRC-S ObsID 19838
counts/bin
15
10
5
0
5.0×10^3
1.0×10^4
1.5×10^4
2.0×10^4
Time [+607652204.2 s]
δt = 0.5 s


μ=3.26 ct
0.25
0.20
0.15
0.10
0.05
0.00
p(X=k)
0
2
4
6
8
10
12
14
k
bin size=0.50 sec

4. *p*-values and Hypothesis Tests

- Compare distributions by setting up competing hypotheses
- Null hypothesis H_0 is that both samples are drawn from the same distribution
- Calculate a statistic from the data and compare to the expected distribution of the statistic. If calculated value *exceeds a critical threshold*, reject the null hypothesis.

5. Basics of Bayesian Analysis

- Axioms
- Bayes' Theorem
- Example — hardness ratio

5.1 Axioms of Probability Theory

1. p(A or not A) = p(A) + p(not A) = 1

2. p(A and B) = p(B) p(A given B) ≡ p(A) p(B given A)

5.1 Axioms of Probability Theory

1. p(A or not A) = p(A) + p(not A) = 1

   $p(A + \bar{A}) = p(A) + p(\bar{A}) = 1$

2. p(A and B) = p(B) p(A given B) ≡ p(A) p(B given A)

   $p(A\,B) = p(B)\,p(A \mid B) \equiv p(A)\,p(B \mid A)$

5.2 Bayes' Theorem

$$p(AB|C) = p(A|BC)\, p(B|C) = p(B|AC)\, p(A|C)$$

$$\mathbf{p(A|BC) = p(B|AC)\, p(A|C) \,/\, p(B|C)}$$

$$p(\theta|D\ I) = p(D|\theta\ I)\, p(\theta|I) \,/\, p(D|I)$$

5.2 prior, likelihood, posterior

$p(\theta|D\ I) = p(D|\theta\ I)\ p(\theta|I)\ /\ p(D|I)$

a priori distribution: $p(\theta|I)$

likelihood: $p(D|\theta\ I)$

a posteriori distribution: $p(\theta|D\ I)$

5.3 Example: Hardness Ratios

- Measure counts in Soft (*S:* lower energies, longer wavelengths; e.g., 1/2-2 keV) and Hard (*H:* higher energies, shorter wavelengths; e.g., 2-8 keV) passbands
- HR := (H-S)/(H+S), R := S/H, C := log(S/H)
- Problem: Gaussian error propagation fails for low counts, or when HR is close to ±1, or because ratios are not distributed as Gaussians
- Need to compute p(*hr*|H,S), p(*r*|H,S), p(*c*|H,S)

5.3 Example: Hardness Ratios

- For all the details, see Park et al. 2006 (ApJ 652, 610)
- $p(H \mid h)$ and $p(S \mid s)$ are Poisson likelihoods, $p(s)$ and $p(h)$ are usually chosen as Gamma priors
- $p(h,s \mid H,S) \propto p(H,S \mid h,s)\, p(h,s) \equiv p(H \mid h)\, p(S \mid s)\, p(h)\, p(s)$

$hr=(h-s)/(h+s)$, $w=(h+s) \Rightarrow h,s=(1 \pm hr)\cdot w/2$

$J(h,s;hr,w) = \mid \partial(h,s)/\partial(hr,w) \mid = w/2$

$p(h,s \mid \ldots)\, dh\, ds \equiv p((1+hr)\cdot w/2,(1-hr)\cdot w/2 \mid \ldots)\, J(h,s;hr,w)\, dhr\, dw$

$p(hr)\, dhr = dhr \int dw\, p(hr,w)$

6. Markov Chain Monte Carlo

- What is it?
  - A method to quickly explore high-dimensional parameter spaces and obtain representative measures of parameter values and uncertainties
- Why do it?
  - Robust, insensitive to starting conditions, easy to code
- How does it work?
  - Compute the likelihood for given parameter values, get a new, randomly drawn value, and compare the new likelihood to the old one
  - If it improves the likelihood, accept the new value and repeat the cycle
  - If it does not improve the likelihood, accept with a probability equal to the ratio, else reject and get a new value

7. Fitting

- Non-linear metric minimization
  - χ^2 (any of several varieties) — $\sum_i (D_i - M_i)^2/\sigma_i^2$
    - fit is good if $\chi^2/\text{dof} \sim 1 \pm \sqrt{2/\text{dof}}$
  - `cstat` — $2 \sum_i (M_i - D_i + D_i \cdot (\ln D_i - \ln M_i))$
    - asymptotically χ^2 — otherwise use parametric bootstrap to determine goodness of fit

7.1 Model Comparison

- Model comparison
  - use F-test iff simpler ("null") model is fully contained within complex ("alternate") model
  - otherwise use posterior predictive p-value checks (see Protassov et al. 2002, ApJ 571, 545):
    - simulate fake datasets from best-fit parameters of null model
    - fit with both null and alternate model
    - compute distributions of ratios of the best-fit statistic and compare against the ratio for actual data
    - if ratio from observed data is far in the tail of the simulated distribution, then it is unlikely that the null model is a good descriptor of the data

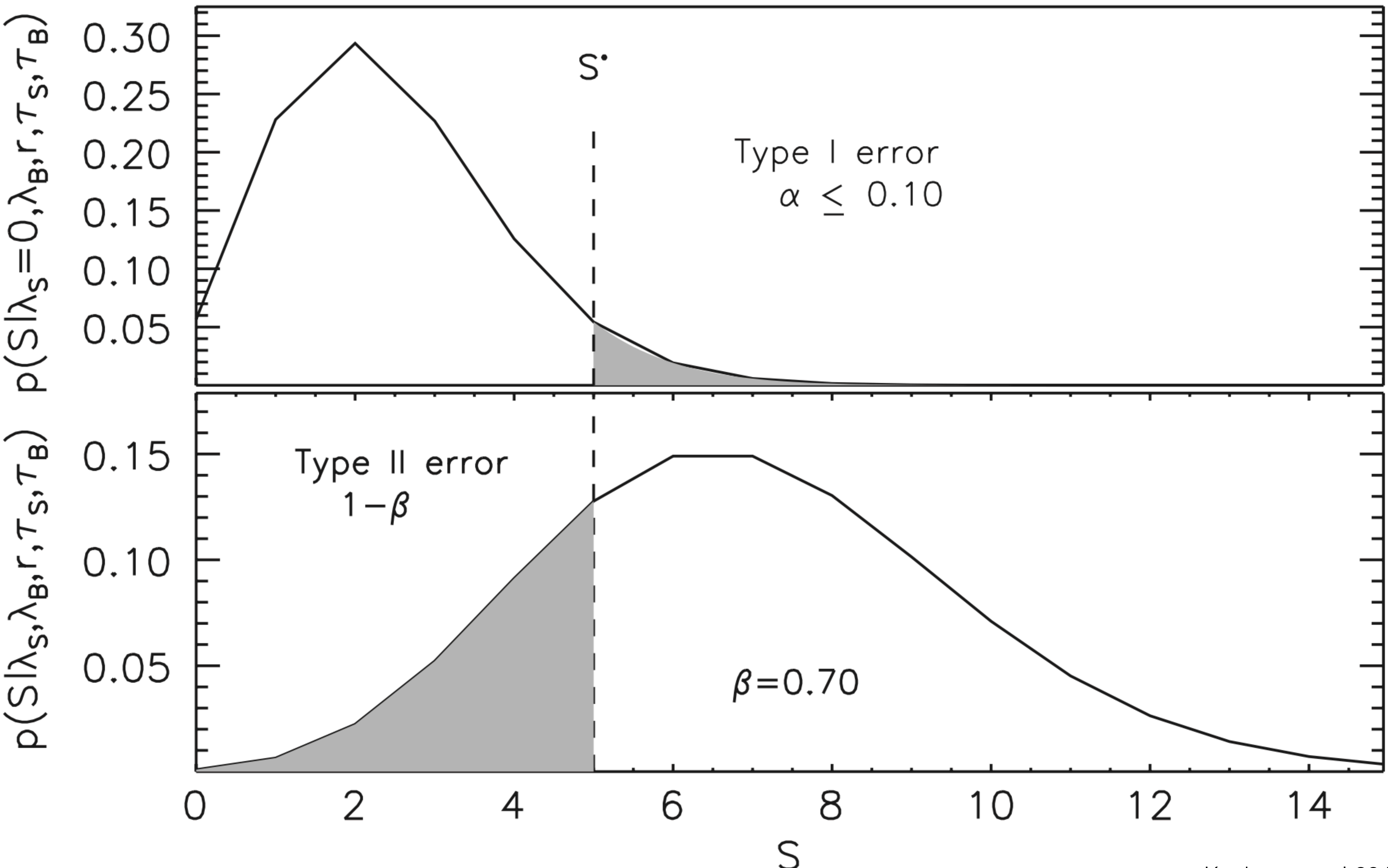
8. Danger Danger

- asymptotic validity — be aware of the assumptions made to get easy analytical results (e.g., *p*-value for F-test, χ^2 as measure of goodness)
- convergence, stopping rules, effect of priors — always do sensitivity tests
- overfitting — to avoid fitting fluctuations in the data, balance bias against variance
- *p*-values — measure of how far in the tail of a distribution the current observation is, not a proof of the validity of an alternative hypothesis, nor of the falsity of the null hypothesis
- Type I, Type II, Type S, Type M errors — false positive, false negatives, sign errors on weak effects, Eddington Bias

8. Types of Bias

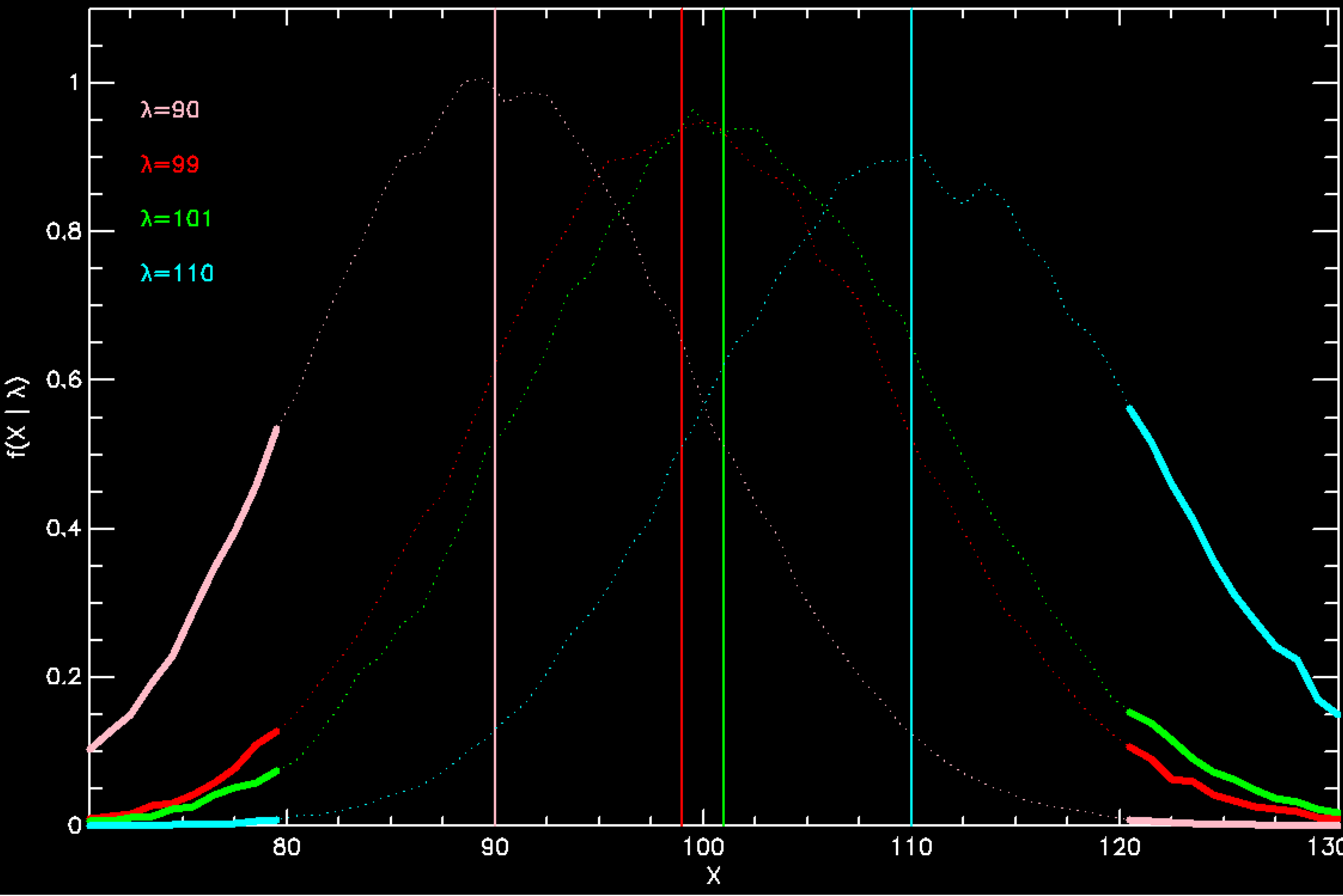
- Type I — false positives, when you claim a detection over a background because of a fluctuation above some threshold
- Type II — false negatives, when you fail to detect an event because its response fell below the detection threshold
- Type M — an incorrect estimation of the *size* of the effect because large fluctuations are preferentially detected (cf. Eddington bias)
- Type S — an incorrect estimation of the *sign* of a weak effect because of fluctuations in the wrong direction

8. Type I and Type II Errors



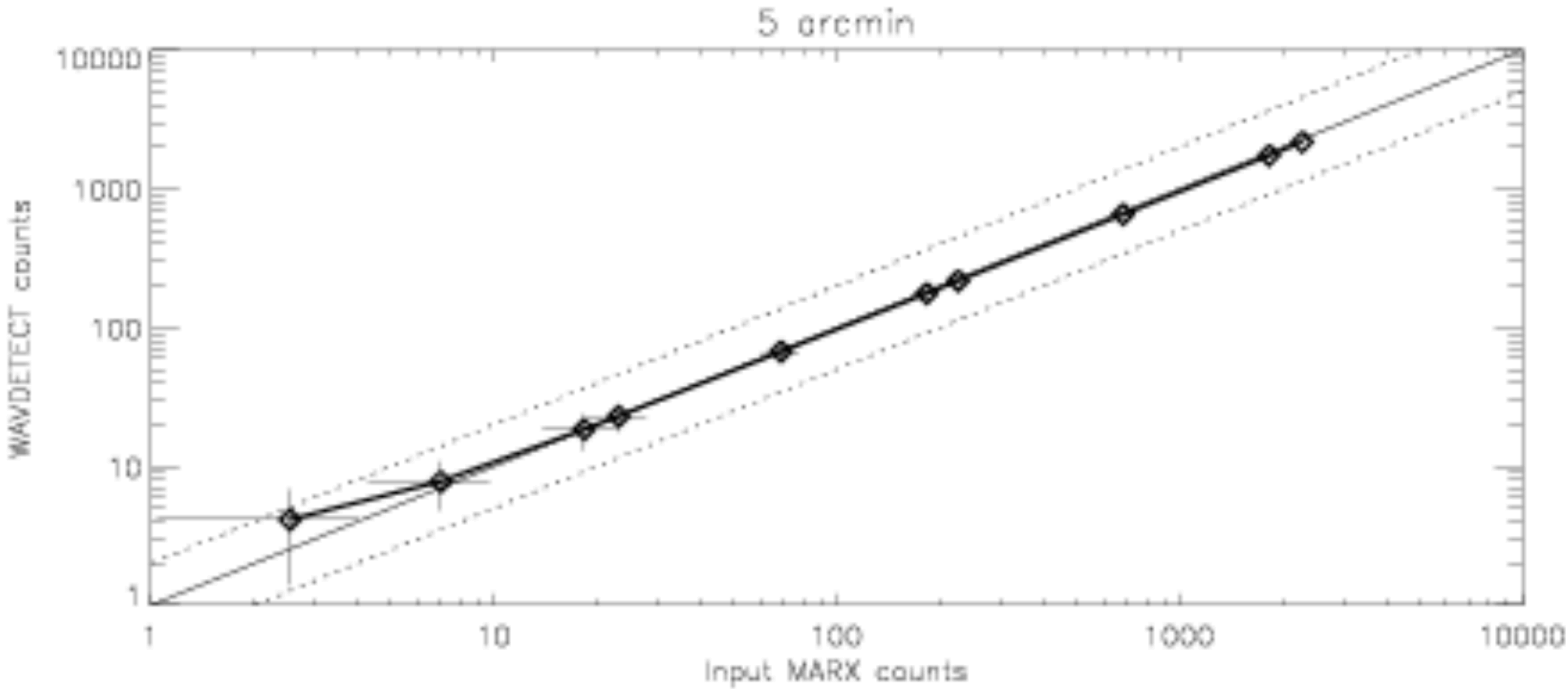
Kashyap et al 2010

8. Type S Error



8. Eddington Bias

Eddington, A.S., 1913, MNRAS, 73, 359, *On a formula for correcting statistics for the effects of a known error of observation*



Kashyap 2001, Power of wavdetect

Statistical Tools in CIAO/Sherpa

- fit/conf/projection: non-linear minimization fitting and uncertainty intervals
- get_draws: MCMC engine (van Dyk et al. 2001, ApJ 548, 224)
- calc_ftest: model comparison via F-test
- plot_pvalue, plot_pvalue_results: to do posterior predictive p-value checks (Protassov et al. 2002, ApJ 571, 545)
- glvary: light curve modeling (Gregory & Loredo 1992, ApJ 398, 146)
- celldetect/wavdetect/vtpdetect: source detection in images
- aprates: Bayesian aperture photometry (Primini & Kashyap 2014, ApJ 796, 24)
- the python interpreter in Sherpa gives access to python libraries, and can be used to call upon libraries in R, which are written by statisticians for statisticians