Previous lecture:

Comparing data with a model:
Least-squares fitting

maximum likelihood method:
Gaussian data

Confidence levels

Outliers!

Today:

“real” maximum likelihood method: Poissonian data

Finding periodicities in data

- Lomb-Scarle diagrams
- Phase dispersion minimisation
- Fourier techniques

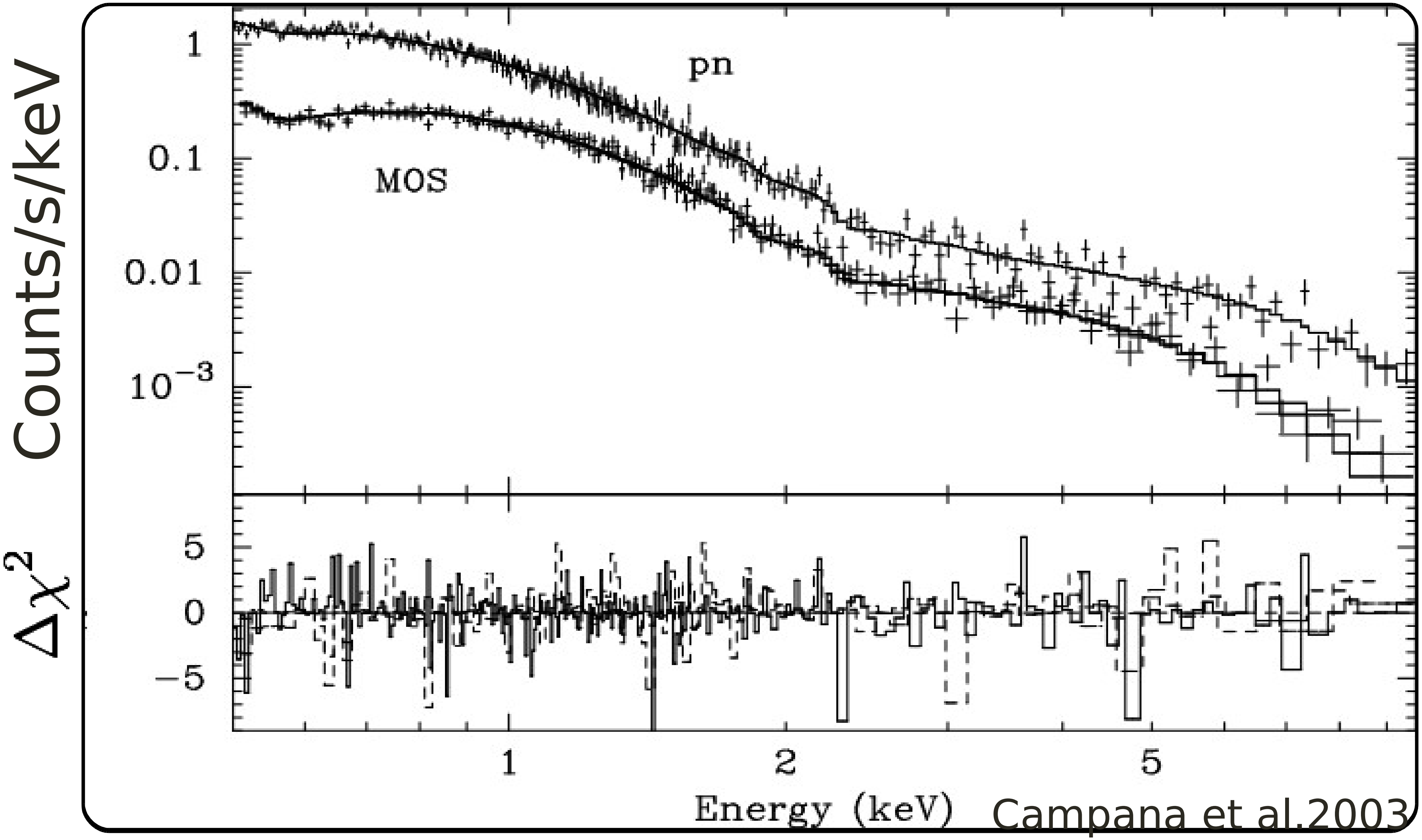
Comparing two distributions

K-S test

OAF2 chapter 6.1 & 6.2

see Num Res Chapter 13.8, 14.3, & 14.7

Example simple Monte Carlo simulation I



Campana et al.2003

errors on data-points Gaussian distributed

Simulation: replace each point with a value

from the Gaussian distribution, redo fit to minimise

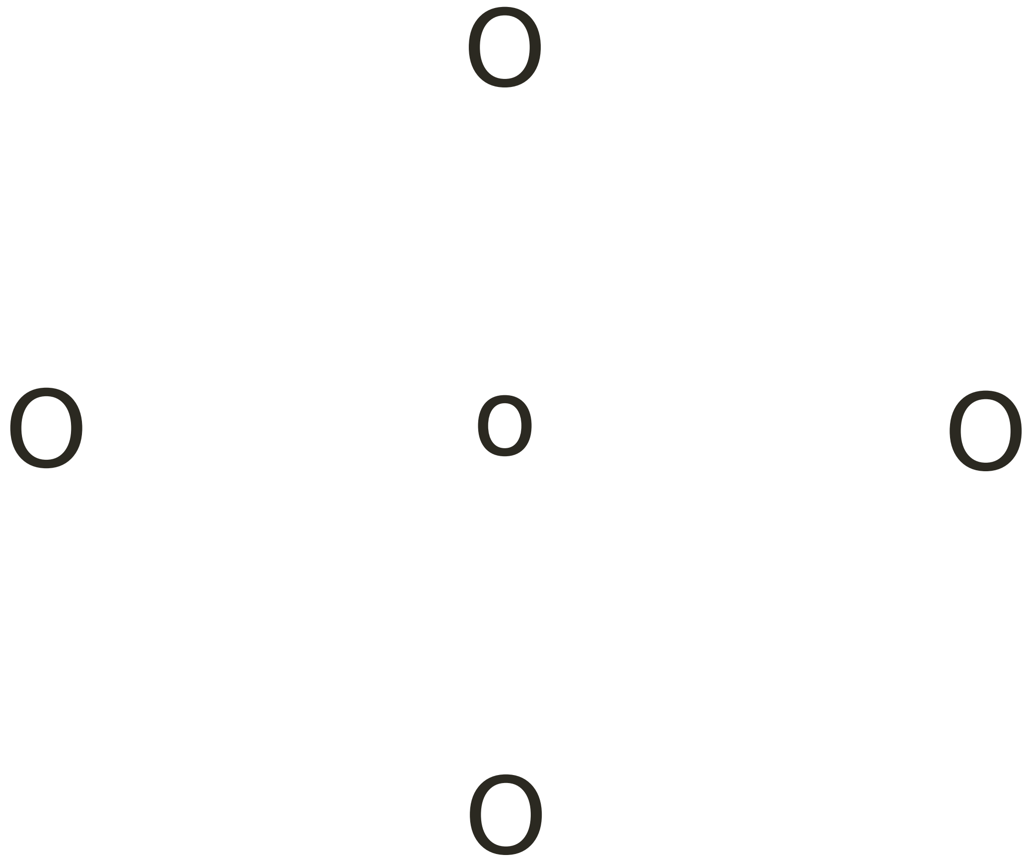
χ^2 repeat often

provide a distribution in χ^2 $\Delta\chi^2 = 1 \Rightarrow 68\%\ \mathrm{confidence}$

Example simple Monte Carlo simulation II

Phase binning

$$\phi_i = \frac{t_i}{P} - \mathrm{INT}(\frac{t_i}{P})$$



How often do we have to observe the system when observing at random times to fill each of 10 phase bins?

Maximum likelihood method

(Poisson noise, unbinned data)

probability to find n_i photons when m_i expected according to your model

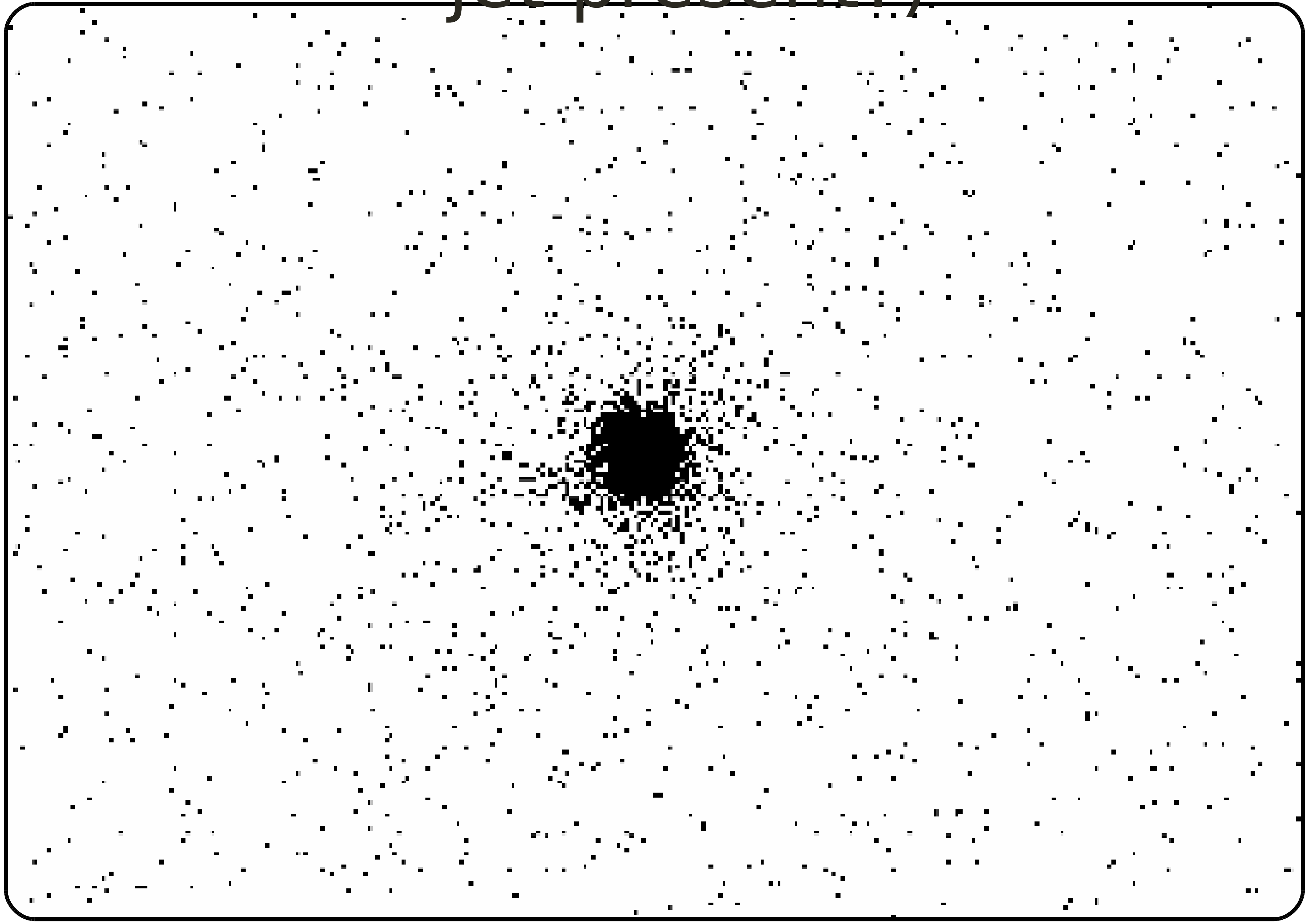
for each pixel i in an image
$$P_i = \frac{m_i^{n_i} e^{-m_i}}{n_i!}$$

total probability
$$L' \equiv \prod_i P_i$$

$$\ln L' \equiv \sum_i \ln P_i = \sum_i n_i \ln m_i - \sum_i m_i - \sum_i \ln n_i!$$

minimise
$$\ln L \equiv -2(\sum_i n_i \ln m_i - \sum_i m_i)$$

Maximum likelihood method
(application X-ray binary Cir X-1, a jet present?)



part of a Chandra HRC observation
model and subsequently subtract PSF
only close to the source the assumption of a constant background is valid

Detection of a constant background, A, plus a source of strength B of which a fraction falls on pixel i

$$-0.5\ln L = \sum_i n_i \ln(A + Bf_i) - \sum_i (A + Bf_i)$$

again search for the minimum of L for variations in A and B

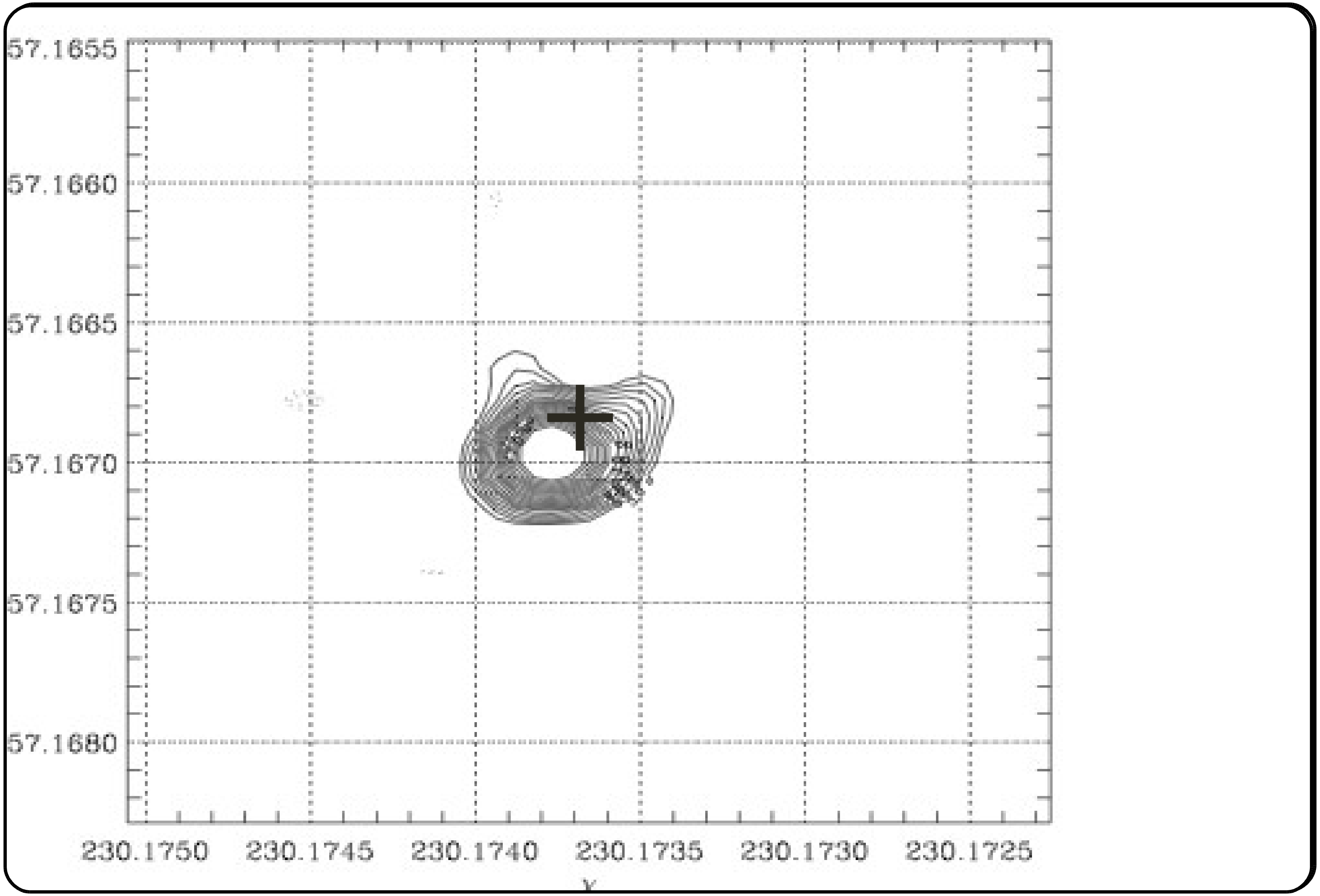
f_i determined independently in some cases

total pixels Z

$$\frac{\partial \ln L}{\partial A} = 0 \Rightarrow \sum_i \frac{n_i}{A + Bf_i} - \sum_i (1) = \sum_i \frac{n_i}{A + Bf_i} - Z = 0$$

$$\frac{\partial \ln L}{\partial B} = 0 \Rightarrow \sum_i \frac{n_i f_i}{A + Bf_i} - \sum_i (f_i) = \sum_i \frac{n_i f_i}{A + Bf_i} - 1 = 0$$

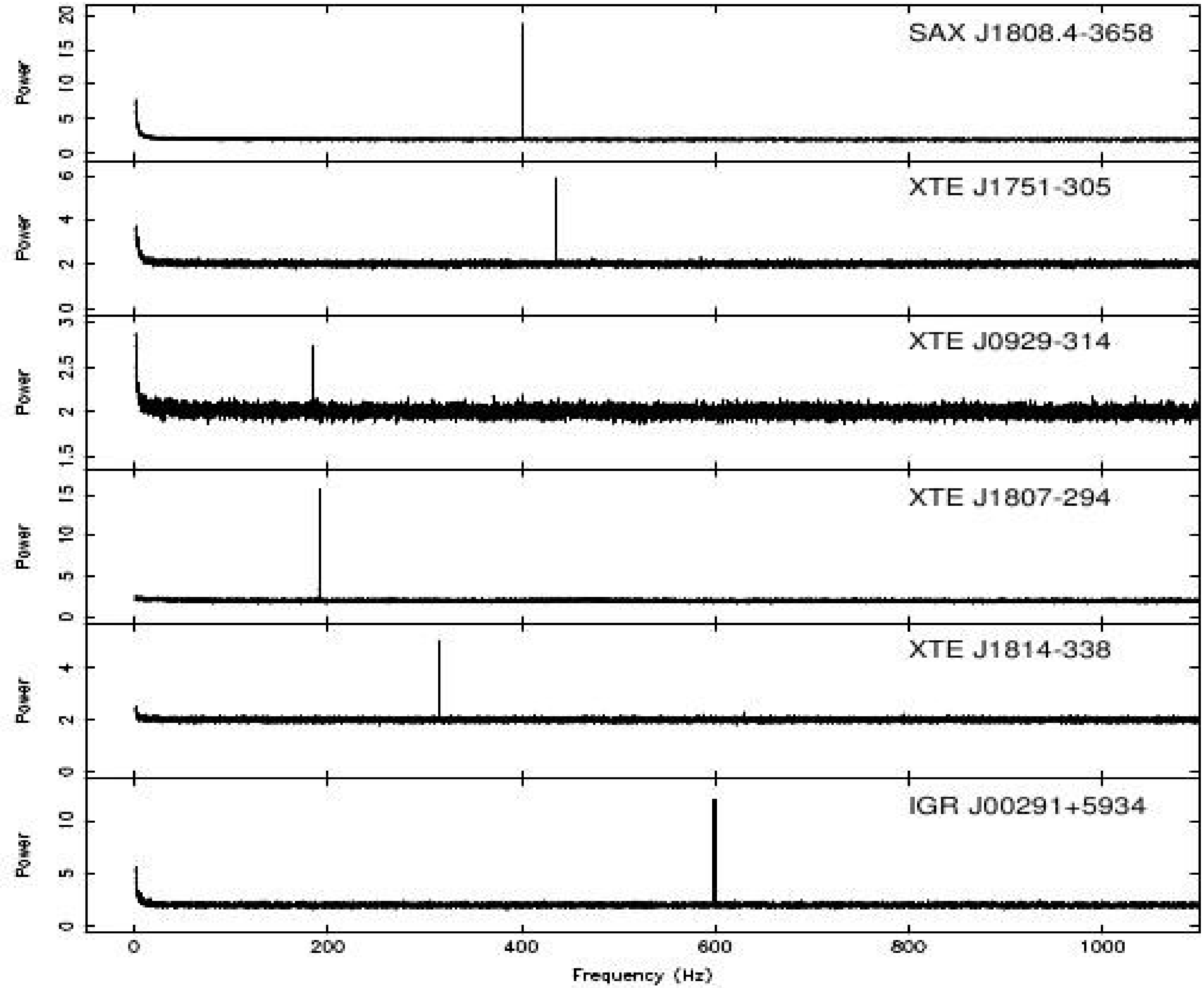
application maximum likelihood method X-ray binary Cir X-1



one source subtracted

Period finding I

Fourier methods

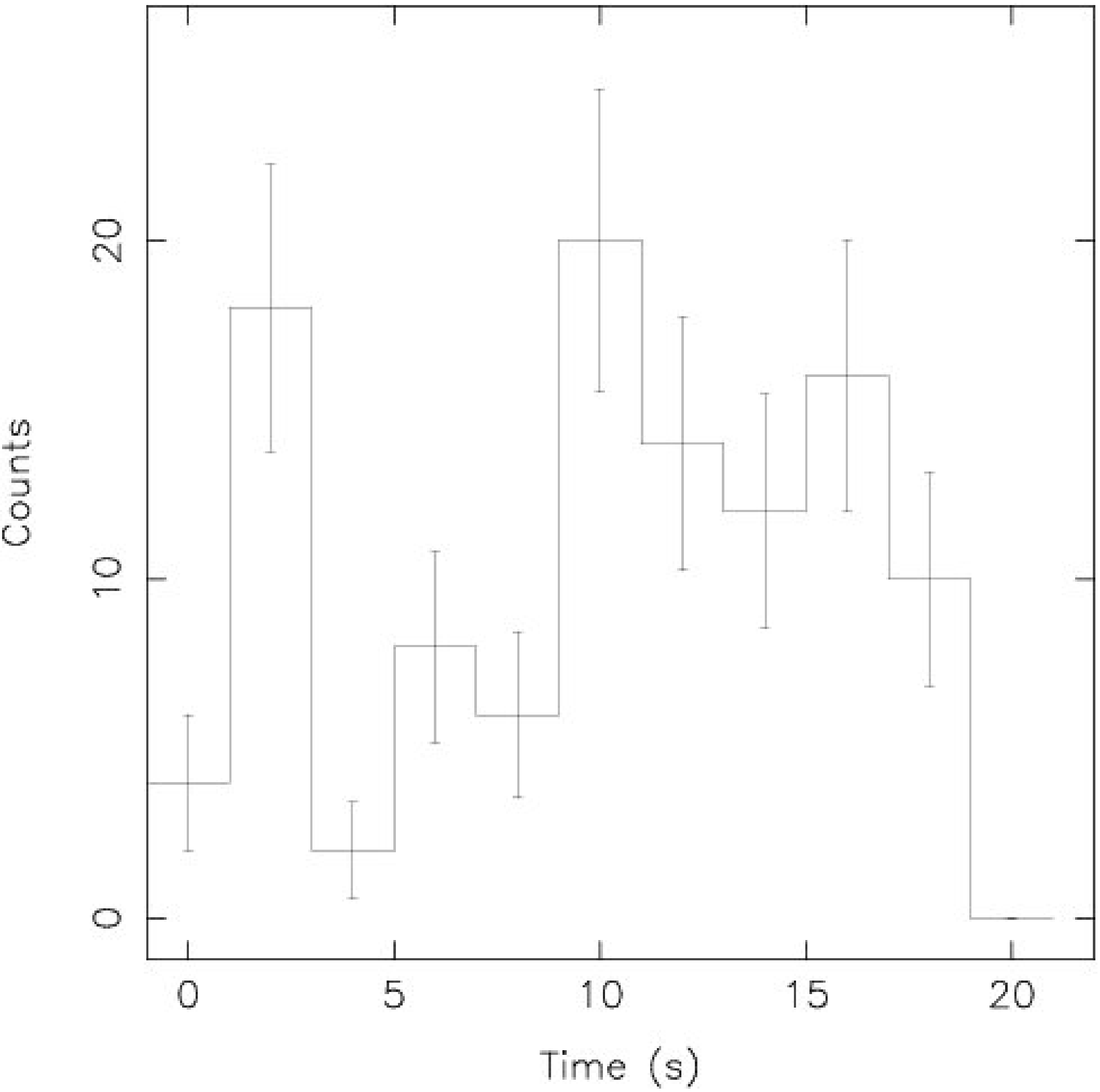


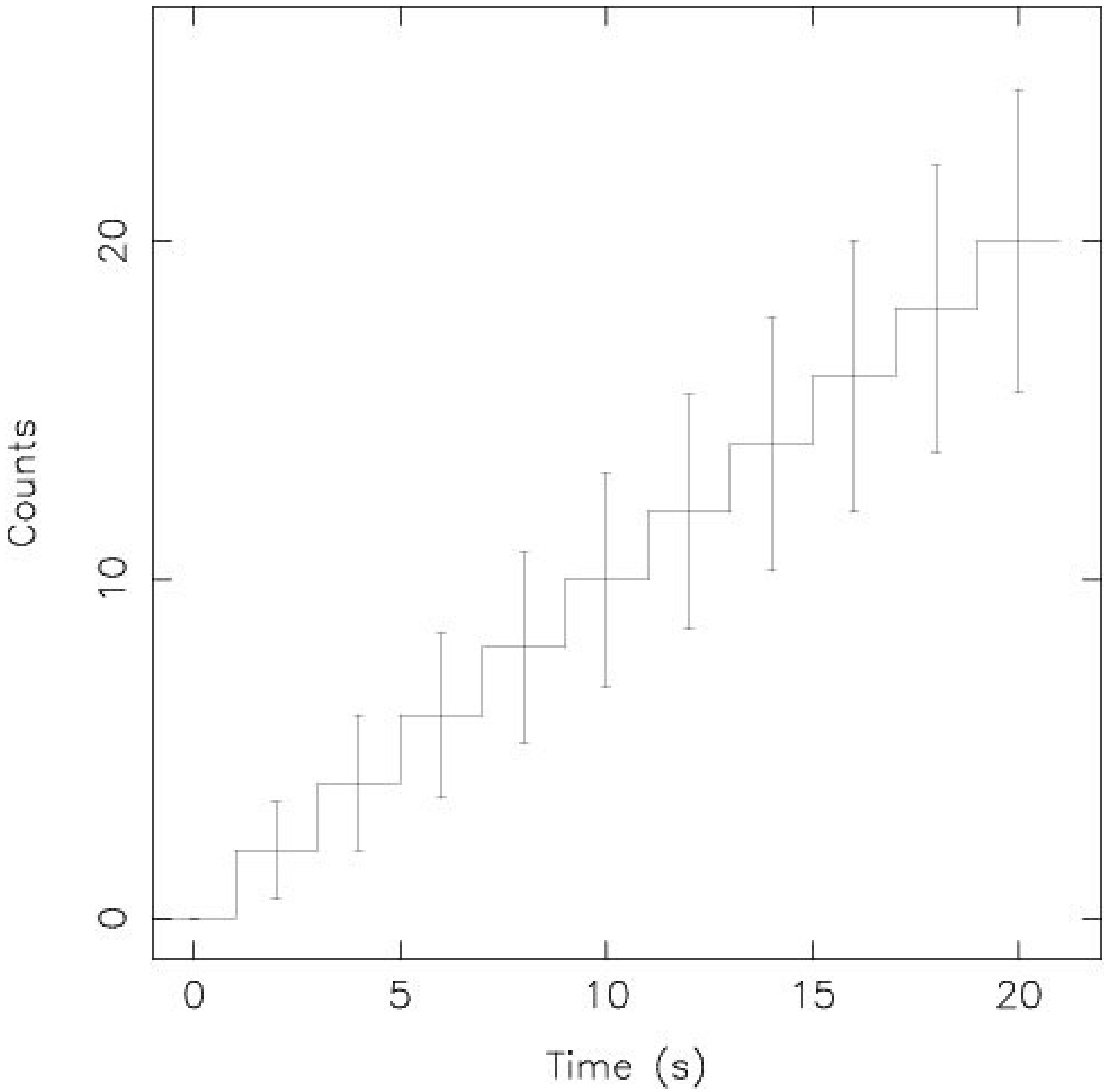
What if data is not evenly sampled

Fill in gaps/resample?

Chi squared fitting?

Chi squared same but data clearly different in terms of variability





Period finding I

Unevenly Sampled Data: Lomb-Scargle

mean $\bar{h} = \frac{1}{N}\sum_i h_i$ variance $\sigma^2 = \frac{1}{N-1}\sum_i (h_i - \bar{h})^2$

Least-squares fitting of

$$h_i = A\cos(\omega t_i) + B\sin(\omega t_i)$$

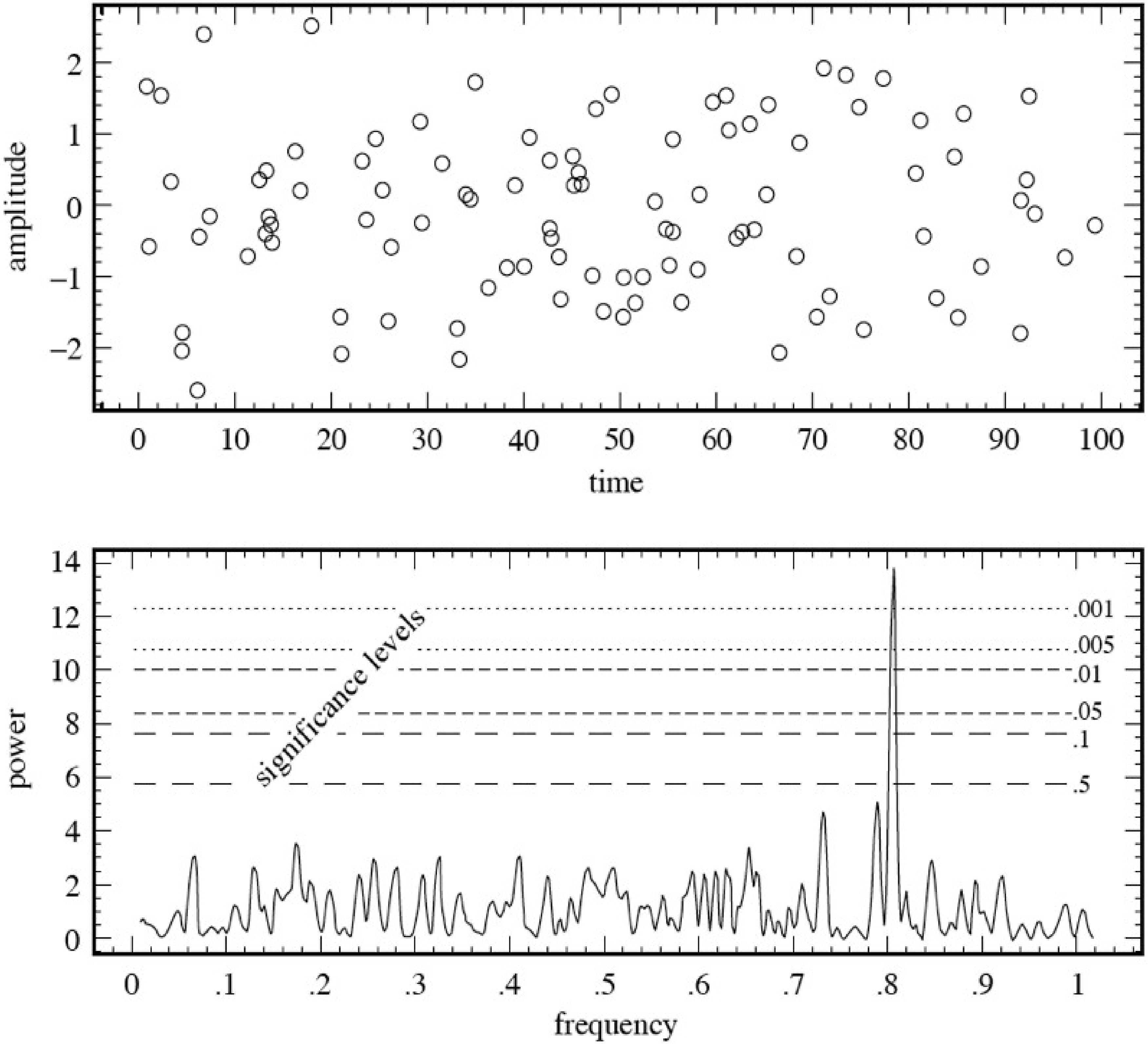
to the data

$$P_N(\omega) \equiv \frac{1}{2\sigma^2}\left\{ \frac{\left[\sum_j (h_j - \bar{h})\cos\omega(t_j - \tau)\right]^2}{\sum_j \cos^2\omega(t_j - \tau)} + \frac{\left[\sum_j (h_j - \bar{h})\sin\omega(t_j - \tau)\right]^2}{\sum_j \sin^2\omega(t_j - \tau)} \right\}$$

spectral power as a function of frequency ω

τ constant= $\tan(2\omega\tau) = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}$

Period finding I (continued)
Normalised Lomb-Scargle periodograms



100 DATA POINTS

Num Res page 571: freq > Nyquist freq if points were evenly distributed!

Period Finding II

Phase-Dispersion Minimisation: PDM

Fold data given a trial period in M bins

Calculate the variance in each bin

Large variance

not the right period

→

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$ variance in the data

$$s_k^2 = \frac{1}{N-1}\sum_{j=1}^{n_k}(x_j - \bar{x})^2$$ variance in one sample

Stellingwerf 1978, ApJ, 224, 953

PDM (continued)

$$s^2 = \frac{\sum_{k=1}^{M}(n_k - 1)s_k^2}{\sum_{k=1}^{M} n_k - M}$$ variance in the samples

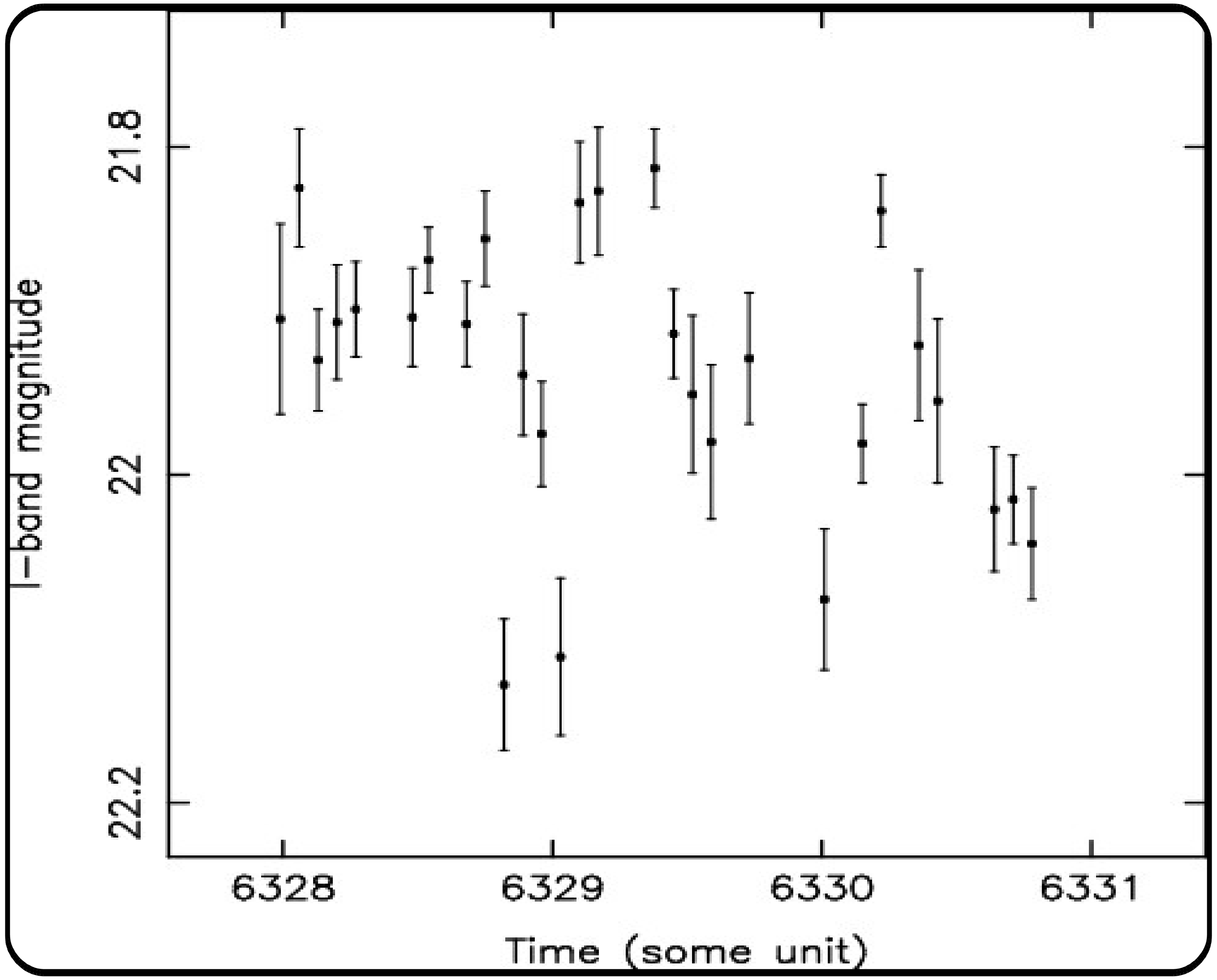
$$\theta = \frac{\sigma^2}{s^2}$$ wrong period Θ≈1

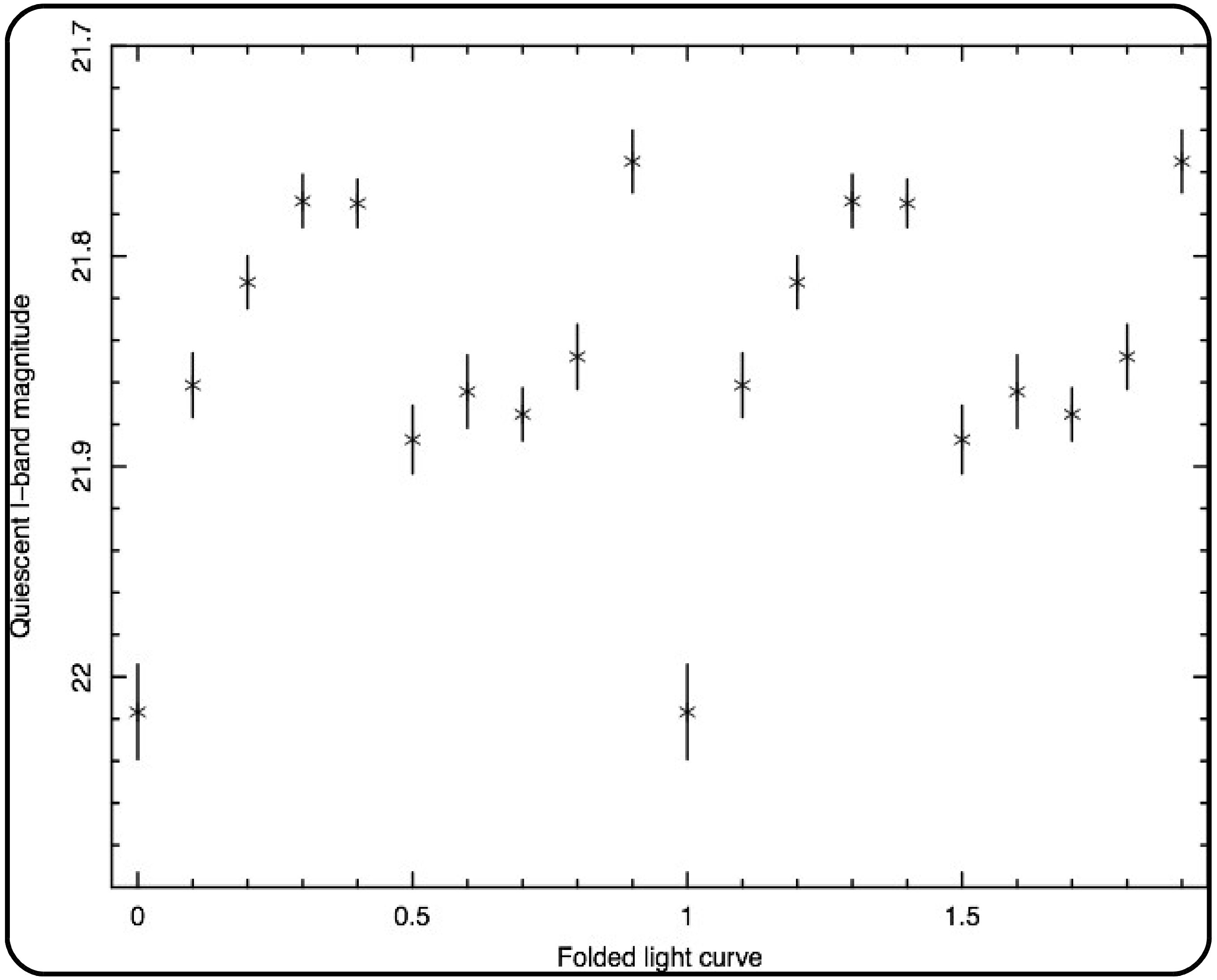
variance in the samples=variance in the data

“right” period Θ≪1

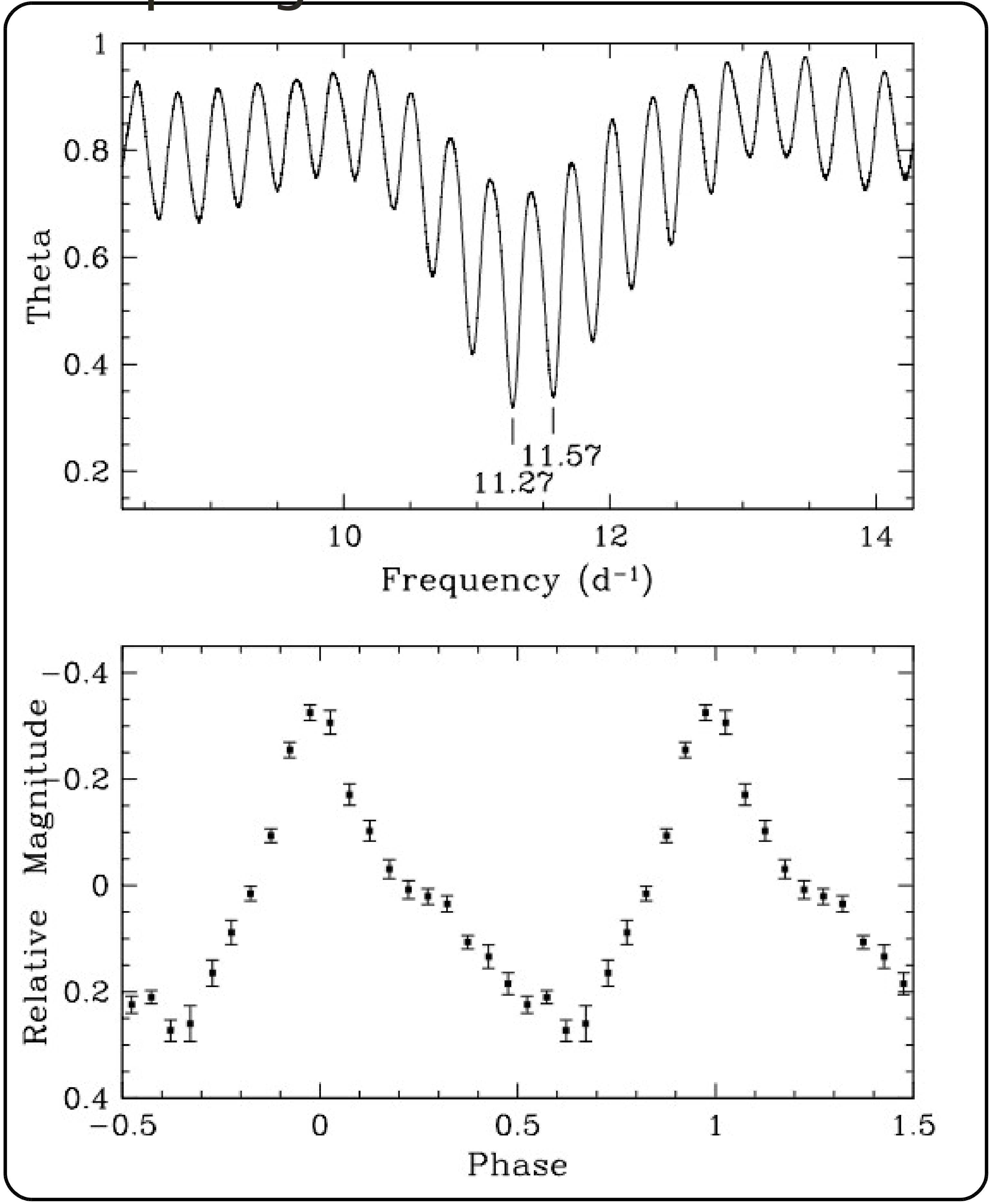
Scramble data in a Monte Carlo simulation to calculate significances

Light curve to scramble



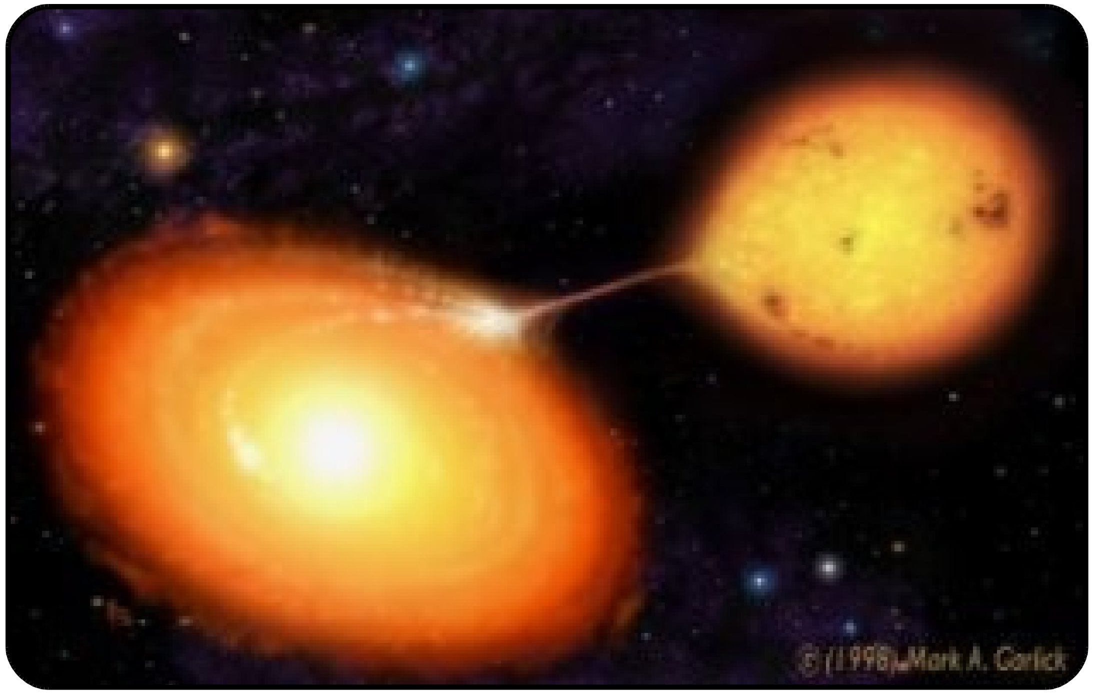


eclipsing SU UMa star: DV Ursae Majoris



Nogami et al. 2001

example use of PDM



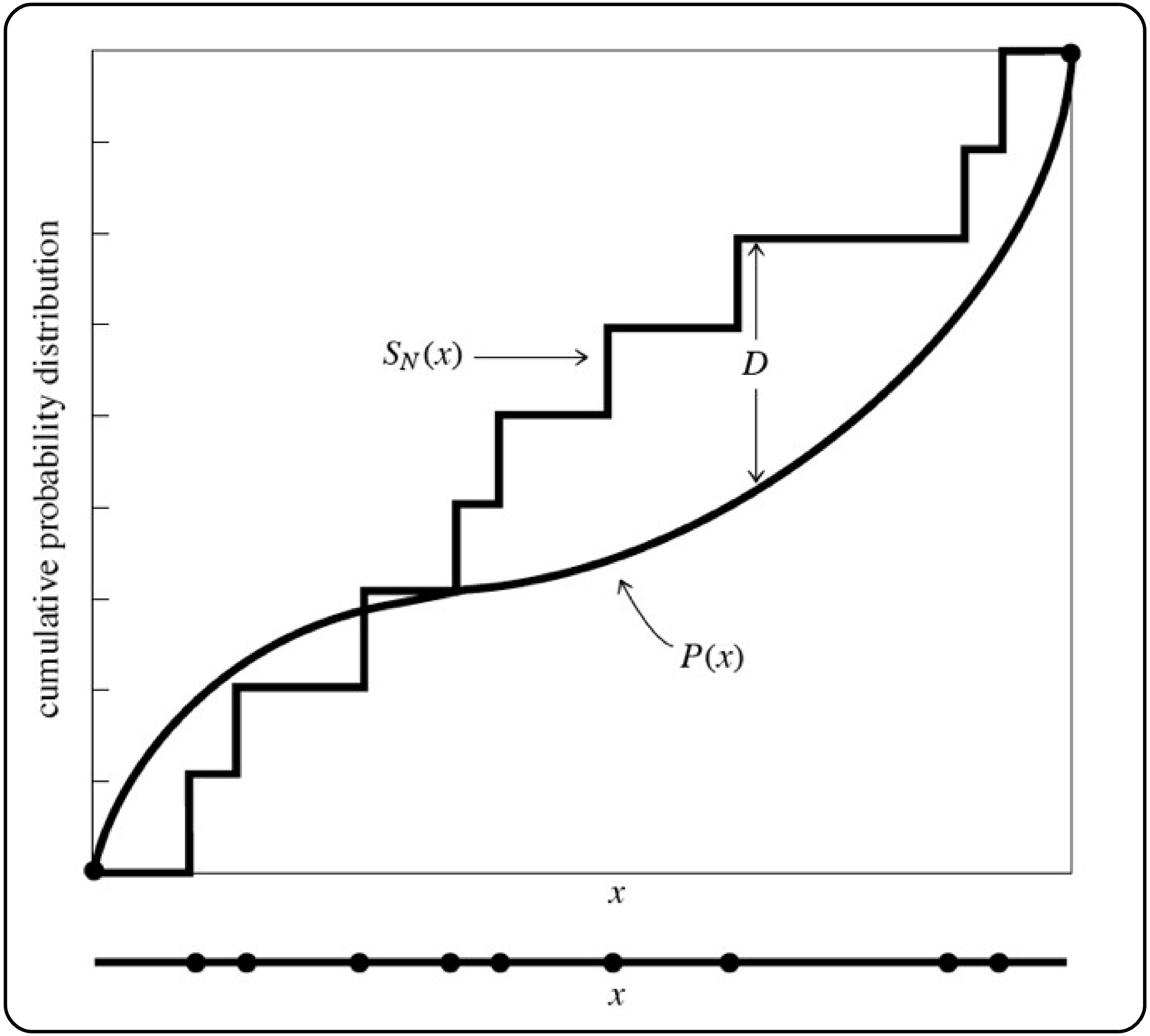
SU UMa artist impression

Data can be variable but not periodic

Example: compare cumulative distribution fie with model of a constant

Comparing a distribution with a theoretical distri or two distributions

Kolmogorov-Smirnov test:



compare two cumulative distribution functions

e.g. 1 observed and 1 theoretical

or

e.g. 2 observed

K-S test

an advantage of using
K-S statistic

the distribution can be calculated in
the case of the null-hypothesis
(data sets from same distri/data
drawn from theoretical curve)

$$Q_{KS}(x) = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2x^2}$$

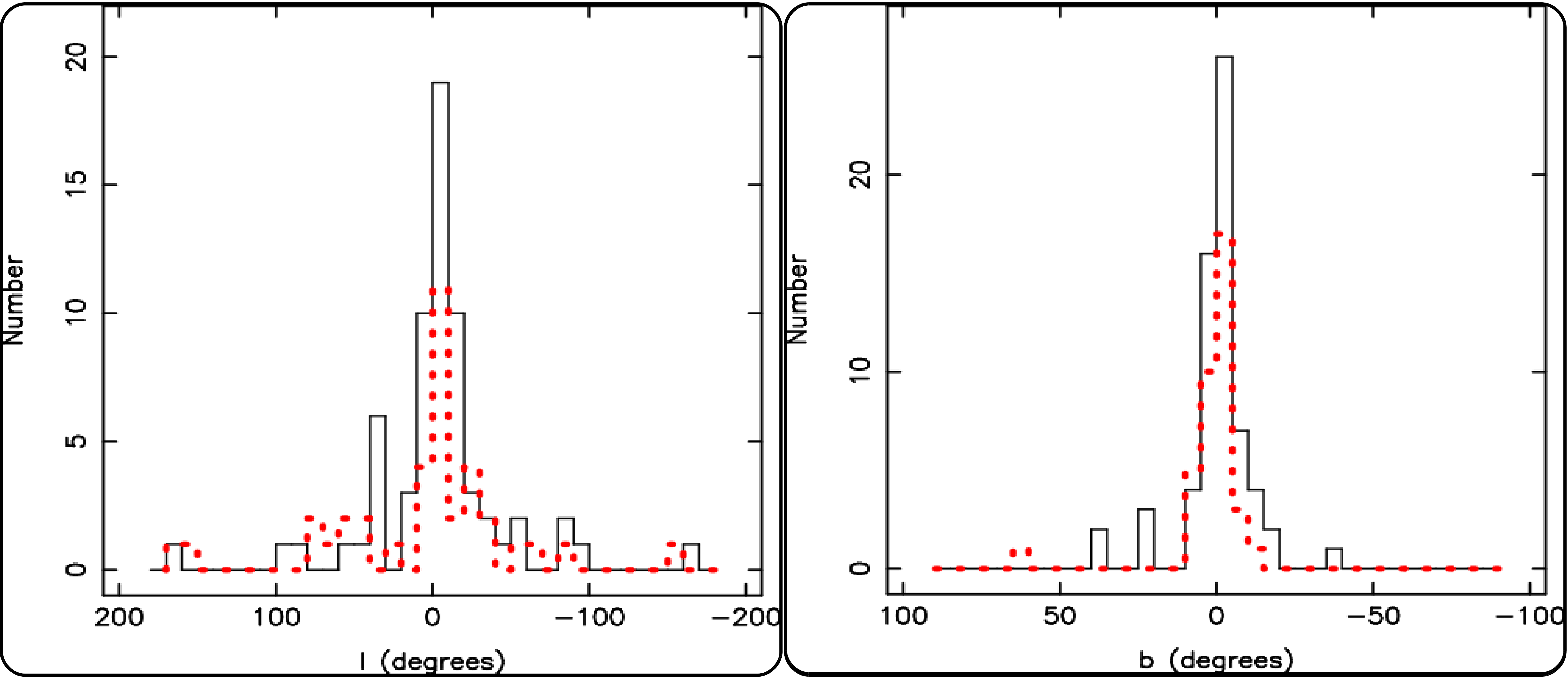
$$\text{Probability } (D > D_{obs}) = Q_{KS}([\sqrt{N_e} + 0.12 + \frac{0.11}{\sqrt{N_e}}D])$$

with $N_e = N$ number of data pnts or $N_e = \frac{N_1 N_2}{N_1 + N_2}$

1 distribution 2 distributions

Example K-S test

distribution of neutron stars and black hole X-ray binaries in our Galaxy



Jonker & Nelemans 2004

Probability that BHs and NSs from the same distribution

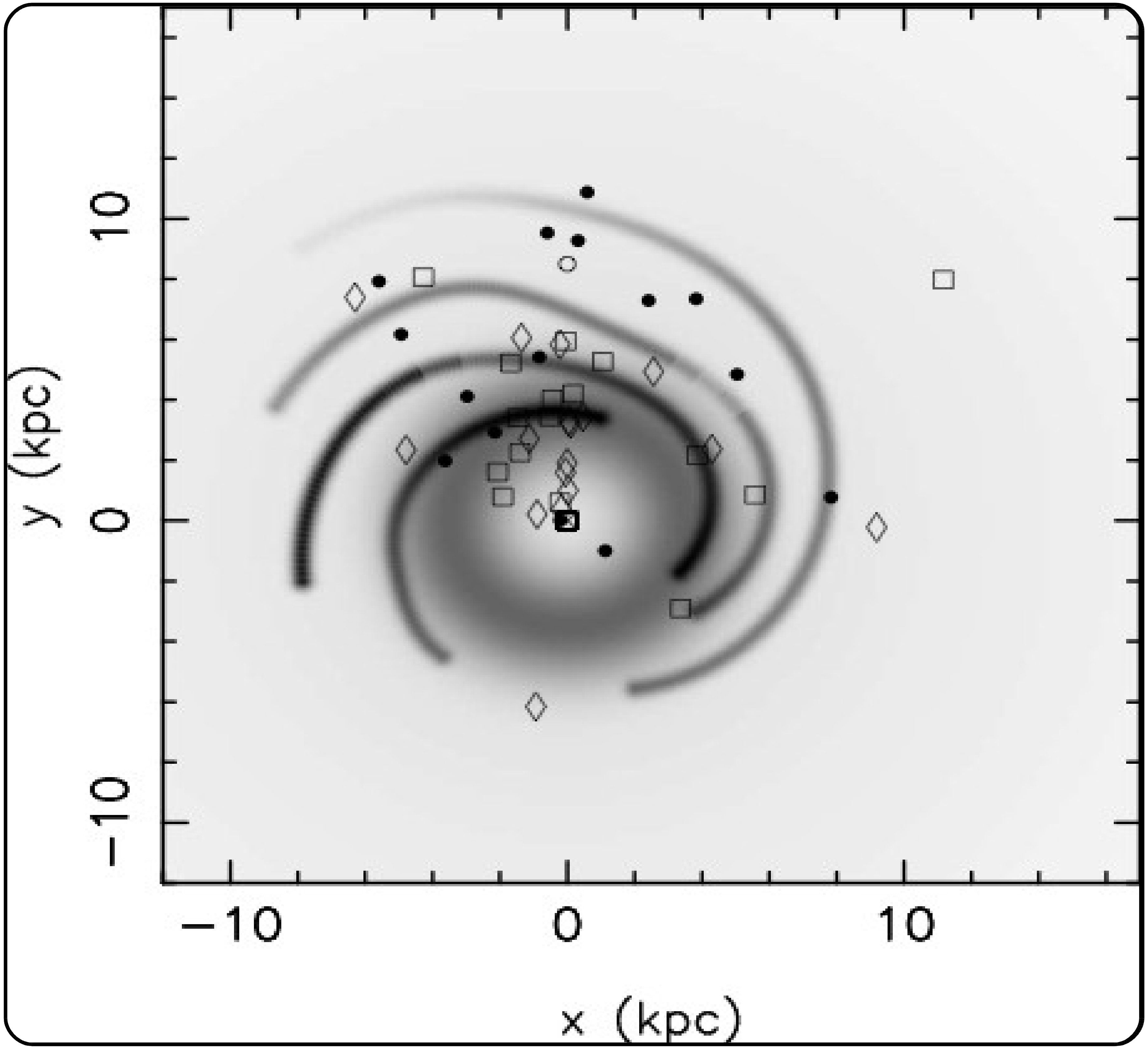
37%, D=0.19

90%, D=0.12

M101 HST image

2D K-S test

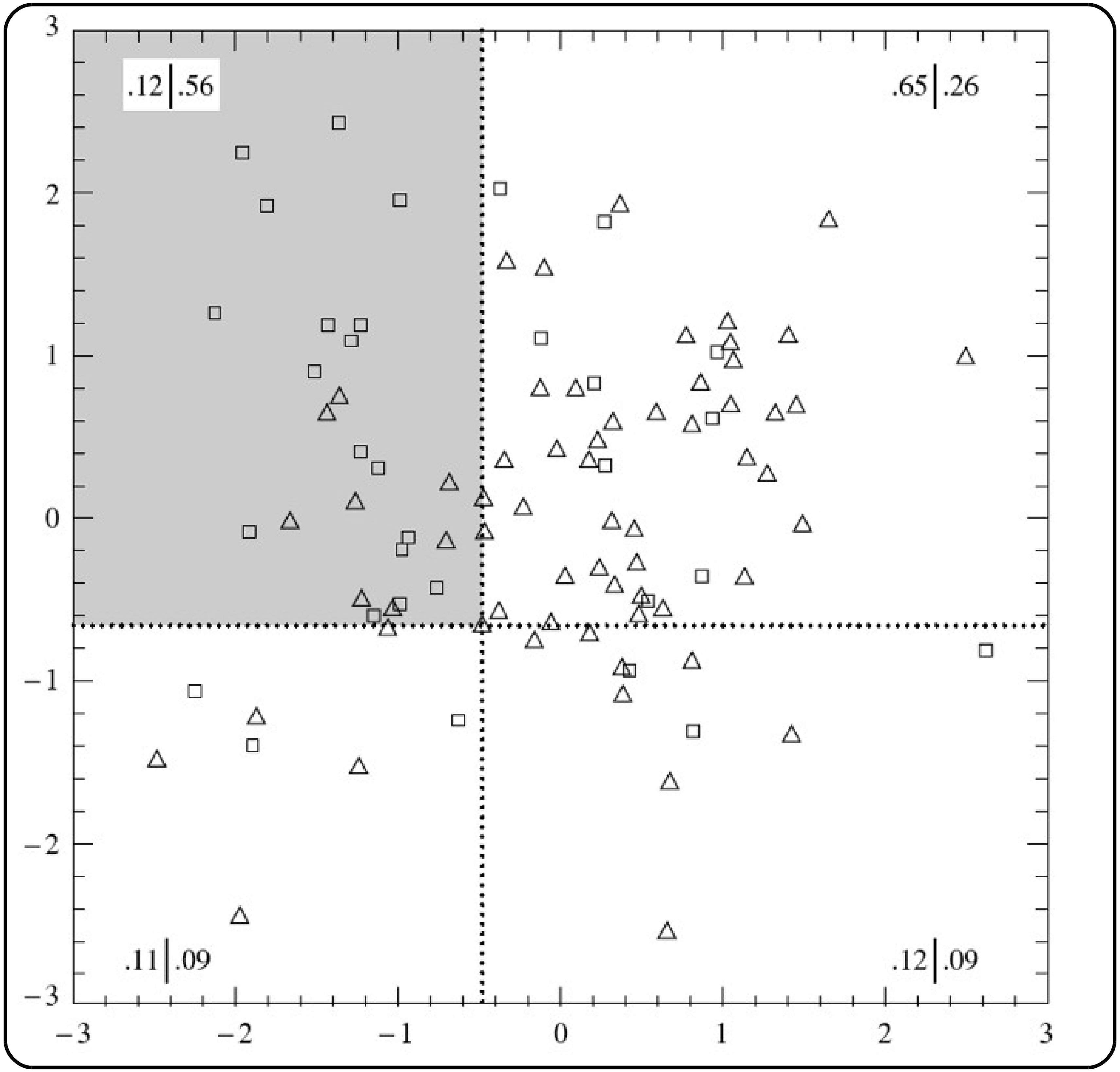
distribution of neutron stars and black hole X-ray binaries in our Galaxy



5.2%
D=0.45

Jonker & Nelemans 2004 Spiral structure Taylor & Cordess 1993

2D K-S test



$$P(D > D_{obs}) = Q_{KS}\left(\frac{\sqrt{N}D}{1+\sqrt{1-r^2(0.25-0.75/\sqrt{N})}}\right)$$

2D K-S test

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})}\sqrt{\sum_i (y_i - \bar{y})}}$$

r=correlation coefficient