Some issues and things

Vera Rubin, 1928 -

- failed to get into Princeton (no women in astro < 1975)

- enrolled at Cornell, studied with Philip Morrison, Richard Feynman and Hans Bethe

- Carnegie Institute of Washington 1965 >

- landmark papers on rotation curves of galaxies showing that they are dominated by dark matter

- many honours. Still publishing.

"In a spiral galaxy, the ratio of dark-to-light matter is about a factor of ten. That's probably a good number for the ratio of our ignorance-to-knowledge. We're out of kindergarten, but only in about third grade."

Where were we?

We considered stats of large-scale structure:

(1) sky projections, Aitoff, Sansom-Flamsteed etc plus contouring

(2) Measures of sky distribution:

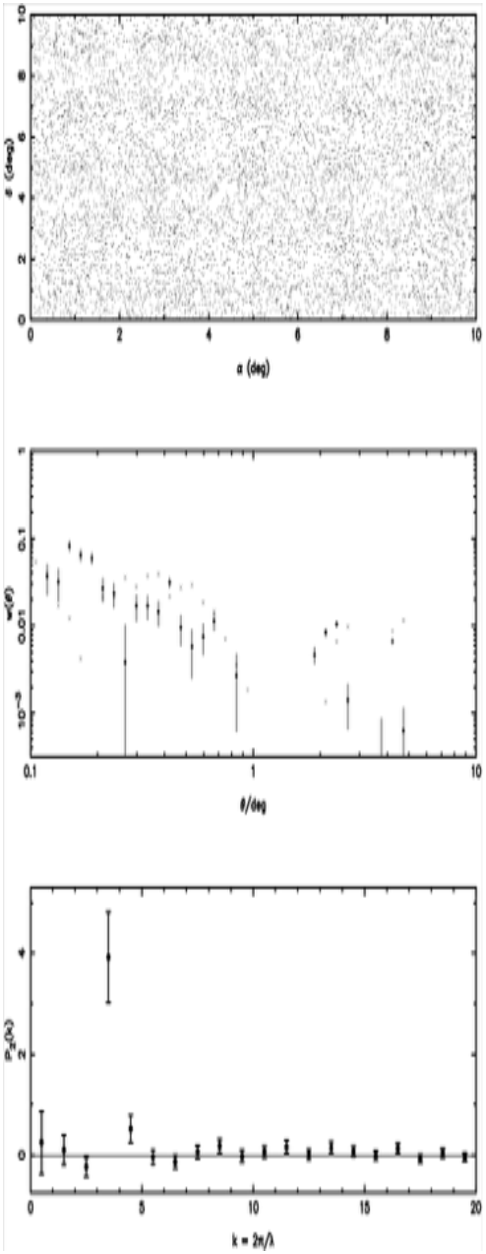
Two-point angular correlation function $w(\theta)$

Counts-in-cells (*C-in-C*)

Angular power spectrum C_l

and examples of how these are related.

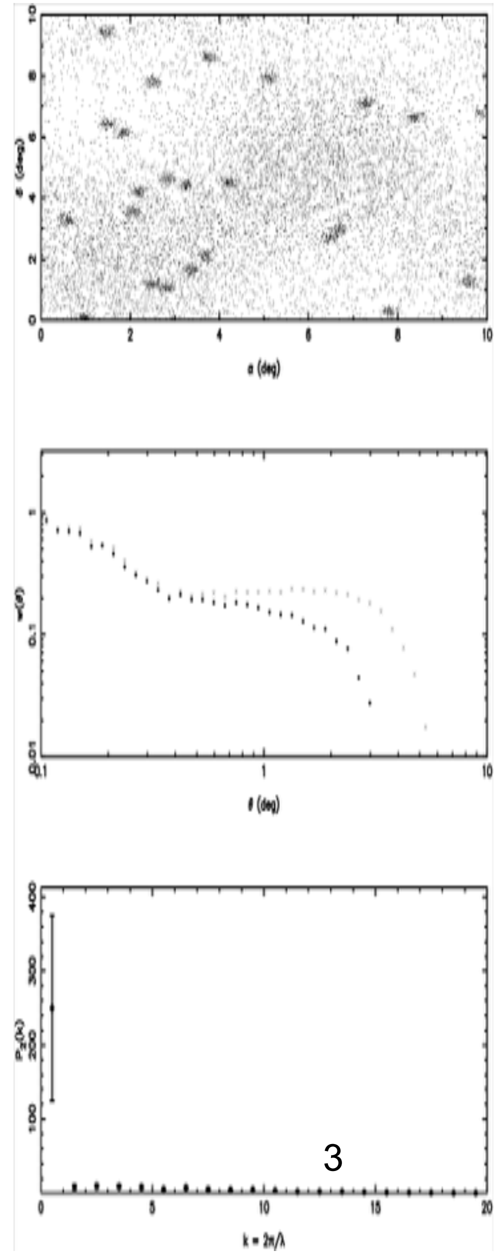
$w(\theta)$ - Example 1 - note angular power spectrum



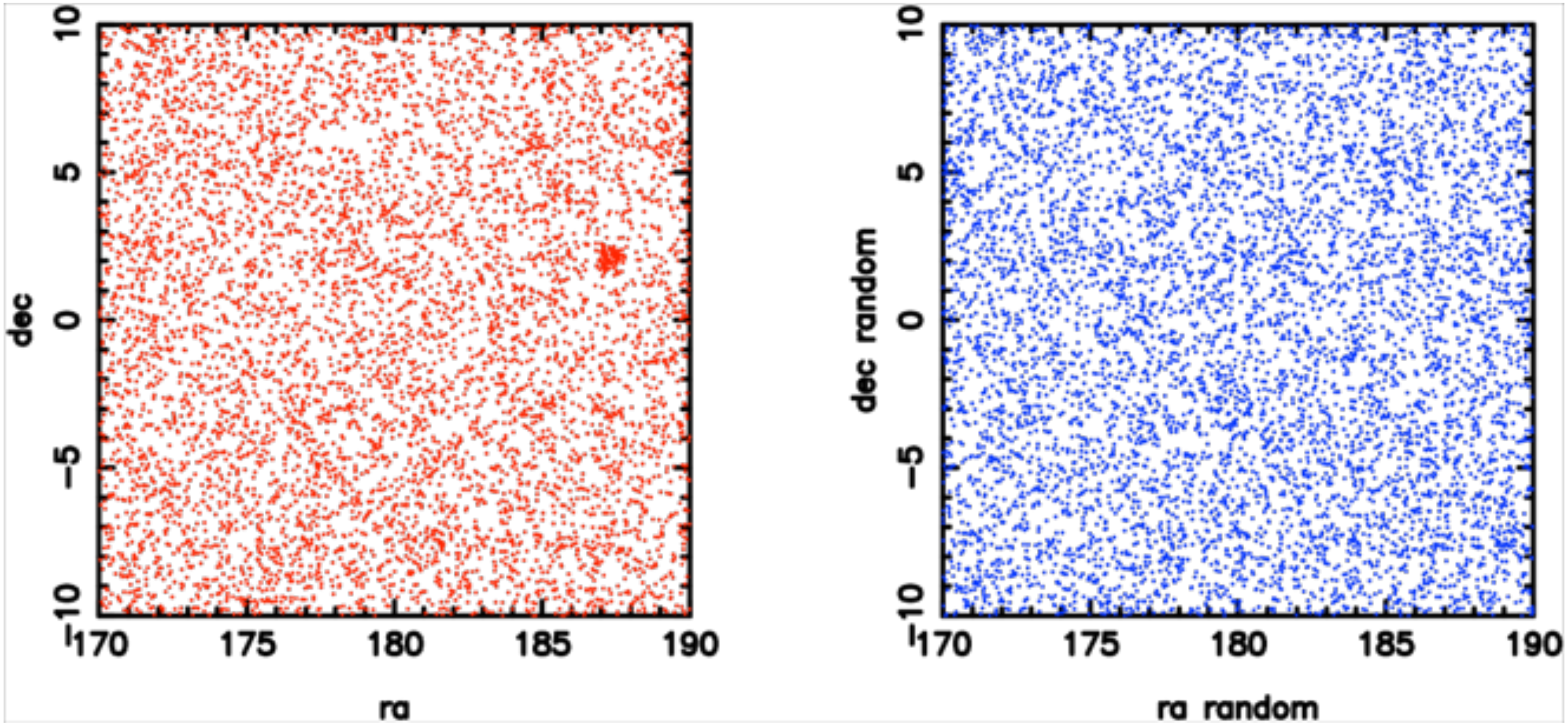
Left -- uniform 2° grid with 25 low-surface-brightness clusters (total 1500 points in Gaussians of width 0.4°) on a random background of 8500 points,

Right -- dipole background 10000 objects background, 2000 in 25 equal clusters of Gaussian width 0.1°, random positions.

Measured **$w(\theta)$** and angular power spectra below each. ***$w(\theta)$*** evaluated with simple estimator (crosses) and the Landy-Szalay estimator (dots with error bars).



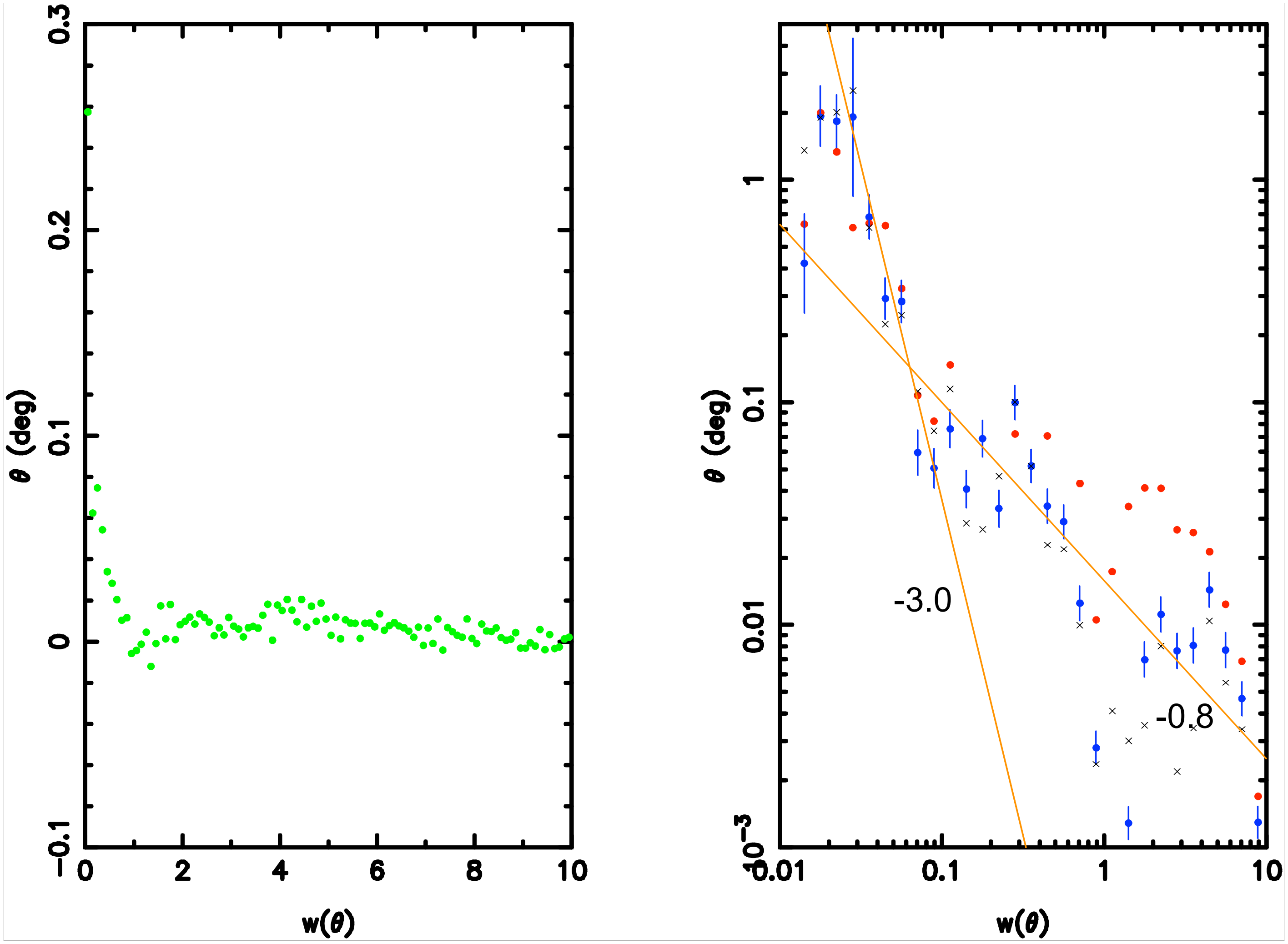
Three More Points on w(θ)



1. What about these skies? Anything strike us?

- Not much signal!

Final points, $w(\theta)$ - plotting and 3D



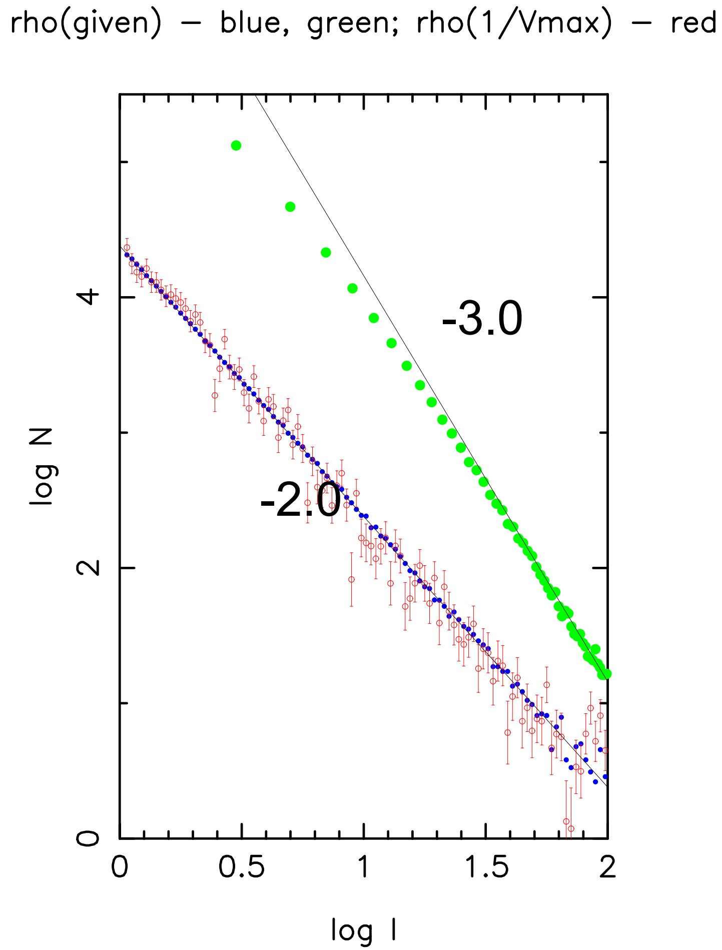
2. Plots and plotting form! – examine your data as fully as possible.

3. Limber's Equation => 3D.

While we're on plotting - the evil power law!

Luminosity function:

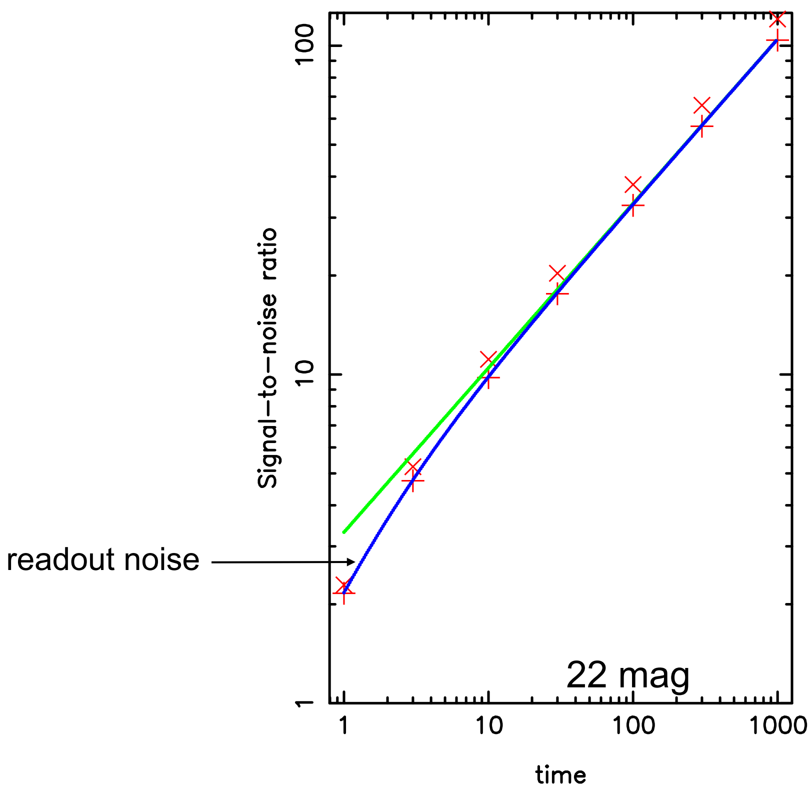
$d\rho/dL = L^{-3}$

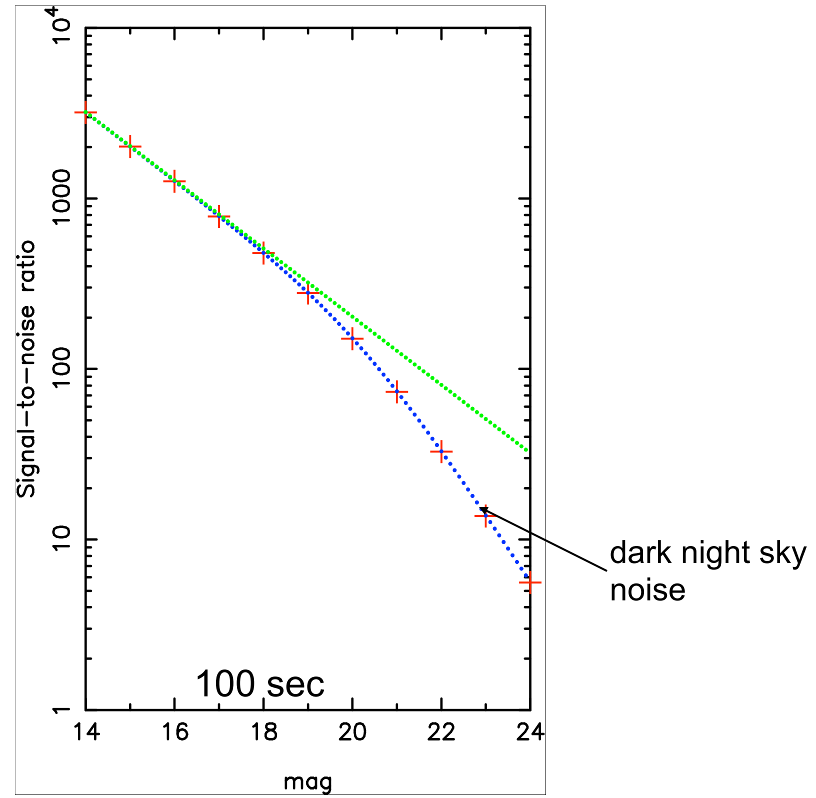


1. NB d(log L)/dL = (1/L)dL, so that dL = L · d(log L)

2. Plotting mid-point of **Δ**L on a log scale is a poor representation.
   To check: make **Δ**L smaller or be cleverer about abscissa position.

Homework answer: you all got it wrong
SIGNAL - s/n for WHT camera (hw4)



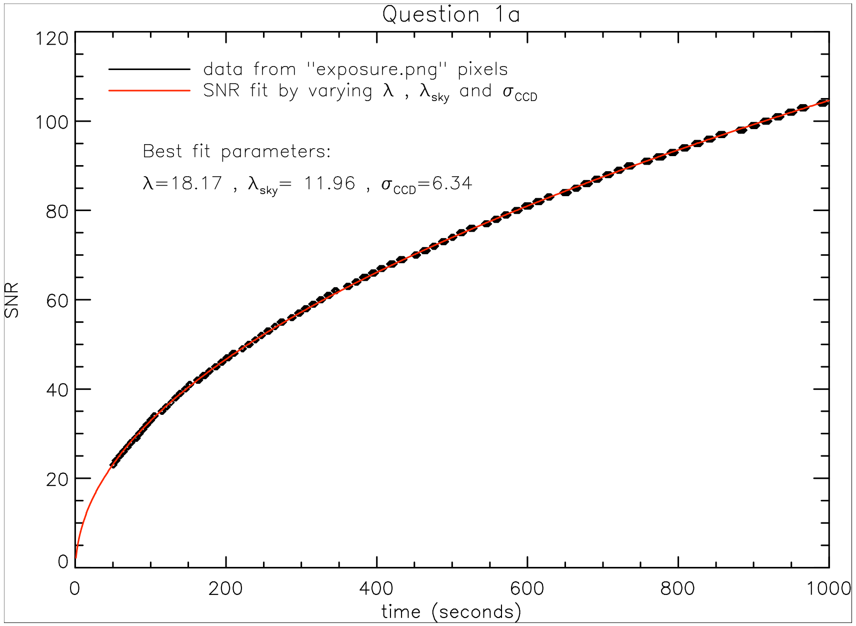


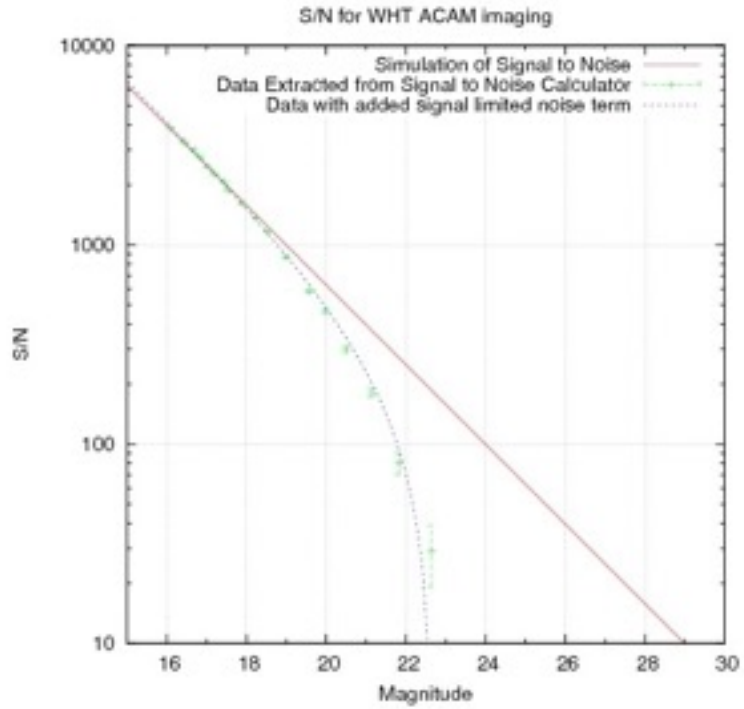
noise = √ {flux from object + 50 x (flux from a night-sky pixel) +50 x (readout noise)^2}

signal = flux from object

The right parameters are flux/sec from a 22.0 mag object = 65 photons, flux/sec from a dark-night-sky pixel = 6.4 photons, readout noise per pixel 3.2 photons. 50 happens to be the number of pixels involved.

SIGNAL - star answers





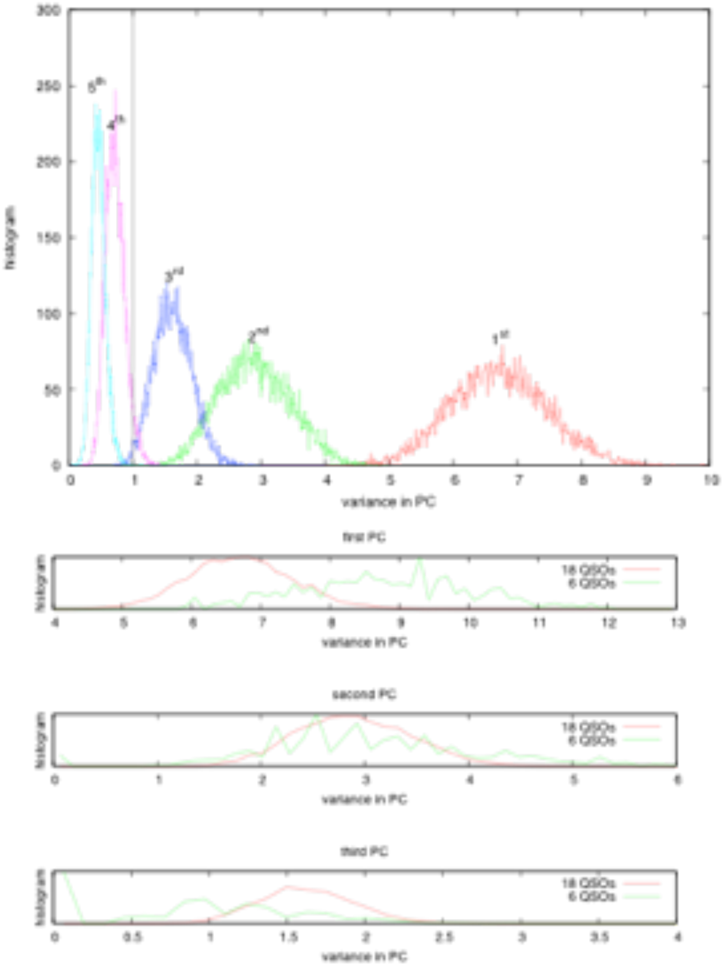
****

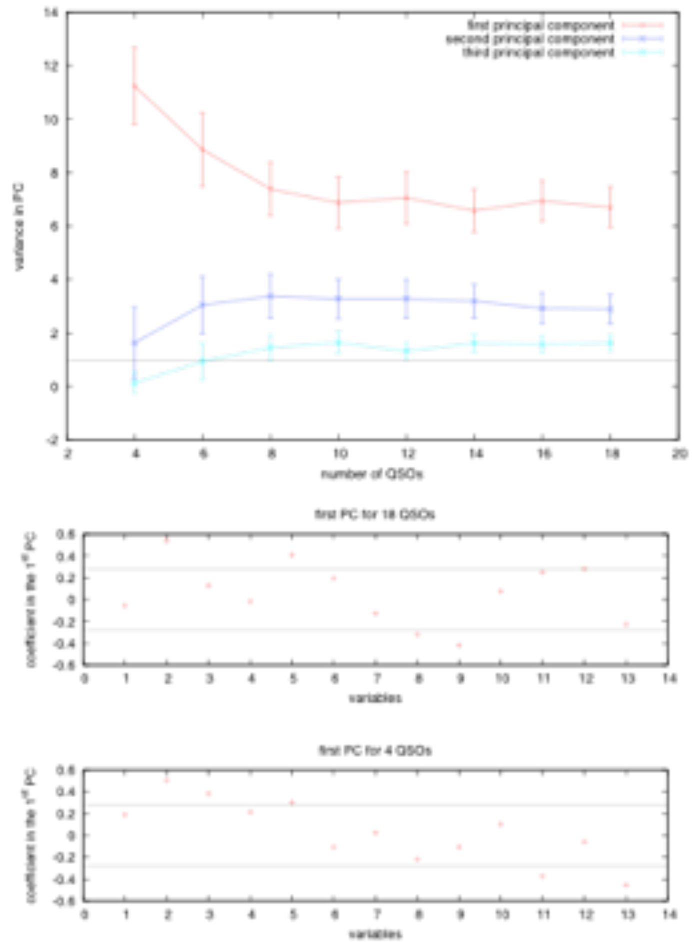
****

noise = √ {flux from object + 50 x (flux from a night-sky pixel) +50 x (readout noise)^2}

signal = flux from object

Principal Component Analysis - hw4-5, model answr

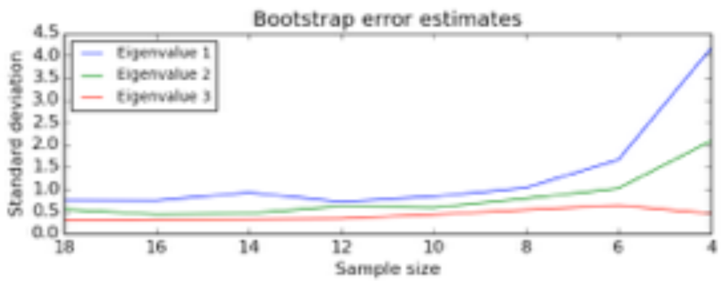


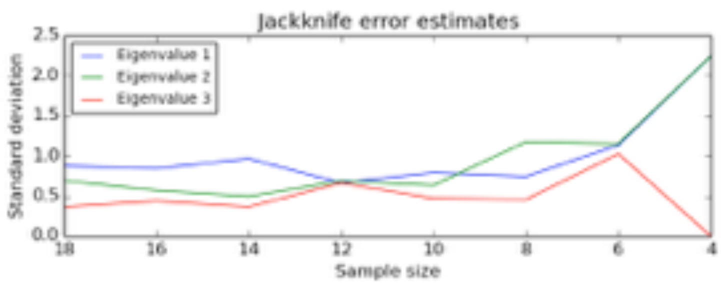


***'s diagrams

Principal Component Analysis - hw4-5

- check variance
- note that sum of eigenvalues has to add up to 13... so if some shrink, others must grow.
- consider the case of a sample of 2!
- TT's answer:





- Checking that there are significant components is only a piece of the puzzle; checking that they correspond to certain combinations of parameters gives you a complete picture and lets you fully discuss when the original results are lost.

Runs Test - hw5

1. You were asked to show if the team has runs of wins or losses which make the number of runs differ from random. This implies that we're looking for **less runs than random, ie a one-tail test**. The answer is that there's no evidence of such a trend; but that's as far as stats will take you.

(The TA in 2011 ran the test for ALL NHL teams and found - a null result. There's no such thing as strings of confidence / lost confidence for NHL teams! There are of course strings of wins/losses, as you would expect from best/worst teams.)

2. **50:50 (a 50% record) has no bearing**; test works for a good team and medium or bad. Number of runs is maximized for **m=n**; in choosing a 50;50 team I only wanted to avoid the situation of an overwhelming number of wins, like 78 out of 82 games, leading to breakdown of any Gaussian approximation.

The Anderson-Darling Test vs the K-S test

(1) For testing if the difference between distributions is significant.

(2) Reputedly better than the K-S Test in most to all circumstances,

(3) Significantly more sensitive to what is happening in the tails of distributions.

(4) Like K-S, good for very small samples.

Data $\{Y_1 < \ldots . < Y_n\}$, put in order. f is a function, F is its integral (Cumulative Distribution Function or CDF). We are testing if Y could be drawn from f.

$$A^2 = -n - S\,, \quad S = \sum_{k=1}^{n} \frac{2k-1}{n} \left[\ln(F(Y_k)) + \ln\left(1 - F(Y_{n+1-k})\right)\right].$$

Compare A against critical values.

See http://src.alionscience.com/pdf/A_DTest.pdf

http://cran.r-project.org/web/packages/nortest/nortest.pdf

http://en.wikipedia.org/wiki/Anderson%E2%80%93Darling_test

HW12 - wrap-up

- MCMC for everyone but for J.. and H.. — something different for them!
- History of astrostats table
- Additional homework - (1) complete TA assessment, and (2) course assessment
- An additional 1 credit Directed Studies option? Any interest?

END