Correlation Functions

Week 8+

Introduction

- An extension to the idea of density estimation and cluster identification is to characterize how far (and on what scales) the distribution of points differs from a random distribution.

For this purpose correlation functions and auto-correlation functions have been introduced.

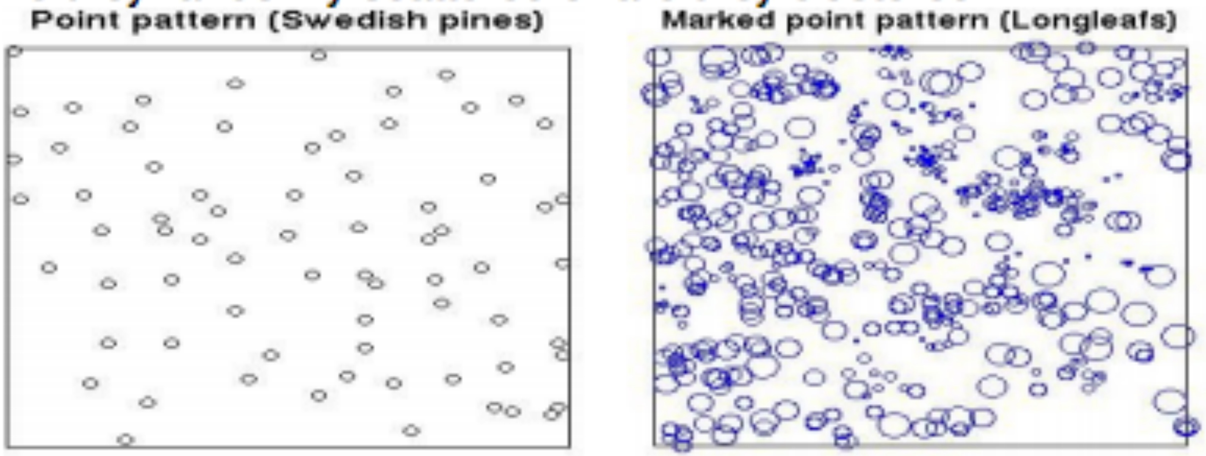
Main goal in astrophysics is to characterize the fluctuations in the densities of galaxies and quasars as a function of luminosity, galaxy type and age of universe. Test it with cosmological models

Spatial Point Process Data examples

Locations of pine saplings in a Swedish forest.

Location, diameter of longleaf pines (*marked* point process).

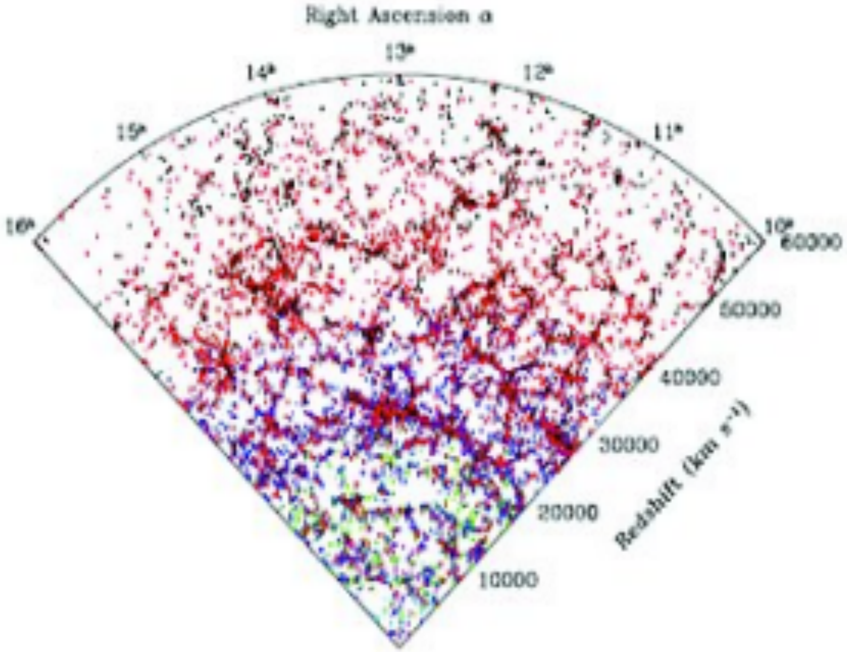
Are they randomly scattered or are they clustered?



(from Baddeley and Turner R package, 2006)

The galaxy distribution: 3D spatial point process

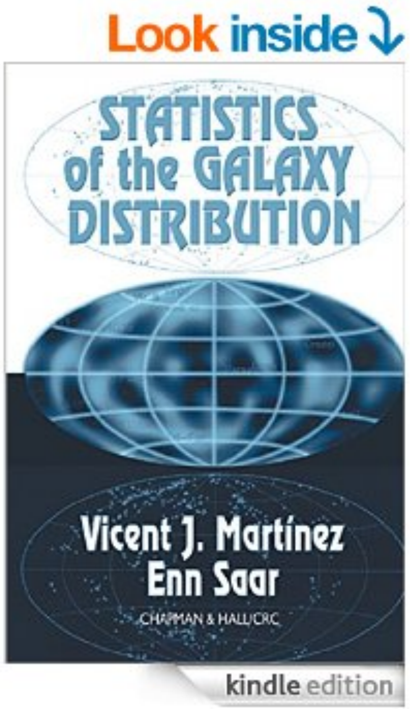
2d location in sky, 1d from redshift as a surrogate for distance.

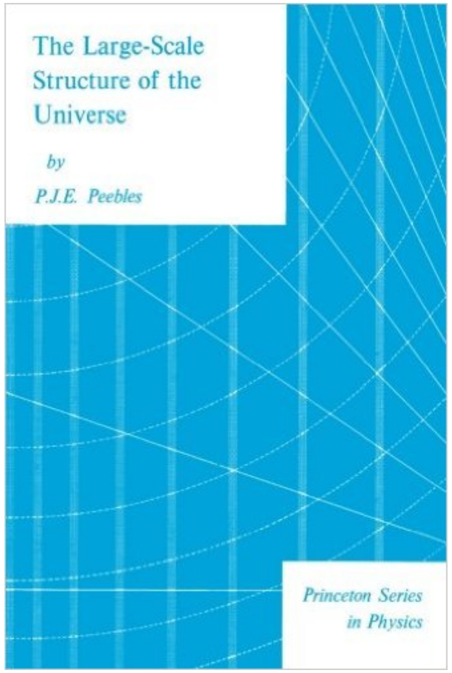


(Tegmark et al. 2004, The three-dimensional power spectrum of galaxies from the Sloan Digital Sky Survey,

Astrophys. J. 606, 702-740)

References





astro-ph/0503603 Power Spectrum Estimation I : The Basics
astro-ph/0406086 Scaling Laws in Distribution of Galaxies (*)
arXiv:0712.3928 by L. Verde
Dragan Huterer lecture notes at Michigan LSS school (2020)

Definition of Two-Point Correlation Function

- Excess probability (compared to a Poisson distribution) of finding a pair of points in two volume elements dV_1 and dV_2 separated by a distance r is given by :

$$dP_{12} = \rho^2 dV_1 dV_2 (1 + \xi(r_{12}))$$

ξ (r) is called *two-point correlation function*

It describes the excess probability of finding a pair of points, as a function of separation compared to a random distribution.

Positive, negative, or zero amplitudes in ξ (r) correspond to distributions which are correlated, anti-correlated or random.

For a homogenous and isotropic universe ξ (r) depends only on lrl and is equal to ξ (lrl).

Conditional probability that a galaxy within volume element dV at a distance r given that

there is a galaxy at the origin of r is given by:

$$dP = n[1 + \xi(r)]dV$$

It measures the clustering in excess or in defect compared with a random Poisson point distribution for which ξ (r) = 0

In statistical mechanics, correlation function used is g(r)= 1+ ξ (r) (called radial Distribution function) and also called pair correlation function in statistics

Integral of the Two-point correlation function is called Ripley K-Function

Relation to Power Spectrum

- Two-point correlation function is the Fourier Transform of the power spectrum.

$$P(k) = \frac{4\pi}{k} \int_0^\infty \sin(kr) \xi(r) r dr$$

$$\xi(r) = \frac{1}{2\pi^2} \int dk \ k^2 P(k) \frac{\sin(kr)}{kr}$$

where k is the wavenumber and is related to scale of wavelength of fluctuation given by $\lambda = 2\pi / k$

Correlation function can be used to define density fluctuation of sources given by:

$$\xi(r) = \langle \frac{\delta\rho(x)}{\bar{\rho}} \frac{\delta\rho(x+r)}{\bar{\rho}} \rangle$$

where $$\frac{\delta\rho(x)}{\bar{\rho}} = \frac{\rho - \bar{\rho}}{\bar{\rho}}$$

is the density contrast , relative to the mean value $\bar{\rho}$ at position x

- In studies of galaxy distributions ξ (r) is parameterized in terms of a power law

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$$

where r_o is the clustering length scale and ɣ the power-law exponent.

For local universe, r_o = 6 Mpc and ɣ ~ 1.38

For cluster r_0 ~ 20 h^{-1} Mpc

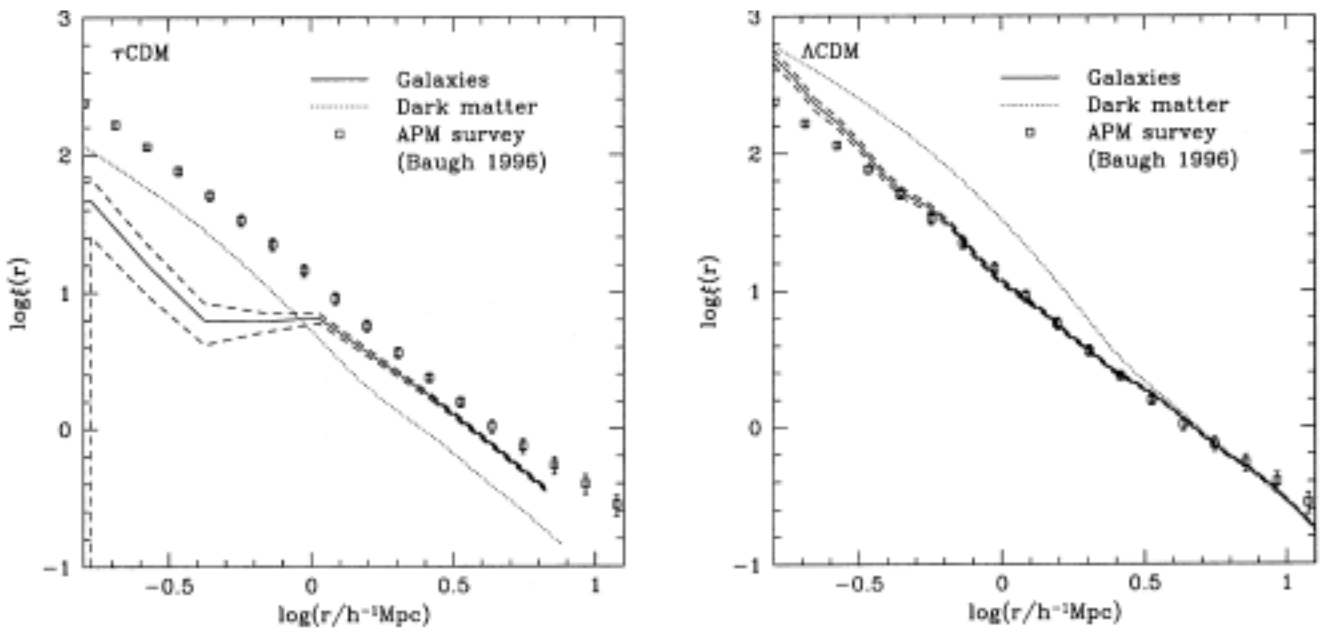


Figure 5. The left hand panel shows the two-point correlation function of galaxies brighter than $M_B - 5\log h = -19.5$ in a τCDM model as a solid line. The dashed lines to either side indicate the Poisson sampling errors. This is compared to the observed APM real-space correlation function (points with error bars) and to the mass correlation functions in the N-body simulations (dotted line). The right hand panel shows the equivalent plot for a ΛCDM model.

astro-ph/9903343 Benson et al

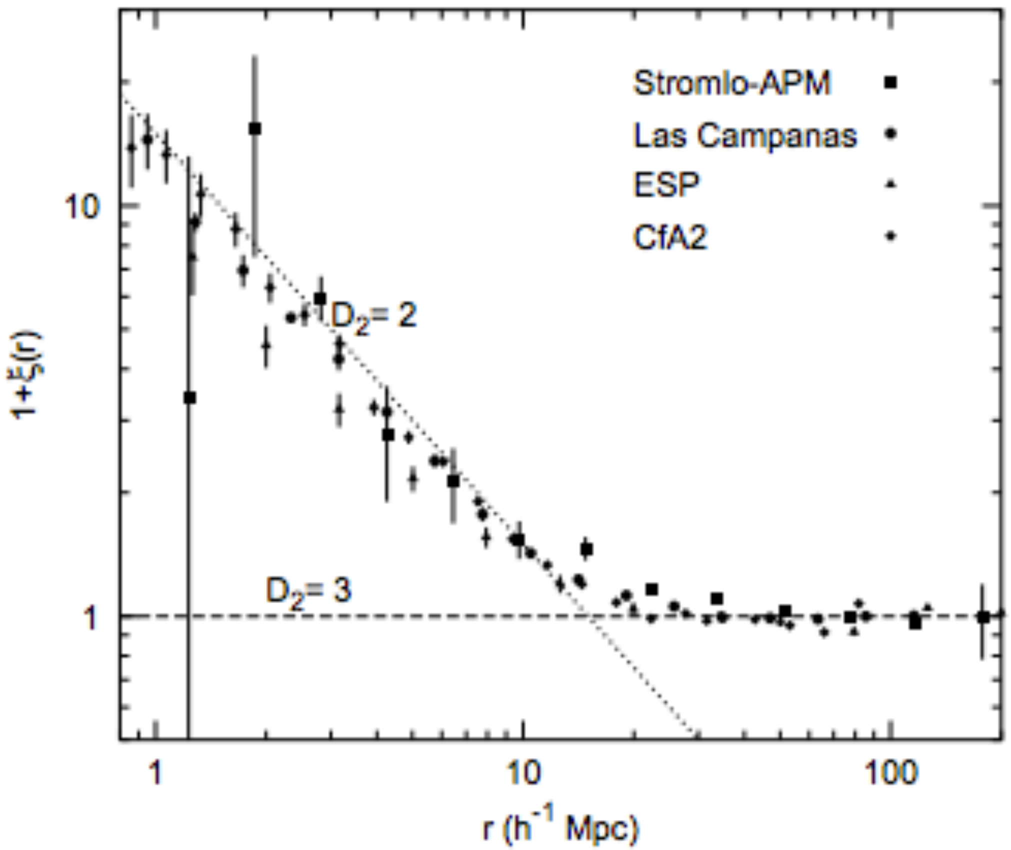


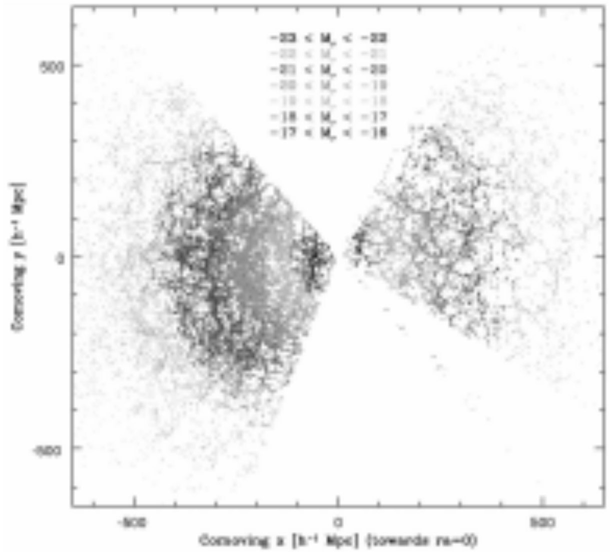
FIG. 10 The correlation function $1 + \xi(r)$ for different samples calculated with different estimators. We can see that the small scale fractal regime is followed by a gradual transition to homogeneity.

$$1 + \xi(r) \propto r^{D_2 - 3}$$

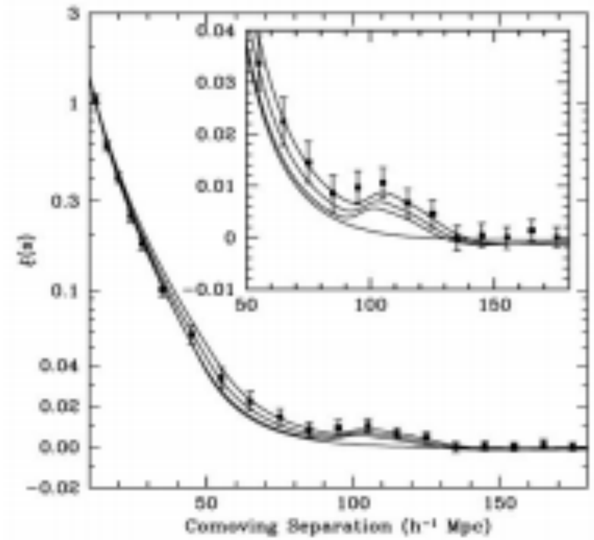
astro-ph/0406086

Example: Galaxy clustering (Sloan Digital Sky Survey)

Galaxy distribution



Two-point correlation function



Distribution of 67,676 galaxies in two slices of the sky showing strong anisotropic clustering (Tegmark et al. 2004).

Bottom: Two-point correlation function showing the faint feature around 100 megaparsec scales revealing cosmological Baryonic Acoustic Oscillations (Eisenstein et al. 2005).

Angular Two-Point Correlation Function

- Angular Two-Point Correlation function $w(\theta)$ can be defined as the conditional probability of finding a galaxy within solid angle $d\Omega$ lying at an angular distance θ from a given galaxy is given by :

$$dP = N[1 + w(\theta)]d\Omega$$

where N is the mean number density of galaxies per unit area in projected space

For galaxies

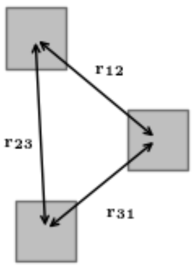
$$w(\theta) = \left(\frac{\theta}{\theta_0}\right)^{\delta}$$

where $\delta = 1 - \gamma$

Three-Point Correlation Function

- 2-point correlation function concept can extended to higher orders by considering configurations of points comprising of triplets, quadruplets, etc

Probability of finding three points in volume elements dV_1, dV_2, dV_3 defined by a triangle with sides r_{12},r_{13}, and r_{23} given by:



(b) 3 point

Three-Point Correlation function

$$dP_{123} = \rho^3 dV_1 dV_2 dV_3 (1 + \xi(r_{12}) + \xi(r_{23}) + \xi(r_{13}) + \zeta(r_{12}, r_{23}, r_{13}))$$

Computing the Two-Point Correlation Function

- 2-point correlation function can be estimated by calculating the excess or deficit of pairs of points within a distance r and r+dr, compared to a random distribution. Random distribution are selected with the same selection function as the data (i.e. within the same volume and within identical masked regions).

$$\hat{\xi}(r) = \frac{DD(r)}{RR(r)} - 1$$

where DD(r) = no of pairs of data points

RR (r) = no of pairs of random points (generated with density higher times than that of the data)

DR (r) = no of data-random pairs

Landy-Szalay estimator

- Edges effects due to interaction between the distribution of sources and irregular survey geometry bias estimates of the correlation function. So therefore this is calculated using Landy-Szalay estimator.

$$\hat{\xi}(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} - 1$$

Landy and Szalay ApJ 412, 64 (1993)

L-S Estimator for Higher-Order Estimators

$$\hat{\xi}(r) = \frac{DDD(r) - 3DDR(r) + DRR(r) - RRR(r)}{RRR(r)}$$

(valid for an equilateral triangle)

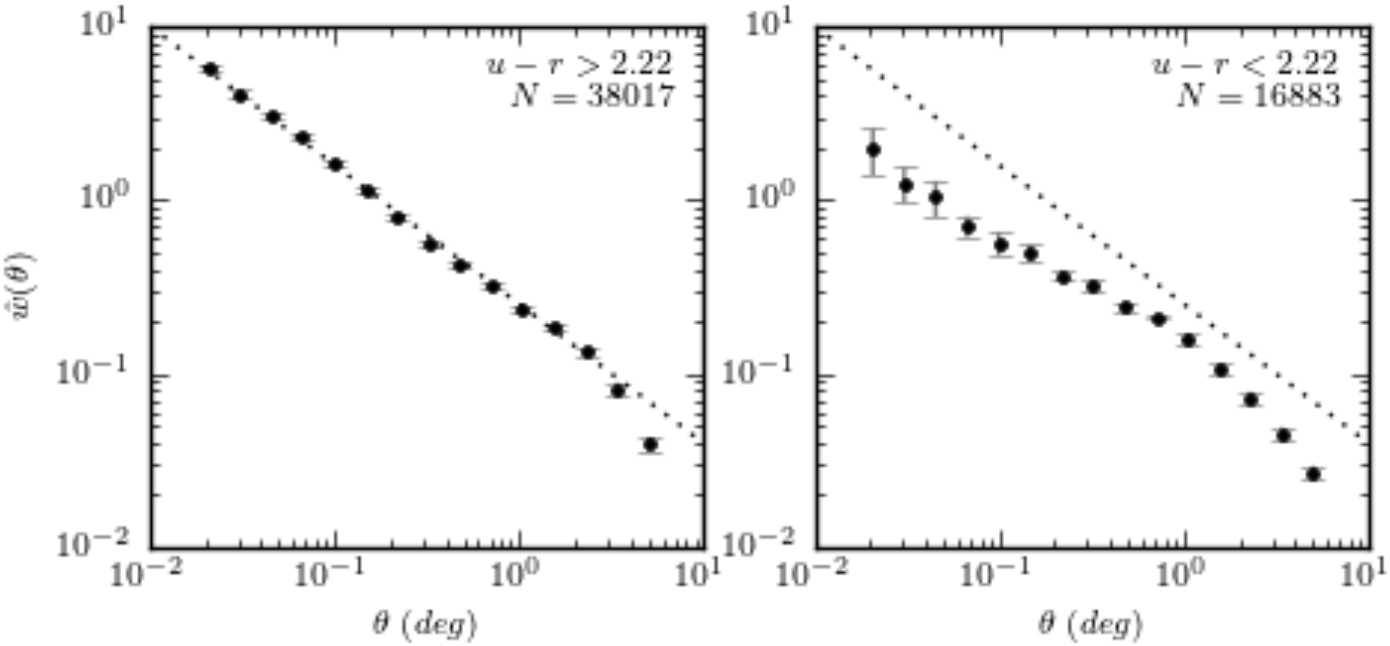
where DDD(r) is the number of data triplets and DDR(r), DRR(r) and RRR(r) are the associated configurations for the data-data-random, data-random-random, and random-random-random triplets.

Python Implementation

```
import  numpy as np
from astroML.correlation import two_point_angular
RA = 40*np.random.random(1000)
DEC = 10*np.random.random(1000)
bins = np.linspace(0.1,10,11) # evaluate in 10 bins with
these edges
corr =  two_point_angular (RA,DEC, bins, method='landy-
szalay')
```

(only other option supported is 'standard')

350 Mpc < d < 8.4 Gpc



The two-point correlation function of SDSS spectroscopic galaxies in the range $0.08 < z < 0.12$, with $m < 17.7$. This is the same sample for which the luminosity function is computed in figure 4.10. Errors are estimated using ten bootstrap samples. Dotted lines are added to guide the eye and correspond to a power law proportional to $\theta^{-0.8}$. Note that the red galaxies (left panel) are clustered more strongly than the blue galaxies (right panel).

AstroML figure 6.14

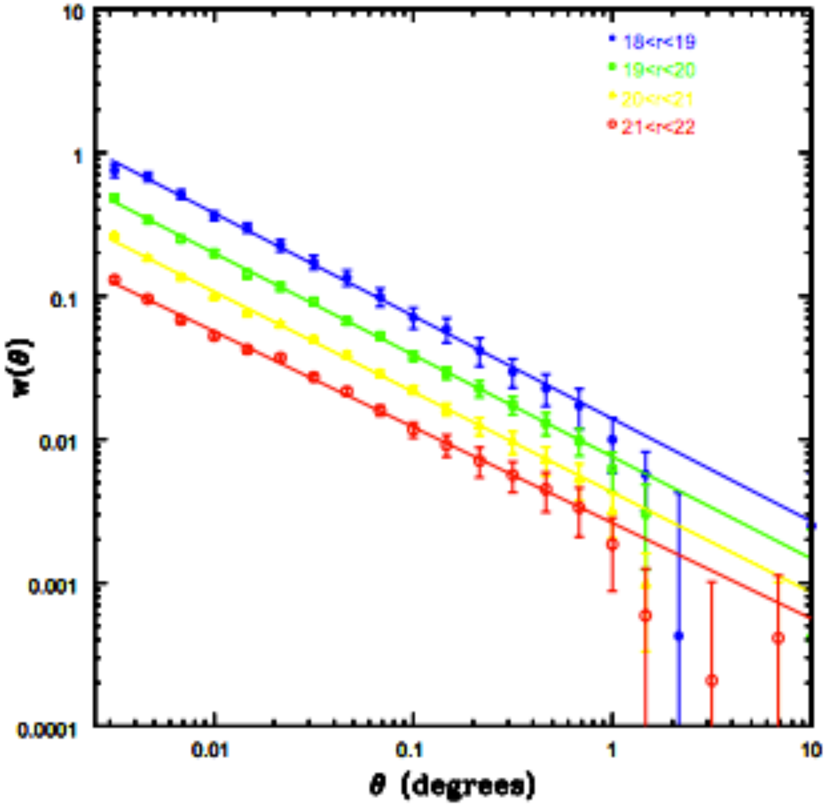


FIG. 9 The angular correlation function from the SDSS as a function of magnitude from Connolly *et al.* (2002). The correlation function is determined for the magnitude intervals $18 < r^* < 19$, $19 < r^* < 20$, $20 < r^* < 21$ and $21 < r^* < 22$. The fits to these data, over angular scales of 1' to 30', are shown by the solid lines.

astro-ph/0406086