Measuring the Two-Point Correlation Function

March 15, 2007

1 the Two-Point Correlation Function

The discrete equivalent of the autocorrelation function $\xi(r)$ is the **two-point correlation function** $\xi_{12}(r)$. If a given point distribution represents a fair sampling of the underlying continuous distribution, the two-point correlation function ξ_{12} should be equal to the autocorrelation function $\xi(r)$.

In cosmology the two-point correlation function $\xi_{12}(r)$ of a homogeneous point process is follows on the basis of the excess probability of finding points at a distance r. For a homogenous Poisson process one knows that if we take two volumes dV_1 and dV_2 at a distance r, the probability dP_{12} (or, rather, number) of points in the two volumes is given by

$$dP_{12} \,=\, \bar{n}^2 \, dV_1 dV_2 \,. \tag{1}$$

For an inhomogeneous point process, i.e. in the case of clustering (due to the existence of underlying density perturbations), there will be an excess with respect to the Poisson distribution. This is encapsulated in the function $\xi_{12}(r)$,

$$\boxed{dP_{12} \,=\, \bar{n}^2 \, \{1 + \xi_{12}(r)\} \, dV_1 dV_2} \tag{2}$$

In other words, the correlation function measures the excess probability. If there is clustering at a distance r, $\xi(r) > 0$. If points are anticorrelated at that distance, ie. tend to avoid each other, then $\xi(r) < 0$. And if there is no clustering at all but a homogeneous distribution we have $\xi(r) = 0$. Note that from now on we simply assume that $\xi_{12}(r) = \xi(r)$. Notice that we assume that because of the **isotropy** of the density fluctuations the two-point correlation function should also be isotropic and only a function of distance r. The significance of the two-point correlation function $\xi(r)$ has

formed the main tool in the study of the large scale galaxy distribution. It has formed the main statistical measure for clustering in the Universe. Every catalogue of galaxy positions, on the sky or in redshift space, has been analyzed to determine the two-point correlation function. The same holds true for catalogs of clusters of galaxies, of active galaxies, etc. There are a variety of reasons for its prominence:

- Clustering of galaxies, clusters of galaxies, radio galaxies, etc. is clearly an important aspect of the cosmic large scale matter distribution. The two-point correlation function is the first order measure for characterizing deviations from a uniform distribution: it forms the first order description of clustering.

- The autocorrelation function is the Fourier transform of the Power Spectrum $P(k)$, and in particular in the linear regime it contains crucial information on the cosmological scenario prevailing in our Universe. Hamilton et al. (1991) even managed to find a relation between the measured nonlinear $\xi(r)$ and the linear power spectrum.

- For highly nonlinear clustering we often find that the two-point correlation function is a power-law function of distance r,

$$\boxed{\xi(r) \;=\; \left(\frac{r}{r_0}\right)^{-\gamma}} \tag{3}$$

The socalled **correlation length** r_0 (the name is a misnomer and often confusing for physicists, who have another definition) is a measure for the amplitude of the clustering process. It is the value of the distance at which $\xi(r) = 1$, and thus the distance at which the clustering strength becomes comparable to the probability of the homogeneous point process. It therefore provides a good measure for the **scale of nonlinearities**: above the correlation scale the point distribution rapidly enters the linear clustering regime.

- The corresponding **power-law slope** γ appears to have a rather universal value of $\gamma \approx 1.8$. In the nonlinear clustering regime γ is closely coupled to the slope n of the power spectrum $P(k)$,

$$n(k) \equiv \frac{d \log P(k)}{d \log k}\,, \tag{4}$$

  and thus contains a wealth of information on the underlying structure formation process.

- The most reliable estimates of the two-point correlation function concern the analysis of (two-dimensional) sky distributions of galaxies. The best galaxy sky catalogues contain millions of galaxies. Statistically this guarantees estimates with small errors. The resulting angular two-point correlation function $\omega(\theta)$ is basically a weighted projection of the spatial two-point correlation function $\xi(r)$ (expressed through the socalled Limber equation). On small scales, the power-law behaviour of the latter thus translates into a power-law angular two-point correlation function,

$$\omega(\theta) = \left(\frac{\theta}{\theta_0}\right)^{1-\gamma} \tag{5}$$

  where γ is the power-law slope of the spatial two-point correlation function. Interesting is the behaviour of the angular correlation scale θ_0. It is very sensitive to the selection of galaxies in the catalogue: it scales with the depth of the sample. The large the apparent magnitude limit m_{lim}, i.e. the deeper we look into the Universe, the smaller θ_0 becomes. This of course is due to the projection of ever more shells on top of each other, as well as in a shift of the angular scale corresponding to a particular physical scale. There is a very precise relation between this angular correlation scale and the depth of the survey on the condition that **we live in a Universe which on the largest scales is homogeneous**. This indeed appears to be true, one of the most convincing arguments for the **Homogeneity of the Universe**, one of the basic tenets of the

Cosmological Principle. Fairness demands to say that this finding has been challenged by a few groups, although none of them came up with convincing evidence for the contrary.

- The two-point correlation function plays an important role in dynamical analysis of structure formation: the measured cosmic flows can be related to the matter distribution through the two-point correlation function ("cosmic virial theorem", although for this we also need the *three-point function*"). From this we may infer cosmological parameters. Most noteworthy in this is the determination of the two-point correlation function in redshift space: the anisotropic distortions induced by the influence of cosmic flows on the measured redshifts can be directly translated into an estimate of Ω_m.

1.1 Measuring the Two-Point Correlation Function

The issue is of course how to measure the correlation function. Again, we are beset by the problem that there is only one realization of our Universe known. Our own cosmos. Luckily, also here we are saved by the *ergodic theorem.* We may measure the function by averaging over many different positions. Thus, we will follow this approach. In essence, it becomes a large counting exercise. We are going to count the number of points within spherical shells around a given point. By adding them all up and averaging them in a proper way we get an estimate of the probability that on average at a distance r we have a certain amount of points and thus it's excess or deficit with respect to a homogeneous Poisson process. From this we may infer $\xi(r)$. Easier said then done ... In this experiment you will be invited to determine the two-point

correlation function of the point process that we encountered in the previous tutorial.

1.2 Task

Evidently, this concerns another computer experiment. We will provide you with a set of three estimators.

Segment Cox	λ_s	μ	length
	1000.	12	0.1
	1000.	12	0.05
	1000.	12	0.01
Matern	λ_c	μ	radius
	1000.	12	0.05
	1000.	12	0.1
SonPee	*eta*	*lambda*	Level
	4	2.5	7
	3	1.5	8

- It is up to you to write a program that computes the correlation functions of the point processess discussed in the previous tutorial. You have to make the corresponding graphs of the function (loglog and linlin).

- As at small distances the two-point correlation function often behaves like a power-law of the distance r between points,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{6}$$

  you will also have to fit a power-law to the correlation functions of the Segment Cox and the Soneira-Peebles Processess. From this fit you should derive the **power-law slope** γ and the **correlation length** r_0.

- All of the above you will have to do for three different estimators $\hat{\xi}$ of the two-point correlation function, the "standard" *Davis-Peebles estimator*, the *Hamilton estimator* and the *Landy-Szalay estimator*. See the following subsection for their specification.

1.3 Correlation Function Estimators

For measuring $\xi(r)$ on the basis of point counts we need to take into account that we cannot always fit in complete spheres of radius r at every position within a survey volume. In other words, one needs a way of dealing with *edge corrections*. The common practice is to deal with this by means of an equivalent Poisson point catalog in exactly the same volume (and with the same selection criteria concerning depth of the survey).

Assume we have therefore two point sets. One is the sample one, designated by the letter "D" (data). It contains N_D points. In addition there is the Poisson point set "R", with N_R points. Position yourself on a number (as large as practically feasible) of the data points and count the number of data points you find in a spherical shell of with radius $[r, r + \Delta r]$. The total sum of points counted is designated as $DD(r)$. One may also count, for the same number of points, the number of Poisson points in the same shells, $DR(r)$.

- The first estimator is that defined by Davis and Peebles (1983), sometimes called the *standard estimator*,

$$\hat{\xi}_{DP}(r) \,=\, \frac{N_R}{N_D}\,\frac{DD(r)}{DR(r)} \,-\, 1 \tag{7}$$

- Hamilton (1993) found systematic biases in this estimator, surpassing the regular uncertainties (due to finite sampling) in $\hat{\xi}_{DP}(r)$. He therefore proposed the socalled *Hamilton estimator*:

$$\hat{\xi}_{HAM}(r) \,=\, \frac{DD(r)\cdot RR(r)}{[DR(r)]^2} \,-\, 1 \tag{8}$$

  in which $RR(r)$ is the number of pairs in the random catalog with separation in the interval $[r, r + \Delta r]$.

- Almost simultaneously another improved estimator was defined by Landy & Szalay (1993). It has similar properties as the Hamilton estimator,

$$\hat{\xi}_{LS}(r) \,=\, 1 \,+\, \left(\frac{N_R}{N_D}\right)^2\,\frac{DD(r)}{RR(r)} \,-\, 2\frac{N_R}{N_D}\,\frac{DR(r)}{RR(r)}\,. \tag{9}$$