The Large-Scale Structure in the Universe: From Power Laws to Acoustic Peaks

V.J. Martínez

Observatori Astronòmic, Universitat de València, Edifici d’Instituts d’Investigació, Polígon La Coma, 46980 Paterna, València, Spain
martinez@uv.es

Abstract. The most popular tools for analysing the large-scale distribution of galaxies are second-order spatial statistics such as the two-point correlation function or its Fourier transform, the power spectrum. In this review, we explain how our knowledge of cosmic structures, encapsulated by these statistical descriptors, has evolved since their first use when applied on the early galaxy catalogues to the present generation of wide and deep redshift surveys.¹

1 Introduction

As the reader can learn from this volume, there are mainly two astronomical observations that provide the most relevant cosmological data needed to probe any cosmological model: the cosmic microwave background radiation and the large-scale structure of the universe. This review deals with the second of these cosmological fossils. The statistical analysis of galaxy clustering has been progressing in parallel with the development of the observations of the galaxy distribution (for a review see, e.g. Jones et al. [26]). Since the pioneering works by Hubble, measuring the distribution of the number counts of galaxies in telescope fields and finding a log-Gaussian distribution [22], many authors have described the best available data at each moment making use of the then well-established statistical tools. For example, Zwicky [64] used the ratio of clumpiness, the quotient between the variance of the number counts and the expected quantity for a Poisson distribution.

The first map of the sky revealing convincing clustering of galaxies was the Lick survey undertaken by Shane and Wirtanen [57]. While the catalogue was in progress, two different approaches to its statistical description were

¹ Being the first editor of this volume gives me the opportunity of updating this review taking into account the most recent developments in the field. I have used this opportunity trying to incorporate the most challenging discovery in the study of the galaxy distribution: the detection of baryon acoustic oscillations.
developed: The Neyman–Scott approach and the Correlation Function school named in this way by Bernard Jones [24].

Jerzy Neyman and Elisabeth Scott were the first to consider the galaxy distribution as a realization of a homogeneous random point process [37]. They formulated a priori statistical models to describe the clustering of galaxies and later they tried to fit the parameters of the model by comparing it with observations. In this way, they modelled the distribution of galaxy clusters as a random superposition of groups following what now is known in spatial statistics as a Neyman–Scott process, i.e. a Poisson cluster process constructed in two steps: first, a homogeneous Poisson process is generated by randomly distributing a set of centres (or parent points); second, a cluster of daughter points are scattered around each of the parent points, according to a given density function. This idea [38, 44] is the basis of the recent halo model [58] that successfully describes the statistics of the matter distribution in structures of different sizes at different scales: at small scales the halo model assumes that the distribution is dominated by the density profiles of the dark matter halos, and therefore correlations come mainly from intra-halo pairs. The most popular density profile is that of Navarro, Frenk and White [36].

The second approach based on the correlation function was envisaged first by Vera Rubin [52] and by Nelson Limber [31]. They thought that the galaxy distribution was in fact a set of points sampled from an underlying continuous density distribution that later was called the Poisson model by Peebles [47]. In spatial statistics this is known as a Cox process [34]. They derived the auto-correlation function from the variance of the number counts of the ongoing Lick survey. Moreover, Limber provided an integral equation relating the angular and the spatial correlation functions valid for small angle separation (a special version of this equation appears also in the paper by Rubin). The correlation function measures the clustering in excess \( \xi(r) > 0 \) or in defect \( \xi(r) < 0 \) compared with a Poisson distribution. It can be defined in terms of the probability \( dP \) of finding a galaxy in a small volume \( dV \) lying at a distance \( r \) of a given galaxy:

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

where \( n \) is the mean number density over the whole sample volume (see Sect. 3 for a more formal definition). Totsuji and Kihara [61] were the first to obtain a power-law behaviour for the spatial correlation function \( \xi(r) = (r/r_0)^{-1.8} \) on the basis of angular data taken from the Lick survey and making use of the Limber equation. Moreover, as we can see in Fig. 1 reproduced from their paper, the observed correlation function of the Lick survey is fitted to an early halo model – the Neyman–Scott process. The evidence of the power-law behaviour was extended by Peebles [45, 46] and demonstrated in Groth and Peebles [16]. The contributions by Jim Peebles and co-workers and, in particular, his influential book published in 1980 “The Large-Scale Structure of the Universe” [47] were crucial to make these statistics widely used in cosmology.