Preface

> You may ask "What can a hard headed statistician offer to a starry eyed astronomer?" The answer is, "Plenty." One normally associates statistics with large numbers, and astronomy is full of large numbers. The number of stars in our galaxy, the so-called Milky Way System, is more than a hundred thousand million. The number of galaxies in the observable universe is upwards of a thousand million. Surely these large numbers justify a *prima-facie* case for the use of statistical techniques! . . . I have every reason to believe that increased interaction between statistics and astronomy will be to the benefit of both the subjects.
>
> — *J. V. Narlikar*

The last half of the twentieth century saw cosmology develop into a very active and diverse field of science. This was largely due to the development of observational techniques that allowed astronomers to observe extremely distant regions of space. This motivated a flow of new theories about the evolution of our universe and the formation of the large-scale structure we found in it. The main tools to compare theoretical results with observations in astronomy are statistical, so the new theories and observations also initiated an active use of spatial statistics in cosmology.

Many of the statistical methods used in the analysis of the large-scale distribution of matter in the universe have been developed by cosmologists and are not too rigorous. In many cases, similar methods, sometimes under different names, had been used for years in mainstream spatial statistics. In the late 1950s, when the Berkeley statisticians J. Neyman and E. Scott carried out an intensive program for the analysis of galaxy catalogs, the connection between spatial statisticians and cosmologists was a fruitful one. However, in the following 30 years cosmologists were not, in general, aware of developments in statistics, and vice versa. Fortunately, recent years have brought the resumption of a dialog between astronomers and mathematicians, led by the Penn State conferences. We hope that this dialog will continue and will be useful. Cosmology is a good field for applications of spatial statistics. Its well-defined and growing data sets represent an important challenge for the statistical analysis, and therefore for the mathematical community.

The very influential book by Peebles (1980), a milestone in the field, gave the first complete description of statistical methods for study of the spatial distribution of galaxies and of the essential cosmological dynamics. In the years that followed new data have been collected, new methods have been developed, and new discoveries have been made.

This book describes the presently available observational data on the distribution of galaxies and the application of spatial statistics in cosmology. It provides a detailed derivation of the basic statistical methods used to study the spatial distribution of galaxies and of the cosmological physics needed to formulate the statistical models.

We have delineated the basic ideas and practical algorithms and have cited original articles for a more detailed description: there is always more information in articles than can be collected in a book. We have tried to select articles that present the subject in the clearest possible way (frequently they are review articles); thus, our selection is not meant to give a full history of the development of the methods.

Cosmological statistics is a rapidly developing field. We have tried to give the most up-to-date results (at least up to the year 2001 at the preprint level). After working through this book, the reader should be able to understand new research articles and to set up her/his own research projects.

The book is meant to appeal to two different communities of scholars:

- graduate students in cosmology and practicing cosmologists
- mathematicians interested in methods of spatial statistics and their applications

Because the prevalent statistical approach in cosmological statistics has been frequentist, a large collection of different statistics has been developed and studied. Many have been applied only in specific problems and models. This book describes the most general and the most widely used methods of this collection. Personal bias is unavoidable in such a selection, although we have tried to be as objective as possible. But we also describe Bayesian techniques that have become popular recently in large-scale structure studies (especially for the estimation of the power spectrum).

The outline of the book is as follows: after a short description of galaxies (the main objects of cosmological statistics), magnitude systems, and distance indicators in astronomy, we provide in Chapter 1 an overview of current knowledge of the structure in the universe, based on the catalogs of galaxies observed thus far. In Chapter 2 we describe briefly the standard hot Big Bang model of the universe, with special emphasis on measuring distances in cosmology. In Chapter 3 we explain how the galaxy distribution can be analyzed using techniques developed for point processes. Correlation functions are explained in detail together with the different estimators used in their evaluation. Since second-order intensity functions have been a topic of intensive research in spatial statistics lately, we have tried in this chapter to break the language barrier between both fields.

Fractal methods have become very popular in the analysis of galaxy clustering and Chapter 4 is devoted to this subject. In Chapter 5 we review the history of the field, introducing the Neyman–Scott processes and other related geometrical models. We also give special attention to the Voronoi models and to a physically motivated Saslaw distribution function. In Chapter 6 we discuss the dynamics of structure formation, which is important background for formulating contemporary statistical models of galaxy clustering. The theory of random density and velocity fields is described in Chapter 7, with emphasis on its cosmological applications. We focus our attention on both Gaussian and non-Gaussian random fields, with special attention to the properties and clustering of peaks. We end this chapter by explaining the mass (intensity) functions predicted by the Press–Schechter theory and the recent halo model of galaxy clustering that is close to the ideas of Neyman and Scott.

Chapter 8 describes the statistical measures of clustering in Fourier space, in particular methods of estimation of the power spectrum from observational data. In the last sections, the properties of lower dimensional fields are discussed, studying what can be inferred from the projected catalogs as well as from one- or two-dimensional surveys (pencil-beams or slices). In Chapter 9 we explain the methods of reconstructing the density field in the nearby universe (cosmography), trying to remove velocity distortions and making use of the statistical knowledge of velocities and positions of galaxies. We briefly review the application of gravitational lensing for studying the large-scale structure of the universe. In the last chapter we present the statistical tools that have been developed to highlight morphological features of the galaxy distribution. The topology of the galaxy distribution, used to judge whether we have a cellular or a sponge-like distribution of galaxies, can be found here together with other morphological descriptors, such as the Minkowski functionals, minimal spanning trees, and wavelets. Algorithms for finding clusters, voids, and possible periodicities in the galaxy distribution are explained in detail. Appendix A gives a short introduction to spherical astronomy and to the coordinate systems used in the catalogs of galaxies, with some practical formulae to perform coordinate transforms. Appendix B provides a brief summary of the basic statistical terminology.

Vicent Martínez and Enn Saar