1 Introduction

1.1 The role of statistics in astronomy

1.1.1 Astronomy and astrophysics

Today, the term "astronomy" is best understood as shorthand for "astronomy and astrophysics". Astronomy (*astro* = star and *nomen* = name in ancient Greek) is the observational study of matter beyond Earth: planets and bodies in the Solar System, stars in the Milky Way Galaxy, galaxies in the Universe, and diffuse matter between these concentrations of mass. The perspective is rooted in our viewpoint on or near Earth, typically using telescopes on mountaintops or robotic satellites to enhance the limited capabilities of our eyes. Astrophysics (*astro* = star and *physis* = nature) is the study of the intrinsic nature of astronomical bodies and the processes by which they interact and evolve. This is an indirect, inferential intellectual effort based on the (apparently valid) assumption that physical processes established to rule terrestrial phenomena – gravity, thermodynamics, electromagnetism, quantum mechanics, plasma physics, chemistry, and so forth – also apply to distant cosmic phenomena. Figure 1.1 gives a broad-stroke outline of the major fields and themes of modern astronomy.

The fields of astronomy are often distinguished by the structures under study. There are planetary astronomers (who study our Solar System and extra-solar planetary systems), solar physicists (who study our Sun), stellar astronomers (who study other stars), Galactic astronomers (who study our Milky Way Galaxy), extragalactic astronomers (who study other galaxies), and cosmologists (who study the Universe as a whole). Astronomers can also be distinguished by the type of telescope used: there are radio astronomers, infrared astronomers, visible-light astronomers, X-ray astronomers, gamma-ray astronomers, and physicists studying cosmic rays, neutrinos and the elusive gravitational waves. Astrophysicists are sometimes classified by the processes they study: astrochemists, atomic and nuclear astrophysicists, general relativists (studying gravity) and cosmologists.

The astronomer might proceed to investigate stellar processes by measuring an ordinary main-sequence star with spectrographs at different wavelengths of light, examining its spectral energy distribution with thousands of absorption lines. The astrophysicist interprets that the emission of a star is produced by a sphere of 10^{57} atoms with a specific mixture of elemental abundances, powered by hydrogen fusion to helium in the core, revealing itself to the Universe as a blackbody surface at several thousand degrees temperature. The development of the observations of normal stars started in the late-nineteenth century,

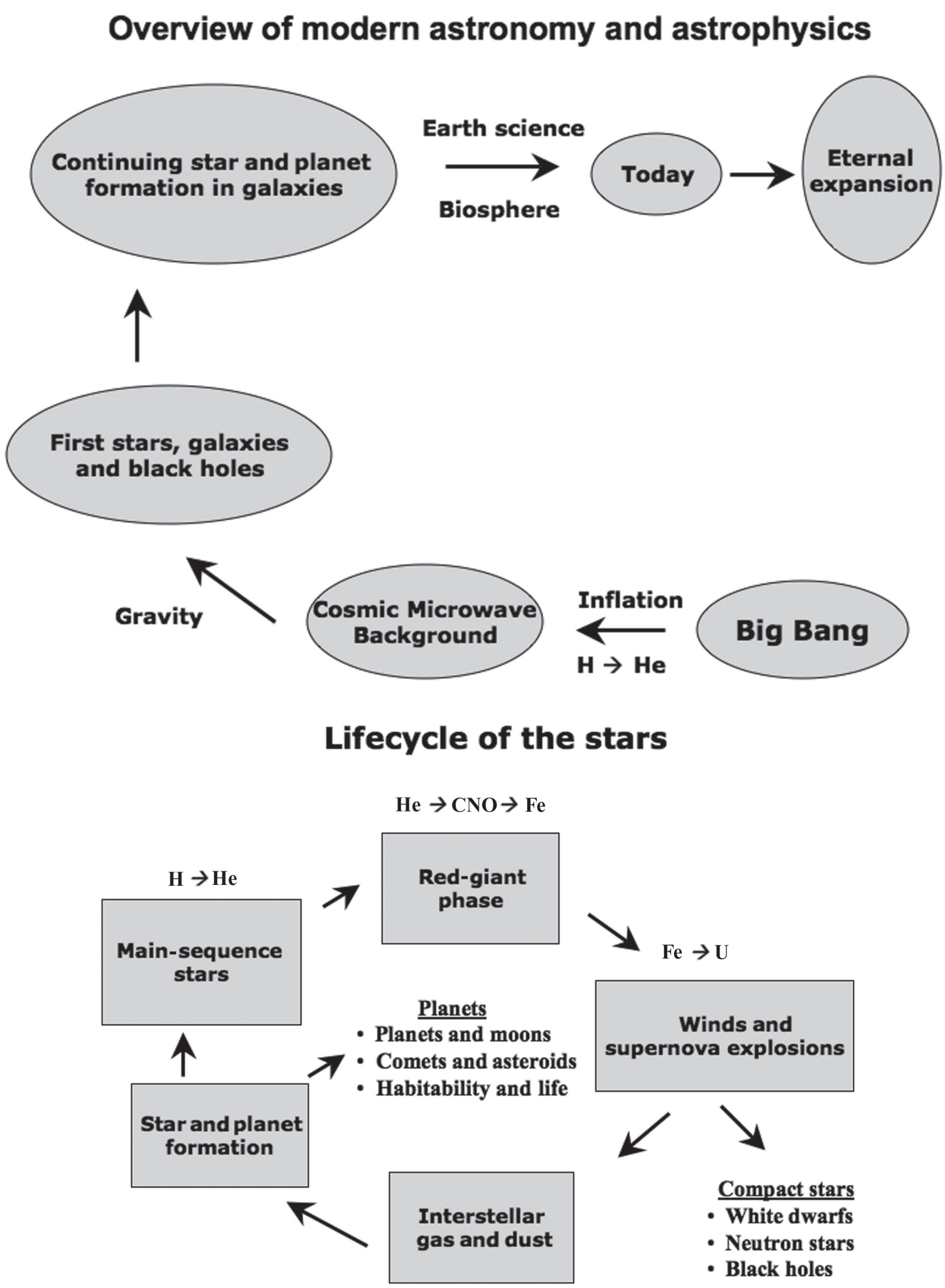

Fig. 1.1 Diagrams summarizing some important fields and themes of modern astronomy. *Top*: the history and growth of structures of the expanding Universe; *bottom*: the evolution of stars with generation of heavy elements and production of long-lived structures.

and the successful astrophysical interpretation emerged gradually throughout the twentieth century. The vibrant interwoven progress of astronomy and astrophysics continues today as many other cosmic phenomena, from molecular clouds to black holes, are investigated. Cosmology, in particular, has emerged with the remarkable inference that the familiar atoms around us comprise only a small fraction of the "stuff" in the Universe which is dominated by mysterious dark matter and dark energy inaccessible to normal telescopes or laboratory instruments.

1.1.2 Probability and statistics

While there is little debate about the meaning and goals of astronomy and astrophysics as intellectual enterprises, the meaning and goals of probability and statistics has been widely debated. In his volume *Statistics and Truth*, C. R. Rao (1997) discusses how the term "statistics" has changed meaning over the centuries. It originally referred to the collection and compilation of data. In the nineteenth century, it accrued the goal of the mathematical interpretation of data, often to assist in making real-world decisions. Rao views contemporary statistics as an amalgam of a science (techniques derived from mathematics), a technology (techniques useful for decision-making in the presence of uncertainty), and an art (incompletely codified techniques based on inductive reasoning).

Barnett (1999) considers various viewpoints on the meaning of statistics. The first group of quotes see statistics as a very useful, but essentially mechanical, technology for treating data. In this sense, it plays a role similar to astronomy's role *vis á vis* astrophysics.

1. "The first task of a statistician is cross-examination of data." (Sir R. A. Fisher, quoted by Rao 1997)
2. "[S]tatistics refers to the methodology for the collection, presentation, and analysis of data, and for the uses of such data." (Neter *et al.* 1978)
3. "Broadly defined, statistics encompasses the theory and methods of collecting, organizing, presenting, analyzing, and interpreting data sets so as to determine their essential characteristics." (Panik 2005)

The following interpretations of statistics emphasize its role in reducing random variations in observations to reveal important effects in the underlying phenomenon under study.

4. "A statistical inference carries us from observations to conclusions about the populations sampled." (Cox 1958)
5. "Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge." (Rao 1997)
6. "My favourite definition [of statistics] is bipartite: statistics is both the science of uncertainty and the technology of extracting information from data." (Hand 2010)
7. "In statistical inference experimental or observational data are modeled as the observed values of random variables, to provide a framework from which inductive conclusions may be drawn about the mechanism giving rise to the data." (Young & Smith 2005)

1.1.3 Statistics and science

Opinions differ widely when considering the relationship between statistical analysis of empirical data and the underlying real phenomena. A group of prominent twentieth-century statisticians express considerable pessimism that statistical models are anything but useful fictions, much as Renaissance Europe debated the meaning of Copernicus' heliocentric cosmological model. These scholars view statistical models as useful but often trivial or even misleading representations of a complex world. Sir D. R. Cox, towards the end of a long career, perceives a barrier between statistical findings and the development or validation of scientific theories.

8. "There is no need for these hypotheses to be true, or even to be at all like the truth; rather one thing is sufficient for them – that they should yield calculations which agree with the observations." (Osiander's preface to Copernicus' *De Revolutionibus*, quoted by Rao 1997)
9. "Essentially, all models are wrong, but some are useful." (Box & Draper 1987)
10. "[Statistical] models can provide us with ideas which we test against data, and about which we build up experience. They can guide our thinking, lead us to propose courses of action, and so on, and if used sensibly, and with an open mind, and if checked frequently with reality, might help us learn something that is true. Some statistical models are helpful in a given context, and some are not. . . . What we do works (when it does) because it can be seen to work, not because it is based on true or even good models of reality." (Speed 1992, addressing a meeting of astronomers)
11. "It is not always convenient to remember that the right model for a population can fit a sample of data worse than a wrong model, even a wrong model with fewer parameters. We cannot rely on statistical diagnostics to save us, especially with small samples. We must think about what our models mean, regardless of fit, or we will promulgate nonsense." (Wilkinson 2005).
12. "The object [of *statistical* inference] is to provide ideas and methods for the critical analysis and, as far as feasible, the interpretation of empirical data . . . The extremely challenging issues of *scientific* inference may be regarded as those of synthesising very different kinds of conclusions if possible into a coherent whole or theory . . . The use, if any, in the process of simple *quantitative* notions of probability and their numerical assessment is unclear . . . " (Cox 2006)

Other scholars quoted below are more optimistic. The older Sir R. A. Fisher bemoans a mechanistic view of statistics without meaning in the world. G. Young and R. Smith imply that statistical modeling can lead to an understanding of the causative mechanisms of variations in the underlying population. I. Hacking, a philosopher, believes statistics can improve our scientific inferences but not lead to new discovery. B. Efron, in an address as President of the American Statistical Association, feels that statistics can propel many sciences towards important results and insights.

13. "To one brought up in the free intellectual atmosphere of an earlier time there is something rather horrifying in the ideological movement represented by the doctrine that

reasoning, properly speaking, cannot be applied to empirical data to lead to inferences valid in the real world." (Fisher 1973)

14. "The quiet statisticians have changed our world, not by discovering new facts or technical developments, but by changing the ways we reason, experiment, and form our opinions." (Hacking 1990)
15. "Statistics has become the primary mode of quantitative thinking in literally dozens of fields, from economics to biomedical research. The statistical tide continues to roll in, now lapping at the previously unreachable shores of the hard sciences. . . . Yes, confidence intervals apply as well to neutrino masses as to disease rates, and raise the same interpretive questions, too." (Efron 2004)
16. "The goal of science is to unlock nature's secrets. . . . Our understanding comes through the development of theoretical models which are capable of explaining the existing observations as well as making testable predictions. . . . Fortunately, a variety of sophisticated mathematical and computational approaches have been developed to help us through this interface, these go under the general heading of statistical inference." (Gregory 2005)

Leading statisticians are thus often more cautious, or at least less self-confident, about the value of their labors for understanding phenomena than are astronomers. Most astronomers believe implicitly that their observations provide a clear window into the physical Universe, and that simple quantitative statistical interpretations of their observations represent an improvement over qualitative examination of the data.

We generally share the optimistic view of statistical methodology in the service of astronomy and astrophysics, as expressed by P. C. Gregory (2005). In the language of the philosophy of science, we are positivists who believe that underlying causal relationships can be discovered through the detection and study of regular patterns of observable phenomena. While quantitative interpretation and models of complex biological and human affairs attempted by many statisticians may be more useful for prediction or decision-making than understanding the underlying behaviors, we feel that quantitative models of many astrophysical phenomena can be very valuable. A social scientist might interview a sample of voters to accurately predict the outcome of an election, yet never understand the beliefs underlying these votes. But an astrostatistician may largely succeed in understanding the orbits of binary stars, or the behavior of an accretion disk around a black hole or the growth of structure in an expanding Universe, that must obey deterministic mathematical laws of physics.

However, we wish to convey throughout this volume that the process of linking statistical analysis to reality is not simple and challenges must be faced at all stages. In setting up the calculation, there are often several related questions that might be asked in a given scientific enterprise, and their statistical evaluation may lead to apparently different conclusions. In performing the calculation, there are often several statistical approaches to a given question asked about a dataset, each mathematically valid under certain conditions, yet again leading to different scientific inferences. In interpreting the result, even a clear statistical finding may give an erroneous scientific interpretation if the mathematical model is mismatched to physical reality.

Astronomers should be flexible and sophisticated in their statistical treatments, and adopt a more cautious view of the results. A "3-sigma" result does not necessarily represent astrophysical reality. Astronomers might first seek consensus about the exact question to be addressed, apply a suite of reasonable statistical approaches to the dataset with clearly stated assumptions, and recognize that the link between the statistical results and the underlying astrophysical truth may not be straightforward.

1.2 History of statistics in astronomy

1.2.1 Antiquity through the Renaissance

Astronomy is the oldest observational science. The effort to understand the mysterious luminous objects in the sky has been an important element of human culture for tens of thousands of years. Quantitative measurements of celestial phenomena were carried out by many ancient civilizations. The classical Greeks were not active observers but were unusually creative in the applications of mathematical principles to astronomy. The geometric models of the Platonists with crystalline spheres spinning around the static Earth were elaborated in detail, and this model endured in Europe for fifteen centuries.

The Greek natural philosopher Hipparchus made one of the first applications of mathematical principles in the realm of statistics, and started a millennium-long discussion on procedures for combining inconsistent measurements of a physical phenomenon (Sheynin 1973, Hald 2003). Finding scatter in Babylonian measurements of the length of a year, defined as the time between solstices, he took the middle of the range – rather than the mean or median – for the best value. Today, this is known as the midrange estimator of location, and is generally not favored due to its sensitivity to erroneous observations. Ptolemy and the eleventh-century Persian astronomer Abu Rayhan Biruni (al-Biruni) similarly recommended the average of extremes. Some medieval scholars advised against the acquisition of repeated measurements, fearing that errors would compound uncertainty rather than compensate for each other. The utility of the mean of discrepant observations to increase precision was promoted in the sixteenth century by Tycho Brahe and Galileo Galilei. Johannes Kepler appears to have inconsistently used arithmetic means, geometric means and middle values in his work. The supremacy of the mean was not settled in astronomy until the eighteenth century (Simpson 1756).

Ancient astronomers were concerned with observational errors, discussing dangers of propagating errors from inaccurate instruments and inattentive observers. In a study of the corrections to astronomical positions from observers in different cities, al-Biruni alludes to three types of errors: ". . . the use of sines engenders errors which become appreciable if they are added to errors caused by the use of small instruments, and errors made by human observers" (quoted by Sheynin 1973). In his 1609 *Dialogue on the Two Great World Views, Ptolemaic and Copernican*, Galileo also gave an early discussion of observational errors concerning the distance to the supernova of 1572. Here he outlined in nonmathematical language many of the properties of errors later incorporated by Gauss into his quantitative theory of errors.

1.2.2 Foundations of statistics in celestial mechanics

Celestial mechanics in the eighteenth century, in which Newton's law of gravity was found to explain even the subtlest motions of heavenly bodies, required the quantification of a few interesting physical quantities from numerous inaccurate observations. Isaac Newton himself had little interest in quantitative probabilistic arguments. In 1726, he wrote concerning discrepant observations of the Comet of 1680 that, "From all this it is plain that these observations agree with theory, in so far as they agree with one another" (quoted by Stigler 1986).

Others tackled the problem of combining observations and estimating physical quantities through celestial mechanics more earnestly. In 1750 while analyzing the libration of the Moon as head of the observatory at Göttingen, Tobias Mayer developed a "method of averages" for parameter estimation involving multiple linear equations. In 1767, British astronomer John Michell similarly used a significance test based on the uniform distribution (though with some technical errors) to show that the Pleiades is a physical, rather than chance, grouping of stars. Johann Lambert presented an elaborate theory of errors, often in astronomical contexts, during the 1760s. Bernouilli and Lambert laid the foundations of the concept of maximum likelihood later developed more thoroughly by Fisher in the early twentieth century.

The Marquis Pierre-Simon de Laplace (1749−1827), the most distinguished French scientist of his time, and his competitor Adrien-Marie Legendre, made seminal contributions both to celestial mechanics and to probability theory, often intertwined. Their generalizations of Mayer's methods for treating multiple parametric equations constrained by many discrepant observations had great impact. In astronomical and geodetical studies during the 1780s and in his huge 1799–1825 opus *Mécanique Céleste*, Laplace proposed parameter estimation for linear models by minimizing the largest absolute residual. In an 1805 appendix to a paper on cometary orbits, Legendre proposed minimizing the sum of the squares of residuals, or the method of least squares. He concluded "that the method of least squares reveals, in a manner of speaking, the center around which the results of observations arrange themselves, so that the deviations from that center are as small as possible" (quoted by Stigler 1986).

Both Carl Friedrich Gauss, also director of the observatory at Göttingen, and Laplace later placed the method of least squares onto a solid mathematical probabilistic foundation. While the method of least squares had been adopted as a practical convenience by Gauss and Legendre, Laplace first treated it as a problem in probabilities in his *Théorie Analytique des Probabilités*. He proved by an intricate and difficult course of reasoning that it was the most advantageous method for finding parameters in orbital models from astronomical observations, the mean of the probabilities of error in the determination of the elements being thereby reduced to a minimum. Least-squares computations rapidly became the principal interpretive tool for astronomical observations and their links to celestial mechanics. These and other approaches to statistical inference are discussed in Chapter 3.

In another portion of the *Théorie*, Laplace rescued from obscurity the postulation of the Central Limit Theorem by the mathematician Abraham De Moivre who, in a remarkable article published in 1733, used the normal distribution to approximate the distribution of the number of heads resulting from many tosses of a fair coin. Laplace expanded De Moivre's finding by approximating the binomial distribution with the normal distribution. Laplace's proof was flawed, and improvements were developed by Siméon-Denis Poisson, an astronomer at Paris' Bureau des Longitudes, and Friedrich Bessel, director of the observatory in Königsberg. Today, the Central Limit Theorem is considered to be one of the foundations of probability theory (Section 2.10).

Gauss established his famous error distribution and related it to Laplace's method of least squares in 1809. Astronomer Friedrich Bessel introduced the concept of "probable error" in a 1816 study of comets, and demonstrated the applicability of Gauss' distribution to empirical stellar astrometric errors in 1818. Gauss also introduced some treatments for observations with different (heteroscedastic) measurement errors and developed the theory for unbiased minimum variance estimation. Throughout the nineteenth century, Gauss' distribution was widely known as the "astronomical error function".

Although the fundamental theory was developed by Laplace and Gauss, other astronomers published important contributions to the theory, accuracy and range of applicability of the normal distribution and least-squares estimation during the latter part of the nineteenth century (Hald 1998). They include Ernst Abbe at the Jena Observatory and the optics firm of Carl Zeiss, Auguste Bravais of the Univerity of Lyons, Johann Encke of the Berlin Observatory, Britain's Sir John Herschel, Simon Newcomb of the U.S. Naval Observatory, Giovanni Schiaparelli of Brera Observatory, and Denmark's Thorvald Thiele. Sir George B. Airy, British Royal Astronomer, wrote an 1865 text on least-squares methods and observational error.

Adolphe Quetelet, founder of the Belgian Royal Observatory, and Francis Galton, director of Britain's Kew Observatory, did little to advance astronomy but were distinguished pioneers extending statistical analysis from astronomy into the human sciences. They particularly laid the groundwork for regression between correlated variables. The application of least-squares techniques to multivariate linear regression emerged in biometrical contexts by Karl Pearson and his colleagues in the early 1900s (Chapter 7).

The intertwined history of astronomy and statistics during the eighteenth and nineteenth centuries is detailed in the monographs by Stigler (1986), Porter (1986) and Hald (1998).

1.2.3 Statistics in twentieth-century astronomy

The connections between astronomy and statistics considerably weakened during the first decades of the twentieth century as statistics turned its attention principally to biological sciences, human attributes, social behavior and statistical methods for industries such as life insurance, agriculture and manufacturing. Advances in astronomy similarly moved away from the problem of evaluating errors in measurements of deterministic processes of celestial mechanics. Major efforts on the equilibrium structure of stars, the geometry of the Galaxy, the discovery of the interstellar medium, the composition of stellar atmospheres, the study of solar magnetic activity and the discovery of extragalactic nebulae generally did not involve statistical theory or application. Two distinguished statisticians wrote series of papers in the astronomical literature – Karl Pearson on correlations between stellar properties around 1907–11, and Jerzy Neyman with Elizabeth Scott on clustering of galaxies around 1952–64 – but neither had a strong influence on further astronomical developments.

The least-squares method was used in many astronomical applications during the first half of the twentieth century, but not in all cases. Schlesinger (1916) admonished astronomers estimating elements of binary-star orbits to use least-squares rather than trial-and-error techniques. The stellar luminosity function derived by Jacobus Kapteyn, and thereby the inferred structure of the Galaxy, were based on subjective curve fitting (Kapteyn & van Rhijn 1920), although Kapteyn had made some controversial contributions to the mathematics of skewed distributions and correlation. An important study on dark matter in the Coma Cluster fits the radial distribution of galaxies by eye and does not quantify its similarity to an isothermal sphere (Zwicky 1937). In contrast, Edwin Hubble's seminal studies on galaxies were often based on least-squares fits (e.g. the redshift–magnitude relationship in Hubble & Humason 1931), although an early study reports a nonstandard symmetrical average of two regression lines (Hubble 1926, Section 7.3.2). Applications of statistical methods based on the normal error law were particularly strong in studies involving positional astronomy and star counts (Trumpler & Weaver 1953). Astronomical applications of least-squares estimation were strongly promoted by the advent of computers and Bevington's (1969) useful volume with FORTRAN code. Fourier analysis was also commonly used for time series analysis in the latter part of the twentieth century.

Despite its formulation by Fisher in the 1920s, maximum likelihood estimation emerged only slowly in astronomy. Early applications included studies of stellar cluster convergent points (Brown 1950), statistical parallaxes from the Hertzsprung–Russell diagram (Jung 1970), and some early work in radio and X-ray astronomy. Crawford *et al.* (1970) advocated use of maximum likelihood for estimating power-law slopes, a message we reiterate in this volume (Section 4.4). Maximum likelihood studies with truly broad impact did not emerge until the 1970s. Innovative and widely accepted methods include Lynden-Bell's (1971) luminosity function estimator for flux-limited samples, Lucy's (1974) algorithm for restoring blurry images, and Cash's (1979) algorithm for parameter estimation involving photon counting data. Maximum likelihood estimators became increasingly important in extragalactic astronomy; they were crucial for the discovery of galaxy streaming towards the Great Attractor (Lynden-Bell *et al.* 1988) and calculating the galaxy luminosity function from flux-limited surveys (Efstathiou *et al.* 1988). The 1970s also witnessed the first use and rapid acceptance of the nonparametric Kolmogorov–Smirnov statistic for two-sample and goodness-of-fit tests.

The development of inverse probability and Bayes' theorem by Thomas Bayes and Laplace in the late eighteenth century took place largely without applications to astronomy. Despite the prominence of the leading Bayesian proponent Sir Harold Jeffreys, who won the Gold Medal of the Royal Astronomical Society in 1937 and served as Society President in the 1950s, Bayesian methods did not emerge in astronomy until the latter part of the twentieth century. Bayesian classifiers for discriminating stars and galaxies (based on the 2001 text written for engineers by Duda *et al.*) were used to construct large automated sky survey catalogs (Valdes 1982), and maximum entropy image restoration gained some interest (Narayan & Nityananda 1986). But it was not until the 1990s that Bayesian methods became widespread in important studies, particularly in extragalactic astronomy and cosmology.

The modern field of astrostatistics grew suddenly and rapidly starting in the late 1990s. This was stimulated in part by monographs on statistical aspects of astronomical image processing (Starck *et al.* 1998, Starck & Murtagh 2006), galaxy clustering (Martínez & Saar 2001), Bayesian data analyses (Gregory 2005) and Bayesian cosmology (Hobson *et al.* 2010). Babu & Feigelson (1996) wrote a brief overview of astrostatistics. The continuing conference series *Statistical Challenges in Modern Astronomy* organized by us since 1991 brought together astronomers and statisticians interested in forefront methodological issues (Feigelson & Babu 2012). Collaborations between astronomers and statisticians emerged, such as the California–Harvard Astro-Statistical Collaboration (http://hea-www.harvard.edu/AstroStat), the International Computational Astrostatistics Group centered in Pittsburgh (http://www.incagroup.org), and the Center for Astrostatistics at Penn State (http://astrostatistics.psu.edu). However, the education of astronomers in statistical methodology remains weak. Penn State's Center and other institutes operate week-long summer schools in statistics for young astronomers to partially address this problem.

1.3 Recommended reading

We offer here a number of volumes with broad coverage in statistics. Stigler's monograph reviews the history of statistics and astronomy. Rice, Hogg & Tanis, and Hogg *et al.* are well-respected textbooks in statistical inference at undergraduate and graduate levels, and Wasserman gives a modern viewpoint. Lupton, James, and Wall & Jenkins are written by and for physical scientists. Ghosh *et al.* and Gregory introduce Bayesian inference.

Ghosh, J. K., Delampady, M. & Samanta, T. (2006) *An Introduction to Bayesian Analysis: Theory and Methods*, Springer, Berlin
A graduate-level textbook in Bayesian inference with coverage of the Bayesian approach, objective and reference priors, convergence and large-sample approximations, model selection and testing criteria, Markov chain Monte Carlo computations, hierarchical Bayesian models, empirical Bayesian models and applications to regression and high-dimensional problems.

Gregory, P. (2005) *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press
This monograph treats probability theory and sciences, practical Bayesian inference, frequentists approaches, maximum entropy, linear and nonlinear model fitting, Markov