Draft version September 27, 2025
Typeset using LaTeX **preprint2** style in AASTeX631

Bayesian Inference and MCMC Methods in Astrophysics

Agastya Gaur
University of Illinois at Urbana-Champaign

ABSTRACT

This is the abstract of the paper. It summarizes the work in a concise form.

Contents

1. INTRODUCTION

1.1. *The History of Astrostatistics*

In the 4th century BC, Hipparchus, attempting to estimate the length of a year, found the middle of the range of a scattered set of Babylonian solstice measurements. An acheivement for the time, Hipparchus's measurement marked the beginning of what would become a long standing marriage between astronomy and statistics. In the centuries to come, a number of breakthroughs in astrostatistics would continue to occur. Notably Brahe, who in the late 1500s employed repeated positional mesurements of stars and used the mean of the data to map them out. His work was so precise, it became the foundation of Kepler's laws of planetary motion, and it took astronomers generations to produce better measurements. (Leavesley & Tárnok 2018). Furthermore, in the 1770s, Laplace rediscovered Bayesian statistics, and over the next decade he continued to expand upon his work, using it in a colossal effort to complete Newton's theory of gravity, work that would have him hailed as a monumental genius (Stigler 1975).

The biggest advancement in astrostatistics before the era of computing came in 1805 when Legendre published the method of least squares regression to model the orbit of comets (Feigelson & Babu 2004). He theorized that the model best fit to a set of data was one that minimized the sum of the squares of the errors. Though Legendre did not provide a formal proof of the method, regarding it only as a convenient trick, later works by Adrian developed formal mathematical proofs of the method. (Merriman 1877). In 1809, Gauss published his own work on least squares, showing it was used to calculate the orbit of the dwarf planet Ceres, even when observing it was impossible due to solar glare. Less elegantly, he also insisted that he had discovered the method years before Legendre (Stigler 1981). As controvertial as the development of least squares regression ended up being, it has cemented itself in history as one of the most important leaps in astrostatistics.

The recurring theme was clear: progress in astronomy often hinged on solving problems of statistical estimation. By the end of the century, astronomy had firmly established itself as a quantitative science, driven by the refinement of statistical methods to identify regularities in scattered measurements, fitting orbital models, and quantifying uncertainty in the presence of noise.

1.2. *Resolving the Identity Crisis*

The next 100 years brought two developments that reshaped the relationship between astronomy and statistics: the rise of physics as the foundation of astronomy, and the advent of computing, which enabled unprecedented scales of data analysis. As astronomy grew increasingly intertwined with the theories of physics, the field transformed into what we now call astrophysics. As more astronomers began to call themselves astrophysicists, the role of statistics began to fade. Though a niche subset called statistical astronomy still remained, the majority of astronomers had little use for statistics in their work (Feigelson & Babu 2004). The focus shifted to deriving physical models from first principles, and statistical methods were often seen as secondary or even unnecessary. In 1930, Hubble determined the fit for the light curve of elliptical galaxies by trial-and-error instead of regression. In 1937, Zwicky first observed dark matter using a curve fitted only by eye (Feigelson et al. 2021).

However, statistics would not be kept away from astronomy for long. As computers became more widely used and accessible, astronomical interest in statistical methods was reignited. Astronomers could now work with much larger sets of data than before. Advances in computing increased the scale of the statistical analysis that was feasible to perform, and since then it has only been rising. While the early history of the field was dominated by statistical reasoning, the growth of physics and computation broadened this into what we now call quantitative analysis (QA): a synthesis of statistical inference, numerical modeling, and data-driven computation.

1.3. *The Data Deluge*

Today, astrophysics sits in the middle of a universe of complex statistical problems that demand new quantitative approaches and more computing power by the day. In many respects, QA has become the backbone of research in modern astrophysics, and at a pivotal moment, as the 21st century has ushered in an unprecedented era of astronomical data generation. Sky surveys like Gaia DR3 alone provide astrometry and photometry for nearly two billion stars, plus more than ten million variable sources across dozens of types (Gaia Collaboration et al. 2023). The nineteenth data release of the Sloan Digital Sky Survey collected robust spectra data from over 6 million objects (Collaboration et al. 2025). Advances in CCD detectors will see data from sky surveys continue to grow in the next decade from an order of gigabytes to terabytes, and possibly petabytes in the future. The same trend can be seen in data from the Rubin Observatory LSST and NASA's Solar Dynamics Observatory, which now generates over a terabyte of data per day (Borne 2009).

This data deluge makes QA indispensable. It brings not only more volume, but also qualitatively harder problems such as disentangling individual frequencies from complex signals and modeling nonlinear, degenerate parameter spaces. The ability to extract meaningful insights from these massive datasets in an organized manner is crucial for advancing our understanding of the universe. QA provides a number of powerful tools spanning statistical inference, computational algorithms, and machine learning methodologies to analyze, interpret, and model this data effectively.

1.4. *Statistical Challenges in Modern Astrophysics*

Across astrophysics, there are two common structures of challenges. The first challenge is that noisy, incomplete, and often degenerate data has different estimated distributions from multiple competing theories. Though theoretical astrophysics has given us the tools to reduce these problems to estimations of physical constants, the number and complexity of parameters still poses a large challenge (Schafer 2015). The second challenge is that the large volume of data creates significant problems with computing time and power. Efficient algorithms and scalable statistical methods are required to make analysis computationally tractable (Huijse et al. 2014). Together, these issues create a need for QA frameworks that can both handle uncertainty in complex parameter spaces and scale efficiently with massive datasets.

Within this landscape, Bayesian inferencing via Monte Carlo Markov Chain (MCMC) methods naturally emerges as potential solution. Bayesian inference offers a principled framework for parameter estimation in complex systems, and MCMC methods provide an effective way to explore the parameter spaces by sampling from posterior distributions. For astrophysicists, this has become one of the most widely used and versatile approaches. Von Toussaint (2011) notes the growing applicability of Bayesian inferencing in physics. Computational models are becoming far more complex, and the data being analyzed is often noisy and incomplete. Bayesian methods, with their ability to incorporate prior knowledge and handle uncertainty, are well-suited to these challenges. MCMC methods, in particular, provide a practical way to sample from complex posterior distributions that arise in Bayesian analysis. This makes them invaluable for parameter estimation, model comparison, and uncertainty quantification in almost any astrophysical problem.

The rest of the paper will have the following structure: Sec. II will provide a foundational explanation of Bayesian statistics as well as the mathematical and computational methodology behind MCMC methods. Next, Sec. III will introduce three case studies within astrophysics where Bayesian inferencing and MCMC methods are being used to push research forward. These concepts include the direct detection of exoplanets, CMB parameter estimation, and gravity wave fitting. Each case study will include current challenges in the field, how Bayesian inferencing is being used to address it, the pros and cons of the approach, as well as future advancements that could be made. Finally, Sec. IV will include a discussion on how Bayesian inferencing is being used throughout astrophysics overall and how it can address problems in other fields as well.

2. METHODOLOGY

REFERENCES

Borne, K. D. 2009, Astroinformatics: A 21st Century Approach to Astronomy, arXiv, doi: 10.48550/arXiv.0909.3892

Collaboration, SDSS., Pallathadka, G. A., Aghakhanloo, M., et al. 2025, The Nineteenth Data Release of the Sloan Digital Sky Survey, arXiv, doi: 10.48550/arXiv.2507.07093

Feigelson, E. D., & Babu, G. J. 2004, Statistical Challenges in Modern Astronomy, arXiv, doi: 10.48550/arXiv.astro-ph/0401404

Feigelson, E. D., de Souza, R. S., Ishida, E. E. O., & Babu, G. J. 2021, Annual Review of Statistics and Its Application, 8, 493, doi: 10.1146/annurev-statistics-042720-112045

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, Astronomy & Astrophysics, 674, A1, doi: 10.1051/0004-6361/202243940

Huijse, P., Estevez, P. A., Protopapas, P., Principe, J. C., & Zegers, P. 2014, IEEE Computational Intelligence Magazine, 9, 27, doi: 10.1109/MCI.2014.2326100

Leavesley, S., & Tárnok, A. 2018, Cytometry Part A, 93, 977, doi: 10.1002/cyto.a.23637

Merriman, M. 1877, The Analyst, 4, 33, doi: 10.2307/2635472

Schafer, C. M. 2015, Annual Review of Statistics and Its Application, 2, 141, doi: 10.1146/annurev-statistics-022513-115538

Stigler, S. M. 1975, Biometrika, 62, 503, doi: 10.1093/biomet/62.2.503

—. 1981, The Annals of Statistics, 9, 465

Von Toussaint, U. 2011, Reviews of Modern Physics, 83, 943, doi: 10.1103/RevModPhys.83.943