A comparative study on theoretical mechanisms and applicability of major parameter estimation methods

Hongyi Guo

Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo, China

Smyhg3@nottingham.edu.cn

Abstract. In recent years, with the rapid growth of data scale and the increasing complexity of statistical models, traditional parameter estimation methods have encountered new challenges. The continuous development of parameter estimation techniques aims to improve accuracy and computational efficiency to meet practical needs in complex environments. This paper investigates the fundamental theories and major methods of parameter estimation, with particular emphasis on the underlying concepts, evaluation standards, and application frameworks in statistical models. By presenting the three core methods, namely Maximum Likelihood Estimation (MLE), Method of Moments (MoM), and Bayesian Estimation, this study analyzes their derivation logic, theoretical properties, and applicable scenarios. Furthermore, it explores computational bottlenecks in high-dimensional Bayesian methods, the trade-off between subjective and objective prior selection, and the emerging trend of hybrid approaches based on empirical Bayes and regularization strategies. The results reveal both the commonalities and distinctions among the methods with respect to consistency, efficiency, and computational complexity. Besides, the potential of artificial intelligence to boost computational efficiency and enable more flexible, high-dimensional modeling in parameter estimation is emphasized, providing useful insights for both theoretical research and practical applications.

Keywords: parameter estimation, Maximum Likelihood Estimation (MLE), Method of Moments (MoM), Bayesian Estimation, method selection

1. Introduction

Parameter estimation, a fundamental element of statistical inference, finds broad applications in econometrics, biostatistics, and machine learning, with the selection of methods exerting a critical influence on model performance and interpretability. For decades, Maximum Likelihood Estimation (MLE), Method of Moments (MoM), and Bayesian Estimation have played central roles in statistical modeling, each offering unique strengths in theoretical research and practical applications. Despite extensive research on the theoretical properties and empirical performance of these methods, comparative analyses are still limited. In particular, their applicability, robustness, and computational efficiency under practical scenarios, such as small sample sizes, high-dimensional data, and complex nonlinear models, remain subjects of considerable debate and challenge [1]. This paper aims to comprehensively review and compare the three parameter estimation methods from both theoretical and practical perspectives, with a focus on statistical consistency, estimation efficiency, computational complexity, as well as applicability. And it further identifies the advantages and limitations of each method across different contexts. Based on an analysis of relevant literature that covers both classical theoretical studies and recent applied advances, it synthesizes methodological characteristics, compares performance outcomes, and summarizes representative cases, to provide statistical theorists and empirical researchers with a reference for selecting parameter estimation methods [2]. Meanwhile, it also offers guidance for future research directions in multi-method integration, algorithm optimization, and AI-assisted estimation.

2. Basic theories and classifications of parameter estimation

2.1. Concept and evaluation criteria of parameter estimation

Widely applied in empirical research, parameter estimation serves as a fundamental method in mathematical statistics for inferring unknown characteristics of a population. Depending on the mode of inference, parameter estimation is generally classified into point estimation and interval estimation. Point estimation computes a single value from sample observations as the

estimate of the unknown parameter, whereas interval estimation constructs a confidence interval based on sample information, providing a plausible range of values for the parameter under a given confidence level. And the two approaches complement each other, providing descriptions of parameters from the perspectives of determinacy and uncertainty. In evaluating the quality of an estimator, statistical theory generally relies on three core properties. Unbiasedness means that the expected value of the estimator equals the parameter being estimated, ensuring that it does not systematically deviate from the true value in the long run. In addition, consistency refers to the property that as the sample size approaches infinity, the estimator converges in probability to the true parameter, thereby ensuring reliability in large samples. Besides, efficiency means that among all unbiased estimators, the one with the smallest variance is considered most efficient, representing the highest level of statistical precision [3]. These properties jointly underpin the evaluation of parameter estimation methods. On this basis, the Cramér-Rao lower bound sets the theoretical minimum variance of unbiased estimators under regularity conditions. When an estimator attains this bound, it is regarded as optimally efficient, representing the limit of efficiency [4]. Consequently, the Cramér-Rao bound serves both as a benchmark for efficiency evaluation and as a guide for constructing estimators.

2.2. Model assumptions in parameter estimation

The theoretical foundation of parameter estimation depends on the proper specification of statistical models. In practice, this begins with establishing an appropriate probability model that reflects the data-generating mechanism and characterizes the mathematical relationship between sample data and the underlying population. In this process, clarifying the structure of the parameter space and specifying the basic distributional assumptions is key to ensuring the validity of estimation methods. The model form, underlying assumptions, and mathematical representation of observed data together constitute the basic framework of the parameter estimation problem.

Specifically, let the observed sample $X_1, X_2, ..., X_n$ consist of independent and identically distributed random variables drawn from a population with probability density or mass function $f(x; \theta)$, where θ is an unknown but fixed parameter belonging to a parameter space Θ. This setup forms the classical framework of parametric statistical models. The central task of parameter estimation is to make valid inferences about the unknown parameter θ based on the sample data under this model assumption [5]. Within this framework, defining the parameter space becomes particularly important. It not only restricts the possible values that parameters can take but also directly influences the choice of estimation methods and the theoretical properties that can be established. At the same time, assumptions of independence and identical distribution in the model ensure the statistical additivity among samples, simplifying the mathematical treatment of the likelihood function and the construction of estimators. Furthermore, the mathematical characterization of the observed data, such as following a normal distribution, an exponential family distribution, or other specific forms, further determines the solvability of the estimation problem and the form of its solution. Common methods like MLE and MoM are all developed based on these modeling assumptions.

2.3. Classification of common methods and core ideas

In the theory and practice of parameter estimation, various estimation methods are often used to address different statistical problems, depending on the model structure, sample size, and availability of prior information. The three approaches (MLE, MoM, and Bayesian estimation) are based on their respective theoretical foundations and inferential mechanisms, making them suitable for different data characteristics and application scenarios.

MLE is one of the most classical and widely used estimation methods. This approach constructs a likelihood function based on the sample data and determines the estimates of unknown parameters by maximizing this function. Under certain regularity conditions, MLE estimators exhibit asymptotic unbiasedness and consistency and achieve the Cramér-Rao lower bound in large samples, thereby attaining asymptotic efficiency [6]. Due to its favorable theoretical properties, MLE is particularly well-suited for classical statistical models with clearly specified parameter forms and sufficiently large sample sizes. In contrast, the MoM derives parameter estimates by matching sample moments with the corresponding theoretical moments of the population. This approach is simple and intuitive, making it particularly suitable when the form of the population distribution is unknown but the low-order moment structure is available. Although its estimation efficiency may be lower than that of MLE in some cases, the MoM shows robustness in practical applications due to its weaker dependence on the model, making it particularly valuable for preliminary estimation or model identification [7]. Bayesian Estimation adopts an inferential paradigm that is fundamentally different from classical methods. In this approach, unknown parameters are treated as random variables, and a posterior distribution is constructed by combining prior distributions with sample information (likelihood function), enabling comprehensive inference about the parameters. Bayesian estimation can operate effectively even with small sample sizes and offers flexibility in handling complex models and uncertain environments, making it particularly suitable for decision analysis in fields where prior knowledge or expert judgment is strong [8]. Its inferential results have a clear probabilistic interpretation, and the method has seen widespread application in cutting-edge areas such as data science and artificial intelligence in recent years.

3. Derivation logic and methodological characteristics of the three estimation approaches

3.1. Construction and properties of Maximum Likelihood Estimation (MLE)

MLE is a central method in modern parameter estimation theory. Its fundamental idea is to construct a likelihood function based on the observed data and obtain estimates of the unknown parameters by maximizing this function. The method has a clear theoretical foundation and exhibits a series of desirable statistical properties in large-sample scenarios, hence making it widely applicable across various statistical modeling and inference contexts.

Formally, let $X_1, X_2, \ldots, X_n$ be independently and identically distributed according to the probability density function $f(x; \theta)$. The corresponding log-likelihood function can then be expressed as Equation (1):

$$l(\theta) = \sum_{i=1}^n \log f(X_i \mid \theta) \tag{1}$$

The MLE corresponds to the parameter values that maximize this function, as shown in Equation (2):

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_\theta l(\theta) \tag{2}$$

The extrema of the log-likelihood function are generally difficult to determine analytically due to its nonlinear nature. In practice, numerical optimization algorithms, such as the Newton-Raphson method or quasi-Newton methods (e.g., the BFGS algorithm), are commonly employed. These methods iteratively utilize gradient information and approximate the Hessian matrix, improving computational efficiency and enhancing the practicality of MLE in high-dimensional and complex models [9]. In terms of theoretical properties, MLE exhibits a series of desirable asymptotic characteristics under large-sample conditions. Specifically, MLE is consistent, meaning that as the sample size approaches infinity, the estimates converge in probability to the true parameter values. Besides, it is asymptotically normal, with its large-sample distribution approximated by a normal distribution and the asymptotic covariance matrix determined by the inverse of the negative information matrix. In addition, MLE attains the Cramér-Rao lower bound, achieving asymptotic optimal information efficiency among unbiased estimators and possessing the property of minimum variance [10].

3.2. Construction and characteristics of the Method of Moments (MoM)

The MoM originates from the moment-matching principle proposed by Karl Pearson in 1894. Its core idea is to approximate and substitute population moments with sample moments. Population moments, as functions of the underlying distribution, uniquely determine the parameters, while sample moments are directly computed from a finite sample. By equating the sample moments to the theoretical moments, a system of moment equations is constructed, allowing the unknown parameters to be estimated indirectly. This method is intuitive and computationally simple, and it is particularly practical when the likelihood function is difficult to construct explicitly or when its computation is costly.

The estimation procedure begins with the specification of the population moments, which are typically expressed as functions of the underlying parameters. The corresponding sample moments are then computed from the data, forming a system of moment equations. Parameter estimates are obtained as the solutions to this system. However, the existence and uniqueness of solutions to the moment equations are not always guaranteed, particularly when the model is complex or the sample size is limited [11]. Thus, the applicability of the MeoM relies in part on the well-posedness of the moment equations and the adequacy of the sample information. The MoM presents significant practical advantages, particularly when the distributional form is known but the likelihood function is analytically intractable or computationally demanding. Compared to MLE, MoM is computationally simpler, bypassing complex numerical optimization procedures, and serves as an effective tool for preliminary parameter estimation or model validation. However, its statistical efficiency is generally lower than that of MLE, with larger variances and reduced precision [12]. Moreover, MoM lacks the asymptotic optimality properties of MLE, which somewhat limits its use in precise inference.

3.3. Inference and computation in bayesian estimation

In Bayesian estimation, unknown parameters are modeled as random variables, and inference is conducted by integrating prior distributions with the likelihood function from the observed data to derive the posterior distribution. The specification of the prior is critical, as it influences both the form of the posterior and the characterization of parameter uncertainty, with different priors potentially yielding distinct inferential outcomes. For instance, the use of conjugate priors allows the posterior distribution to be obtained in an analytical form when combined with the likelihood function, thereby greatly facilitating computation. Noninformative priors, on the other hand, aim to minimize prior influence, rendering the inference closer to a frequentist

approach. In practice, the choice of prior must balance the incorporation of domain knowledge with the avoidance of subjective bias. Computing the posterior distribution is a central component of Bayesian analysis, and with advances in modern computational techniques, Bayesian inference for complex models has become increasingly feasible and efficient, greatly expanding its range of applications.

In practical applications, the posterior distribution is usually not analytically tractable, requiring numerical methods for its approximation. Markov Chain Monte Carlo (MCMC) methods approximate the posterior by constructing a Markov chain whose stationary distribution matches the target posterior, enabling effective sampling. Moreover, Variational Inference (VI) approximates the posterior by optimizing over a family of tractable distributions to find the closest match. These approaches substantially enhance the computational efficiency and practicality of Bayesian estimation in high-dimensional and complex models [13]. In addition, Bayesian methods exhibit clear advantages in small-sample settings, as they can fully incorporate prior information to compensate for data limitations. However, the choice of prior introduces sensitivity: different priors can lead to markedly different posterior inferences. Consequently, prior specification and sensitivity analysis are key to robust Bayesian inference [14]. Effective Bayesian estimation depends on incorporating domain knowledge and data characteristics, and on assessing the plausibility and influence of prior assumptions on the inference.

3.4. Performance comparison of parameter estimation methods

To compare the three main parameter estimation methods, this paper analyzes them from multiple perspectives, including statistical properties such as consistency and asymptotic efficiency, computational complexity, and applicable scenarios. By highlighting the strengths and weaknesses of each method, this multidimensional comparison framework offers a theoretical foundation for choosing suitable estimation techniques according to different data characteristics and model settings. The evaluation of parameter estimation methods, in terms of statistical properties, computational demands, and applicability, provides the basis for comparing the three main approaches, MLE, MoM, and Bayesian Estimation, each demonstrating unique strengths and limitations across different contexts.

From the perspective of theoretical performance, MLE exhibits superior asymptotic properties, including consistency, asymptotic normality, and information efficiency, and can achieve the Cramér-Rao lower bound in the large-sample limit, making it widely regarded as a classical optimal estimator. The MoM estimates parameters by matching sample moments with theoretical moments; although it is consistent, its asymptotic efficiency is generally lower than that of MLE, resulting in a certain disadvantage in statistical efficiency. Bayesian estimation combines prior distributions with sample information to perform posterior inference, making it particularly suitable for small-sample or complex model scenarios. It allows the integration of subjective or objective prior information, but its theoretical properties largely depend on the appropriateness of the chosen prior and the model specification [15].

In terms of computational complexity, MLE typically relies on numerical optimization of the likelihood function, with its computational burden influenced by the parameter dimension and the chosen optimization algorithm, resulting in a moderate overall cost. MoM involves solving linear or nonlinear moment equations, making the computation relatively simple and intuitive. And Bayesian estimation entails a considerably higher computational load, particularly when employing MCMC or VI methods, which significantly increase the computational cost and pose a major practical challenge [16]. With respect to applicability, MLE is suitable for classical statistical models with correctly specified structures and well-defined likelihood functions. MoM is more flexible with respect to distributional assumptions, making it appropriate for complex models or situations where the likelihood is difficult to specify. Bayesian estimation provides the highest flexibility, accommodating complex hierarchical structures and prior uncertainty, and is particularly advantageous in small-sample or non-standard data scenarios [17]. Accordingly, Table 1 presents the main performance metrics and applicability of the three methods.

Table 1. Comparative summary of three parameter estimation methods

Performance Dimension	MLE	MoM	Bayesian Estimation
Consistency	√	√	√
Asymptotic Efficiency	√	○	Depends on prior
Computational Complexity	Moderate	Low	High
Prior Dependence	No	No	Yes
Typical Applications	Large samples, well-specified likelihoods	Complex models, moments available	Small samples, complex models

It can be observed that MLE is generally the preferred method under large-sample conditions with accurately specified models. When the model is complex and the likelihood function is difficult to specify explicitly but moment information is readily available, the Method of Moments and its extension, the Generalized Method of Moments (GMM), demonstrate

advantages. In small-sample scenarios, or when prior information is abundant or the model is highly non-standard, Bayesian estimation provides a more comprehensive characterization of parameter uncertainty. The choice of estimation method should consider sample size, model structure, availability of prior information, and computational resources, with the potential to combine multiple approaches to balance estimation accuracy and computational feasibility.

4. Current challenges and hybrid strategies

4.1. Computational bottlenecks and sampling efficiency

Bayesian parameter estimation relies on numerical approximation of the posterior distribution, with MCMC methods being the most widely used computational tool. However, when the number of parameters increases, the efficiency of MCMC sampling drops sharply, giving rise to the so-called "curse of dimensionality." The chains tend to get trapped in local regions, slowing convergence and thus affecting the accuracy and stability of the estimates. In addition, for high-dimensional models, the posterior distribution frequently displays multimodality and intricate structures, which hinders traditional random-walk and Metropolis-Hastings algorithms from effectively exploring the full parameter space.

To address these challenges, various improved algorithms have been proposed, such as Hamiltonian Monte Carlo (HMC), unbiased estimation sampling methods, and gradient-based variational inference. These approaches enhance sampling efficiency by leveraging gradient information and local geometric structure, yet they still face high computational costs and complex tuning requirements [18]. Thus, improving the efficiency and stability of high-dimensional MCMC remains a key direction for future research. On one hand, it is necessary to develop intelligent sampling algorithms that can adaptively adjust step sizes and proposal distributions. On the other hand, integrating distributed computing and hardware acceleration (e.g., GPU/TPU) holds promise for substantially reducing convergence time and facilitating the widespread application of Bayesian inference in large-scale, complex models.

4.2. Prior selection and information trade-offs

The prior distribution serves as the cornerstone of Bayesian inference, fundamentally influencing both the characteristics of the posterior distribution and the robustness of the resulting estimates. In practice, highly subjective priors may introduce significant human bias, whereas noninformative or weakly informative priors, while reducing subjectivity, can sometimes yield posterior distributions that are unreasonable or difficult to interpret. Recently, the theory of Penalized Complexity (PC) priors has emerged, constructing normalized priors based on the Kullback-Leibler divergence to balance model complexity and goodness-of-fit, offering strong theoretical support and practical effectiveness [19]. Besides, empirical Bayes methods use data-driven approaches to automatically adjust prior parameters, partially mitigating the subjectivity of prior choice, though they can introduce potential overfitting risks. Designing priors that objectively reflect domain knowledge while maintaining adaptability remains a critical challenge in Bayesian theory and applications, particularly in high-dimensional or complex model settings, where this issue profoundly affects the accuracy and generalizability of inference.

4.3. Hybrid strategies and intelligent modeling

Given the limitations of traditional estimation methods in high-dimensional, nonlinear, or non-normal settings, integrating multiple methodologies has become a crucial strategy for enhancing estimation performance. Empirical Bayes methods combine the strengths of the frequentist paradigm by estimating prior distribution parameters from observed data, allowing for semi-adaptive adjustment of prior information. Regularized MLE, incorporating penalty functions like Lasso or Ridge, effectively mitigates overfitting and enhances model generalization. Such hybrid strategies not only improve the robustness of estimation methods but provide a theoretical foundation for handling large-scale, complex datasets [20]. Simultaneously, artificial intelligence techniques, particularly deep learning, have been introduced into parameter estimation, demonstrating strong nonlinear modeling and automatic feature extraction capabilities. Neural network-based variational inference models, along with frameworks like Generative Adversarial Networks (GANs), enable efficient approximation of complex posterior distributions, addressing the computational and expressive limitations of traditional methods [21]. In addition, reinforcement learning-based optimization algorithms can dynamically adjust hyperparameters during estimation, enhancing convergence efficiency and stability. And AI-assisted parameter estimation integrates statistical inference with advanced computational techniques, effectively addressing the challenges posed by high-dimensional and complex data.

5. Conclusion

This study has reviewed the basic theoretical framework of parameter estimation and three major methods, comparing their differences in terms of consistency, efficiency, and computational complexity. The results indicate that MLE demonstrates prominent advantages in asymptotic properties and statistical efficiency, MoM provides practical value via computational simplicity and tolerance for model assumptions, and Bayesian Estimation offers a flexible approach to small-sample and complex modeling problems by integrating prior knowledge with observed data. However, this study has certain limitations, as it is primarily based on a literature review, lacking systematic empirical and simulation-based validation, and it pays limited attention to emerging estimation approaches. Future research may employ large-scale simulations to evaluate the applicability of different methods under complex data environments and further explore their integration with artificial intelligence and hybrid approaches, advancing parameter estimation toward greater efficiency, intelligence, and adaptability.

References

[1] Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer.
[2] Robert, C.P. (2007). *The Bayesian Choice* (2nd ed.). Springer.
[3] Casella, G., & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Duxbury.
[4] Kullback, S., & Le Cam, L. (2022). Fisher information and the Cramér-Rao inequality: A modern review. *Annual Review of Statistics and Its Application, 9*, 1-26.
[5] Efron, B., & Hastie, T. (2016). *Computer Age Statistical Inference.* Cambridge University Press.
[6] Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*(1), 90-100.
[7] Ackerberg, D. A., Geweke, J., Hahn, J., & Liao, Z. (2022). Generalized method of moments: A gentle introduction and recent developments. *Annual Review of Economics, 14*, 1-32.
[8] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.
[9] Nocedal, J., & Wright, S. (2006). N*umerical Optimization* (2nd ed.). Springer.
[10] Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.
[11] Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.
[12] Cattaneo, M. D., & Jansson, M. (2023). Large sample estimation and inference in econometrics: A modern perspective. *Journal of Economic Literature, 61*(2), 345–389.
[13] Brooks, S., Gelman, A., Jones, G., & Meng, X. L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press.
[14] Berger, J. O. (2013). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). Springer.
[15] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association, 112*(518), 859-877.
[16] Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations.* CRC Press.
[17] Murphy, K. P. (2022). *Probabilistic Machine Learning: An Introduction*. MIT Press.
[18] Betancourt, M. (2017). *A conceptual introduction to Hamiltonian Monte Carlo*. arXiv preprint arXiv: 1701.02434.
[19] Simpson, D. (2017). Penalising model complexity. *Statistical Science, 32*(1), 1-28.
[20] Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413-1432.
[21] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.