

Research and Analysis of MLE Based on Maximum Likelihood Estimation

Anxing Xu

Wuxi No.1 High School, Wuxi, China
8177005009@jiangnan.edu.cn

Abstract: This paper presents an in-depth exploration of Maximum Likelihood Estimation (MLE), a paramount method in the domains of statistics and machine learning for parameter estimation. MLE's theoretical underpinnings are discussed, detailing both likelihood and log-likelihood functions and their integral mathematical properties. The estimation process is outlined, emphasizing MLE's applications across various fields such as biostatistics and machine learning through concrete examples. Noteworthy is MLE's dual role in both parameter estimation and model validation, supported by its consistency and asymptotic normality which enhances its applicability in analytical scenarios. However, the study also addresses MLE's limitations, particularly the challenges posed by computational complexity and small sample sizes. Solutions to mitigate these issues are proposed, enhancing MLE's robustness and utility in practical applications. This comprehensive analysis not only clarifies MLE's operational mechanics but also underscores its versatility and pivotal role in advancing empirical research across multiple scientific disciplines.

Keywords: Maximum likelihood estimation, Parameter estimation, Machine learning, Biostatistics

1. Introduction

Maximum likelihood Estimation (MLE) is a method of parameter estimation based on probabilistic model [1]. Through maximizing likelihood function to get the model parameter, the observed data can be maximized under the given parameter. MLE has a solid theoretical foundation and excellent statistical properties, such as consistency and asymptotic normality, and is widely applied in fields such as statistics, machine learning, and biostatistics [2-3].

There are two general methods for parameter estimation. They are Least Squares Estimation (LSE) and Maximum Likelihood Estimation (MLE). Unlike MLE, LSE does not require or requires only minimal distribution assumptions, which can be used to obtain descriptive measures to summarize the observed data [4]. At the same time, MLE is still not fully accepted or recognized, that is, there is a lack of a complete and clear understanding of the application and analysis of maximum likelihood estimation in the academic community, which will be answered in this article. By deepening the understanding of the application of MLE, it can be better utilized in various fields, including statistics, machine learning, biostatistics, economics, and engineering. Whether in parameter estimation, model selection, or hypothesis testing, MLE is a commonly used and effective method. MLE has several advantages: first, its computational feasibility; although the computation of MLE may be complex,

the development of modern computational techniques and optimization algorithms (such as gradient descent, EM algorithm, etc.) has made MLE feasible in practical applications [5].

Additionally, the use of the log-likelihood function simplifies the computation process [6]. Second, it possesses excellent statistical properties; under certain conditions, MLE estimators have consistency, asymptotic normality, and efficiency. This means that as the sample size increases, MLE estimators will converge to the true parameter value and have the smallest variance. Finally, MLE has achieved significant success in many practical applications [7-8]. For example, in machine learning, MLE is used to train logistic regression, Gaussian mixture models, etc.; in biostatistics, MLE is used for gene expression data analysis, epidemiological models, etc. This further demonstrates the feasibility of MLE.

The article first introduces the theoretical foundation of maximum likelihood estimation, including the definitions of the likelihood function and the log-likelihood function, as well as their mathematical properties. It then elaborates on the estimation process of maximum likelihood estimation in detail and demonstrates its applications in statistics, machine learning, and biostatistics through specific examples. Finally, the article discusses the limitations of maximum likelihood estimation, particularly regarding computational complexity and small sample issues, and proposes possible solutions.

2. Theoretical Foundation

In this section, the Maximum likelihood estimation and the basics of mathematics (likelihood functions, log-likelihood functions, estimation processes) will be introduced.

2.1. Principle of Maximum Likelihood Estimation

The maximum likelihood estimation is a probability-based statistical method used to estimate model parameters from observed data. The core idea is to find a set of parameter values that maximizes the probability of the observed data occurring under those parameters. Assume that there exists a probability model whose distribution is determined by the parameter θ (θ can be a scalar or a vector). Given a set of observed data $X=(x_1, x_2, \dots, x_n)$ and a parameterized probability model $P(X|\theta)$, the objective of MLE is to find a value of the parameter θ that maximizes the probability of this data occurring under that parameter.

2.2. Mathematical Foundation

2.2.1. Likelihood function

The likelihood function $L(\theta, X)$ represents the probability of observing the data X given the parameter θ . Let $X = [X_1, \dots, X_n]^T$ be a random vector drawn from a joint PDF $f_X(X, \theta)$, and let $x = [x_1, \dots, x_n]^T$ be the realizations. The likelihood function is a function of the parameter θ given the realizations x :

$$L(\theta, X) = f_X(X, \theta) \quad (1)$$

$L(\theta, X)$ can be viewed as a function of θ . This function changes its shape according to the observed data x .

2.2.2. Log-likelihood Function

In practical operations, the likelihood function is often processed by taking the logarithm before further calculations are performed. Define the resulting function as the Log-likelihood Function [9].

$$\log L(\theta, X) = \log f_X(X, \theta) = \sum_{n=1}^N f_{X_n}(x_n; \theta) \quad (2)$$

The log-likelihood function, without altering the properties of the function, is superior to the likelihood function in terms of computation, mathematical properties, and theoretical analysis. Firstly, it simplifies calculations; the likelihood function is often expressed as a product of probabilities, which transforms into a summation upon taking the logarithm, thereby reducing computational difficulty. The product of probabilities may lead to extremely small values, while the logarithmic transformation compresses the range of values, minimizing numerical issues in calculations, resulting in more stable values after taking the logarithm. Secondly, its mathematical properties are more favorable; the derivative form of the logarithmic function is simple, facilitating differentiation and optimization. Thirdly, it aids in theoretical analysis; the log-likelihood function possesses good properties under large samples, making it easier to derive the asymptotic distribution of estimators. Additionally, its Hessian matrix is related to the information matrix, which can be used to calculate standard errors and confidence intervals. Therefore, the logarithmic transformation is necessary.

2.2.3. Estimation Process

In order to obtain the parameters that best meet the requirements, the most commonly used method is to treat the likelihood function as a function of the parameter θ and to solve for its maximum value [10]. The parameter value obtained at this time is the one that maximizes the probability of the observed data. The specific steps are as follows:

First, derive the likelihood function of the random variable X and the parameter θ .

$$L(\theta, X) = f_X(X, \theta) \quad (3)$$

Then take the logarithm on both sides simultaneously and simplify.

$$\log L(\theta, X) = \log f_X(X, \theta) = \sum_{n=1}^N \log f_{X_n}(x_n; \theta) \quad (4)$$

Differentiate with respect to θ and set the derivative to zero, solving for θ .

$$\hat{\theta} = \operatorname{argmax}_{\theta} L(\theta, X) \quad (5)$$

This calculation method can solve most parameter estimation problems. By substituting the obtained parameters into the original probability equation, a better description of the model's probability can be made, allowing it to be applied in various fields.

3. Analysis and Applications

3.1. Parameter Estimation in Statistics

The maximum likelihood estimation is one of the most commonly used parameter estimation methods in statistics. Its core idea is to find the model parameters that are most likely to generate the observed data by maximizing the likelihood function of the observed data. The application of maximum likelihood estimation in statistics is very extensive, especially in the fields of parameter estimation, hypothesis testing, and model selection. The following will take the normal distribution as an example.

Assume there exists a set of data $X=(x_1, x_2, \dots, x_n)$, and that this data follows a normal distribution $N(\mu, \sigma^2)$. The purpose is to estimate the mean μ and variance σ^2 .

According to the formula of the normal distribution, its likelihood function can be derived.

$$L(\mu, \sigma^2; X) = \prod_{i=1}^n f(x_i | \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i - \mu)^2}{2\sigma^2}} \quad (6)$$

After taking the logarithm, the likelihood function becomes like this:

$$\log L(\mu, \sigma^2; X) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \quad (7)$$

First, take the partial derivative of μ and set it to zero to solve.

$$\frac{\partial L}{\partial \mu} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \quad (8)$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (9)$$

Take the partial derivative of σ^2 then, set it to zero, and substitute the value of μ .

$$\frac{\partial L}{\partial \sigma^2} \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \right] = 0 \quad (10)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (11)$$

It can be seen that the value of parameters can be resolved in statistics through likelihood estimation. This enables the model to better fit and interpret observational data [11]. These parameter values are not only used for model fitting and prediction but also support statistical inference, such as hypothesis testing and the construction of confidence intervals.

Furthermore, MLE estimators possess good statistical properties, such as consistency and asymptotic normality, ensuring their reliability in large sample situations.

3.2. Parameter Estimation in Statistics

3.2.1. Model Training

In machine learning, maximum likelihood estimation is commonly used for model training. Specifically, given a probabilistic model $P(X|\theta)$, maximum likelihood estimation estimates the model parameters θ by maximizing the likelihood function of the observed data. For instance, in supervised learning, given input data X and labels Y , maximum likelihood estimation can be used to estimate the conditional probability $P(Y|X, \theta)$.

3.2.2. Maximum a posteriori estimation

In machine learning, the combination of Bayesian estimation and maximum likelihood estimation provides a powerful framework for better addressing issues such as parameter estimation, model selection, and uncertainty quantification. Maximum a posteriori estimation is the most direct form of the combination of Bayesian estimation and MLE. It introduces prior distributions based on MLE, thereby integrating prior knowledge and observational data in parameter estimation. It is used in machine learning to introduce regularization, preventing overfitting, and performs exceptionally well, especially in small sample situations. MAP also offers a natural framework for combining prior knowledge with observational data, enhancing model robustness, and is widely applied in tasks such as regression, classification, and Bayesian neural networks [12].

3.3. Application in Biostatistics

Maximum likelihood estimation (MLE) is widely used in biostatistics for parameter estimation and model fitting, which maximizes the likelihood function to determine model parameters so that the model can best interpret experimental data. For example, in the HIV-1 viral dynamics model, MLE is used to estimate parameters such as infection rate, clearance, etc., to help understand the dynamics

of viral load. In addition, MLE is used to guide experimental design by quantifying the impact of experimental data on parameter estimation and identifying which experimental measurements are most useful for determining the best fit parameters, thereby reducing parameter uncertainty. In dynamic biological systems, MLE is able to process experimental data over time and update model parameters in real time, such as during the COVID-19 pandemic, to recalibrate models to predict trends. MLE also combines sensitivity analysis and contour likelihood function to evaluate the sensitivity of model parameters to the data, and solve the problem of parameter uncertainty and non-identifiability. These applications make MLE an important tool in biostatistics and mathematical modeling [13].

4. Disadvantages

4.1. Computational Complexity

In the high-dimensional parameter space, the optimization problem of likelihood functions can become very complex and usually requires more computational resources and time. For some complex models, the likelihood function may be non-convex, meaning that there may be multiple local maxima. In this case, it can be very difficult to find the global maximum, and the optimization algorithm may get stuck in the local optimal solution. In real-world calculations, likelihood functions can involve a large number of products or exponential operations, which can lead to numerical instability or computational overflow problems. For example, when calculating the likelihood function of a high-dimensional Gaussian distribution, the inverse matrix calculation of the covariance matrix can be time-consuming.

4.2. Issues with Small Sample Sizes

When the sample size is small, the maximum likelihood estimates may deviate from the true parameter values, leading to estimation bias. In small sample situations, the model may overfit the training data, resulting in decreased generalization performance. Maximum likelihood estimation tends to select parameter values that best fit the training data, but this may lead to poor performance of the model on unseen data. In small sample cases, the variance of the estimates may increase, resulting in unstable estimation results. This means that the maximum likelihood estimates may exhibit significant fluctuations across different sample sets.

5. Conclusion

The maximum likelihood estimation, as a classic parameter estimation method, has a wide range of applications in fields such as statistics, machine learning, and biostatistics. By analyzing sample data, MLE can effectively estimate model parameters, thereby providing a solid foundation for subsequent data analysis and model construction. This article summarizes its importance in modern data analysis through a discussion of the theoretical basis, application scenarios, and its advantages and disadvantages of MLE.

First, the core idea of MLE is to estimate parameters by maximizing the likelihood function, which has a solid mathematical foundation and can provide consistent and efficient estimates in large sample situations. In statistics, MLE is widely used in fields such as regression analysis and time series analysis; in machine learning, MLE provides theoretical support for parameter estimation in many models (such as logistic regression, Gaussian mixture models, etc.); in biostatistics, MLE helps researchers extract useful information from complex biological data.

MLE also has some limitations. First, its computational complexity is relatively high, especially in high-dimensional data or complex models, where the computation process can be very time-

consuming. Second, MLE performs poorly in small sample situations, making it prone to overfitting issues. Therefore, in practical applications, researchers need to choose appropriate estimation methods based on specific problems or combine other techniques (such as regularization) to enhance estimation effectiveness.

With the rapid development of data science and artificial intelligence, Maximum Likelihood Estimation (MLE) will continue to play an important role in the field of parameter estimation. Future research can further explore the combination of MLE with other optimization techniques to enhance its computational efficiency in large-scale and high-dimensional data. Additionally, regarding small sample issues, research can investigate how to improve the robustness of MLE by introducing prior knowledge or refining estimation methods. In summary, as a classic estimation method, MLE still has vast potential for theoretical research and practical application.

References

- [1] Mahgoub A. Salih, Taysir M. Elmahdi. (2024) *Exploring the Implications of the Deformation Parameter and Minimal Length in the Generalized Uncertainty Principle*. *Journal of Quantum Information Science*, 14(1), 1-14.
- [2] Seifert J, Shao Y, van Dam R, Bouchet D, van Leeuwen T, Mosk AP. (2023) *Maximum-likelihood estimation in ptychography in the presence of Poisson-Gaussian noise statistics: publisher's note*. *Opt Lett*, 48(23), 6291. doi: 10.1364/OL.513661. PMID: 38039249.
- [3] Zhu, X., Liu, Z., Cambria, E., Yu, X., Fan, X., Chen, H., & Wang, R. (2025). *A client-server based recognition system: Non-contact single/multiple emotional and behavioral state assessment methods*. *Computer Methods and Programs in Biomedicine*, 260, 108564.
- [4] Wang, R., Zhu, J., Wang, S., Wang, T., Huang, J., & Zhu, X. (2024). *Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking*. *International Journal of Multimedia Information Retrieval*, 13(4), 39.
- [5] Wang, F., Ju, M., Zhu, X., Zhu, Q., Wang, H., Qian, C., & Wang, R. (2025). *A Geometric algebra-enhanced network for skin lesion detection with diagnostic prior*. *The Journal of Supercomputing*, 81(1), 1-24.
- [6] Zhao, Z., Zhu, X., Wei, X., Wang, X., & Zuo, J. (2021, June). *Application of Workflow Technology in the Integrated Management Platform of Smart Park*. In *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (Vol. 4, pp. 1433-1437)*. IEEE.
- [7] Zhang, Y., Zhao, H., Zhu, X., Zhao, Z., & Zuo, J. (2019, October). *Strain Measurement Quantization Technology based on DAS System*. In *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (pp. 214-218)*. IEEE.
- [8] Paparas Alex, Fotopoulos Stergios B., Jandhyala Venkata K., Paparas Dimitris. (2023) *Maximum likelihood estimation of a change point for Poisson distributed data*. *Model Assisted Statistics and Applications*, 4, 347-358.
- [9] Ren, L., Rui, Z., & Lei, C. (2014). *Immune Clone Maximum Likelihood Estimation of Improved Non-homogeneous Poisson Process Model Parameters*. *Journal of Donghua University (English Edition)*, 31(6), 801-804.
- [10] Wei, J., Luo, Y., Di, W., Lan, H., & Cao, H. (2024). *Sparse Representation Method under Mixed Gaussian Noise Conditions and Its Application in Impact Fault Feature Extraction*. *Mechanical Science and Technology*, 43(6), 917-924.
- [11] Seifert J, Shao Y, van Dam R, Bouchet D, van Leeuwen T, Mosk AP. (2023) *Maximum-likelihood estimation in ptychography in the presence of Poisson-Gaussian noise statistics: publisher's note*. *Opt Lett*, 48(23), 6291. doi: 10.1364/OL.513661. PMID: 38039249.
- [12] Squartini, T., & Garlaschelli, D. (2011). *Analytical maximum-likelihood method to detect patterns in real networks*. *New Journal of Physics*, 13(8), 083001.
- [13] Yang, Y., & Ma, C. (2024). *Random pairing MLE for estimation of item parameters in Rasch model*. *arXiv preprint arXiv:2406.13989*.