

The State-Of-Art Applications for Maximum Likelihood Theory

Maohua Zhou *

University of Science and Technology Beijing Beijing, China

* Corresponding Author Email: 41921172@xs.ustb.edu.cn

Abstract. Maximum Likelihood theory is widely applied in the scientific research. Generally, it can send the message that the relevant parameters of the population distribution can be estimated by the corresponding samples and provide the ideology to obtain the value of the estimators. This paper aims at demonstrating the development of the ML theory in order to have a relatively clear cognition about the theory. The paper generalizes the principle of the statistics theory and reviews several theoretical and empirical improvement and application based on the ML theory, including distribution parameter estimation in statistics, uncertainty regression in econometrics, and the assimilation of the precipitation radar observation, forming an outline about the current advance of the theory and algorithms and having some outlooks about the future development orientations in potential. In addition, some instructions are presented to approach ML theory and do further research as the basis. These results shed light on guiding further exploration of ML theory.

Keywords: Maximum Likelihood, Estimation Experiment, Uncertainty Regression, Data Assimilation.

1. Introduction

Contemporarily, statistic methods are getting increasingly significant in many scientific research fields. When it comes to statistics, sampling and estimation are obviously the inevitable parts [1]. For statistics requirement, Maximum Likelihood theory can be a prominent guiding ideology in estimation. It is clear that MLE is powerful to estimate the required parameters with a great dataset and give a prompt to utilize the data effectively. Once the distribution of the random variable is given, it is easy to deal with it relatively and work out the results by fundamental knowledge of mathematics analysis [2]. With ML theory, it is able to confirm the orientation of sampling and calculating process. Besides, there are complete proofs and precise methods about the consequences of MLE and its estimating error, hence the veracity and accuracy can be assured. So far, ML theory has been devoted into varieties of disciplines, such as statistics, econometrics, engineering and geography [3]. Frontier functions, radar observation about raining amount are all closely related to ML theory. These empirical applications imply that ML theory can do benefits on the exploration about the specific and detailed distribution [4].

As a matter of fact, ML theory has seen a long-term development and progress. It was proposed first by Gauss in his discussion about the estimation error in 1821, but the achievement usually owes to Fisher in England for the theory was declared relatively completely and had this name [5]. Afterwards, Wilks proved that the MLE Equation for the logarithms is of chi-square distribution in 1938, providing a theoretical basis for confidence interval estimation of maximum likelihood parameter estimation [6]. Furthermore, with the advance of ML theory, some algorithms based on the idea extended MLE. 39 years later, the Expectation Maximization algorithm was explained and named by Dempster, Laird, and Rubin in 1977 [7], but the idea of this algorithm had already existed, and it is applicable to the case of missing data or the introduction of "potential data" to simplify the maximum likelihood estimation of mixed normal distributions. Owing to the simplicity of the Expectation step and Maximization step, there is lower computational complexity for EM algorithm. Moreover, Hunt and Lange summarize and present the framework of MM algorithm including MLE as a special circumstance in 2000, which make the algorithm based on MLE more complete. Besides, modern estimation methods (e.g., pseudo-likelihood) estimate is also valuable in statistics. It guarantees that when the likelihood equation is not accurately described, the parameter estimation is still consistent and asymptotically normal, but certainly its accuracy would decrease. However, due to this characteristic of the algorithm, complex likelihood equation can be replaced by simplified

equation to reduce the computational complexity. At present, the development of maximum likelihood estimation has been very mature after more than 100 years, and most of the recent research mainly focuses on improving the calculating speed of the algorithm.

The primary objective of this paper is going to review the previous achievement and have a future look at the potential advance of the ML theory. This separation part of the paper is organized as follows. In Section II, the fundamental principle and the theoretical analysis will be restated. Then Section III will discuss several classic examples about the relevant applications, showing the developing consequences of the ML theory. Section IV presents the advantages and limitation of the ML theory and makes some outlooks about the theory advance. Finally, a brief summary is given in Sec. V.

2. The Basic Principle of ML Theory

Setting the population distribution is $f(z_1; \theta)$, $\{z_i\}_{i=1}^n$ is the set of the selected samples from the population, then one obtains the joint distribution of $\{z_i\}_{i=1}^n$:

$$L(z_1, z_2, \dots, z_n; \theta) = f(z_1; \theta)f(z_2; \theta) \cdots f(z_n; \theta) \quad (1)$$

When θ is fixed, L is the function of the set $\{z_i\}_{i=1}^n$, L is a possibility density function. If $\{z_i\}_{i=1}^n$ is fixed, L is a function of θ , since θ has a certain value but is not known as a random variable, it is not called probability, but called likelihood. The point maximizing the likelihood is:

$$\theta^* = \arg \max L(z_1, z_2, \dots, z_n; \theta) \quad (2)$$

and this value is the estimated value of the parameter θ . With the collected samples, θ^* is called the maximum likelihood estimation of θ . To make the process of maximizing simple, it is typical to take the logarithm of L , so

$$\log L = \sum_{i=1}^n f(z_i; \theta) \quad (3)$$

Subsequently, the consistent derivations of the function make it easy to obtain the value of θ^* .

3. Applications

3.1. Classical Application in MLE

Since ML theory plays an important role in the estimation, it is not rare to see the ideology applied in the research on statistics, econometrics, geography and many other scientific fields.

In statistics, the MLE to estimate the parameter for a given distribution is very familiar. Here, an investigation is carried out on the students' learning time on a daily basis in the university. Given 7 hours for students to arrange freely, how much time would students like to spend on study is an interesting question. Therefore, the author wants to have a look at the random variable t , the students' time for study in one day. The investigation gave students a monotonic number series to choose one as x : 0.5, 1.0, 1.5, ..., 6.0, and each x stands for the interval $x - 0.5 < t \leq x$. The data was collected mainly from two majors in the university, Mathematics and Automations. Finally, there are 495 pieces of feedback and the raw data is as shown in Table I.

Table. 1 The raw data of the time for learning

Time Range (Hours)	The Number of Students
0.5	10
1.0	16
1.5	32
2.0	51
2.5	68
3.0	80
3.5	70
4.0	55
4.5	42
5.0	28
5.5	23
6.0	20

Table. 2 Serial Number of Students i and Time t_i (Hours)

Serial Number of Students i	Time t_i (Hours)
1	0.05
2	0.13
3	0.15
4	0.02
5	0.36
6	0.26
7	0.06
8	0.05
9	0.37
10	0.08

The basic assumption of the experiment is that in every interval, t is uniform distribution so it is reasonable to give a number of uniform pseudo-random number as the samples of the random variable according to how many samples are in every interval. The sectional processed samples are shown in Table II.

With the sampling values of t , the distribution of the random variable should be tested. Based on Q-Q Testing, it is intuitionistic that the Q-Q figure is roughly an asymptotic straight line, so t can be regarded as normal distribution [8]. Based on the knowledge about the normal distribution, it is clear that the possibility density function $f(t; \mu, \sigma)$ of t is

$$f(t; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \quad (4)$$

In the equation, μ is the expectation of the distribution, and σ is the standard derivation. The maximum likelihood function l can be written as:

$$l = \prod_{i=1}^n f(t_i; \mu, \sigma) \quad (5)$$

To maximize the function, it is necessary to pay attention to the partial derivations respectively on the parameter μ and σ . Let the one-order partial derivations be 0:

$$\begin{cases} \frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (t_i - \mu) = 0; \\ \frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (t_i - \mu)^2 = 0; \end{cases} \quad (6)$$

Besides, it comes to results that the expectation of the distribution is the average value of samples, and the standard derivation of it is that of samples. Furthermore, it should not be ignored that the values of two-order partial derivations when the parameters take the value above, for they can be nonnegative. Fortunately, the results come out that both of them are strictly negative, assuring that the values of the parameters are the maximum. By MATLAB, the computing results of the two parameters are respectively about 3.02h and 1.31h. Then, the specific distribution of the random variable t can be confirmed, and it gets easier to learn about the situation about student management for study. In the case above, it is obvious that ML theory provides a reasonable methodology to do the parameter estimation.

3.2. The Modification in Uncertain Regression Analysis

ML Theory is certain to play a significant role in the regression analysis in econometrics, which providing a way to get each estimated values of regression coefficients. The method is based on the assumption that the disturbance is normal distribution with the parameter expectation and standard derivation which take constant 0 and σ as their values respectively. Therefore, on account of the fact that the part of explanatory variables is observable and decided, the distribution of the explained variable as well as the possibility density function can be obtained. The remaining work is to construct the maximum likelihood function, analysis the maximum condition to obtain the relationship between the coefficients and the samples, and finally calculate them. However, in uncertain regression, the samples are not always precise, and the distribution may be unobservable, which leads to the difficulty of constructing maximum likelihood function. In this case, the traditional ML theory may not be proper or effective. Thus, it is necessary to introduce measure and the limitation into the model, i.e., it can be simplified to obtain the likelihood function. Attributed to the fact that an estimated parameter with high likelihood is usually considered, with definition of derivation and the subadditivity of the uncertain measure, the infimum is proper to measure the likelihood and define a likelihood function with respect to a set of unknown parameters θ .

Moreover, according to Baoding Liu's research, it can be proved that the likelihood function of the uncertainty distribution is uniformly continuously differentiable, although the sample data is imprecise and the distribution is uncertain [9]. Therefore, the method pays attention to minimizing the absolute residual and maximize the likelihood function. With the sampling data of the random variable, the coefficients can be calculated.

3.3. MLE in Precipitation Radar Data Assimilation

ML theory is widely applied in diverse scientific fields, which can even be met in the precipitation. The estimate is formulated based on the MLE, called 1D-MLE. In Ikuta's research [10], they consider the case when typhoon GONI passed Japan. Figure 1 shows the RH retrieved using 1D-MLE from GPM/DPR around GONI from 23 to 26 August 2015. Figure 2 illustrates that many observations were assimilated between 23 and 25 August as depicted in the coverage map. The average ETS difference, and shaded area indicates the 95% confidence interval.

The 1D-MLE was developed for this indirect assimilation method and confirmed that the 1D-MLE performs correctly if 1D-Bay underestimates RH near supersaturation and overestimate RH in the case of bimodal-distribution. Additionally, on account of the fact that the 1D-MLE method is not activated without background precipitation, the chances of DPR assimilation are restricted. Besides, the RH is also estimated by creating a PDF from a pseudo-ensemble made by the output of the forecast model. According to the analysis, the indirect assimilation using 1D-MLE is a very useful method that can use reflectivity observations. When the error of the forecast model is large, the model is unable to accept large changes imposed by the assimilation, which will disrupt the atmospheric balance.

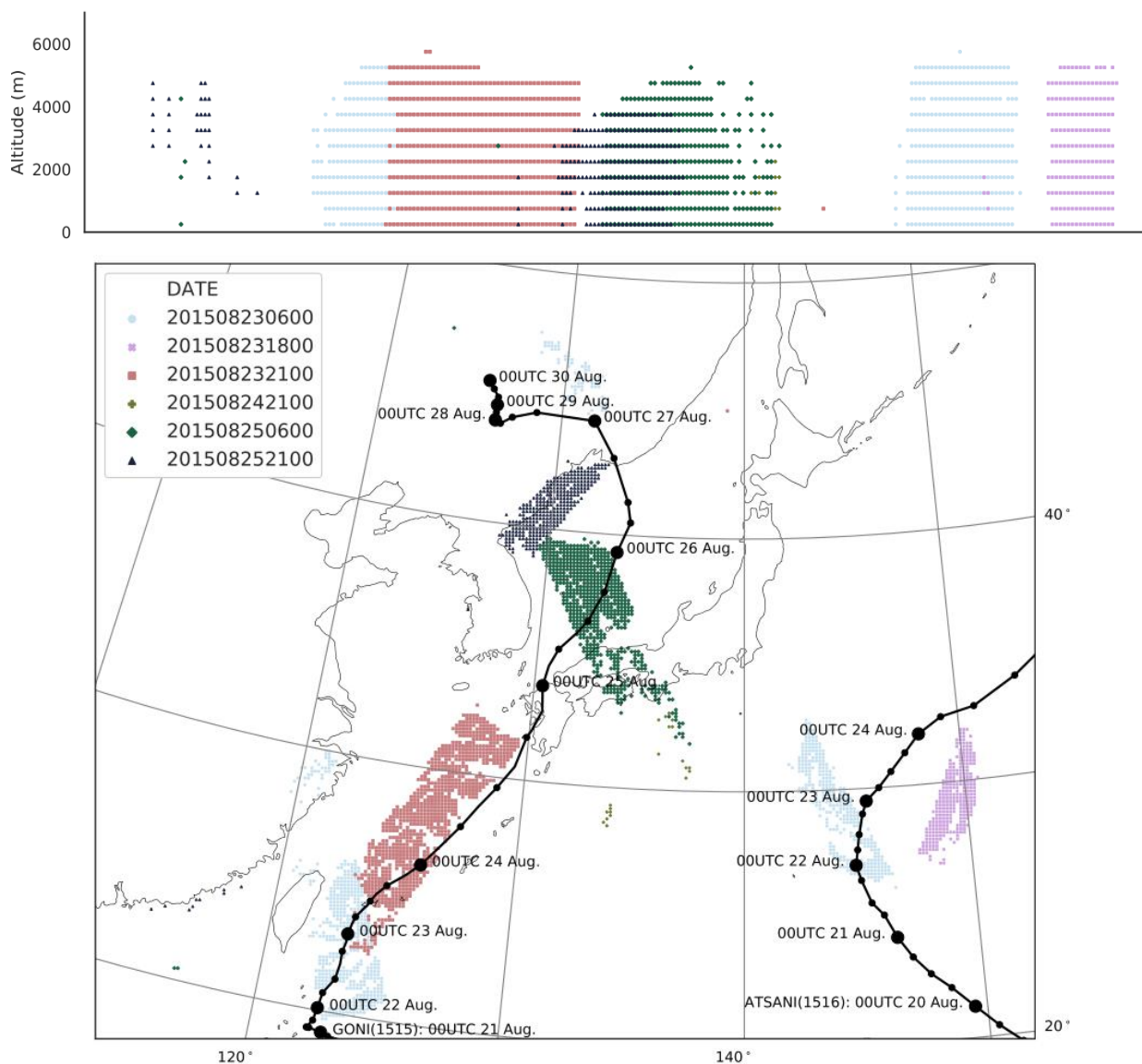


Figure. 1 Pseudo-observing map Coverage map in various colors corresponding to beginning time prediction [10].

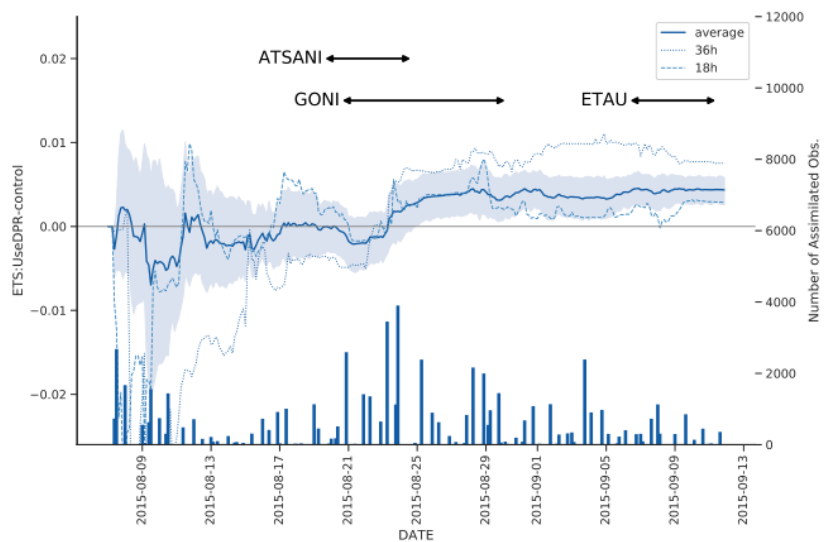


Figure. 2 The time series of ETS [10].

4. Limitations & Future outlook

From the diverse applications of the ML theory illustrated above, it is widely believed that ML theory makes great sense for the estimation work. It reveals the relation between the unknown parameters and the collected samples, bonding between the distribution and the specific data. However, ML theory is only capable on some distribution. In other word, it was based on the distribution of the estimated random variable and may be unable to deal with some more complex distributions. Besides, the existence of error when utilizing ML theory is certain. In the case of great dataset, the number of the data is big enough to assure the confidence interval can be small, which means the estimated values of the parameters can be very precise, owing to the Central Limitation Theory. Conversely, if the dataset is very limited, the estimating error can be hard to ignore, leading to the difficulties in the factual decisions and measures.

Therefore, in the future, there are many orientations to do the further research. Taking a look at the 1D-MLE, the dimension of the method will probably be extended to the higher level in order to satisfy more complex climate and topography measurement requirement. It is certain to claim that ML theory itself has completely developed for nearly a century. Nevertheless, combined with the modern computer technology, the derived algorithms are still to improve, of which the calculating speed and the space complexity can be in consideration.

5. Conclusion

In summary, this paper discusses the wide application in ML theory from the perspectives of classical estimation, regression with uncertainty and the data assimilation. ML theory provides a relatively scientific kind of ideology for estimation, linking sampling consequences to the general distribution for an explained random variable. It makes the relationship between the collected data and theoretical distribution more clear and uncovers the principle of sampling and estimation by maximizing the possibility of the event that the samples being selected. Because of its powerful function of estimation, it can be utilized in various fields, promoting the advance of statistics, modifying the regression models in econometrics, and instructing the assimilation work on the precipitation observation, etc. However, the ML theory is limited by the scale of the datasets and its extension on the uncertainty statistics are not comprehensive yet. Therefore, the further research can focus on the improvement of the relevant algorithms and the wider application on the equations, time series and etc. The paper has summarized about the ideology of the theory and gives a clear summary of the recent academic achievement about it, discussing the potential orientations of the ML future research. Overall, these results offer a guideline for the further innovational orientations about the ML theory, including the breakthrough in uncertainty and modern algorithms in geographic measurement and forecast.

References

- [1] J. Orear, Notes on Statistics for Physicists, University of California rep. UCRL-8417, 1958.
- [2] A. Hald, "On the history of maximum likelihood in relation to inverse probability and least squares." Statistical Science vol. 14 (2), 1999, pp. 214–222.
- [3] N. Richard, "BLUE: Combining correlated estimates of physics observables within ROOT using the Best Linear Unbiased Estimate method." SoftwareX Volume 11, January–June 2020, 100468.
- [4] D. Witten, et al, An Introduction to Statistical Learning with Applications in R, vol. 133, 2017.
- [5] Fisher, Ronald A. "On the mathematical foundations of theoretical statistics." Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character vol. 222.594-604, 1922, pp. 309-368.
- [6] R. Barlow, "Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment" Volume 297, Issue 3, 1990, 501.

- [7] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 1977.
- [8] G. Bohm, and G. Zech. *Introduction to statistics and data analysis for physicists*. Vol. 1. Hamburg: Desy, 2010.
- [9] W. Lio, and B. Liu. "Uncertain maximum likelihood estimation with application to uncertain regression analysis." *Soft Computing* vol. 24.13 2020, pp. 9351-9360.
- [10] Y. Ikuta, K. Okamoto, and T. Kubota. "One-dimensional maximum-likelihood estimation for spaceborne precipitation radar data assimilation." *Quarterly Journal of the Royal Meteorological Society* vol. 147.735, 2021, pp. 858-875.