International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 8, August - 2023, Pages: 20-23

Review of Maximum Likelihood Estimation Method

Nahla Hadi Abdul Sahib

Department of Automobiles Engineering/College of Engineering Al-Musayab/ University of Babylon, Iraq
nahla.hadiabdul@uobabylon.edu.iq

Abstract- *This article is review about maximum likelihood estimation method. We study some its characteristics.*

Keywords- Likelihood Function, Efficiency, Consistency, Normal Distribution.

1. INTRODUTION AND PRIMILINARIES

Definition: The method of estimating the maximum probability is one of the most common methods for estimating statistical parameters, as the method was proposed by the mathematician Fisher as he derived the optimal properties of the estimate.

1.1 Technique the Method

1. Likelihood Function: represents the probability of observing the given data as a function of the model parameters. The likelihood function is denoted by L(θ|X), where X represents the observed data and θ represents the parameter vector. It is derived from the probability distribution function (pdf) or probability mass function (pmf) of the statistical model.
2. Log-Likelihood Function: To simplify computations, the Log- L(θ|X) is often used instead of the likelihood function. It denoted by ℓ(θ|X), is the normal logarithm of the likelihood function. Taking logarithm does not alter the location of the maximum, but it simplifies the calculations by converting products into sums.
3. The Max. Likelihood Estimation: The MLE estimator seeks to find the parameter values that maximize the likelihood (or equivalently, the Log- L(θ/X)). Mathematically, it can be express as follows:
$$\hat{\theta} = argmax\ \theta[l(\theta/X)]$$
The symbol $\hat{\theta}$ represents the MLE estimate of the parameter vector θ that maximizes the likelihood function given the observed data X.
4. Optimization: Finding the maximum likelihood estimate involves solving an optimization problem. Depending on the complexity of the likelihood function and the model, analytical solutions may not be available, requiring the use of numerical optimization algorithms. Common optimization methods include gradient descent, Newton-Raphson, and the Expectation-Maximization (EM) algorithm.

1.2 Properties of MLE Estimators

1. Under certain regularity conditions, the MLE estimators possess desirable properties, including consistency, asymptotic normality, and efficiency. Consistency implies that as the sample size increases, the MLE estimator converges to the true parameter value. Asymptotic normality means that the MLE estimator follows a normal distribution with mean equal to the true parameter value and a variance that approaches zero as the sample size increases. Efficiency indicates that the MLE estimator has the smallest asymptotic variance among all unbiased estimators.
2. Inference and Hypothesis Testing: MLE estimation allows for statistical inference and hypothesis testing. Once the MLE estimates are obtained, confidence intervals and hypothesis tests can be constructed based on the asymptotic properties of the estimators.
3. Extensions and Applications: The method of maximum likelihood can be applied to a wide range of statistical models, including linear regression, logistic regression, survival analysis, mixture models, and more. It provides a flexible framework for estimating parameters in both simple and complex models.

2. Compare the Maximum Likelihood Method to Other Methods of Estimation

The method of maximum likelihood (MLE) estimation possesses several distinct features and advantages when compared to other methods of estimation. Here are some key features of the maximum likelihood method:

2.1 Efficiency: MLE estimators have desirable properties in terms of efficiency. Under certain regularity conditions, the MLE estimator is asymptotically efficient, meaning that it has the smallest asymptotic variance among all unbiased estimators. This efficiency property makes MLE estimation particularly attractive when aiming to obtain highly precise parameter estimates.

2.2 Asymptotic Normality: MLE estimators exhibit asymptotic normality. As the sample size increases, the distribution of the MLE estimator approaches a normal distribution centered around the true parameter value. This property enables the use of statistical inference techniques, such as constructing confidence intervals and conducting hypothesis tests, based on the asymptotic normality of the MLE estimators.

2.3 Consistency: MLE estimators are consistent, meaning that as the sample size increases, the MLE estimator converges to the true parameter value. Consistency ensures that given sufficient data, the MLE estimator will provide increasingly accurate estimates.

2.4 Flexibility: The MLE method is applicable to a wide range of statistical models. It can be used to estimate parameters in both simple and complex models, including linear regression, logistic regression, generalized linear models, survival analysis, mixture models, and more. This versatility makes MLE estimation a valuable tool in various fields of statistics and data analysis.

2.5 Model-Based Inference: MLE estimation is closely tied to the underlying statistical model. By assuming a specific probability distribution or functional form, the MLE method allows for model-based inference. This means that in addition to estimating parameters, the MLE framework provides a foundation for making probabilistic statements about the data and model assumptions.

2.6 Information-Theoretic Interpretation: The maximization of the likelihood function in MLE estimation can be interpreted from an information-theoretic perspective. The log-likelihood function is proportional to the Kullback-Leibler (KL) divergence between the true data-generating distribution and the estimated distribution. Maximizing the likelihood is equivalent to minimizing the KL divergence, which is a measure of information gain or loss.

2.7 Computational Optimization: While analytical solutions for MLE estimation may not always be available, numerical optimization techniques can be employed to find the maximum likelihood estimates. Various optimization algorithms, such as gradient descent, Newton-Raphson, and the Expectation-Maximization (EM) algorithm, can be utilized to maximize the likelihood function efficiently.

It is important to note that while the maximum likelihood method offers numerous advantages, it also has limitations. MLE estimation assumes a specific parametric form for the likelihood function and relies on the regularity conditions for its desirable properties. In cases where the underlying assumptions are violated or the model is misspecified, alternative estimation methods, such as nonparametric or robust methods, may be more appropriate.

3. Applications

The maximum likelihood (MLE) method has numerous applications across various fields of statistics and data analysis. Here are some common areas where the MLE method is frequently applied:

3.1 Regression Analysis: MLE estimation is widely used in linear regression models, where the goal is to estimate the regression coefficients that best describe the relationship between predictor variables and a continuous outcome variable. The MLE method allows for efficient estimation of the regression parameters, including the intercept and slope coefficients.

3.2 Logistic Regression: Logistic regression is employed when the outcome variable is binary or categorical. MLE estimation is used to estimate the regression coefficients in logistic regression models, which provide insights into the probability of an event occurring based on predictor variables.

3.3 Survival Analysis: In survival analysis, MLE estimation is used to estimate the parameters of survival models, such as the Cox proportional hazards model. These models analyze time-to-event data, where the event of interest could be, for example, the failure of a device or the occurrence of a medical event.

3.4 Mixture Models: Mixture models involve modeling a population as a mixture of several subpopulations or components. MLE estimation is utilized to estimate the parameters that define the mixture proportions and the distribution parameters for each component. Mixture models find applications in clustering, image segmentation, and population genetics, among other areas.

3.5 Gaussian Mixture Models (GMM): GMM is a specific type of mixture model where each component follows a Gaussian distribution. MLE estimation is employed to estimate the mean, variance, and mixture proportions of the Gaussian components. GMM finds applications in areas such as image processing, pattern recognition, and data clustering.

3.6 Hidden Markov Models (HMM): HMMs are widely used in sequential data modeling, such as speech recognition and natural language processing. MLE estimation is used to estimate the transition probabilities and emission probabilities in HMMs, allowing for accurate modeling and prediction of sequential data.

3.7 Item Response Theory (IRT): IRT is utilized in educational and psychological testing to model individuals' responses to test items. MLE estimation is used to estimate the item parameters, such as difficulty and discrimination, in order to assess individual abilities and make reliable inferences about their performance.

The maximum likelihood method is used because it is characterized by its flexibility, efficiency, and statistical properties that make it a valuable tool for parameter estimation in many statistical models.

4. Applied to Estimate the Parameters of Various Probability Distributions

Here are some examples of popular distributions where the MLE method is commonly used:

4.1 Normal Distribution: MLE estimation is used to estimate the mean and variance parameters of the normal distribution. This distribution is extensively used in many fields, such as finance, engineering, and social sciences.

4.2 Binomial Distribution: MLE estimation is employed to estimate the success probability parameter for this distribution. This distribution models the number of successes in a fixed number of independent Bernoulli trials. It's used for quality control and genetics applications.

4.3 Poisson Distribution: The MLE method is used to estimate the rate parameter of the Poisson distribution. This distribution models the number of events occurring in a fixed interval of time or space and is commonly used in areas such as insurance claims analysis and queue theory.

4.4 Exponential Distribution: MLE estimation is applied to estimate the rate parameter for this distribution. This distribution models the time between events in a Poisson process and is commonly used in reliability analysis , fail functions and survival analysis.

4.5 Gamma Distribution: MLE estimation is used to estimate the scale and shape parameters for this distribution. the distribution model is flexible and has applications in areas such as modeling income, waiting times, etc.

4.6 Weibull Distribution: MLE estimation is applied to estimate the shape parameters for this distribution. the distribution model is commonly used in reliability engineering to model the time to failure of systems.

4.7 Beta Distribution: The MLE method is used to estimate the scale and shape parameters for this distribution. The distribution model is commonly used as a prior distribution in Bayesian analysis and is also employed in modeling proportions and rates.

The MLE method of estimating parameters can be applied to many other probability distributions, including geometric distribution, binomial negative distribution, Pareto distribution, etc. The MLE method is an essential estimation tool in statistical modeling and inference.

Founded the maximum likelihood estimation of the parameters (mean and variance) of a normal distribution.

Assuming we have a sample of n independent and identically distributed observations $X = \{x_1, x_2, ..., x_n\}$, and we want to estimate the parameters μ (mean) and σ^2 (variance) of the normal distribution.

The likelihood function, denoted as $l(\mu,\sigma^2/X)$, is the probability of observing the given sample X, given the parameters μ and σ^2. For a normal distribution, the likelihood function is given by:

$$l(\mu,\sigma^2/X) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n}\exp\left(-\sum\frac{(x_i-\mu)^2}{2\sigma^2}\right) \qquad (1)$$

To simplify calculations, we take the logarithm of the likelihood function:

$$logl(\mu,\sigma^2/X) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \sum\left(\frac{(x_i-\mu)}{(2\sigma^2)}\right) \qquad (2)$$

The maximum likelihood estimates (MLEs) of μ and σ^2 are the values that maximize the log-likelihood function.

To find the MLE for μ, we differentiate eq.(2) with respect to μ and set it equal to zero:

$$\frac{\partial logl(\mu,\sigma^2./X)}{\partial\mu} = \sum\left(\frac{(x_i-\mu)}{(\sigma^2)}\right) = 0$$

Simplifying the above equation, we get:

∑(xi - μ) = 0

This equation implies that the MLE of μ is the sample mean:

$$\hat{\mu} = \frac{\sum x_i}{n}$$

To find the MLE for σ^2, we differentiate the log-likelihood function with respect to σ^2 and set it equal to zero:

$$\frac{\partial logl(\mu, \sigma^2./X)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \sum \frac{((x_i - \mu)^2)}{2\sigma^4} = 0$$

Simplifying the above equation, we get:

$$\sum((x_i - \mu)^2) = n\sigma^2$$

This equation implies that the MLE of σ^2 is the sample variance:

$$\widehat{\sigma^2} = \frac{\sum((x_i - \mu)^2)}{n}$$

Therefore, the MLEs of the parameters of the normal distribution are the sample mean $\hat{\mu}$ and the sample variance $\widehat{\sigma^2}$.

It's important to note that these estimators have desirable properties such as consistency and asymptotic normality, meaning that as the sample size increases, the estimators converge to the true parameter values and follow an approximate normal distribution, respectively.

5. References

1. Casella, G., & Berger, R. L. (2002). Statistical Inference, Second Edition,. Duxbury Press.
2. Lehmann, E. L., & Casella, G. (1998). Theory of Point Estimation, Second Edition, Springer.
3. Rao, C. R. (1973). Linear Statistical Inference and Its Applications, Second Edition, John Wiley & Sons.
4. Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Edition,. Springer.
5. Wasserman, L. (2004). All of Statistics: A Concise Course in Statistical Inference. Springer.
6. Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). Statistical Inference Under Order Restrictions, 92-388.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second edition, Springer.