

Maximum Likelihood Estimation and Likelihood Ratio test revisited

1. Introduction

Maximum likelihood Estimation is an important aspect of frequentist approach which was introduced by RA Fisher [1]. Maximum Likelihood estimation method helps us to find the estimator for the unknown population parameter. There are other methods of estimation also available such as Least Square Estimation and Bayesian Estimation methods but Maximum Likelihood Estimation is the widely used method to estimate the parameters. This paper provides an overview of Maximum Likelihood Method with example to calculate a Maximum Likelihood Estimator from a sample data set.

2. Maximum Likelihood Estimation Method

Let X is a random variable with probability mass function $P(X/\theta)$ where θ is the parameter of the distribution. Let X_1, X_2, \dots, X_n be the observation from the given sample. Then the joint probability or likelihood function is defined as

$$P(X_1, \dots, X_n / \theta) = P(X_1 / \theta) \times P(X_2 / \theta) \dots P(X_n / \theta) \dots \dots \dots (1)$$

Equation (1) is also a likelihood function and can be written as

$$L(\theta) = \prod_{i=1}^n P(X_i; \theta) = P(X_1, \theta) \cdot P(X_2, \theta) \dots P(X_n, \theta) \dots \dots \dots (2)$$

The maximum likelihood estimator $\hat{\theta}$ is defined the value of the parameter θ which maximizes the likelihood function. It will be easier to maximize the log of the likelihood function instead of the likelihood function directly

$$\text{Log } L(\theta) = \sum_{i=1}^n \log P(X_i, \theta) \dots \dots \dots (3)$$

Let us take the case of a binomial variable X with one instance x_1 and one unknown parameter P . Then the above equations (1) to (3) can be written as follows

$$P(X_1 / \theta) = P(X_1 / \theta) \dots \dots \dots (4)$$

$$L(\theta) = \prod_{i=1}^n P(X_i; \theta) = P(X_1, \theta) \dots \dots \dots (5)$$

$$\text{Log } L(\theta) = \sum_{i=1}^n \log P(X_i, \theta) \dots \dots \dots (6)$$

Here the Binomial variable X takes value 0,1 (Success or failure in a trail) and n is the total number of trails. For example if there are 10 trails and we get 3 success out of that 10 trails then the probability of observing 3 successes in 10 trails is given by

$$P(X=3,p) = \binom{10}{3} p^3(1-p)^7 \dots\dots\dots(7)$$

We need to find out the value of p which maximizes the above equation (7) (Likelihood function L(p,3))

$$L(p,3) = \binom{10}{3} p^3(1-p)^7 \dots\dots\dots(8)$$

$$\text{Log } L(P,3) = \text{Log} \left(\binom{10}{3} p^3(1-p)^7 \right) \dots\dots\dots(9)$$

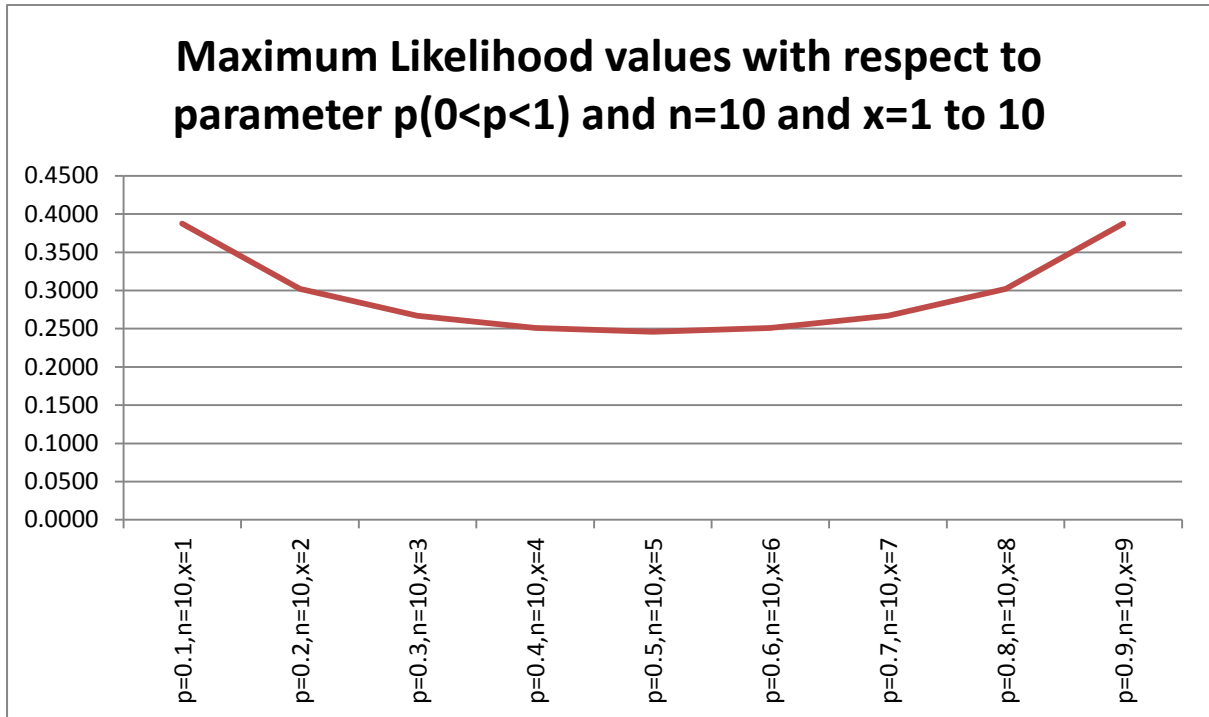
$$= 3 \log p + 7 \log(1-p) + \log \binom{10}{3} \dots\dots\dots(10)$$

We need the maximum for the equation (8) or (9) which will be the maximum likelihood estimate for the parameter p. The [table-1](#) below includes the maximum value of p for each value of X= 1, 2,..10 and n=10. For X=3 and n=10 and p=0.3, the likelihood equation (9) attains maximum i.e. 0.267.

Hence the value of p=0.3 is the maximum likelihood estimate for the parameter p and $\hat{p}=X/n$ is the maximum likelihood estimator for p.

Table-1: Likelihood values for the (Binomial Parameter n=10 and variable x= 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) at different values for the parameter p [2]

x(Total number of success in n trials)= np	n(nu mber of trails)	n!	x!	n-x!	n!/x!(n-x)!	Probability of success in a single trial(p ranges from 0 <p< 1)										
						0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	10	3628800	1	362880	10	0.0000	0.3874	0.27	0.12	0.04	0.01	0.00	0.00	0.00	0.00	0.00
2	10	3628800	2	40320	45	0.0000	0.19	0.3020	0.23	0.12	0.04	0.01	0.00	0.00	0.00	0.00
3	10	3628800	6	5040	120	0.0000	0.06	0.20	0.267	0.21	0.12	0.04	0.01	0.00	0.00	0.00
4	10	3628800	24	720	210	0.0000	0.01	0.09	0.20	0.2508	0.21	0.11	0.04	0.01	0.00	0.00
5	10	3628800	120	120	252	0.0000	0.00	0.03	0.10	0.20	0.2461	0.20	0.10	0.03	0.00	0.00
6	10	3628800	720	24	210	0.0000	0.00	0.01	0.04	0.11	0.21	0.2508	0.20	0.09	0.01	0.00
7	10	3628800	5040	6	120	0.0000	0.00	0.00	0.01	0.04	0.12	0.21	0.27	0.20	0.06	0.00
8	10	3628800	40320	2	45	0.0000	0.00	0.00	0.00	0.01	0.04	0.12	0.23	0.3020	0.19	0.00
9	10	3628800	362880	1	10	0.0000	0.00	0.00	0.00	0.00	0.01	0.04	0.12	0.27	0.3874	0.00
10	10	3628800	3628800	1	1	0.0000	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.11	0.35	0.0000



From the above we can also find out the likelihood curve reaches a minimum at $p=0.5$ (success or failure) , $n=10$ and $x=5$!!

3. Asymptotic (when the sample size is large) Properties of Maximum Likelihood estimators [3]

1. Sufficiency

Maximum Likelihood Estimators (MLE) has sufficient information about the unknown population parameter.

2. Consistency [4]

When the sample size is sufficiently large, probability of $p = \hat{p}$ will be 1 (when n tends to infinity)

3. Asymptotic normality

4. Efficiency

MLE attains the Cramer Rao lower bound because of the fact that the MLE is consistent and asymptotically normal

Likelihood ratio test [5]

If we would like to test the hypothesis that the x follows a distribution with parameter θ_1 against an alternative hypothesis θ_2 then the likelihood ratio will help to us to test whether the parameters θ_1 and θ_2 are similar or not

$$L = \frac{L(\theta_1)}{L(\theta_2)}$$

Where L is the likelihood ratio and L will take value between 0 and 1.

We need to compute the χ^2 to test the Likelihood ratio using the formula $\chi^2 = 2 \ln L$. If the theoretical value of the χ^2 is greater than the calculated value i.e $p > 0.05$, then we accept the null hypothesis that θ_1 and θ_2 are similar.

Conclusion

This paper revisited the maximum likelihood estimation test with examples and also description the likelihood ratio test.

References

- [1]. Fisher, R. A. (1925, July). Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 22, No. 05, pp. 700-725). Cambridge University Press.
- [2]. Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1), 90-100.
- [3]. Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605-610.
- [4]. Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, 887-906.
- [5]. Woolf, B. (1957). THE LOG LIKELIHOOD RATIO TEST (THE G-TEST). *Annals of human genetics*, 21(4), 397-409.