ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

A note on Likelihood and Likelihood Based Inference

Dr. James Kurian

Associate Professor, Department of Statistics, Maharaja's College, Ernakulam, Kerala, India

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150600082

Received: 21 June 2026; Accepted: 26 June 2026; Published: 07 July 2026

ABSTRACT

Statistical inference aims to draw conclusions about unknown parameters using observed data. Among the various inferential frameworks, likelihood-based inference occupies a central position because it provides a direct measure of the support that observed data give to competing parameter values. This paper presents a concise review of the likelihood function and its role in statistical inference. The concept of likelihood is introduced for both discrete and continuous probability models, emphasizing its interpretation as a measure of relative evidence rather than a probability distribution for the parameter. Several illustrative examples, including binomial, normal, and exponential models, demonstrate the construction and interpretation of likelihood functions. The paper discusses important properties of likelihood, including invariance under multiplication by positive constants and the combination of independent likelihoods through multiplication. Connections between likelihood and Bayesian inference are examined, showing how posterior information can be viewed as the product of prior and current likelihood information. Fundamental principles of likelihood-based inference, including the Likelihood Principle and the Stopping Rule Principle, are reviewed with classical examples. The role of maximum likelihood estimation as a practical method for parameter estimation is also highlighted. The discussion illustrates how the likelihood function serves as a compact and informative summary of statistical evidence, providing a unified framework for estimation and inference.

Key words: Likelihood Function; Likelihood-Based Inference; Likelihood Principle; Maximum Likelihood Estimation; Bayesian Inference; Stopping Rule Principle; Statistical Evidence; Parameter Estimation.

INTRODUCTION

Statistical inference is concerned with drawing conclusions about unknown population characteristics on the basis of observed data. Over the past century, several inferential paradigms have been developed, among which likelihood-based inference has emerged as one of the most influential and widely applicable frameworks. The likelihood approach provides a direct mechanism for quantifying the support that observed data offer to competing values of an unknown parameter and forms the theoretical foundation for many modern statistical methods (Fisher, 1922; Cox, 2006). The concept of likelihood was introduced and systematically developed by Fisher (1922), who distinguished likelihood from probability and emphasized its role in parameter estimation. While probability describes the chance of observing future data under a specified model, likelihood measures the relative plausibility of parameter values given the observed data. This distinction has profound implications for statistical reasoning and has led to the development of a coherent inferential framework based solely on the information contained in the observed sample (Edwards, 1992). Furthermore, many desirable large-sample properties of estimators, including consistency, asymptotic normality, and efficiency, arise naturally within the likelihood framework (Casella & Berger, 2002; Lehmann & Casella, 1998). Likelihood also plays a central role in Bayesian inference. Bayes' theorem combines prior information with the likelihood derived from observed data to produce the posterior distribution of the parameter. From this perspective, the likelihood function acts as the mechanism through which new information is incorporated into prior beliefs, highlighting the close relationship between Bayesian and likelihood-based approaches (Royall, 1997; Cox, 2006).

The objective of this paper is therefore to provide a concise exposition of the likelihood function and its role in statistical inference. Through a series of illustrative examples, the paper discusses the construction and interpretation of likelihood functions, the combination of evidence from independent sources, the relationship

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

between likelihood and Bayesian inference, and the implications of the Likelihood Principle and Stopping Rule Principle. The discussion highlights how the likelihood function for an observed data serves as a compact and informative summary of statistical evidence and provides a unifying framework for statistical inference.

Likelihood

The likelihood function is one of the most basic concepts in statistical inference. Entire theories of inference have been constructed based on it. We use scalar data using the notation x, vector data using the notation **x**, a scalar parameter using the notation □ and a vector parameter using the notation □. But it makes no difference in likelihood inference if the data x is a vector. Nor does it make a difference in the fundamental definitions if the parameter □ is a vector. Moreover, the term 'probability density' is used to denote both discrete and continuous models in this document. The purpose of the likelihood function is to convey the information about the unknown quantities. Likelihood inferences are based only on the data x and the model $\{□_□: □ \in \Omega\}$ which represents the entire set of possible probability measure under consideration.

Having observed x, the likelihood function for the given sample with a probability density $f_\theta(x)$ is denoted by $L(\theta)$ *or* $L(\theta|x)$. Now likelihood $L(\theta)$, defined on the parameter space Ω taking values in R^1, defined by

$$L(\theta) = f_\theta(x)$$

Since, $f_\theta(x)$ is just the probability of obtaining the data x when the true value of the parameter is θ, on Ω, we prefer θ_1 as the true value of the parameter θ over θ_2 whenever $f_{\theta_1}(x) > f_{\theta_2}(x)$.

Assuming a *discrete* model parameterized by a fixed unknown θ, the likelihood $L(\theta)$ is the probability of the observed x considered as a function of θ. Uncertainty in the data is conveyed through the likelihood function and provides relative preferences for various parameter values. It describes how likely the observed data are for each value of θ. The interpretation of the likelihood is the probability of x assuming the true value of the parameter is θ but *not* the probability of θ for given x. Thus, likelihood is considered as a function of θ, for fixed x, whereas the density is considered as a function of x for fixed θ.

For example, consider a binomial experiment with n=5, the likelihood function is

$$L(\theta) = f_\theta(x) = P_\theta(X = x)$$

$$= \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

and for x=0,1,2,3,4 and 5, the plot of $L(\theta)$ are given below.

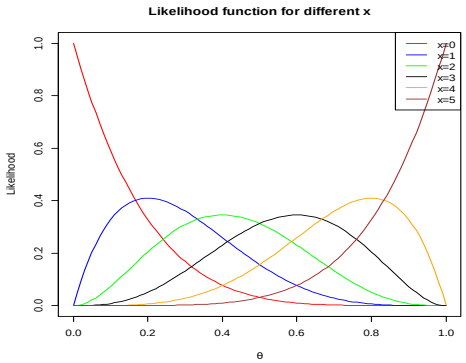


Fig-1

Obvious interpretation of the graph is that, when x=0, the likelihood is concentrated near zero indicating that the true value of θ is close to zero. When x=1, the likelihood is concentrated near 0.2 indicating that the true value of θ is close to 0.2 and so on. Hence, likelihood provides a relative preference for various parameter values. Also note that, for a typical value of x=1 out of n=5 tosses, the appropriate model for the observed data is the Binomial $(5, \theta)$ model with $\theta \in \Omega = (0, 1)$ with the corresponding likelihood

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

$$L(\theta|x=1) = \binom{5}{1}\theta^1(1-\theta)^4$$

which represented by the blue colour in Fig-1. It shows that the likelihood is peaked at $\theta = 0.2$ and takes a maximum of 0.4.

In continuous case, we can define

$$L(\theta) = P_\theta\left(X \in \left(x - \frac{\delta x}{2}, x + \frac{\delta x}{2}\right)\right) = \int_{x-\frac{\delta x}{2}}^{x+\frac{\delta x}{2}} f_\theta(\varkappa)\, d\varkappa$$

$$\approx \delta x f_\theta(x)$$

Or, if $f_{\theta_1}(x) > f_{\theta_2}(x)$, assuming the continuity of f

$$\int_{x-\frac{\delta x}{2}}^{x+\frac{\delta x}{2}} f_{\theta_1}(\varkappa)\, d\varkappa > \int_{x-\frac{\delta x}{2}}^{x+\frac{\delta x}{2}} f_{\theta_2}(\varkappa)\, d\varkappa$$

The above result shows that for the comparison of θ within $P_\theta(x)$, the likelihood is required up to $f_\theta(x)$, and hence we can ignore δx.

In general, the likelihood for a model can be expressed as

$$L(\theta) = h(x) f_\theta(x)$$

where h(x) is any strictly positive valued function of x that does not contain the parameter θ and hence, for the purpose of making inference about θ, it does not matter if multiplicative terms not containing unknown parameters are dropped from the likelihood function. In other words, a likelihood function and any positive multiple of it are equivalent as far as likelihood-based inferences are concerned.

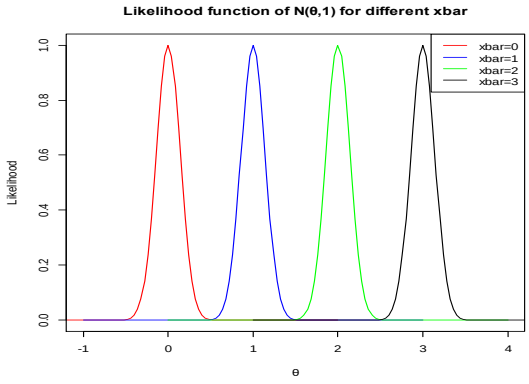
Example -1: Suppose that $(x_1, x_2, \ldots, x_n)$ are i.i.d. samples from $N(\theta, \sigma^2)$ distribution where, $\theta \in \Omega$ is unknown and $\sigma^2 > 0$ is known. The likelihood function is given by

$$L(\theta|\, x_1, x_2, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i)$$

$$= (2\pi\sigma^2)^{-n/2}\, exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right) exp\left(\frac{n-1}{2\sigma^2}s^2\right)$$

Therefore, we need only to consider the part $L(\theta|\, x_1, x_2, \ldots, x_n) \propto exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right)$

For example, suppose n= 50, $\sigma^2 = 1$, the plot of the likelihood function for different values of $\bar{x}$ = 0, 1, 2 and 3 are shown below.



We can observe from the above figure that the likelihood function concentrates near to the mean value. Hence, the entire likelihood function is the carrier of information on θ under x and not the maximiser of

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

the parameter. Further, if x_1 and x_2 are two independent data sets with probabilities $f_{1,\theta}(x_1)$ and $f_{2,\theta}(x_2)$, then likelihood of the combined data set is

$L(\theta) = f_{1,\theta}(x_1) * f_{2,\theta}(x_2) = L_1(\theta) * L_2(\theta)$, where $L_1(\theta)$ and $L_2(\theta)$ are the likelihoods of the individual data sets.

Example-2:Suppose that a coin is tossed 10 times and the head occurred less than 4 times. Then the likelihood function is

$$L(\theta) = \sum_{x=0}^{3} 10C_x \theta^x (1-\theta)^{10-x}$$

Example-3: The medical team identifies the 50th sampled person having the first AB blood group. The likelihood for observing an AB blood group is

$$L(\theta) = (1-\theta)^{49}\theta$$

Example-4: Based on a random sample of size 4 from exponential distribution with parameter θ, it is only reported the maximum of the sample value $x_{(4)}$ =10 and the second sample of size n=3 reported the sum s=8, then

For the maximum, the likelihood function is given by

$$L_1(\theta) = n[F(x)]^{n-1} f(x)$$

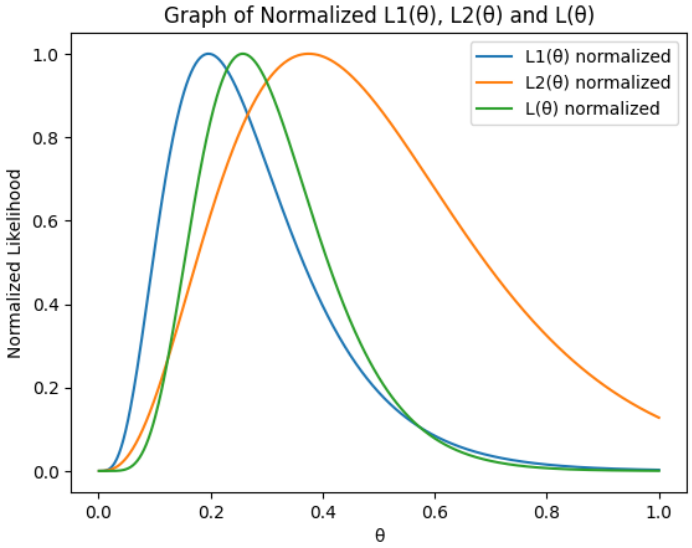
For n=4 and $x_{(4)}$ =10, we have $L_1(\theta) = 4\theta e^{-10\theta}\left(1 - e^{-10\theta}\right)^3$

For the sum, the distribution is Gamma (3, θ), then

$$L_2(\theta) = 64\frac{\theta^3}{\Gamma(3)}e^{-8\theta}$$

Therefore, the likelihood of the combined data

$$L(\theta) = L_1(\theta) * L_2(\theta) = 128\theta^4 e^{-18\theta}\left(1 - e^{-10\theta}\right)^3$$



likelihood curves have been normalized to have maximum value 1, so their shapes can be compared directly without scale effects.

likelihood curves have been normalized to have maximum value 1, so their shapes can be compared directly without scale effects.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

$L_1(\theta)$ is more left-shifted and relatively sharper, indicating that the extreme observation favors smaller θ. $L_2(\theta)$ is broader and peaks near θ=3/8, reflecting information from the total exposure time. $L(\theta)$, is the most concentrated, illustrating how combining independent summaries reduces uncertainty.

Likelihood function and Bayesian

Bayes rule is stated as $f(\theta|x) = c * f(\theta) * f(x|\theta)$, where c is the normalizing constant and the likelihood of the sample is $f(x|\theta) = f_\theta(x)$. The distribution of $f(\theta)$ is known as the prior distribution, which we can call as prior likelihood.

This means that the effect of Bayesian method is the same as likelihood method as we can consider posterior likelihood as the combining of prior likelihood and current likelihood. That is

Posterior Likelihood $\propto$ Prior likelihood * Current likelihood

If we do not know anything about θ, we can take $f(\theta)$ =1, the likelihood function expresses the current information on θ, after observing x, and in such cases posterior density and likelihood functions would be the equivalent.

Likelihood Principle (Alan Birnbaum (1962))

If two experiments yield data such that their likelihood functions are proportional, then those two sets of data represent equivalent instances of statistical evidence. Likelihood principle asserts that the likelihood function contains all the necessary information and hence if two model and data combinations yield equivalent likelihood functions, then inferences about the unknown parameter must be the same. According to this principle, the likelihood is the most compact summary of the data without loose of information and, any inference regarding the unknown θ must be based on $L(\theta|x)$ alone. In other words, when two likelihood functions are equal or even proportional to each other, both are equivalent and must yield same parameter and variance estimates. Principle has the effect of concluding equivalent strength of evidence for proportional likelihood functions.

Example-1

Case-1: Suppose that a coin is tossed 10 times and that x=4 heads observed resulting in x is a Binomial(10, θ) with $\theta \in \Omega = (0, 1)$. The likelihood function is given by

$$L(\theta|4) = \binom{10}{4} \theta^4(1-\theta)^6.$$

Case-2: Suppose a coin is tossed independently until four heads are obtained and the number of tails observed until the fourth head is x=6. Then x is a Negative-Binomial(4, θ) with $\theta \in \Omega = (0, 1)$. The likelihood function is given by $L(\theta|6) = \binom{9}{6} \theta^4(1-\theta)^6$.

Note that in both cases the likelihoods are proportional. So, the likelihood principle asserts that we must ignore the fact that the data were obtained in entirely different ways and these two model and data combinations must yield the same inferences about the unknown θ. But there are many situations in which we might consider additional information regarding the data beyond the likelihood function and might take different inferences for the two situations. For example, while using a moment method of estimation or testing the null hypothesis of $H_0: \theta = \theta_0$ we make use of the sampling method and additional model features beyond the observed data and we may reach entirely different inference about θ for the two situations.

Example-2 (Basu,1975)

Consider four different coin tossing experiments

(1) Toss the coin exactly10 times;
(2) Continue tossing until 6 heads appear;
(3) Continue tossing until 3 consecutive heads appear;

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

(4) Continue tossing until the accumulated number of heads exceeds that of tails by exactly 2

Suppose that all four experiments have the same outcome x=(T, H, T, T, H, H, T, H, H, H).

We may feel that the evidence for θ, the probability of heads, is the same in every case. Once the sequence of heads and tails is known, the intentions of the original experimenter are immaterial to inference about the probability of heads.

Stopping Rule Principle

The intentions of the experimenter, represented by τ, are irrelevant for making inferences about θ, once the observations (x_1 , x_2 , … , x_n) are known. This means, once the data is observed, we can ignore the sampling plan.

CONCLUSION

The likelihood function provides a fundamental framework for statistical inference by summarizing the information contained in observed data regarding unknown parameters. Unlike probability distributions, which describe uncertainty in future observations, likelihood functions quantify the relative support that observed data provide for different parameter values. Through a series of examples, this paper has demonstrated how likelihood functions can be constructed, interpreted, and combined across independent sources of information. The discussion highlights that the entire shape of the likelihood function carries inferential information and not merely its maximum point. The Likelihood Principle emphasizes that all relevant evidence concerning a parameter is contained in the likelihood function, while the Stopping Rule Principle suggests that inference should depend only on the observed data and not on the sampling intentions. Furthermore, the close relationship between likelihood and Bayesian inference illustrates how prior and current information can be integrated within a unified inferential framework. Overall, likelihood-based inference offers a coherent and powerful approach for estimation, hypothesis testing, and evidence evaluation, making it one of the cornerstones of modern statistical inference.

REFERENCES

1. Basu, D. (1975). *Statistical Information and Likelihood*. Sankhya Series A, 37, 1–71
2. Birnbaum, A. (1972). *On the Foundations of Statistical Inference*. Journal of the American Statistical Association, 57(298), 269–306.
3. Birnbaum, A. (1972). More concepts of statistical evidence. Journal of the American Statistical Association 67, 858–861.
4. Casella G. and Berger R. L. (2002). Statistical Inference. Duxbury Press.
5. Cox D. R. (2006). Principles of Statistical Inference. Cambridge University Press.
6. Edwards A. W. F. (1992). Likelihood. Johns Hopkins University Press.
7. Fisher R. A. (1922). *On the Mathematical Foundations of Theoretical Statistics*. Philosophical Transactions of the Royal Society A, 222, 309–368.
8. Lehmann E. L. and Casella G. (1998). Theory of Point Estimation. Springer.
9. Pawitan Y. (2001) In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press.
10. Royall R.(1997) Statistical Evidence: A Likelihood Paradigm. Chapman & Hall.