Frequently Asked Questions

1. What do you mean by method of maximum likelihood?

Answer:

In statistics, maximum-likelihood estimation (MLE) is a method of estimating the parameters of a statistical model. When applied to a data set and given a statistical model, maximum-likelihood estimation provides estimates for the model's parameters.

The method of maximum likelihood corresponds to many well-known estimation methods in statistics. For example, one may be interested in the heights of adult female giraffes, but be unable to measure the height of every single giraffe in a population due to cost or time constraints.

Assuming that the heights are normally (Gaussian) distributed with some unknown mean and variance, the mean and variance can be estimated with MLE while only knowing the heights of some sample of the overall population. MLE would accomplish this by taking the mean and variance as parameters and finding particular parametric values that make the observed results the most probable. The method aims at obtaining the values of unknown parameters of the population by maximizing the likelihood function

The principle of maximum likelihood estimation (MLE), originally developed by R.A. Fisher in the 1920s, states that the desired probability distribution is the one that makes the observed data "most likely", which means that one must seek the value of the parameter vector that maximizes the likelihood function .

2. Write a note on Likelihood Function.

Answer:

A likelihood function of a number of sample observations is defined to be their joint density function. To use the method of maximum likelihood, one first specifies the joint density function for all observations. For an i.i.d. sample, this joint density function is

$F(x_1, x_2, \ldots, x_n, \theta) = f(x1,\theta)\, f(x2,\theta)\, f(x3,\theta) \quad \ldots\ldots\ldots\ f(xn,\theta)$

Now we look at this function from a different perspective by considering the observed values x1, x2... xn to be fixed "parameters" of this function, whereas θ will be the function's variable and allowed to vary freely; this function will be called the likelihood.

In case the sample observations are independent the likelihood function happens to be the product of the density functions of the random observations. If $x_1, x_2, \ldots, x_n$ are n independent and identically distributed observations, from a population with an unknown parameter θ then the likelihood function of the random observations denoted by $L = L(x_1, x_2, \ldots, x_n, \theta)$ is given by

$L = L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$ where f(xi,θ) is the p.d.f of the population.

L gives relative likelihood or chance that the random observations assume a particular set of values. For a particular sample(x_1, x_2, …, x_n) L becomes a function of (x_1, x_2, …, x_n) and an unknown parameter θ and p.d.f. f (x,θ).

3. Explain briefly how we find the MLE of the parameters of the population under consideration.

Answer:
The method of maximum likelihood is the one in which for a given set of values (x_1, x_2… x_n) an estimator of θ is found that maximizes L. Thus if there exists $\hat{\theta}$, a function of (x_1, x_2… x_n) for which L is maximum for variations in θ . Then $\hat{\theta}$ is the MLE of θ then

$\frac{\partial L}{\partial \theta} = 0$ and $\frac{\partial^2 L}{\partial \theta^2} < 0$

In practice it is often more convenient to work with the logarithm of the likelihood function, called the **log-likelihood**.

Since logL attains maximum when L attains maximum the solution of the equation

Log L (x_1, x_2, …, x_n ,θ)= $\sum_{i=1}^{n} \log f(xi, \theta)$

$\frac{\partial \log L}{\partial \theta} = 0$ with $\frac{\partial^2 \log L}{\partial \theta^2} < 0$ also gives the MLE of θ

An MLE estimate is the same regardless of whether we maximize the likelihood or the log-likelihood function, since log is a monotone transformation.

4. What are the advantages of method of maximum likelihood?

Answer:

The method of maximum likelihood provides a basis for many of the techniques.

- The method has a good intuitive foundation. The underlying concept is that the best estimate of a parameter is that giving the highest probability that the observed set of measurements will be obtained.
- The least-squares method and various approaches to combining errors or calculating weighted averages, etc. can be derived or justified in terms of the maximum likelihood approach.
- The method is of sufficient generality that most problems are amenable to a straightforward application of this method, even in cases where other techniques become difficult. Inelegant but conceptually simple approaches often provide useful results where there is no easy alternative.

5. What are the applications of the maximum likelihood method?

Answer

Maximum likelihood estimation is used for a wide range of statistical models, including:

- Linear Models and generalized linear models
- Exploratory and confirmatory factor analysis
- Structural Equation Modelling
- Many situations in the context of hypothesis testing and confidence interval formation
- Discrete Choice models

These uses arise across applications in widespread set of fields, including:

- Communication Systems
- Psychometrics
- Econometrics
- Time-delay of arrival (TDOA) in acoustic or electromagnetic detection
- Data modelling in nuclear and particle physics
- Magnetic resonance imaging
- Computational Phylogenetics
- Origin/destination and path-choice modelling in transport networks
- Geographical satellite-image classification

6. Describe the principle behind the method of maximum likelihood.

Answer:

Suppose there is a sample $x1, x2, \ldots, xn$ of n independent and identically distributed observations, coming from a distribution with an unknown pdf $f(\cdot)$. It is however surmised that the function f belongs to a certain family of distributions $\{ f(\cdot,\theta), \theta \in \Theta \}$, called the parametric model, so that $f = f(\cdot,\theta)$.

The value θ is unknown and is referred to as the "true value" of the parameter. It is desirable to find an estimator $\hat{\theta}$, which would be as close to the true value θ as possible. Both the observed variables xi and the parameter θ can be vectors.

To use the method of maximum likelihood, one first specifies the joint density function for all observations. For an i.i.d. sample, this joint density function is

$$f(x1,x2,\ldots,xn, \theta)= f(x1, \theta)f(x2, \theta)\ldots\ldots f(xn, \theta)$$

Now we look at this function from a different perspective by considering the observed values x1, x2, ..., xn to be fixed "parameters" of this function, whereas θ will be the function's variable and allowed to vary freely; this function will be called the likelihood:

$L(x1,x2,\ldots,xn, \theta) = f(x1,,x2,\ldots,xn, \theta) = \prod_{i=1}^{n} f(xi,\theta)$

In practice it is often more convenient to work with the logarithm of the likelihood function, called the **log-likelihood**:

$\ln \; L(x1,x2,\ldots,xn, \theta) = \sum_{i=1}^{n} \ln f(xi,\theta)$

7. Briefly explain the properties of maximum likelihood estimators.

Answer:

A maximum-likelihood estimator is an extremum estimator obtained by maximizing, as a function of θ, the objective function

Maximum-likelihood estimators have no optimum properties for finite samples, in the sense that (when evaluated on finite samples) other estimators have greater concentration around the true parameter-value.

However, like other estimation methods, maximum-likelihood estimation possesses a number of attractive limiting properties: As the sample-size increases to infinity, sequences of maximum-likelihood estimators have these properties:

- Consistency: a subsequence of the sequence of MLEs converges in probability to the value being estimated.
- Asymptotic Normality: as the sample size increases, the distribution of the MLE tends to the Gaussian distribution with mean θ and covariance matrix equal to the inverse of the Fisher information matrix.
- Efficiency: i.e., it achieves the Cramér–Rao lower bound when the sample size tends to infinity. This means that no asymptotically unbiased estimator has lower asymptotic mean squared error than the MLE (or other estimators attaining this bound).

8. Obtain the M.L.E of mean λ of the Poisson population.

Answer:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from a given population

$L = L(x_1, x_2, \ldots, x_n, \lambda) = \prod_{i=1}^{n} f(x_{i,}\lambda)$

$$= \frac{e^{-n\lambda} \lambda^{\sum xi}}{\prod xi!}$$

Ln L = -nλ+∑xi ln λ – ln $\prod xi!$

$$\frac{\partial \ln L}{\partial \lambda} = 0 \Rightarrow -n + \frac{\sum xi}{\lambda} = 0 \Rightarrow \hat{\lambda} = \frac{\sum xi}{n} = \bar{x}$$

With $\frac{\partial^2 \ln L}{\partial \lambda^2} < 0$

Therefore, sample mean is the MLE of the population mean of the Poisson population.

9. Let x1, x2, …, xn be a random sample of size n drawn from an exponential population with density function

$F(x,\theta) = (1/\theta)\, e^{-xi/\theta}$, xi >0. Find the m.l.e of θ

Answer:

Given $f(x,\theta) = (1/\theta)\, e^{-xi/\theta}$

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

$$= \left(\frac{1}{\theta}\right)^n e^{-\frac{\sum xi}{\theta}}$$

Ln L= -n ln θ - ∑xi/θ

$$\frac{\partial \ln L}{\partial \theta} = 0 \Rightarrow \frac{-n}{\theta} + \frac{\sum xi}{\theta^2} = 0 \Rightarrow -n + \frac{\sum xi}{\theta} = 0 \Rightarrow \theta = \frac{\sum xi}{n} = \bar{x}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2\sum xi}{\theta^3} = \frac{-n}{\theta^2} < 0$$

Therefore, MLE of θ is $\hat{\theta} = \frac{\sum xi}{n} = \bar{x}$

10. Derive the MLE of the parameter θ of the Geometric population.

Answer:

Let $x_1, x_2 \ldots x_n$ be a random sample of size n drawn from the Geometric population with probability mass function

$F(x,\theta) = \theta\,(1-\theta)^x$

The likelihood function

$L = L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_{i,}\theta) = \theta^n (1-\theta)^{\Sigma xi}$

$\text{Log } L = n \log\theta + \sum xi \log(1-\theta)$

$$\frac{\partial \log L}{\partial\theta} = 0 \Rightarrow \frac{n}{\theta} - \frac{\sum xi}{(1-\theta)} = 0 \Rightarrow n(1-\theta) - \theta\sum xi = 0 \Rightarrow \theta = \frac{1}{1+\overline{x}}$$

Hence, MLE of θ of Geometeic population is $\widehat{\theta} = \frac{1}{1+\overline{x}}$

Again, this estimator is the method of moments estimator, and it agrees with the intuition because, in n observations of a geometric random variable, there are **n** successes in the summation xi trials. Thus, the estimate of θ is the number of successes divided by the total number of trials.

11. Find the MLE of θ in the following Uniform distribution U(0,θ)

$F(x,\theta) = 1/\theta,\ 0<x<\theta$

Answer:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from an Uniform population with probability density function

$F(x,\theta) = 1/\theta,\ 0<x<\theta$

$L = L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_{i,}\theta) = (1/\theta)^n$

$\text{Log } L = -n \log\theta$

$$\frac{\partial \log L}{\partial \theta} = 0 \Rightarrow -\frac{n}{\theta} = 0 \Rightarrow \widehat{\theta} = \infty$$ or n=0

This is meaningless

Hence, we use the basic principle of the maximum likelihood estimation

MLE is that value of θ for which L is maximum

L is maximum when $1/\theta^n$ is maximum

That is when θ^n is minimum

That is when θ is minimum

But 0 ≤ x ≤ θ $\Rightarrow 0 \le x_{(1)} \le x_{(2)} \le ...... \le x_{(n)} \le \theta$

Minimum value of θ is $x_{(n)}$, the maximum of the observations

Therefore MLE of θ is $\widehat{\theta} = x_{(n)}$, the maximum of the observations

12. Obtain the M.L.E of α and β for the following distribution.
F(x) = 1/(β-α), α ≤ x ≤β

Answer:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from the given population with probability density function

f(x) = 1/(β-α), α ≤ x ≤β

L = L($x_1, x_2, \ldots, x_n$,α,β)= $\prod_{i=1}^{n} f(x_{i,}\alpha.\beta)$ = $(1/\beta\text{-}\alpha)^n$

Log L = -n log (β-α)

$$\frac{\partial \log L}{\partial \beta} = 0 \Rightarrow -\frac{n}{\beta - \alpha} = 0 \Rightarrow \beta - \alpha = \infty$$ Which is meaningless

Hence the usual method of finding m.L.E fails. Nowwe use the basic principle of the maximum likelihood estimation

M.L.E's of α and β is that values for which L is maximum

L is maximum when $1/(\beta-\alpha)^n$ is maximum

That is when $(\beta-\alpha)^n$ is minimum

That is when (β-α) is minimum

(β-α) is minimum when α is maximum and β is minimum

But $\alpha \le x \le \beta \Rightarrow \alpha \le x_{(1)} \le x_{(2)} \le ...... \le x_{(n)} \le \beta$

Minimum value of β is $x_{(n)}$, the maximum of the observations and the maximum value of α is $x_{(1)}$, the minimum of the observations

Therefore, MLE of α is $x_{(1)}$ and β is $x_{(n)}$,

13. For a random sampling from a Normal population N(θ,σ2) find the MLE for θ when σ2 is known

Answer:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from a given population

$$L = L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_{i,}\theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}(xi-\theta)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(xi-\theta)^2}$$

$$\text{Ln L} = \frac{-n}{2}\ln\sigma^2 - n\ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum(xi-\theta)^2$$

Maximum likelihood estimate of θ when σ^2 is known

$$\frac{\partial \ln L}{\partial \theta} = 0 \Rightarrow \frac{\sum(xi-\theta)}{\sigma^2} = 0$$

$$\Rightarrow \sum xi - n\theta = 0 \Rightarrow \hat{\theta} = \frac{\sum xi}{n} = \bar{x}$$

Therefore, MLE of θ is $\hat{\theta} = \frac{\sum xi}{n} = \bar{x}$

With $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$

14. For a random sampling from a Normal population N(θ,σ2) find the MLE for θ and σ2 when both are unknown

Answer:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from a given population

L = L($x_1, x_2, \ldots, x_n$,θ)= $\prod_{i=1}^{n} f(x_{i,}\theta)$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}(xi-\theta)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(xi-\theta)^2}$$

Ln L = $\frac{-n}{2}\ln\sigma^2 - n\ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum(xi-\theta)^2$

Maximum likelihood estimate of θ and σ^2 when both are unknown

$$\frac{\partial \ln L}{\partial \theta} = 0 \Rightarrow \frac{\sum(xi-\theta)}{\sigma^2} = 0 \Rightarrow \hat{\theta} = \frac{\sum xi}{n} = \bar{x} \text{ -------(1)}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = 0 \Rightarrow \frac{-n}{} + \frac{\sum (xi-\theta)^2}{\sigma^2} = 0 \Rightarrow \frac{-n}{} + \frac{\sum (xi-\hat{\theta})^2}{\sigma^2} = \frac{-n}{} + \frac{\sum (xi-\bar{x})^2}{\sigma^2} 0$$

From equation (1)

$$\Rightarrow \hat{\sigma}^2 = \frac{\sum (xi-\bar{x})^2}{n}$$

Therefore, MLE of θ is $\hat{\theta} = \frac{\sum xi}{n} = \bar{x}$ and the MLE of σ^2 when θ is unknown is

$$\hat{\sigma}^2 = \frac{\sum (xi-\bar{x})^2}{n}$$

15. When the Maximum-likelihood estimators can lack asymptotic normality and can be inconsistent?

Answer:

Maximum-likelihood estimators can lack asymptotic normality and can be inconsistent if there is a failure of one (or more) of the below regularity conditions:

Estimate on boundary

Sometimes the maximum likelihood estimate lies on the boundary of the set of possible parameters, or (if the boundary is not, strictly speaking, allowed) the likelihood gets larger and larger as the parameter approaches the boundary. Standard asymptotic theory needs the assumption that the true parameter value lies away from the boundary. If we have enough data, the maximum likelihood estimate will keep away from the boundary too.

Data boundary parameter-dependent

For the theory to apply in a simple way, the set of data values, which has positive probability (or positive probability density), should not depend on the unknown parameter. A simple example where such parameter-dependence does hold is the case of estimating θ from a set of independent identically distributed when the common distribution is uniform on the range (0,θ). For estimation purposes, the relevant range of θ is such that θ cannot be less than the largest observation. Because the interval (0,θ) is not compact, there exists no maximum for the likelihood function.

Nuisance parameters

For maximum likelihood estimations, a model may have a number of nuisance parameters. For the asymptotic behaviour outlined to hold, the number of nuisance parameters should not increase with the number of observations (the sample size).

Increasing information

For the asymptotic to hold in cases where the assumption of independent identically distributed observations does not hold, a basic requirement is that the amount of information in the data increases indefinitely as the sample size increases. Such a requirement may not be met if either there is too much dependence in the data (for example, if new observations are essentially identical to existing observations), or if new independent observations are subject to an increasing observation error.