STATISTICS FOR DATA SCIENCE

Principles of Point Estimation

D. Uma
Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Maximum Likelihood Estimation

D. Uma
Department of Computer Science and Engineering

Topics to be covered...

- ✓The Method of Maximum Likelihood for

  Bernoulli, Binomial and Poisson Distributions.

Maximum Likelihood Estimate (MLE)

- Maximum Likelihood Estimate (MLE) is the good method that can be applied for estimating parameters.

- It can be obtained from any given distribution using the observed data.

- The suggestion is to estimate the parameter with the value that makes the observed data most likely.

Method to Construct Good Estimator

Method of Maximum Likelihood(MLE) is regarded as the best method to point estimation.

Likelihood function is the probability of obtaining observed value.

If single observation is made, Likelihood function is just the probability of obtaining that value.

MLE is that value of the estimator which when substituted in place of the parameter, maximizes the Likelihood function.

General Steps to proceed with MLE.

Step 1: Write down the likelihood function.

Step 2: Take natural log of likelihood function.
(Reason: the quantity that maximizes log of a function is always the same quantity that maximizes the function itself)

Step 3: Differentiate log-likelihood function with respect to the parameter being estimated.

Step 4: Set the derivative equal to 0 to get MLE.

Bernoulli Distribution – Estimate Likelihood Function for (p)

Let $X_1, \ldots, X_n$ be a random sample from the population with Bernoulli (p)distribution.

The probability mass function is given by,

$$P(X = x_i) = p^{x_i} (1-p)^{1-x_i}$$

The likelihood function is,

$$L = f(p \mid x_1, \ldots, x_n) = p^{x_1}(1-p)^{1-x_1} * p^{x_2}(1-p)^{1-x_2} * \ldots * p^{x_n}(1-p)^{1-x_n}$$

$$L = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} \sim p^{\sum x_i}(1-p)^{n-\sum x_i}$$

Take log of the likelihood function,

$$\ln L = \sum x_i \ln p + \left(n - \sum x_i\right) \ln(1-p)$$

Bernoulli Distribution – Estimate Likelihood Function for (p)

Taking derivative with respect to p and set it to 0,

$$\frac{d\ln L}{dp} = 0 \rightarrow \sum x_i \frac{d\ln p}{dp} + \left(n - \sum x_i\right) + \frac{d\ln(1-p)}{dp} = 0$$

$$\frac{\sum x_i}{p} + \left(n - \sum x_i\right) * \frac{-1}{1-p} = 0$$

$$\sum x_i - p\sum x_i + p\sum x_i = 0$$

$$\sum x_i - np = 0$$

$$p = \frac{\sum x_i}{n} = \bar{X}$$

The MLE of p is $\hat{p} = \bar{X}$

Binomial Distribution – Estimate Likelihood Function

Let $X \sim \text{Bin}(n, p)$ where n is known and p is unknown

The probability mass function of X is $f(x; n, p) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$

The likelihood function is given by,

$$L(p; n, x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

Take log of the likelihood function,

$$\ln L(p; n, x) = \ln\left[\left(\frac{n!}{x!\,(n-x)!}\right) p^x (1-p)^{n-x}\right]$$

$$= \ln\left(\frac{n!}{x!\,(n-x)!}\right) + x \ln p + (n-x)\ln(1-p)$$

Binomial Distribution – Estimate Likelihood Function

Taking derivative with respect to p and set it to 0,

$$\frac{d}{dp}L(p;n,x) = \frac{d}{dp}\left(\ln n! - \ln x! - \ln(n-x)! + x\ln p + (n-x)\ln(1-p)\right)$$

$$= 0 + \frac{x}{p} - \frac{x}{1-p} - \frac{n}{1-p}$$

Apply chain rule,

Multiply both sides by $p(1-p)$

$$0 = x - xp + xp - np$$

$$p = \frac{x}{n}$$

The MLE of p is $\hat{p} = \frac{X}{n}$

Example – Binomial Distribution – Estimate Likelihood Function

$X \sim \mathrm{Bin}(20, p)$, p is unknown

Suppose we observe $X = 7$. The probability mass function is,

$$f(7;p) = \frac{20!}{7!\,13!}\, p^7\,(1-p)^{13}$$

In the probability mass function it is as written f (7; p) rather than f (7). Here the data value 7 is constant.

When a probability mass function or probability density function is considered to be a function of parameters, it is called a **likelihood function.**

How can we maximise this likelihood function?

When $\hat{p}$ is substituted for p, it maximizes the likelihood function

Let's compute maximum likelihood function for f(7; p)

We could maximize this function by taking the derivative with respect to p

and setting it equal to 0.

$$ln\, f(7;p) = \ln 20! - \ln 7! - \ln 13! + 7 \ln p + 13 \ln(1-p)$$

We take the derivative with respect to p and set it equal to 0.

$$\frac{d}{dp} \ln f(7;p) = \frac{7}{p} - \frac{13}{1-p} = 0$$

The maximizing value is $\frac{7}{20}$

Therefore the maximum likelihood estimate is $\hat{p} = \frac{7}{20}$

MLE for Poisson Distribution (λ) – Estimating Likelihood

Let $X_1, \ldots, X_n$ be an independent random sample from an Exp(λ), where λ is unknown.

The likelihood function is the joint probability density function of $X_1, \ldots, X_n$

considered as a function of the parameter λ.

The probability function is given by,

$$f(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$$

The likelihood function is,

$$L(x_i; \lambda) = \sum_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x!}$$

MLE for Poisson Distribution (λ) – Estimating Likelihood

Take log of the likelihood function,

$$\ln L(x_i;\lambda) = \sum_{i=1}^{n} \ln e^{-\lambda} + \sum_{i=1}^{n} x_i \ln \lambda - \sum_{i=1}^{n} \ln x_i$$

Taking derivative with respect to λ and set it to 0,

$$\frac{d}{d\lambda} L(x_i;\lambda) = 0$$

$$\frac{d}{d\lambda}\left(\sum_{i=1}^{n}(-\lambda) + \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \ln x_i\right) = 0$$

MLE for Poisson Distribution (λ) – Estimating Likelihood

$$= -n + \frac{1}{\lambda}\sum_{i=0}^{n} x_i + 0 = 0$$

$$\lambda = \frac{1}{\lambda}\sum_{i=1}^{n} x_i = \bar{X}$$

The MLE of λ is $\hat{\lambda} = \overline{\mathrm{X}}$

Example - Poisson Distribution

Problem:

The following data are the observed frequencies of occurrence of domestic accidents: we have n = 647 data as follows

Number of Accidents	Frequency
0	447
1	132
2	42
3	21
4	3
5	2

Problem Source - http://wwwf.imperial.ac.uk/

Example - Poisson Distribution

Solution:

$$\lambda = \frac{1}{\lambda}\sum_{i=1}^{n} x_i = \bar{X}$$

The MLE of λ is $\hat{\lambda} = \bar{X}$

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$$

$$= \frac{(447 * 0) + (132 * 1) + (42 * 2) + (21 * 3) + (3 * 4) + (2 * 5)}{674}$$

$$= 0.465$$

Problem Source - http://wwwf.imperial.ac.uk/

Maximum Likelihood (MLE) – Desirable Properties

Maximum likelihood is the most commonly used method of estimation.

The main reason for this is that in most cases that arise in practice, MLEs have two very desirable properties,

1. In most cases, as the sample size *n* increases, the bias of the MLE converges to 0.

2. In most cases, as the sample size *n* increases, the variance of the MLE converges to a theoretical minimum.

Pitfalls of Point Estimators

- A **point estimator** is a single number which may vary from sample to sample.
- Certainly the point estimators are slightly different from true population parameter .
- It cannot be confidently claimed to be close to the actual parameter.
- This can be solved by estimating population parameters in the given intervals of values where point estimator can be centered. This interval is called **confidence interval.**

THANK YOU

D. Uma
Department of Computer Science and Engineering

umaprabha@pes.edu

+91 99 7251 5335