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
Computer Science and Engineering

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.

Definition

Let $X_1, \dots, X_n$ have joint probability density or probability mass function

$f(x_1, \dots\dots, x_n; \theta_1, \dots\dots, \theta_k)$, where $\theta_1, \dots\dots, \theta_k$ are parameters and

$x_1, \dots\dots, x_n$ are values observed from $X_1, \dots, X_n$.

The values $\hat{\theta}_1, \dots, \hat{\theta}_k$ that maximize f are the

maximum likelihood estimates of $\theta_1, \dots\dots, \theta_k$

If the random variables $X_1, \dots, X_n$ are substituted for $x_1, \dots\dots, x_n$, then

$\hat{\theta}_1, \dots, \hat{\theta}_k$ are called maximum likelihood estimators

The abbreviation MLE is often used for both maximum likelihood estimate

and maximum likelihood estimator

Points to Remember

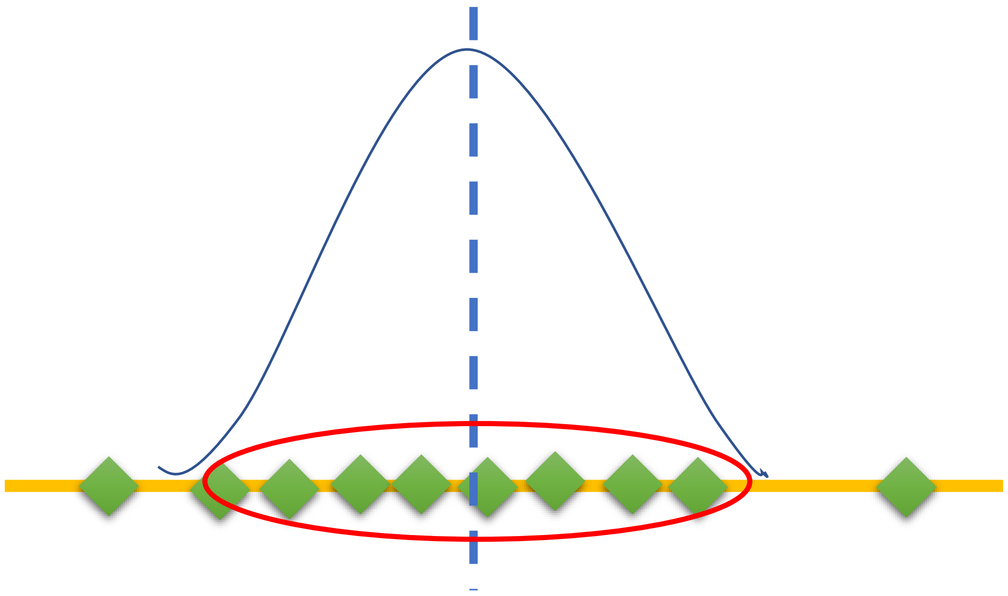
- The **maximum likelihood estimate** is the value of the estimators that when substituted in for the parameters maximizes the likelihood function.
- The likelihood function can be a probability density function or a probability mass function.
- It can also be a joint probability density or mass function, and is often the joint density or mass function of independent random variables.

Obesity in Women – Understanding MLE

Annie is a post graduate student who want study on the growing health problems with women due to obesity. She decided to collect data from the samples. She had chosen between age of 20 – 25 years. She surveys with the questionnaire of the diet and exercise habits of her 10 class mates to start with and collects their weights and plots it from low to high.

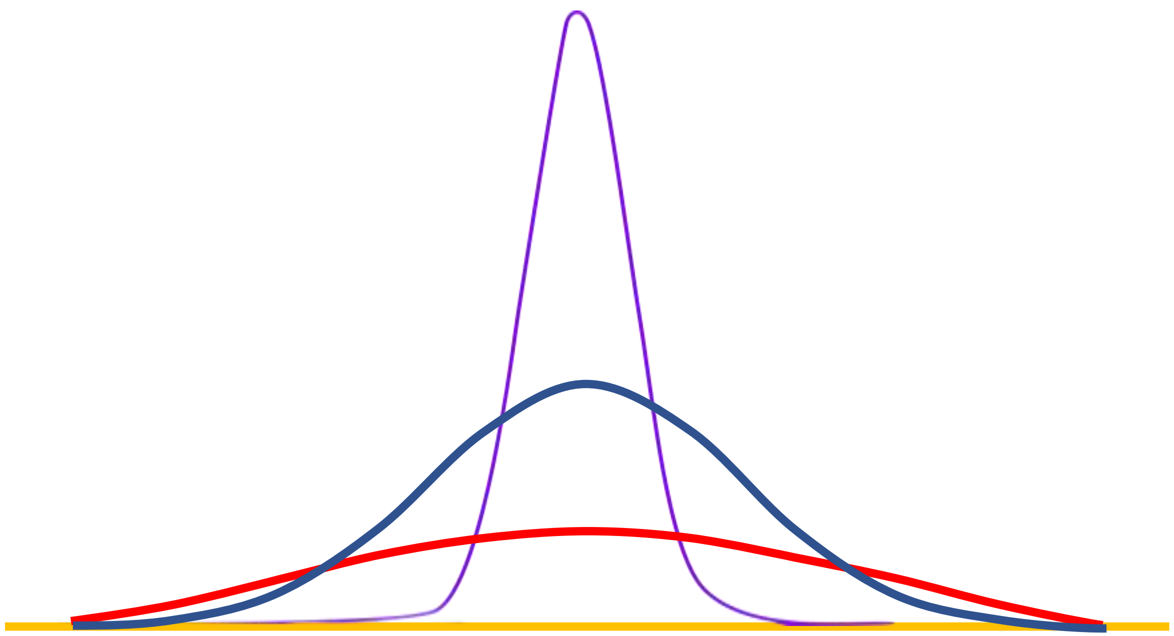
From the collected sample estimates, she intends to make inferences for the population parameters.

Obesity in Women – MLE for Normal Distribution



Obesity in Women – How do we go about the distribution?

MLE is the method that would help us in finding the value of μ and σ that will result in the bell curve that fits our data best in.



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.

Normal Distribution – Estimate Likelihood Function

The probability of observing data points that can be generated by Normal distribution is,

$$f(x_i;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

The joint probability function of $X_1, \dots, X_n$ is,

$$f(x_1,\dots\dots,x_n;\mu,\sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

The likelihood function is,

$$f(x_1,\dots\dots,x_n;\mu,\sigma) = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\sum_{i=1}^{n}\frac{x_i^2}{2\sigma^2}}$$

Taking logarithm of the likelihood,

$$\ln f(x_1,\dots\dots,x_n;\mu,\sigma) =$$

$$\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/2\sigma^2}\right) * \dots\dots * \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/2\sigma^2}\right)$$

Normal Distribution – Estimate Likelihood Function

Let's simpify to get the likelihood function by solving its part

$$= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \ln\left(e^{-(x_1-\mu)^2/2\sigma^2}\right)$$

$$= \ln\left[(2\pi\sigma^2)^{-\frac{1}{2}}\right] - \frac{(x_1-\mu)^2}{2\sigma^2}\ln e$$

$$\ln e = 1$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{2}{2}\ln(\sigma) - \frac{(x_1-\mu)^2}{2\sigma^2}$$

With all the occurences, $x_1, \ldots\ldots, x_n,$

the log of the likelihood function is given by

$$= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \sum_{i=1}^{n}\frac{(x_n-\mu)^2}{2\sigma^2}$$

Estimate Likelihood Function for mean (μ)

Now taking derivative with respect to μ and set to 0,

$$\frac{d}{d\mu}\ln f(x_1,\ldots\ldots,x_n;\mu,\sigma) = \frac{d}{d\mu}\left(-\frac{n}{2}\ln(2\pi) - n\ln\sigma - \sum_{i=1}^{n}\frac{(x_n-\mu)^2}{2\sigma^2}\right)$$

$$= 0 - 0 - \frac{2(x_n-\mu)}{2\sigma^2}$$ Apply chain rule,

$$= \frac{1}{\sigma^2}\,(x_1-\mu+\cdots+x_n-\mu)$$

$$= \frac{1}{\sigma^2}\,[(x_1+\cdots+x_n)-n\mu]$$

Estimate Likelihood Function for mean (μ)

$$0 = \frac{1}{\sigma^2}\left[(x_1 + \cdots + x_n) - n\mu\right]$$

Solving for μ,

$$\mu = \frac{(x_1 + \cdots + x_n)}{n}$$

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

The MLE of μ is $\hat{\mu} = \bar{X}$

Estimate Likelihood Function for standard deviation (σ)

Now taking derivative with respect to σ and set to 0,

$$\frac{d}{d\mu}\ln f(x_1,\ldots\ldots,x_n;\mu,\sigma) = \frac{d}{d\mu}\left(-\frac{n}{2}\ln(2\pi) - n\ln\sigma - \sum_{i=1}^{n}\frac{(x_n-\mu)^2}{2\sigma^2}\right)$$

$$0 = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum\nolimits_{i=1}^{n}(x_i^2 - \mu^2)$$

$$0 = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left[(x_1-\mu)^2 + \ldots + (x_n-\mu)^2\right]$$

Solving for σ,

$$\sigma^2 = \frac{\left[(x_1-\mu)^2 + \ldots + (x_n-\mu)^2\right]}{n}$$

Estimate Likelihood Function for standard deviation (σ)

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Take square root on both sides ,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

After substituting $\mu = \bar{X}$, we obtain

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$$

The MLE of σ is $\widehat{\sigma} = \sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$

Example – Maximum Likelihood Estimate for Mean (μ)

Problem:

A random sample of 10 weights (in pounds) of Annie's class mates are given as 115 122 130 127 149 160 152 138 149 180.

Solution:

We know that, the MLE of μ is $\hat{\mu} = \bar{X}$

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

$$\hat{\mu} = \frac{1}{10}(115 + 122 + 130 + 127 + 149 + 160 + 152 + 138 + 149 + 180)$$

$$= 142.2$$

THANK YOU

D. Uma
Department of Computer Science and Engineering

umaprabha@pes.edu

+91 99 7251 5335