Lecture 3

Optimal Estimation Theory and Its Applications

Prof. Yan LIANG

liangyan@nwpu.edu.cn

School of Automation,

Northwestern Polytechnical University

Autumn,2019

Outline

Maximum Likelihood Estimation

RA Fisher (1890-1962)

maximum likelihood estimation 1906

one of the leading scientists of the 20th century; making major contributions to Statistics, Evolutionary Biology and Genetics.

- Maximum likelihood
- Fisher information
- Analysis of variance

Maximum Likelihood Estimation

A Basic Commonsense

The matter with larger probability happens more frequently in statistics.

A Basic Experience

To a happened matter, it would be with biggest probability by the fact that such assertion has the least risk in statistics.

The derived rule

Given a measurement $Z = z$ related to X, The best estimate of X should guarantee that the probability "$Z = z$" is maximum.

Problem Formulation

Consider two variables/vectors X and Z, define the probability (density) function $p\{Z|X\}$:

- In the case that X is deterministic, $p\{Z|X\}$ represents the probability (density) function of Z with the parameter X
- In the case that X is stochastic, $p\{Z|X\}$ represents the conditional the probability (density) function of Z conditioned that X is known. $p\{Z|X\} = p\{Z, X\}/p\{X\}$

$p\{Z|X\}$ measures the level that "$X = x$" supports "$Z = z$".

Here $p\{Z|X\}$ represents $p_{Z|X}\{Z = z|X = x\}$

Problem Formulation

- In the case that the measurement Z is unknown, $p\{Z|X\}$ is the probability measure of Z.
- In the case that the measurement is available as $Z = z$ but X is unknown, $p\{Z = z|X\}$ is the function of X. It measures the level that "$X = x$" supports "$Z = z$". In such case, $p\{Z|X\}$ is called likelihood function.

The maximum-likelihood criterion

Given the measurement $Z = z$(observation information)and $p\{Z|X\}$(a priori information), The best estimate of $X = x$ should guarantee the largest support on the matter "$Z = z$".

Method Presentation

$$\hat{X}_{ML} = \arg\max_x p(Z = z|X = x)$$

If $p(z|x)$ is derivable with respect to x, then the MLE satisfies

$$\left.\frac{\partial p(z|x)}{\partial x}\right|_{x=\hat{x}_{ML}(z)} = 0 \quad \text{or} \quad \left.\frac{\partial \ln p(z|x)}{\partial x}\right|_{x=\hat{x}_{ML}(z)} = 0$$

Likelihood Equation

it belongs to the set of boundary points of $p(z|x)$

The "ln" function is monotonically increasing without changing the extreme value.

Method Presentation

$$\hat{X}_{ML} = \arg\max_x p(Z = z | X = x)$$

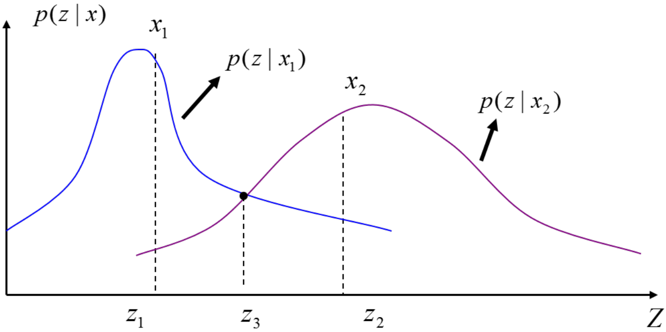
If $p(z|x)$ is derivable with respect to x, then the MLE satisfies

$$\left.\frac{\partial p(z|x)}{\partial x}\right|_{x=\hat{x}_{ML}(z)} = 0 \quad \text{or} \quad \left.\frac{\partial \ln p(z|x)}{\partial x}\right|_{x=\hat{x}_{ML}(z)} = 0$$

and $\left.\frac{\partial^2 \ln p(z|x)}{\partial x^2}\right|_{x=\hat{x}_{ML}(z)} < 0$
or its boundary point if the likelihood equation has no solution.

Method Presentation

Example2.1: Consider the unknown X may be x_1 or x_2. Please determine the MLE of X in the case that Z is z_1, z_2, or z_3, respectively.

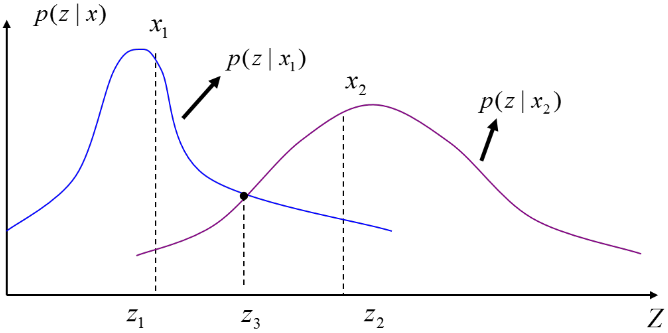


Method Presentation

To $Z = z1$, $p(z1|x1) > p(z1|x2)$ and thus $\hat{X}_{ML} = x_1$

To $Z = z2$, $p(z1|x1) < p(z1|x2)$ and thus $\hat{X}_{ML} = x_2$

To $Z = z3$, $p(z1|x1) = p(z1|x2)$ the best estimate is not unique.



Method Presentation

Example2.2: Here X is uniformly distributed

$$f_X(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1 \\ 0, & \text{otherwise} \end{cases}$$

Suppose a Z is measured, which is related to X by

$$Z = \ln(\frac{1}{X}) + V$$

where V is noise with exponential distribution

$$f_V(v) = \begin{cases} e^{-v}, & v \geqslant 0 \\ 0, & v < 0 \end{cases}$$

and X and V are independent of each other. Find the ML estimate.

Method Presentation

Solution

$$f_{Z|X}(z|x) = f_V(z - \ln(\frac{1}{x}))$$

$$= \begin{cases} e^{-(z-\ln(1/x))} & z - \ln(\frac{1}{x}) \geqslant 0 \\ 0, & z - \ln(\frac{1}{x}) < 0 \end{cases} = \begin{cases} \frac{1}{x}e^{-z} & x \geqslant e^{-z} \\ 0, & x < e^{-z} \end{cases}$$

$$\Rightarrow \hat{X}_{ML} = e^{-z}$$

Method Presentation

Example2.3: Consider one-dimensional normal random variable Z, i.e., its PDF is

$$p\{z,\sigma\} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-(z-m)^2}{2\sigma^2} \right\}$$

Given k independent observations, $z_1, \cdots, z_k$, determine

1. the MLE of the standard error σ in the case that the mean m is known;
2. the MLE of the standard error σ and the mean m.

Method Presentation

(1)

For multiple independent observations, the joint density function

$$p\{Z|X\} = \prod_{i=1}^{k} p\{z_i|X\}$$

we have

$$\text{ln } p\{z_1, \cdots, z_k|\sigma\} = \sum_{i=1}^{k} \ln \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{-(z_i - m)^2}{2\sigma^2} \right\} \right]$$

$$= \sum_{i=1}^{k} \frac{-(z_i - m)^2}{2\sigma^2} - k \ln \sqrt{2\pi} - k \ln \sigma$$

Method Presentation

$$\left.\frac{\partial \ln p\{z_1, \cdots z_k | m, \sigma\}}{\partial \sigma}\right|_{\sigma=\hat{\sigma}_{ML}} = 0$$

$$\Rightarrow \left.\frac{1}{\sigma^3}\sum_{i=1}^{k}(z_i - m)^2 - \frac{k}{\sigma}\right|_{\sigma=\hat{\sigma}_{ML}} = 0$$

$$\Rightarrow \hat{\sigma}_{ML} = \sqrt{\frac{\sum\limits_{i=1}^{k}(z_i - \hat{m})^2}{k}}$$

Method Presentation

(2)

$$\begin{cases} \left.\frac{\partial \ln p\{z_1,\cdots z_k|m,\sigma\}}{\partial m}\right|_{m=\hat{m}_{ML}} = 0 \\ \left.\frac{\partial \ln p\{z_1,\cdots z_k|m,\sigma\}}{\partial \sigma}\right|_{\sigma=\hat{\sigma}_{ML}} = 0 \end{cases} \Rightarrow \begin{array}{c} \hat{m}_{ML} = \frac{\sum\limits_{i=1}^{k} z_i}{k} \\ \hat{\sigma}_{ML} = \sqrt{\frac{\sum\limits_{i=1}^{k} (z_i-\hat{m})^2}{k}} \end{array}$$

Consideration

- Can the MLE estimate the unknown but deterministic parameters?
- Can the MLE estimate random variables?
- Is the MLE is a linear estimation scheme?
- Is the MLE suitable to discrete random variables?
- Does the MLE require the observation independence?
- Is the MLE unique?
- Why observation number is always large?

Homework

(1)Let the data Z be related to an unknown X by

$$Z = \ln X + V$$

with scalar random variables X and V independent. Suppose X is uniformly distributed between 0 and b, and noise V is Laplacian [i.e. $f_V(v) = (a/2)e^{-a|v|}$] with $a = 1$.

Find the ML estimate.

Homework

(2) To examine the effects of inaccurately known noise parameters, suppose we have

$$Z = \begin{bmatrix} 1 \\ 1 \end{bmatrix} X + V, \quad V \sim N\left(0, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

If we obtain one measurement sample

$$z = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

Find the ML estimate.

Homework

(3) A stochastic unknown X and data Z are jointly distributed according to

$$f_{XZ}(x, z) = \frac{6}{7}(x + z)^2, 0 \leqslant x \leqslant 1 \text{ and } 0 \leqslant z \leqslant 1$$

Show that

$$f_X(x) = \frac{6}{7}(x^2 + x + \frac{1}{3}), 0 \leqslant x \leqslant 1$$

$$f_{X|Z}(x|z) = \frac{x^2 + 2xz + z^2}{\frac{1}{3} + z + z^2}, 0 \leqslant x \leqslant 1$$