Advanced Signal Processing Seminar

Estimation Theory 2

Vedran DIZDAREVIC

Overview

- Linear models
- Maximum Likelihood Estimation
- Least Squares Estimation

Linear Data Model

$$x_0 = a + 0b + w_0$$
$$x_1 = a + 1b + w_1$$

$$\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

$$\mathbf{x} = \mathbf{H}\theta + \mathbf{w}$$

Linear Estimator

CRLB -> Unbiased estimator can be found iff

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \mathbf{I}(\theta)(g(\mathbf{x}) - \theta)$$

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{\mathbf{H}^T\mathbf{H}}{\sigma^2}\left[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \theta\right]$$

MVU ESTIMATOR

$$\hat{\theta} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

$$\mathbf{C}_{\hat{\theta}} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$$

Generalized Linear Estimator

Generalization for arbitrary noise conditions

If we know the correlation structure of noise

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$$

$$\hat{\theta} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}$$

Example 1 – Curve Fitting

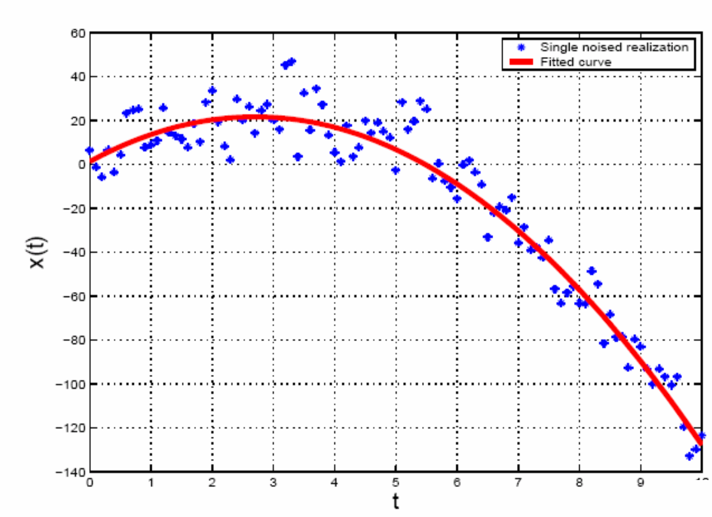
$$x(t_n) = \theta_0 + \theta_1 t_n + \theta_2 t_n^2 + w(t_n)$$

$$\mathbf{x} = [x(t_0)x(t_1)\cdots x(t_{N-1})]^T$$

$$\theta = [\theta_0\theta_1\theta_2]^T$$

$$\mathbf{H} = \begin{bmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{N-1} & t_{N-1}^2 \end{bmatrix}$$

Example 1 (2)



5000 realisations

$$\theta = [1.2 \; 15.1 \; -2.8]^T$$

$$\hat{\theta} = [1.1939 \; 15.107 \; -2.8003]^T$$

$$\sigma = 10$$

Maximum Likelihood Estimation (MLE)

- PROBABILITY $p(\mathbf{x};\theta)$

  Known parameters -> prediction of outcome

- LIKELIHOOD $L(\theta;\mathbf{x})$

  Observation of data -> estimation of parameters

Likelihood Function

$L(\theta;\mathbf{x})$ is the probability of sample occurence of a specific sample configuration x_1,...,x_n given that the probability density is known

$$L(\theta;\mathbf{x}) = p(x_1,\theta)\dots p(x_n,\theta)$$

Basic Properties

$$\hat{\theta} = \arg\max_{\theta} p(\mathbf{x};\theta)$$

$$\lim_{N\to\infty} E(\hat{\theta}) = \theta$$

$$\lim_{N\to\infty} var(\hat{\theta}) = CRLB$$

Asymptotically unbiased and efficient (MVU)

MLE Example 1 – Binomial distribution

- Problem: Given set of binary data, find the underlying probabilities p1 and p2

$$\frac{n!}{h!(n-h)!}p^h(1-p)^{n-h}$$

h ... total number of realisations (H1 + H2)
n ... outcomes of H1
p ... probability of H1

MLE Example 1 (2)

n = 1000; h=658 -> find optimal p1 and p2

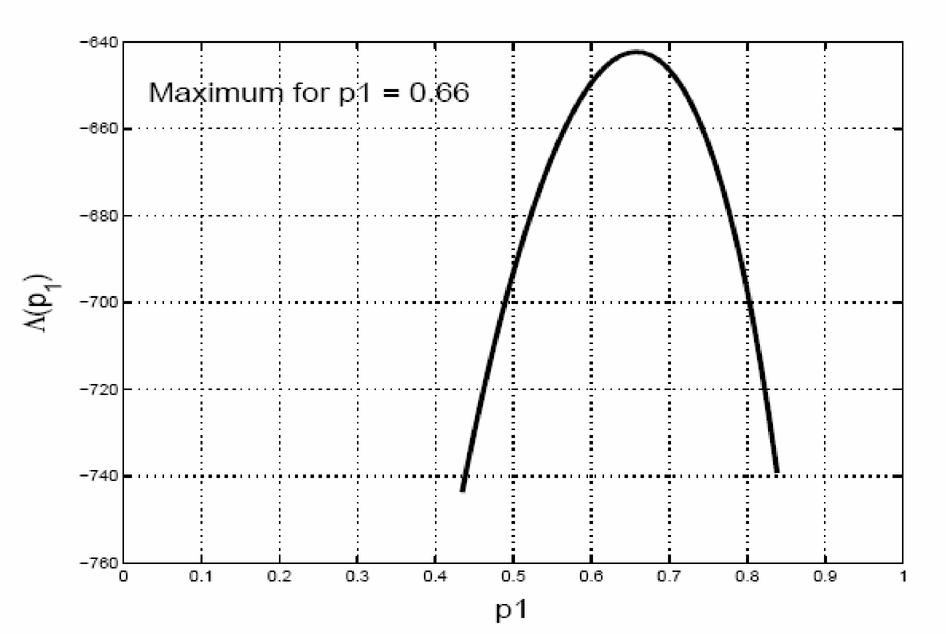
assume p1 = 0.5

$$\frac{1000!}{658!342!}0.5^{658}(1-0.5)^{342}$$

BETTER: optimize the natural logarithm of the Likelihood function!

$$\ln(0.5^{658}0.5^{342}) \approx -693.15$$

MLE Example 1 (3)



MLE Example 2 – Normal Distribution

$$L(\theta;\mathbf{x}) = L(\mu,\sigma;x_1,...,x_n) = \prod_i \frac{1}{\sigma\sqrt{2\pi}} exp\Big(\frac{-(x_i-\mu)^2}{2\sigma^2}\Big)$$

$$L(\theta;\mathbf{x}) = \frac{(2\pi)^{-n/2}}{\sigma^n} exp\Big(-\frac{\sum_i (x_i-\mu)^2}{2\sigma^2}\Big)$$

$$\Lambda(\mu,\sigma;\mathbf{x}) = \ln(L(\mu,\sigma;\mathbf{x})) = \frac{-1}{2} n\ln(2\pi) - n\ln(\sigma) - \frac{\sum_i (x_i-\mu)^2}{2\sigma^2}$$

MLE Example 2 (2) - Results

mean

$$\frac{\partial \Lambda(\mu,\sigma;\mathbf{x})}{\partial \mu} = \frac{\sum_i (x_i - \mu)}{\sigma^2} = 0 \rightarrow \hat{\mu} = \frac{\sum_i x_i}{n}$$

standard deviation

$$\frac{\partial \Lambda(\mu,\sigma;\mathbf{x})}{\partial \sigma} = \frac{-n}{\sigma} + \frac{\sum_i (x_i - \mu)^2}{\sigma^3} = 0 \rightarrow \hat{\sigma} = \sqrt{\frac{\sum_i (x_i - \hat{\mu})^2}{n}}$$

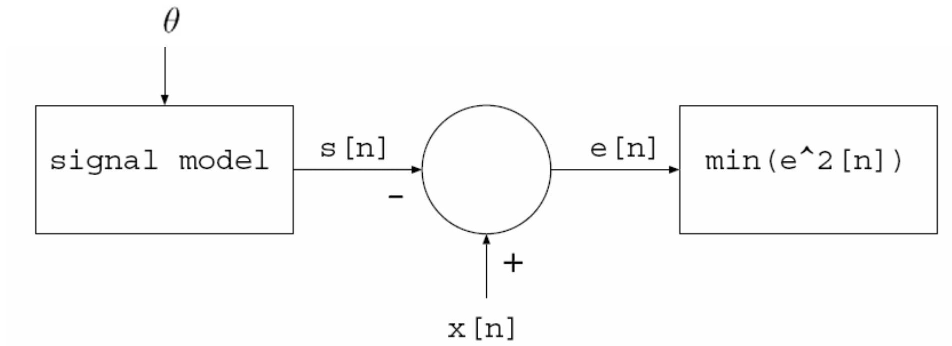
Least Squares Estimation

- generally no optimality criterion
- no probabilistic assumptions on data
- a signal model assumed

Advantage: broader range of applications

Minimization of squared difference

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - s[n])^2 = (\mathbf{x} - \mathbf{s}(\theta))^T (\mathbf{x} - \mathbf{s}(\theta))$$



s[n] deterministic
x[n] pertubated version of s[n] (noise is not explicitely modelled)

Practical Considerations

works well if:

1. we have zero-mean noise
2. some idea of model generating our data.

Classes of LS Estimators

- Class will depend on the relationship of parameter vector and our signal model

1. Linear models
2. Nonlinear models
3. Separable LSE problem

$$J(A, f_0) = \sum_{n=0}^{N-1} (x[n] - A\cos(2\pi f_0 n))^2$$

GRAZ UNIVERSITY OF TECHNOLOGY TUG

Types of Linear Estimators

- Depending on the application

1. Linear estimation (batch approach)
2. Weighted linear estimation
3. Order-recursive estimation
4. Sequential estimation

(Weighted) Linear Estimators

$$\mathbf{x} = \mathbf{H}\theta$$

$$J(\theta) = \sum_{n=0}^{N-1}(x[n] - s[n])^2 = (\mathbf{x} - \mathbf{H}\theta)^T(\mathbf{x} - \mathbf{H}\theta)$$

$$\hat{\theta} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

$$J(\theta) = (\mathbf{x} - \mathbf{H}\theta)^T\mathbf{W}(\mathbf{x} - \mathbf{H}\theta)$$

$$\hat{\theta} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{x}$$

Order Recursive LS

- Unknown order of the data model
  - Assume a data model k -> $J_{min,k}$
  - Increase model order k+1 -> $J_{min,k+1}$

$$J_{min,k+1} \leq J_{min,k}$$

Adding orthogonal vectors to the observation matrix

$$\mathbf{H}_{k+1} = [\mathbf{H}_k \ \mathbf{h}_{k+1}]$$

Sequential Least Squares

- processing data sequentialy in time
- here statistical knowledge of the noise process is essential

$$\hat{\theta}[n] = \hat{\theta}[n-1] + \mathbf{K}[n](x[n] - \mathbf{h}^T\hat{\theta}[n-1])$$

- update gain vector K[n] and covariance matrix after every iteration
- no matrix inversion needed
- estimation of initial values

Nonlinear Least Squares

$$J(\theta) = \sum_{n=0}^{N-1}(x[n] - s[n])^2 = (\mathbf{x} - \mathbf{s}(\theta))^T(\mathbf{x} - \mathbf{s}(\theta))$$

$$\frac{\partial J}{\partial \theta} = \frac{\partial \mathbf{s}(\theta)^{\mathbf{T}}}{\partial \theta}(\mathbf{x} - \mathbf{s}(\theta)) = 0$$

nonlinear regression problem
requires solution of N simultaneous nonlinear functions!

Nonlinear Least Squares (2)

- Newton-Raphson method
  - derivative of J is linearized at the current iterate

- Gauss-Newton method
  - linearizes the signal model around current parameter vector