Computer Controlled Systems

Elements of mathematical statistics
Parameter estimation

Katalin Hangos

University of Pannonia
Faculty of Information Technology
Department of Electrical Engineering and Information Systems

hangos.katalin@virt.uni-pannon.hu

April 2017

Overview

1. Random variables - a repetition
   - Vector-valued random variables
2. Elements of mathematical statistics
3. Parameter estimation of discrete time LTI stochastic system models
4. How to carry out parameter estimation in practice?

Scalar-valued random variables

The random variable ξ has a *normal or Gaussian distribution*, in notation

$$\xi \sim \mathbb{N}(m, \sigma^2) \tag{1}$$

where m is its *mean value* and σ^2 is its *variance*.
The *mean value* and *variance* of the random variable ξ with its p.d.f. f_ξ

$$E\{\xi\} = \int x f_\xi(x) dx \quad , \quad \sigma^2\{\xi\} = \int (x - E\{\xi\})^2 f_\xi(x) dx$$

The *covariance* of two scalar-valued random variables ξ and θ

$$COV\{\xi, \theta\} = E\{(\xi - E\{\xi\})(\theta - E\{\theta\})\}$$

Correlation (normed covariance):

$$\rho\{\xi, \theta\} = \frac{E\{(\xi - E\{\xi\})(\theta - E\{\theta\})\}}{\sigma\{\xi\}\sigma\{\theta\}}$$

Vector-valued random variables

Given a vector valued random variable ξ

$$\xi \ : \ \xi(\omega), \quad \omega \in \Omega, \quad \xi(\omega) \in \mathbb{R}^{\mu}$$

Its ***mean value*** *$m \in R^{\mu}$ is a real vector.*

Its ***variance*** $COV\{\xi\}$ is a square real matrix, the *covariance matrix*:

$$COV\{\xi\} = E\{(\xi - E\{\xi\})(\xi - E\{\xi\})^T\}$$

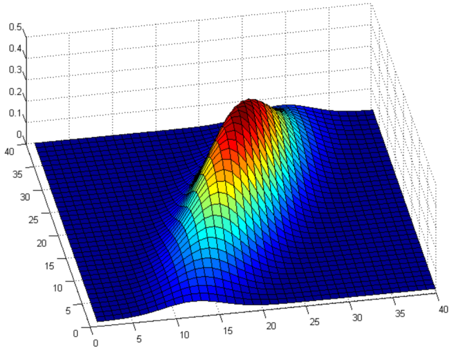
Covariance matrices are positive definite symmetric matrices:

$$z^T COV\{\xi\} z \geq 0 \quad , \quad \forall z \in \mathbb{R}^{\mu}$$

Two dimensional Gaussian distribution

Probability density function:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} - 2r\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)}$$



Linearly transformed random variables

Let us transform the vector-valued random variable $\xi(\omega) \in R^n$ using the non-singular square transformation matrix $T \in R^{n \times n}$:

$$\eta = T\xi$$

The properties of the vector-valued random variable η:

$$E\{\eta\} = TE\{\xi\} \quad , \quad COV\{\eta\} = TCOV\{\xi\}T^T$$

If the random variable ξ has a Gaussian distribution $N(m_\xi, \Delta_\xi)$ with mean value m_ξ and covariance matrix Δ_ξ, then the transformed random variable η will also be Gaussian $N(m_\eta, \Delta_\eta)$, where

$$m_\eta = Tm_\xi \quad , \quad \Delta_\eta = T\Delta_\xi T^T$$

Overview

Sample, statistics

Consider a (scalar valued) random variable ξ *with probability density function* $f_\xi(x)$.

Sample

is a collection (set) of n independent random variables

$$S(\xi) = \{\xi_1, \xi_2, ..., \xi_n\}$$

where every ξ_i has the same distribution as ξ.

- the sample corresponds to a set of *measurements* about ξ

Statistics

is a (deterministic) function of the sample elements (a random variable itself)

$$s(S) = F(\xi_1, \xi_2, ..., \xi_n)$$

- a statistics is used to construct an *estimate*

Measured data set, estimate

Consider a (scalar valued) random variable ξ *with a sample* $S(\xi) = \{\xi_1, \xi_2, ..., \xi_n\}$.

Measured data set

is a collection (set) of n measurements of the sample elements $\{\xi_1, \xi_2, ..., \xi_n\}$

$$D(\xi, n) = \{x_1, x_2, ..., x_n\}$$

D is a realization of S.

- the measured data set contains an *actual set of measurements* about ξ

Estimate

is a realization of a statistics $s(S) = F(\xi_1, \xi_2, ..., \xi_n)$

$$\hat{s}(D) = F(x_1, x_2, ..., x_n)$$

- an estimate is computed from the *actual measurement values in the data set D*

Estimation of the mean value and the variance

Assume that the underlying random variable ξ has a mean value m and the variance σ^2

- Mean value

  statistics is the *sample mean*

$$\mu(S) = \frac{1}{n}(\xi_1 + \xi_2 + ... + \xi_n) \quad , \quad \hat{m}(D) = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

  Property: $E[\mu] = m$

- Variance

  statistics is the *corrected empirical variance*

$$\theta(S) = \frac{1}{n-1}\left((\xi_1 - \mu)^2 + (\xi_2 - \mu)^2 + ... + (\xi_n - \mu)^2\right)$$

  Property: $E[\theta] = \sigma^2$

- Unbiased estimate

  if the *mean value of the statistics is the real value of the parameter to be estimated*

Estimation of the parameters of stochastic processes

Consider (scalar valued) random variables ξ_i from the same distribution but ***not independent****. They form a "generalized" sample* $S(\xi) = \{\xi_1, \xi_2, ..., \xi_n\}$.

Estimation of the mean value m

- Estimate

$$\hat{m}(D) = \frac{1}{n}(x_1 + ... + x_n)$$

- It may be a biased estimate

Estimation of the auto-covariances $r_{\xi\xi}(s),\ s = 0, 1, ...$

- Estimate for $s << n$

$$\hat{r}(D) = \frac{1}{n-s}((x_1 - \hat{m})(x_{s+1} - \hat{m}) + ... + (x_{n-s} - \hat{m})(x_n - \hat{m}))$$

- It may be a biased estimate

Parameter estimation of static models – 1

Given:

- A model that is linear in parameters $p \in \mathbb{R}^n$

$$y^{(M)} = x^T p = \sum_{i=1}^{n} x_i p_i$$

where $x \in \mathbb{R}^n$ are deterministic independent variables (measured) and $y^{(M)} \in \mathbb{R}^\mu$ is a random variable with ***measurement error***.

- From m ($m \geq \mu$) measurements we form

$$y = \begin{bmatrix} y_1 \\ y_2 \\ .. \\ y_m \end{bmatrix} \quad , \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Parameter estimation of static models – 2

- Consider a ***quadratic loss function*** L

$$L = r^T W r = \sum_{i=1}^{m} \sum_{j=1}^{m} r_i W_{ij} r_j$$
$$r_i = y_i - y_i^{(M)} = y_i - x^{(i)T} p$$

where r is the *residual vector* and W is a weighting matrix (often $W = I$)

Least squares (LS) estimate:
The $\hat{p}$ estimate of the parameters p minimizes L.

The minimum of L: $\frac{\partial L}{\partial p} = 0$

$$\hat{p} = (X^T W X)^{-1} X^T W y \quad , \quad COV\{\hat{p}\} = (X^T W X)^{-1} \sigma_r^2$$

Overview

DT-LTI SISO I/O system models

Discrete difference equation models: for SISO systems

- Deterministic model in backward difference form

$$y(k)+a_1 y(k-1)+\ldots+a_n y(k-n) = b_0 u(k-d)+\ldots+b_m u(k-d-m)$$

where $d = n_a - n_b \geq 0$ is the *pole excess (time delay)*.

  - *Compact form*

$$A^*(q^{-1})y(k) = B^*(q^{-1})u(k-d)$$

- Stochastic version with $d = 0$ and a white noise process $\{e(k)\}_{k=0}^{\infty}$

$$A^*(q^{-1})y(k) = B^*(q^{-1})u(k) + C^*(q^{-1})e(k)$$

where $C^*(q^{-1}) = c_0 + \ldots + c_n q^{-n}$

DT-LTI stochastic SISO ARX model

ARX model: DT-LTI stochastic SISO model with $C^*(q^{-1}) = 1$

$$y(k) + a_1 y(k-1) + ... + a_n y(k-n) = b_0 u(k) + ... + b_m u(k-m) + e(k)$$

Measured data set:

$$D(N) = \{(y(k), u(k)),\ k = 1, 2, ..., N\}$$

Model linear in parameters:

$$y^{(M)}(k) = -a_1 y(k-1) - ... - a_n y(k-n) + b_0 u(k) + ... + b_m u(k-m)$$

Parameters to be estimated:

$$p = [-a_1,\ ..., -a_n;\ b_0,\ ..., b_m]^T\ ,\ x^{(k)} = [y(k-1),\ ..., y(k-n);\ u(k),\ ..., u(k-$$

Estimation with least squares (linear regression)

Overview

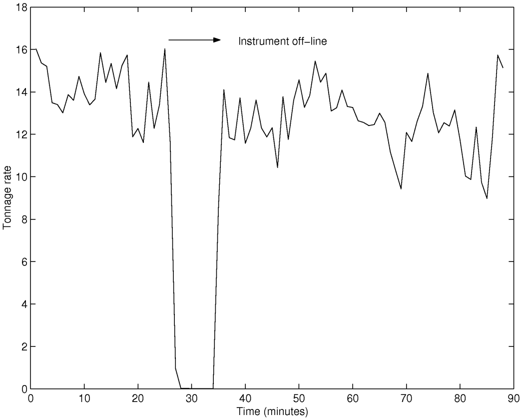
Overview of the measured data

Problems to be detected in a single time-dependent measurement

- drift
- outliers
- gross errors

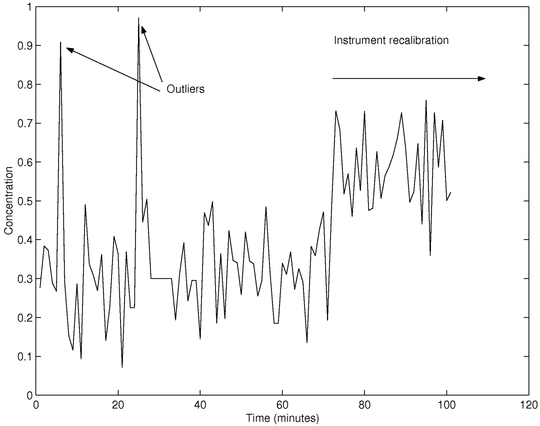
Visual overview of the measured data – 1

Dataset with gross error



Visual overview of the measured data – 2

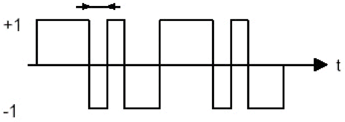
Dataset with outliers and gross error



Experiment design

Elements of the experiments to be designed

- sampling time
- magnitude of the input signals
- sufficient excitation (*white noise input is needed in the dynamic case*)



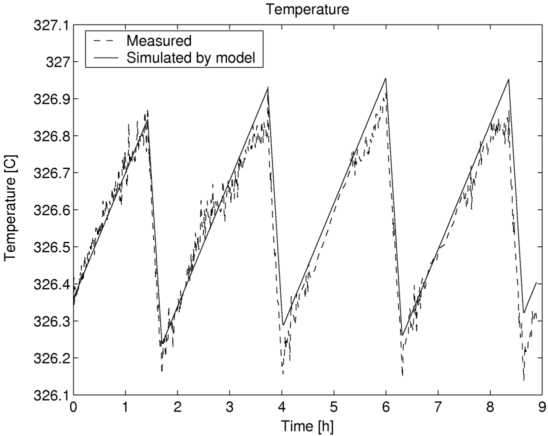
Evaluating the quality of the estimates

The evaluation can be and should be done using

- the properties of the **prediction errors** (mean values, covariances)
- the properties of the **estimates** itself (mean values, covariances, confidence regions)

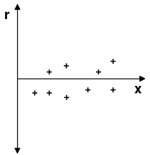
Evaluation of the prediction errors

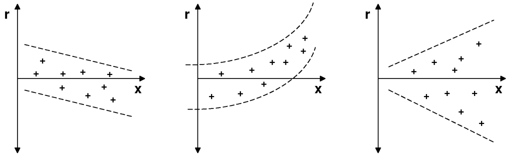
Predicted values against measured ones



Evaluation of the prediction errors

For unbiased estimates the prediction errors (residuals) are white noise processes
(independent identically distributed random variables)





Estimating the confidence region

Using the **the level set of the loss function**