PARAMETER ESTIMATION – 1

Basic notions
Elements of random variables and mathematical statistics

Katalin Hangos
Department of Electrical Engineering and Information Systems
University of Pannonia

Sept 2020

Lecture overview

1. What does "parameter estimation" mean?
   - Model, variables and parameters
   - Why estimation?
   - Course content and requirements
2. Random variables
   - Scalar-valued random variables
   - Vector-valued random variables
3. Elements of mathematical statistics
   - Sample and statistics
   - Constructing an estimate from a measured data set
   - Estimation of the mean value and the covariances
   - Estimating the probability density function - histogram
4. Tutorial

Model, variables and parameters

Relationships between data are described by a so called model

$$y = \mathcal{M}(x, p)$$

where

- the vector x is the *measurable independent variable* that we can manipulate/set in an error-free way
- the vector y is the *measurable dependent variable* (*subject to measurement errors*)
- the vector p of *constant parameters*

Important

The aim of parameter estimation is to estimate the unknown parameters p from measured sets of dependent and independent variable values $(y_i, x_i), \; i = 1, ..., n$ and given model form $\mathcal{M}$.

NOTES

Variable: a quantity that can vary or we can change it over time or from one measurement setting to another.
Variables can be set (and then their values are known) or measured.

Parameter: an unknown quantity that is normally constant over time or from one measurement setting to another.
We aim at determining the value of the parameter from measurements of the variables.

Model types

$$y = \mathcal{M}(x, p)$$

- **linear in parameters**

$$\mathcal{M}(x, p) = p^T \mathcal{F}(x)$$

  where $\mathcal{F}(x)$ is a possibly nonlinear function of the independent variable vector x

- **dynamic**
  discrete time index $k = 0, 1, ..., K, ...$ such that

$$y(k) = \mathcal{M}(x(k), x(k-1), ..., x(k-K); p) \quad , \quad k = K, K+1, ..., n$$

Examples:

- *Static, linear in parameters and variables*

$$y = ax_1 + bx_2 + c$$

  where y, x_1 and x_2 are scalar valued variables, and a, b and c are scalar parameters.

- *Static nonlinear in parameters*

$$y = ae^{bx+c}$$

  where y and x are scalar valued variables, and a, b and c are scalar parameters.

- *Dynamic linear in parameters*

$$y(k) = a_1x^2(k) + a_2x(k-1)$$

  where y and x are scalar valued time-dependent variables, and a_1 and a_2 ar
  scalar parameters.

Errors and estimates

The measurable dependent variable vector y is subject to *measurement errors*, and the model is also often not precise (*modelling errors* are also present):

$$y = \mathcal{M}(x, p) + \varepsilon \quad \left(y^{(M)} = \mathcal{M}(x, p)\right)$$

where ε, and thus also y are (vector valued) *random variables*.

Important

The result of a parameter estimation procedure can only be an "estimate" of the true model parameter vector p (denoted by $\hat{p}$), such that $\hat{p}$ is a vector valued random variable in itself.

Course structure

The course is given weakly in the form of a

- Lecture and tutorial, or a
- Laboratory with the use of MATLAB

Important (Course web page)

https://virt.uni-pannon.hu/index.php/en/education/courses/160-parameterbecsles-vemivim133p

Contents (Lectures and tutorials)

- Basic notions, Elements of random variables and mathematical statistics
- The properties of the estimates, Linear regression
- Stochastic processes, Discrete time stochastic dynamic models
- Least squares (LS) estimation by minimizing the prediction error, The properties of the LS estimation
- Special methods for LS estimation of dynamic model parameters: Instrumental variable (IV) method, Parameter estimation of dynamic nonlinear models
- Practical implementation of parameter estimation: Data checking and preparation, Evaluation of the results of parameter estimation

Evaluation

The pre-requisite of the course signature is

- to submit in the given deadline at least 90% of the homework specified on the lectures-tutorials-laboratories,
- to submit the project results and documentation to the given deadline, and
- to achieve at least 50% on the closed-book exam (the results of the homework are added to the points of the exam).

Important

The evaluation is based upon a mid-semester closed-book exam and on a parameter estimation project work to be implemented in MATLAB.

Overview

1. What does "parameter estimation" mean?
2. Random variables
   - Scalar-valued random variables
   - Vector-valued random variables
3. Elements of mathematical statistics
4. Tutorial

Scalar-valued random variables

A scalar-valued random variable ξ is characterized by its *probability density function (p.d.f.)* $f_\xi : \mathbb{R} \mapsto \mathbb{R}_{\geq 0}$.

Properties

The *mean value* and *variance* of the random variable ξ with its p.d.f. f_ξ are

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

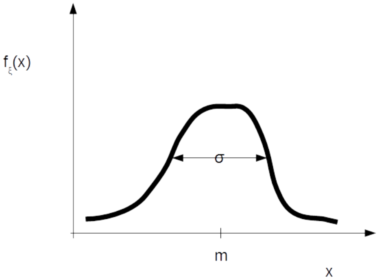
Normally distributed scalar-valued random variables

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

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

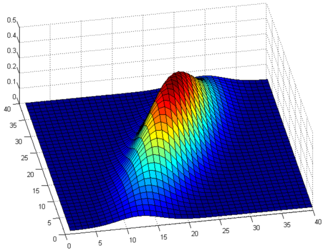
where m is its *mean value* and σ^2 is its *variance*, when

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{(x-m)^2}{\sigma^2}\right)}$$



Independence of two scalar-valued random variables

The joint distribution of two scalar-valued random variables ξ and θ is characterized by their *joint p.d.f* $f_{\xi,\theta}(x, y) : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}_{\geq 0}$.



Important

Two scalar-valued random variables ξ and θ are called independent, if $f_{\xi,\theta}(x, y) = f_\xi(x) \cdot f_\theta(y)$.

Covariance and correlation of random variables

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

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

where $\overline{\xi} = (\xi - E\{\xi\})$ is a *centered random variable*.

Correlation (normed covariance): $\rho\{\xi, \theta\} = \frac{E\{(\xi - E\{\xi\})(\theta - E\{\theta\})\}}{\sigma\{\xi\}\sigma\{\theta\}}$

Important

Independence implies $\rho\{\xi, \theta\} = 0$, but $\rho\{\xi, \theta\} = 0$ implies independence only in case of Gaussian joint distribution.

Important

The covariance of a scalar-valued random variables ξ with itself is its variance , i.e. $COV\{\xi, \xi\} = \sigma^2\{\xi\}$

Vector valued random variables – 1

Given a vector valued random variable ξ

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

Scalar valued entries of vector valued random variables

$$\xi = \begin{bmatrix} \xi_1 \\ \dots \\ \xi_\mu \end{bmatrix}$$

where each entry ξ_i is a scalar valued random variable

Vector valued random variables – 2

Given a vector valued random variable $\xi \in \mathbb{R}^\mu$

- Its **mean value** $m \in R^\mu$ *is a real vector.*
- Its **variance** $COV\{\xi\}$ is a square real matrix, the *covariance matrix*:

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

Important

Covariance matrices are positive definite symmetric matrices:

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

Covariance matrix and covariances

Consider a two dimensional vector valued random variable

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$$

with $E\{\xi_i\} = m_i$, and its centered version $\overline{\xi}_i = \xi_i - m_i$

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

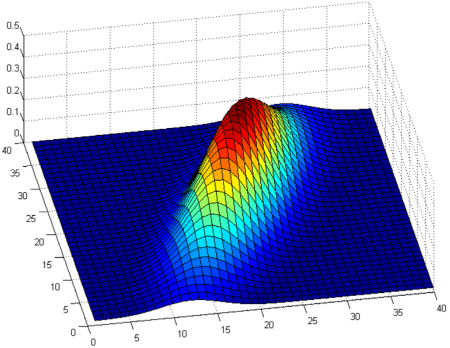
$$E\{\overline{\xi\xi}^T\} = E\left\{\begin{bmatrix} \overline{\xi}_1^2 & \overline{\xi}_1\overline{\xi}_2 \\ \overline{\xi}_1\overline{\xi}_2 & \overline{\xi}_2^2 \end{bmatrix}\right\} = \begin{bmatrix} \sigma^2\{\xi_1\} & COV\{\xi_1,\xi_2\} \\ COV\{\xi_1,\xi_2\} & \sigma^2\{\xi_1\} \end{bmatrix}$$

diagonal: variances, off-diagonal: covariances

Two dimensional Gaussian distribution

Probability density function:

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



Two dimensional Gaussian distribution - 1

Probability density function:

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

Assume non-correlated elements ξ_1 and ξ_2 with $r = 0$. Then

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)} = f_1(x_1) \cdot f_2(x_2)$$

with $f_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{(x_i-m_i)^2}{\sigma_i^2}\right)}$

Therefore ξ_1 and ξ_2 are independent.

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:

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

Important (Gaussian case)

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

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 that are **not** random variables but deterministic values (a realization).

Estimates

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

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

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*

Important

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

Estimation of the mean value – 1

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

- Statistics for the mean value: *sample mean*

$$\mu(S) = \frac{1}{n}(\xi_1 + \xi_2 + ... + \xi_n)$$

  Property: $E[\mu] = m$ $===>$ unbiased

- Estimate of the mean value

$$\hat{m}(D) = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

Estimation of the variance

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

- Statistics for the variance: *corrected empirical variance*

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

  with $\mu(S) = \frac{1}{n}(\xi_1 + \xi_2 + ... + \xi_n)$
  Property: $E[\theta] = \sigma^2$ ===> unbiased

- Estimate of the variance

$$\hat{\sigma}^2(D) = \frac{1}{n-1}\left((x_1 - \hat{m}(D))^2 + ... + (x_n - \hat{m}(D))^2\right)$$

  with $\hat{m}(D) = \frac{1}{n}(x_1 + x_2 + ... + x_n)$

Estimation of the mean value – 2

Assume that the underlying μ-dimensional vector valued random variable ξ has a mean value $m \in \mathbb{R}^\mu$. The sample is a collection of independent vector valued random variables

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

where $\xi_i = [\xi_{i,1}, \ ..., \xi_{i,\nu}]^T$ and the independence is considered entry-wise.

- Statistics for the mean value: *sample mean*

$$\mu(S) = \frac{1}{n}(\xi_1 + \xi_2 + ... + \xi_n)$$

  Property: $E[\mu] = m \quad ===> \quad$ unbiased

- Estimate of the mean value

$$\hat{m}(D) = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

  is computed entry-wise.

Estimation of the covariances

Assume that the underlying μ-dimensional vector valued random variable ξ has a mean value $m \in \mathbb{R}^\mu$. The sample is $S(\xi) = \{\xi_1, ..., \xi_n\}$ and $\overline{\xi}_i = \xi_i - m_i$ is the centered version of the sample element ξ_i.

- Statistics for the covariances of the entries (i, j)

$$\rho_{ij}(S) = \frac{1}{n-1} \sum_{k=1}^{n} (\overline{\xi}_{k,i} \cdot \overline{\xi}_{k,j})$$

- Estimate of the covariance $r_{ij} = COV\{\xi_{\cdot,i}, \xi_{\cdot,j}\}$

$$\hat{r}_{ij}(D) = \frac{1}{n-1} \sum_{k=1}^{n} (x_{k,i} - \hat{m}_i) \cdot (x_{k,j} - \hat{m}_j)$$

with $\hat{m}_i = \frac{1}{n-1} \sum_{k=1}^{n} x_{k,i}$

Estimation of the auto-covariances – 1

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 using the sample mean as statistics

- Estimate

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

- This is an unbiased estimate

Estimation of the auto-covariances – 2

Consider (scalar valued) random variables ξ_i *from the same distribution and with a* **pairwise constant covariance** $r = COV\{\xi_i, \xi_{i+1}\}$ and with a "generalized" sample $S(\xi) = \{\xi_1, \xi_2, ..., \xi_n\}$.

The estimate of the mean value m is

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

The estimate of the autocovariance r is

$$\hat{r}(D) = \frac{1}{n-1} \sum_{i=1}^{n-1} ((x_i - \hat{m})(x_{i+1} - \hat{m}))$$

It may be a biased estimate

Histogram construction

Consider a (scalar valued) random variable ξ with the probability density function $f_\xi(z)$ and a sample $S(\xi) = \{\xi_1, \xi_2, ..., \xi_n\}$.

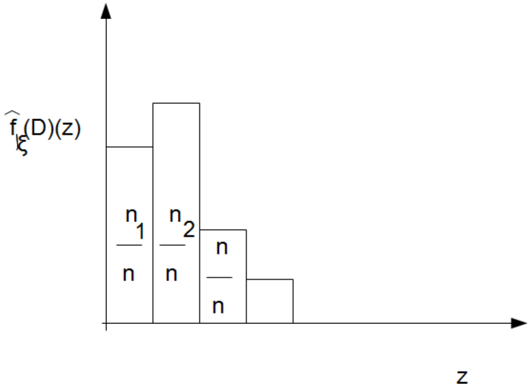
Histogram construction

- Let x_M the maximal and x_m the minimal element of the data set D with n elements.
- Divide the interval $[x_m, x_M]$ into ℓ sub-intervals ($\delta = \frac{x_M - x_m}{\ell}$) such that $z_i = x_m + (i-1)\delta$
- Denote by n_i the number of data set elements in the interval $[z_i, z_{i+1}]$

Estimate of $f_\xi(z)$:
the piece-wise constant function $\hat{f}_\xi(D)(z)$ such that

$$\hat{f}_\xi(D)(z) = \frac{n_i}{n} \text{ for } z \in [z_i, z_{i+1}] \ , \ i = 1, ..., \ell$$

A simple histogram



Tutorial problems

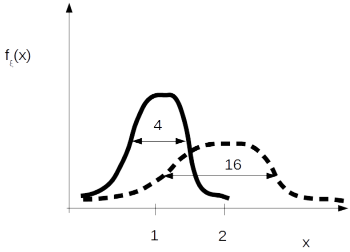
- A. Vector valued random variables
- B. Mean value and covariance estimation

Tutorial problems – A

Example (Vector valued random variables – 1)

Given two scalar-valued Gaussian random variables $\eta_1 \sim \mathbb{N}(1,4)$, *and* $\eta_2 \sim \mathbb{N}(2,16)$.

- *Plot the probability density functions* f_{η_1} *and* f_{η_2} *of random variables* η_1 *and* η_2 *in the same coordinate system!*



Tutorial problems – A

Example (Vector valued random variables – 2)

Given two scalar-valued Gaussian random variables $\eta_1 \sim \mathbb{N}(1, 4)$, *and* $\eta_2 \sim \mathbb{N}(2, 16)$.
Assume that the random variables η_1 *and* η_2 *are independent and form a vector valued random variable* $\eta = [\eta_1, \eta_2]^T$ *from them.*

- *Which type of distribution does the vector valued random variable* η *have?*
  vector valued Gaussian (2 dimensional)

- *Compute the mean value and the variance (covariance matrix) of the vector valued random variable* η.

$$m_\eta = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad , \quad COV\{\eta\} = \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix}$$

Tutorial problems – A

Example (Vector valued random variables – 3)

Given two scalar-valued Gaussian random variables $\eta_1 \sim \mathbb{N}(1,4)$*, and* $\eta_2 \sim \mathbb{N}(2,16)$*.*
Assume that the random variables η_1 *and* η_2 *have a covariance* $COV(\eta_1,\eta_2) = 2.3$ *and form a vector valued random variable* $\eta = [\eta_1,\eta_2]^T$ *from them.*

- *Which type of distribution does the vector valued random variable* η *have?*
  vector valued Gaussian (2 dimensional)
- *Compute the mean value and the variance of* η*.*

$$m_\eta = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \;, \quad COV\{\eta\} = \begin{bmatrix} 4 & 2.3 \\ 2.3 & 16 \end{bmatrix}$$

Tutorial problems – B

Example (Mean value and covariance estimation – 1)

Consider a scalar valued random variable ξ with a measured data set

$$D(5) = \{0.5,\ -0.6,\ 0.3,\ -0.2,\ 0.0\}$$

- *Compute an estimate of the mean value of ξ.*
  statistics: sample mean
  $\hat{m} = \frac{0.5-0.6+0.3-0.2+0.0}{5} = 0$
- *Compute an estimate of the variance of ξ.*
  statistics: corrected empirical variance
  $\hat{\sigma}^2 = \frac{0.5^2-0.6^2+0.3^2-0.2^2+0.0^2}{4}$
- *Could the measured data be independent? Compute an estimate of $r = COV\{\xi_i, \xi_{i+1}\}$.* $\hat{r} = \frac{-0.5\cdot 0.6-0.6\cdot 0.3-0.2\cdot 0.3-o.2\cdot 0.0}{4} << 0$
  $==>$ *NOT independent*

HOMEWORK

Consider a scalar valued random variable ξ and with a measured data set

$$D(5) = \{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5\}$$

- Compute an estimate of the mean value of ξ.
- Compute an estimate of the variance of ξ.
- Could the measured data be independent? Compute an estimate of $r = COV\{\xi_i, \xi_{i+1}\}$.