# Mathematics for chemical engineers

Drahoslava Janovská

## 2. Linear and nonlinear regression

# Outline

# Evaluation of experimental data

## Evaluation of experimental data

- Solving of chemical-engineering problems usually let us derive the model of process or phenomenon taking place in the device.
- Frequently, we are not able to identify numerical values of model parameters.

Function $\eta(x) = E(Y(x))$ defined on the domain $A \subset \mathbb{R}$ is called regression function. Regression is the relationship between $E(Y(x))$ – the mean value of random variable $Y(x)$ – and the independent variable $x$.

Assume that we know the form of the regression function. Based on random selection, we estimate its unknown parameters:
We choose $n$ values of independent variable $x_j \in A,\ j = 1, \ldots, n$, and for each $x_j$ observe (measure) the realization (value) $y_j$ of random variable $Y_j$:

$$x_j \in A,\ j = 1, \ldots, n \quad \longrightarrow \quad y_j = Y(x_j)\,.$$

Obtained pairs of values $(x_1, y_1), \ldots, (x_n, y_n)$ are used for the estimation of unknown parameters.

# Basic model of linear regression

Model of linear regression must fulfil:

1. $\eta(x)$ is a linear function of the form

$$\eta(x) = \sum_{k=1}^{p} \beta_k f_k(x)\,,$$

where $f_k(x)$ are known functions, and $\beta_k,\ k = 1, \ldots, p$, unknown parameters. The function $\eta$ is linear in parameters.

2. Value $x_j$ is assigned random variable $Y_j$, for which it reads

$$E(Y_j) = \eta(x_j)\,, \quad D(Y_j) = \sigma^2\,, \quad j = 1, \ldots, n\,,$$

The second equation means that the variance is independent of $x_j$, and thus it is constant. For example, it corresponds to the case that all realizations $y_1, \ldots, y_n$ of random variables $Y_1, \ldots, Y_n$ are measured with the same precision.

3. Matrix $F = (f_{ij})$, where $f_{ij} = f_i(x_j)$, $i = 1, \ldots, p$, $j = 1, \ldots, n$, has the rank $p$. Note that the number $n$ of pairs $(x_j, y_j)$ must be greater than the number of unknown parameters $p$, precisely, it should hold true $n - p > 2$.
4. Random variables $Y_1, \ldots, Y_n$ are not correlated, i.e.

$$\operatorname{cov}(Y_i, Y_j) = 0, \quad i, j = 1, \ldots, n, \quad i \neq j.$$

Matrix notation

$$C_y = \sigma^2 E_n,$$

where $E_n$ is the identity matrix of order $n$, $C_y$ is the covariance matrix of variables $Y_1, \ldots, Y_n$.

**Example** A regression line, i.e. the regression function of the form $\eta(x) = \alpha + \beta x$ has the number of unknown parameters $p = 2$, and $\beta_1 = \alpha$, $f_1 = 1$, $\beta_2 = \beta$, $f_2 = x$.

Described model in the equivalent form is

$$Y_j = \eta(x_j) + \varepsilon_j = \sum_{k=1}^{p} \beta_k \, f_{kj} + \varepsilon_j \,, \quad j = 1, \ldots, n \,, \tag{1}$$

where values $x_1, \ldots, x_n$ are values of nonrandom variables, values $f_{kj} = f_k(x_j)$ fulfil the 3. condition of the model. Random errors $\varepsilon_j$ , $j = 1, \ldots, n$ , and the covariance matrix $C_\varepsilon$ of the random vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ fulfil

$$E(\varepsilon_j) = 0 \,, \quad j = 1, \ldots, n \,, \qquad C_\varepsilon = \sigma^2 \, E_n = C_y \,.$$

Equation (1) reads in the matrix form

$$\vec{Y} = F^{\mathrm{T}} \vec{\beta} + \vec{\varepsilon} \,.$$

# Least squares method

We find unknown parameters $\beta_1, \ldots, \beta_p$ in described model of linear regression by least squares method. Let these estimations are $b_1, \ldots, b_p$, which are selection functions of random choice $Y_1, \ldots, Y_n$. We minimize the sum of squares of deviations from observed values $y_j$ and their mean values $\eta_j = \eta(x_j)$, thus the sum of squares is

$$Q(\beta_1, \ldots, \beta_p) = \sum_{j=1}^{n} (y_j - \eta_j)^2 = \sum_{j=1}^{n} \left( y_j - \sum_{k=1}^{p} \beta_k f_{kj} \right)^2 .$$

Thus estimations $b_1, \ldots, b_p$ are found as a solution of the set of equations

$$\frac{\partial Q}{\partial \beta_k} = 0, \quad k = 1, \ldots, p.$$

This system is called the system of normal equations.

We can rewrite the system of normal equations for searched estimations $b_1, \ldots, b_p$ in a lucid form

$$\begin{array}{ccccccccc} b_1 S_{11} & + & b_2 S_{12} & + & \cdots & + & b_p S_{1p} & = & S_{1y} \\ b_1 S_{21} & + & b_2 S_{22} & + & \cdots & + & b_p S_{2p} & = & S_{2y} \\ & & & & & & & \vdots & \\ b_1 S_{p1} & + & b_2 S_{p2} & + & \cdots & + & b_p S_{pp} & = & S_{py}\,, \end{array}$$

where

$$S_{ki} = \sum_{j=1}^{n} f_{kj}\, f_{ij}\,, \quad i,k = 1,\ldots,p\,,$$

$$S_{ky} = \sum_{j=1}^{n} f_{kj}\, y_j\,, \quad k = 1,\ldots,p\,.$$

Clearly $S_{ik} = S_{ki}$ for $i,k = 1,\ldots,p\,.$

Matrix notation:
If $\overrightarrow{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, $\overrightarrow{b} = (b_1, \ldots, b_p)^{\mathrm{T}}$, then normal equations can be written in the form

$$F\,F^{\mathrm{T}}\,\overrightarrow{b} = F\,\overrightarrow{y}\,. \tag{2}$$

Assume that $h(F) = p$, then also $h(F\,F^{\mathrm{T}}) = p$ and $F\,F^{\mathrm{T}}$ is of the type $p \times p$, regular $\Longrightarrow$ exists $(F\,F^{\mathrm{T}})^{-1}$, and thus from equation (2) we can express vector $\overrightarrow{b}$:

$$\overrightarrow{b} = (F\,F^{\mathrm{T}})^{-1}\,F\overrightarrow{y}\,,$$

the vector $\overrightarrow{b}$ is uniquely determined and its each component is a linear combination of values $y_1, \ldots, y_n$.

Attention! Calculation is extremely numerical unstable, see the lecture "Linear algebra".

# ★ Example

Let a regression line goes through the beginning, $\eta(x) = a\,x$, then $f_j = x_j$ and $\beta_1 = a$. Denote $\overrightarrow{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$, $F = (x_1, \ldots, x_n)$. Then

$$F\,F^{\mathrm{T}} = (x_1, \ldots, x_n) \cdot (x_1, \ldots, x_n)^{\mathrm{T}} = \sum_{j=1}^{n} x_j^2 \quad \Longrightarrow \quad (FF^{\mathrm{T}})^{-1} = \frac{1}{\sum_{j=1}^{n} x_j^2}\,.$$

Estimation of the parameter $a$ is

$$a = (F\,F^{\mathrm{T}})^{-1} F\overrightarrow{y} = \left(\frac{1}{\sum_{j=1}^{n} x_j^2}\right) \overrightarrow{x}^{\mathrm{T}}\overrightarrow{y} = \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} x_j^2}\,.$$

# ★ Example

Experiments were conducted in which ice crystals were placed into a compartment at a constant temperature (-5°C). In order to analyze the growth of the ice crystals as a function of time, the saturation of the air by water was kept constant. The experimental data points were randomized over time. The experimental data are presented in the following table, where $y$ is the axial length of the crystals in microns and $x$ is the time in seconds. Repeated measurements were also performed in order to examine the lack of fit. Use a straight–line model, $y = \beta_0 + \beta_1 x$ and fit it to the data.

| $x[s]$ | $y[mm]$ | $x[s]$ | $y[mm]$ |
|---|---|---|---|
| 50 | 19 | 125 | 28 |
| 60 | 20, 21 | 130 | 31, 32 |
| 70 | 17, 22 | 135 | 34, 25 |
| 80 | 25, 28 | 140 | 26, 33 |
| 90 | 21, 25, 31 | 145 | 31 |
| 95 | 25 | 150 | 36, 33 |
| 100 | 30, 29, 33 | 155 | 41, 33 |
| 105 | 35, 32 | 160 | 40, 30, 37 |
| 110 | 30, 28, 30 | 165 | 32 |
| 115 | 31, 36, 30 | 170 | 35 |
| 120 | 36, 25, 28 | 180 | 38 |

# ★ Solution

The experimental data are in vector and matrix notation

$$y = \begin{bmatrix} 19 \\ 20 \\ 21 \\ 17 \\ \vdots \\ 35 \\ 38 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 50 \\ 1 & 60 \\ 1 & 60 \\ 1 & 70 \\ \vdots & \vdots \\ 1 & 170 \\ 1 & 180 \end{bmatrix},$$

where $y$ is $n \times 1$ vector and $X$ is an $n \times p$ matrix; $n = 43$ is the total number of experimental points, and $p$ represents the number of parameters, in this case two: $\beta_0, \beta_1$. The first column of $X$ should only contain 1 in each position. The number of different $x-$positions is $m = 22$. As there are many repeated experiments, $m$ is significantly lower than the total number of experimental points $n$. We denote $n_i$ the number of observations for each $x_i$, $i = 1, \ldots, m$, and $n = \sum_{i=1}^{m} n_i$.The model parameters can be calculated as follows:

$$b = \begin{bmatrix} \beta_0, \beta_1 \end{bmatrix} = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} y = \begin{bmatrix} 14.19 \\ 0.1346 \end{bmatrix}.$$

We obtained the model $y = 14.19 + 0.1346\, x$ .

# ★ Unbiased estimation of linear parametric function

**Task** We find estimation of the linear function of parameters $\overrightarrow{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$. Assume a parametric function

$$\gamma = \sum_{k=1}^{p} c_k \beta_k = \overrightarrow{c}^{\mathrm{T}} \cdot \overrightarrow{\beta} \,,$$

where $\overrightarrow{c} = (c_1, \ldots, c_p)^{\mathrm{T}}$ is known nonzero vector ($\overrightarrow{c} \neq 0$).

**Statement** The best estimation of linear parametric function $\overrightarrow{c}^{\mathrm{T}} \cdot \overrightarrow{\beta}$ is a function (statistics) $g = \overrightarrow{c}^{\mathrm{T}} \cdot \overrightarrow{b}$, where $\overrightarrow{b}$ is the solution of normal equations. $E(g) = \gamma$, and "the best" means that the variance $D(g)$ is minimal in the class of unbiased estimations.

# Nonlinear regression

**Goal:** Estimation of parameters $a_1, \ldots, a_n$ in nonlinear empirical formulae

$$y = f(\mathbf{x}, \mathbf{a}) .$$

We will minimize the sum of squares of deviations

$$S(\mathbf{a}) = \sum_{j=1}^{m} \left( f(x^j, a) - y^j \right)^2 = \sum_{j=1}^{m} q_j^2(a) ,$$

where $q_j$ is the residuum of $j$th measured point. Denote by $a^+$ the point in which the sum of squares $S(a)$ has its minimum. The value $a^+$ is obtained as a limit of so called minimizing sequence $a^k$ in such a way that

$$S(a^{k+1}) < S(a^k) .$$

Taylor series of function $f$ (neglect terms of higher order than 1):

$$f(x,a) \approx f(x,a^k) + \mathrm{grad}_a^{\mathrm{T}} f(x,a^k)\,(a-a^k) \quad \Longleftrightarrow$$

$$f(x,a) \approx f(x,a^k) + \sum_{j=1}^{n} \frac{\partial f(x,a^k)}{\partial a_j}(a_j - a_j^k)\,.$$

Evaluate the approximation formulae

$$y - f(x,a^k) = \sum_{j=1}^{n} \frac{\partial f(x,a^k)}{\partial a_j} \triangle a_j^k\,.$$

Let $\Gamma(a)$ be the Jacobi matrix,

$$\Gamma(a) = \begin{pmatrix} \dfrac{\partial f(x^1,a)}{\partial a_1} & \dfrac{\partial f(x^1,a)}{\partial a_2} & \cdots & \dfrac{\partial f(x^1,a)}{\partial a_n} \\ \vdots & & & \vdots \\ \dfrac{\partial f(x^m,a)}{\partial a_1} & \dfrac{\partial f(x^1,a)}{\partial a_2} & \cdots & \dfrac{\partial f(x^m,a)}{\partial a_n} \end{pmatrix}\,.$$

We search for a solution:

$$\triangle^{+} a^{k} = -\left(\Gamma^{\mathrm{T}}(a^{k})\,\Gamma(a^{k})\right)^{-1}\Gamma^{\mathrm{T}}(a^{k})q(a^{k})\,,$$

where $q = (q_1, \ldots, q_m)$. Using $\triangle^{+} a^{k}$ we calculate the next iteration

$$a^{k+1} = a^{k} + \lambda \triangle^{+} a^{k}\,, \quad \lambda \in (0, 1\rangle\,.$$

The starting value: $\lambda = 1$. If $S(a^{k+1}) \geq S(a^{k})$, then we decrease $\lambda$.

The calculation is performed for

$$\underbrace{\Gamma^{\mathrm{T}}(a^{k})\,\Gamma(a^{k})}_{\text{matrix } n \times n}\triangle^{+} a^{k} = -\Gamma^{\mathrm{T}}(a^{k})\,q(a^{k})\,.$$

Process is stopped, if $||\triangle^{+} a^{k}||$ is less than a required precision.

# Recommended literature


- Cox D.R., Donnelly C. A.: Principles of Applied Statistics. Cambridge University Press, 2011.
- Motulsky H., Christopoulos A.: Fitting Models to Biological Data using Linear and Nonlinear Regression. A practical guide to curve fitting. 2003, GraohPad Software Inc. San Diego CA, www.graphpad.com.
- Rasmuson A., Andersson B., Olsson L., Andersson R.: Mathematical Modeling in Chemical Engineering. Cambridge University Press, 2014.