Applied Data Analysis

Exercise Sheet 3

Exercise 11

Consider a linear model

$$\boldsymbol{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

as given in Definition I.4.2 with (unknown) parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ and design matrix $B = [B_1| \ldots |B_p] \in \mathbb{R}^{n\times p}$. I.e., $B_1, \ldots, B_p$ denote the column vectors of the matrix B.

For $i \in \{1, \ldots, p\}$, show that the individual parameter β_i of the parameter vector $\boldsymbol{\beta}$ is identifiable if and only if $B_i \notin \operatorname{span}\big(\{B_j \,\big|\, j \in \{1, \ldots, p\}\backslash\{i\}\}\big)$, i.e. B_i is not representable as linear combination of B_j, $j \in \{1, \ldots, p\}\backslash\{i\}$.

Definition:

(i) Let $A \in \mathbb{R}^{n\times n}$ be a real *square* matrix. A representation of A as matrix product

$$A = Q\,R$$

with an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ and an upper triangular matrix $R \in \mathbb{R}^{n\times n}$ is called a *QR-decomposition* of A.

(ii) More generally, let $A \in \mathbb{R}^{n\times m}$ be a real *rectangular* matrix with $n \geq m$. Then, a factorization

$$A = Q\,R$$

with a matrix $Q \in \mathbb{R}^{n\times m}$ consisting of m orthogonal column vectors and an upper triangular matrix $R \in \mathbb{R}^{m\times m}$ is called a *QR-decomposition* of A.

Within the appendix of this Exercise Sheet (on page 3), a constructive way of deriving a QR-decomposition of a rectangular matrix using the *Gram-Schmidt-procedure* is described.

Exercise 12

(a) Consider a linear model

$$\boldsymbol{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

as in I.4.2 with $\operatorname{rank}(B) = p \leq n$ and let $B = Q\,R$ be a QR-decomposition of B.

Then, show that the least-square-estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = R^{-1}\,Q'\,\boldsymbol{Y}\ .$$

(b) Use the *Gram-Schmidt-procedure* to derive a QR-decomposition of $A := \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}$.

(c) Now, consider the linear model

$$\boldsymbol{Y} = A\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with parameter vector $\boldsymbol{\beta} \in \mathbb{R}^2$ and $\boldsymbol{\varepsilon}$ given as in Definition I.4.2.

Use the results of (a) and (b) to derive a least-square-estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ for the observation $\boldsymbol{y} = (y_1, y_2, y_3)' = (1, 1, 1)'$.

In Definition I.4.2 it is assumed (as an implication of the stochastical independence of $\varepsilon_1, \ldots, \varepsilon_n$) that $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 I_n$ with some $\sigma > 0$.

That is, the error terms $\varepsilon_1, \ldots, \varepsilon_n$ and, thus, also the observations $Y_1, \ldots, Y_n$ are supposed to be uncorrelated and are supposed to have same finite variance σ^2 (*homoscedasticity*).

In practice, this assumption is often violated. If, for example, one assumes a regression model for a time series, one may see that the scatter around the regression line increases with time. In other situations, the variance may increase with the magnitude of the observations.

In these cases, the method of ordinary least squares (OLS) – as given in I.4.2 – may lead to statistically inefficient procedures.

Definition:

Consider the linear model

$$\text{(1)} \qquad \boldsymbol{Y} \;=\; X\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with a unknown parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \Theta \subseteq \mathbb{R}^p$, a known matrix $X \in \mathbb{R}^{n\times p}$ and error terms $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$ assuming $\text{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\,\Sigma$, with $\sigma > 0$ unknown and a known variance-covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$.

Then, if Σ is regular (and thus positive definite), the method of *generalized least squares* (GLS) estimates the parameter vector $\boldsymbol{\beta}$ by minimizing the function

$$\boldsymbol{\beta} \mapsto \widetilde{\psi}(\boldsymbol{\beta}) \;:=\; \big|\big|\boldsymbol{y} - X\,\boldsymbol{\beta}\big|\big|_\Sigma^2 \;=\; (\boldsymbol{y} - X\,\boldsymbol{\beta})'\,\Sigma^{-1}\,(\boldsymbol{y} - X\,\boldsymbol{\beta})$$

i.e. the squared Mahalanobis distance of the residual vector for a given realization $\boldsymbol{y}$ of $\boldsymbol{Y}$.

Each solution $\hat{\boldsymbol{\beta}}$ is called *generalized-least-squares-estimate* of $\boldsymbol{\beta}$.

Exercise 13

Consider the model in (1) according to the definition above with $\text{rank}(X) = p \leq n$ and regular matrix Σ. Furthermore, let $\hat{\boldsymbol{\beta}}$ denote the (uniquely determined) GLS estimator of $\boldsymbol{\beta}$.

(a) Find a linear transformation $\boldsymbol{Z} = A\,\boldsymbol{Y}$ with a suitable chosen matrix $A \in \mathbb{R}^{n\times p}$ such that

$$\text{(2)} \qquad \boldsymbol{Z} \;=\; B\,\boldsymbol{\beta} + \boldsymbol{\eta}$$

is a linear model according to Definition I.4.2 with some matrix $B \in \mathbb{R}^{n\times p}$ and error term $\boldsymbol{\eta}$ fulfilling $\text{Cov}(\boldsymbol{\eta}) = \sigma^2\,I_n$.

(b) Derive the normal equations for $\hat{\boldsymbol{\beta}}$ (as solution of the minimization problem given above).

(c) Show the following representation of the GLS estimator $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} \;=\; (X'\,\Sigma^{-1}\,X)^{-1}\,X\,\Sigma^{-1}\,\boldsymbol{Y}\;.$$

(d) Calculate $\text{E}(\hat{\boldsymbol{\beta}})$ and $\text{Cov}(\hat{\boldsymbol{\beta}})$.

Now, for the following last parts, additionally assume $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2\,\Sigma)$.

(e) Find the likelihood equations for $\boldsymbol{\beta}$ considering model (1) and use the results of the Lecture and (a) to sow that the corresponding maximum likelihood estimator coincides with the GLS estimator $\hat{\boldsymbol{\beta}}$.

(f) Analogously to the testing procedure considered in I.4.40, derive a testing procedure for the model (1).

Hint to (a): Exercise 2.

Appendix

Derivation of a QR-decomposition using the Gram-Schmidt-procedure

Consider a matrix $A = [a_1|\dots|a_m] \in \mathbb{R}^{n\times m}$ with $\operatorname{rank}(A) = m \le n$. Then, let

$$\begin{aligned}
\boldsymbol{u_1} &:= \boldsymbol{a_1} \quad \text{and} & \boldsymbol{e_1} &:= \frac{\boldsymbol{u_1}}{||\boldsymbol{u_1}||}\\
\boldsymbol{u_2} &:= \boldsymbol{a_2} - \langle \boldsymbol{a_2}, \boldsymbol{e_1}\rangle \boldsymbol{e_1} \quad \text{and} & \boldsymbol{e_2} &:= \frac{\boldsymbol{u_2}}{||\boldsymbol{u_2}||}\\
&\dots\\
\boldsymbol{u_m} &:= \boldsymbol{a_m} - \langle \boldsymbol{a_m}, \boldsymbol{e_1}\rangle \boldsymbol{e_1} - \dots - \langle \boldsymbol{a_m}, \boldsymbol{e_{m-1}}\rangle \boldsymbol{e_{m-1}} \quad \text{and} & \boldsymbol{e_m} &:= \frac{\boldsymbol{u_m}}{||\boldsymbol{u_m}||}
\end{aligned}$$

Then, the set of vectors $\{\boldsymbol{e_1}, \dots, \boldsymbol{e_m}\}$ is orthonormal and for each subset $I \subseteq \{1, \dots, m\}$, the set $\{\boldsymbol{e_i} \,|\, i \in I\}$ has the same span as the corresponding set $\{\boldsymbol{a_i} \,|\, i \in I\}$. The above algorithm is called the *Gram-Schmidt procedure*.

The Gram-Schmidt procedure results in the following QR-decomposition of the matrix A:

$$A \;=\; Q\,R \quad \text{with} \quad Q := [e_1|\dots|e_n] \quad \text{and} \quad R := \begin{pmatrix} \langle \boldsymbol{a_1}, \boldsymbol{e_1}\rangle & \langle \boldsymbol{a_2}, \boldsymbol{e_1}\rangle & \dots & \langle \boldsymbol{a_m}, \boldsymbol{e_1}\rangle \\ 0 & \langle \boldsymbol{a_2}, \boldsymbol{e_2}\rangle & \dots & \langle \boldsymbol{a_m}, \boldsymbol{e_2}\rangle \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \langle \boldsymbol{a_m}, \boldsymbol{e_m}\rangle \end{pmatrix}.$$

The procedure can be understood as first normalizing the vector $\boldsymbol{a_1}$ to a unit vector $\boldsymbol{e_1}$ and then, for each $k \in \{2, \dots, m\}$, successively first eliminating the components in the directions of the previous vectors $\boldsymbol{a_1}, \dots, \boldsymbol{a_{k-1}}$ (thus ensuring orthogonality) before again normalizing to a unit vector $\boldsymbol{e_m}$. In particular, this yields the following representations:

$$\begin{aligned}
\boldsymbol{a_1} &= \langle \boldsymbol{e_1}, \boldsymbol{a_1}\rangle\, \boldsymbol{e_1}\\
\boldsymbol{a_2} &= \langle \boldsymbol{e_1}, \boldsymbol{a_2}\rangle\, \boldsymbol{e_1} + \langle \boldsymbol{e_2}, \boldsymbol{a_2}\rangle \boldsymbol{e_2}\\
&\dots\\
\boldsymbol{a_m} &= \sum_{j=1}^{m} \langle \boldsymbol{e_j}, \boldsymbol{a_k}\rangle\, \boldsymbol{e_j}
\end{aligned}$$

with respect to the new orthonormal basis $\{\boldsymbol{e_1}, \dots, \boldsymbol{e_m}\}$ of the span of the column vectors.

Remark

Note that for a given rectangular matrix $A \in \mathbb{R}^{n\times m}$ with $n \ge m$, some computer algebra programs use the convention

$$A \;=\; Q\,R \;=\; \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0_{(n-m)\times m} \end{pmatrix} \;=\; Q_1\,R_1$$

where $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix and $R_1 \in \mathbb{R}^{m\times m}$ is a an upper triangular matrix. The submatrix $Q_2 \in \mathbb{R}^{n\times(n-m)}$ is, in general, not unique and completes the column vectors of $Q_1 \in \mathbb{R}^{n\times m}$ to an orthonormal basis of $\mathbb{R}^n$.