Tutorium

Allgemeines Lineares Modell

BSc Psychologie SoSe 2023

7. Termin: (6) Parameterschätzung

Sean Mulready

Inhalte basieren auf Kursmaterialien für ALM von Dirk Ostwald, lizenziert unter CC BY-NC-SA 4.0

Parameterschätzung - Selbstkontrollfragen

1. Geben Sie das Theorem zum Betaparameterschätzer wieder.
2. Warum ist der Betaparameterschätzer ein Maximum-Likelihood Schätzer?
3. Geben Sie das Theorem zum Varianzparameterschätzer wieder.
4. Geben Sie die Parameterschätzer bei n unabhängigen und identisch normalverteilten Zufallsvariablen wieder.
5. Geben Sie die Parameterschätzer bei einfacher linearer Regression wieder.
6. Geben Sie das Theorem zur Verteilung des Betaparameterschätzers wieder.
7. Geben Sie das Theorem zur Verteilung des Varianzparameterschätzers wieder.

1. Geben Sie das Theorem zum Betaparameterschätzer wieder.

Theorem (Betaparameterschätzer)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{1}$$

das ALM und es sei

$$\hat{\beta} := (X^T X)^{-1} X^T \upsilon. \tag{2}$$

der *Betaparameterschätzer*. Dann gilt, dass

$$\hat{\beta} = \arg\min_{\tilde{\beta}} (\upsilon - X\tilde{\beta})^T (\upsilon - X\tilde{\beta}) \tag{3}$$

und dass $\hat{\beta}$ ein unverzerrter Maximum Likelihood Schätzer von $\beta \in \mathbb{R}^p$ ist.

Beispiel für $\hat{\beta}$ bei ALM mit $n = 5$ und $p = 2$

Wir betrachten das ALM

$$\upsilon \sim N(X\beta, \sigma^2 I_5) \text{ mit } X \in \mathbb{R}^{5\times 2}, \beta \in \mathbb{R}^2, \sigma^2 > 0.$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \upsilon_3 \\ \upsilon_4 \\ \upsilon_5 \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \\ x_{51} & x_{52} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix} = \begin{pmatrix} x_{11}\beta_1 + x_{12}\beta_2 + \varepsilon_1 \\ x_{21}\beta_1 + x_{22}\beta_2 + \varepsilon_2 \\ x_{31}\beta_1 + x_{32}\beta_2 + \varepsilon_3 \\ x_{41}\beta_1 + x_{42}\beta_2 + \varepsilon_4 \\ x_{51}\beta_1 + x_{52}\beta_2 + \varepsilon_5 \end{pmatrix}$$

Dann sieht die Betaparameterschätzerformel ausgeschrieben wie folgt aus

$$\hat{\beta} := (X^T X)^{-1} X^T \upsilon = \left(\begin{pmatrix} x_{11} & x_{21} & x_{31} & x_{41} & x_{51} \\ x_{12} & x_{22} & x_{32} & x_{42} & x_{52} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \\ x_{51} & x_{52} \end{pmatrix} \right)^{-1} \times \begin{pmatrix} x_{11} & x_{21} & x_{31} & x_{41} & x_{51} \\ x_{12} & x_{22} & x_{32} & x_{42} & x_{52} \end{pmatrix} \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \upsilon_3 \\ \upsilon_4 \\ \upsilon_5 \end{pmatrix}$$

Allgemeine Theorie - Selbstkontrollfragen

Beispiel für $\hat{\beta}$ bei ALM mit $n = 5$ und $p = 2$ (fortgeführt)

$$
\begin{aligned}
&= \begin{pmatrix} x_{11}^2 + x_{21}^2 + x_{31}^2 + x_{41}^2 + x_{51}^2 & x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} + x_{41}x_{42} + x_{51}x_{52} \\ x_{12}x_{11} + x_{22}x_{21} + x_{32}x_{31} + x_{42}x_{41} + x_{52}x_{51} & x_{12}^2 + x_{22}^2 + x_{32}^2 + x_{42}^2 + x_{52}^2 \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} x_{11} & x_{21} & x_{31} & x_{41} & x_{51} \\ x_{12} & x_{22} & x_{32} & x_{42} & x_{52} \end{pmatrix} \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \upsilon_3 \\ \upsilon_4 \\ \upsilon_5 \end{pmatrix} \\
&= \quad \ldots \\
&= \quad \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \hat{\beta} \in \mathbb{R}^p
\end{aligned}
$$

Wenn wir uns die Dimensionen der einzelnen Terme anschauen, wird klar, dass das Ergebnis am Ende 2×1-dimensional ist. Im Detail:

- Für den ersten Term $(X^T X)$ halten wir fest, dass die Inverse A^{-1} einer Matrix $A \in \mathbb{R}^{n \times n}$ die gleiche Größe hat wie die Matrix, also $A^{-1} \in \mathbb{R}^{n \times n}$. Entsprechend hat das Ergebnis des ersten Terms 2 Zeilen und 2 Spalten, i.e. (2×2)
- Für den zweiten Term $(X^T y)$ gilt $(2 \times 5)(5 \times 1) = (2 \times 1)$.
- Somit ergibt sich insgesamt für $\hat{\beta}$ die Dimension $(2 \times 2)(2 \times 1) = (2 \times 1)$.

2.Warum ist der Betaparameterschätzer ein Maximum-Likelihood Schätzer?

Wir betrachten für ein festes $y \in \mathbb{R}^n$ und ein festes $\sigma^2 > 0$ die Log-Likelihood Funktion

$$
\begin{aligned}
\ln N(y; X\tilde{\beta}, \sigma^2 I_n) &= \ln \left((2\pi)^{-\frac{n}{2}} |\sigma^2 I_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right) \right) \\
&= -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 I_n| - \frac{1}{2\sigma^2} (\upsilon - X\tilde{\beta})^T (\upsilon - X\tilde{\beta})
\end{aligned} \tag{4}
$$

Ein Maximum-Likelihood Schätzer **maximiert** diese Funktion.

Der einzige Term der betrachteten Log-Likelihoodfunktion, der von $\tilde{\beta}$ abhängt ist

$$
-\frac{1}{2\sigma^2} (\upsilon - X\tilde{\beta})^T (\upsilon - X\tilde{\beta}). \tag{5}
$$

Da aber gilt, dass die Summe der quadrierten Abweichungen $(\upsilon - X\tilde{\beta})^T (\upsilon - X\tilde{\beta}) \geq 0$ ist, wird der Term aufgrund des negativen Vorzeichens maximal, wenn die Summe der quadrierten Abweichungen minimal wird. Dies ist genau für $\tilde{\beta} = \hat{\beta}$ der Fall, wie im Beweis (1) gezeigt wurde (Folien 11 & 12).

Siehe auch Lösung SKF Regression, SKF 14 Teil 1 (Folie 18).

3.Geben Sie das Theorem zum Varianzparameterschätzer wieder.

Theorem (Varianzparameterschätzer)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{6}$$

das ALM in generativer Form. Dann ist

$$\hat{\sigma}^2 := \frac{(\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta})}{n-p} \tag{7}$$

ein unverzerrter Schätzer von $\sigma^2 > 0$.

4.Geben Sie die Parameterschätzer bei n unabhängigen und identisch normalverteilten Zufallsvariablen wieder.

Wir betrachten das Szenario von n unabhängig und identisch normalverteilten Zufallsvariablen mit Erwartungswertparameter $\mu \in \mathbb{R}$ und Varianzparameter σ^2. D.h., für jede Komponente des Datenvektors gilt $\upsilon_i \sim N(\mu, \sigma^2)$ für $i = 1, \ldots, n$. Äquivalent dazu können wir das generative Modell schreiben als

$$\upsilon_i = \mu + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \ldots, n, \text{ mit unabhängigen } \varepsilon_i$$

In Matrixschreibweise:

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}^1, \sigma^2 > 0.$$

$$\upsilon = X\beta + \varepsilon = X\mu + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mu + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \mu + \varepsilon_1 \\ \vdots \\ \mu + \varepsilon_n \end{pmatrix}$$

Dann gilt für die Beta- und Varianzparameterschätzer

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \upsilon_i =: \bar{\upsilon} \text{ und } \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\upsilon_i - \bar{\upsilon})^2 =: s_\upsilon^2.$$

In diesem Szenario ist der Betaparameterschätzer mit dem Stichprobenmittel $\bar{\upsilon}$ der $\upsilon_1, \ldots, \upsilon_n$ und der Varianzparameterschätzer mit der Stichprobenvarianz s_υ^2 der $\upsilon_1, \ldots, \upsilon_n$ identisch.

4.Geben Sie die Parameterschätzer bei n unabhängigen und identisch normalverteilten Zufallsvariablen wieder.

Anmerkungen:

- $\hat{\beta} := (X^T X)^{-1} X^T \upsilon$ ist wie wir im Theorem (Betaparameterschätzer) gelernt haben, ein erwartungstreuer Schätzer für $\beta \in \mathbb{R}^p$ des ALM.
- Weiterhin haben wir gelernt, dass im Szenario von n u.i.(normal-)v. Zufallsvariablen gilt, dass $X := 1_n \in \mathbb{R}^n$, und somit $\hat{\beta} =: \bar{\upsilon}$
- Entsprechend ist $\bar{\upsilon}$ ein erwartungstreuer Schätzer für β im ALM Szenario von n unabh. u.i.(normal-)v. Zufallsvariablen.
- Analog ist s_{υ}^2 ein erwartungstreuer Schätzer für σ^2 im ALM Szenario von n u.i.(normal-)v. Zufallsvariablen.

5.Geben Sie die Parameterschätzer bei einfacher linearer Regression wieder.

Wir betrachten das Modell der einfachen linearen Regression

$$\upsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \ldots, n.$$

In Matrixschreibweise

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2, \sigma^2 > 0.$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \beta_0 + x_1\beta_1 + \varepsilon_1 \\ \vdots \\ \beta_0 + x_n\beta_1 + \varepsilon_n \end{pmatrix}$$

Dann gilt für die Beta- und Varianzparameterschätzer

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{\upsilon} - \frac{c_{x\upsilon}}{s_x^2}\bar{x} \\ \frac{c_{x\upsilon}}{s_x^2} \end{pmatrix} \text{ und } \hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(\upsilon_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2,$$

wobei $\bar{x}$ und $\bar{\upsilon}$ die Stichprobenmittel der $x_1, \ldots, x_n$ und $\upsilon_1, \ldots, \upsilon_n$, respektive, bezeichnen, $c_{x\upsilon}$ die Stichprobenkovarianz der $x_1, \ldots, x_n$ und $\upsilon_1, \ldots, \upsilon_n$ bezeichnet, und s_x^2 die Stichprobenvarianz der $x_1, \ldots, x_n$ bezeichnet.

6. Geben Sie das Theorem zur Verteilung des Betaparameterschätzers wieder.

Theorem (Frequentistische Verteilung des Betaparameterschätzers)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{8}$$

das ALM. Weiterhin sei

$$\hat{\beta} := (X^T X)^{-1} X^T \upsilon \tag{9}$$

der Betaparameterschätzer. Dann gilt

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}). \tag{10}$$

7.Geben Sie das Theorem zur Verteilung des Varianzparameterschätzers wieder.

Theorem (Frequentistische Verteilung des Varianzparameterschätzers)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{11}$$

das ALM. Weiterhin sei

$$\hat{\sigma}^2 = \frac{(\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta})}{n-p} \tag{12}$$

der Varianzparameterschätzer. Dann gilt

$$\frac{n-p}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-p) \tag{13}$$