

### week10: Effizienz der Maximum-Likelihood-Schätzer

Es gelten dieselben Bezeichnungen wie in den weeks 8+9. Die Maximum-Likelihood-Schätzer waren durch die folgenden Formeln gegeben:

$$\hat{\beta}_{\text{ML}} = (X^T X)^{-1} X^T \vec{y} \quad (1)$$

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{n} [P_{X^\perp} \vec{y}]^2 \quad (2)$$

Die  $\hat{\beta}_{j,\text{ML}}$  waren erwartungstreu,

$$\mathbb{E}[\hat{\beta}_{j,\text{ML}}] = \beta_j$$

aber das  $\hat{\sigma}_{\text{ML}}^2$  war nur im Limes  $n \rightarrow \infty$  erwartungstreu. Deshalb hatten wir den erwartungstreuen Schätzer  $\hat{s}^2$  für die Varianzen der Fehler  $\varepsilon_i$  definiert, das war also

$$\hat{s}^2 = \frac{1}{n-(p+1)} [P_{X^\perp} \vec{y}]^2 \quad (3)$$

Wir wollen jetzt zeigen, dass diese Schätzer effizient sind. Das heisst, dass sie innerhalb einer grösseren Klasse von Schätzern eine minimale Varianz besitzen. Wir betrachten zunächst die Schätzer  $\hat{\beta}_{j,\text{ML}}$  für die Regressionskoeffizienten  $\beta_j$  und definieren die folgende Menge von linearen, erwartungstreuen Schätzern

$$\mathcal{L} := \left\{ \tilde{\beta} : \mathbb{R}^n \rightarrow \mathbb{R}^{p+1} \mid \tilde{\beta}(\vec{y}) = A\vec{y} + \vec{b}, A \in \mathbb{R}^{(p+1) \times n}, \vec{b} \in \mathbb{R}^{p+1}, \mathbb{E}[\tilde{\beta}(\vec{y})] = \vec{\beta} \right\} \quad (4)$$

wobei die  $\vec{y} = (y_1, \dots, y_n)$  durch

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \quad (5)$$

gegeben sind mit  $N(0, \sigma)$ -normalverteilten unabhängigen Zufallszahlen  $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$  und  $X$  wie üblich die Matrix der Regressoren mit den Spaltenvektoren  $\vec{x}_0, \vec{x}_1, \dots, \vec{x}_p \in \mathbb{R}^n$ . Dann gilt das folgende

**Theorem 10.1:** Der Maximum-Likelihood-Schätzer  $\hat{\beta}_{\text{ML}}$  aus Gleichung (1) ist effizient in der Klasse von linearen, erwartungstreuen Schätzern  $\mathcal{L}$  gegeben durch Gleichung (4). Das heisst genauer, für jeden Schätzer  $\tilde{\beta}$  aus  $\mathcal{L}$  gilt:

$$\mathbb{V}[\tilde{\beta}_j] \geq \sigma^2 (X^T X)^{-1}_{j,j} = \mathbb{V}[\hat{\beta}_j] . \quad (6)$$

**Beweis:** Wegen

$$\begin{aligned} \mathbb{E}[\tilde{\beta}] &= \mathbb{E}[A\vec{y} + \vec{b}] \\ &= \mathbb{E}[A(X\vec{\beta} + \vec{\varepsilon}) + \vec{b}] \\ &= AX\vec{\beta} + A\mathbb{E}[\vec{\varepsilon}] + \vec{b} \\ &= AX\vec{\beta} + \vec{b} \stackrel{!}{=} \vec{\beta} \end{aligned}$$

muss gelten

$$AX = Id \quad (7)$$

$$\vec{b} = 0 \quad (8)$$

Also,

$$\tilde{\beta}(\vec{y}) \stackrel{(8)}{=} A\vec{y} \stackrel{(5)}{=} AX\vec{\beta} + A\vec{\varepsilon} \stackrel{(7)}{=} \vec{\beta} + A\vec{\varepsilon}$$

und wir bekommen

$$\begin{aligned} \mathbb{V}[\tilde{\beta}_j] &= \mathbb{E}[(\tilde{\beta}_j - \beta_j)^2] \\ &= \mathbb{E}[(A\vec{\varepsilon})_j(A\vec{\varepsilon})_j] \\ &= \sigma^2 (AA^T)_{j,j} \end{aligned} \quad (9)$$

wobei wir in der letzten Zeile von (9) das Lemma 8.2 verwendet haben. Wir müssen zeigen:  
Die rechte Seite von (9) wird minimiert durch

$$A = (X^T X)^{-1} X^T \quad (10)$$

Dazu schreiben wir

$$A = (X^T X)^{-1} X^T + D$$

mit  $D := A - (X^T X)^{-1} X^T$ . Wegen

$$[(X^T X)^{-1} X^T]^T = X[(X^T X)^{-1}]^T = X[(X^T X)^T]^{-1} = X(X^T X)^{-1}$$

bekommen wir dann

$$\begin{aligned} AA^T &= [(X^T X)^{-1} X^T + D][(X^T X)^{-1} X^T + D]^T \\ &= [(X^T X)^{-1} X^T + D][X(X^T X)^{-1} + D^T] \\ &= (X^T X)^{-1} X^T X (X^T X)^{-1} + DX(X^T X)^{-1} + (X^T X)^{-1} X^T D^T + DD^T \\ &= (X^T X)^{-1} + DX(X^T X)^{-1} + (X^T X)^{-1} (DX)^T + DD^T \end{aligned}$$

Nun ist aber

$$DX = [A - (X^T X)^{-1} X^T] X = AX - (X^T X)^{-1} X^T X \stackrel{(7)}{=} Id - Id = 0$$

und wir erhalten

$$AA^T = (X^T X)^{-1} + DD^T$$

Wegen

$$(DD^T)_{j,j} = \sum_{k=1}^n D_{j,k} (D^T)_{k,j} = \sum_{k=1}^n D_{j,k} D_{j,k} \geq 0$$

erhalten wir damit aus (9)

$$\mathbb{V}[\tilde{\beta}_j] = \sigma^2 (AA^T)_{j,j} \geq \sigma^2 (X^T X)^{-1}_{j,j}$$

und das Theorem 10.1 ist bewiesen. ■

Jetzt wollen wir ein analoges Theorem für den Schätzer

$$\begin{aligned} \hat{s}^2 &= \hat{s}^2(\vec{y}) = \frac{1}{n-(p+1)} [P_{X^\perp} \vec{y}]^2 \\ &= \frac{1}{n-(p+1)} \langle P_{X^\perp} \vec{y}, P_{X^\perp} \vec{y} \rangle \\ &= \frac{1}{n-(p+1)} \langle \vec{y}, P_{X^\perp}^T P_{X^\perp} \vec{y} \rangle \\ &= \frac{1}{n-(p+1)} \langle \vec{y}, P_{X^\perp} P_{X^\perp} \vec{y} \rangle \\ &= \frac{1}{n-(p+1)} \langle \vec{y}, P_{X^\perp} \vec{y} \rangle \end{aligned} \quad (11)$$

formulieren. Dazu definieren wir die folgende Menge von positiven, quadratischen und erwartungstreuen Schätzern

$$\mathcal{Q} := \left\{ \tilde{s}^2 : \mathbb{R}^n \rightarrow \mathbb{R} \mid \tilde{s}^2(\vec{y}) = \langle \vec{y}, A\vec{y} \rangle + \langle \vec{b}, \vec{y} \rangle + c \geq 0, \right. \\ \left. A \in \mathbb{R}^{n \times n}, \vec{b} \in \mathbb{R}^n, c \in \mathbb{R}, \mathbb{E}[\tilde{s}^2(\vec{y})] = \sigma^2 \right\} \quad (12)$$

wobei die  $\vec{y} = (y_1, \dots, y_n)$  wieder durch (5) gegeben sind mit  $N(0, \sigma)$ -normalverteilten unabhängigen Zufallszahlen  $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ . Dann gilt das folgende

**Theorem 10.2:** Der Schätzer  $\hat{s}^2$  aus Gleichung (11) ist effizient in der Klasse von positiven, quadratischen und erwartungstreuen Schätzern  $\mathcal{Q}$  gegeben durch Gleichung (12). Das heisst genauer, für jeden Schätzer  $\tilde{s}^2$  aus  $\mathcal{Q}$  gilt:

$$\mathbb{V}[\tilde{s}^2] \geq \frac{2\sigma^4}{n-(p+1)} = \mathbb{V}[\hat{s}^2] \quad (13)$$

**Beweis:** Wir zeigen zunächst die folgende

Behauptung 1: Für alle  $\tilde{s}^2 \in \mathcal{Q}$  und mit  $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$  können wir schreiben

$$\tilde{s}^2(\vec{y}) = \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle \quad (14)$$

und wir können o.B.d.A. annehmen, dass  $A$  symmetrisch ist,  $A = A^T$ . Weiterhin gilt

$$AX = 0 \quad (15)$$

Beweis Behauptung 1: Wegen

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon}$$

ist

$$\begin{aligned} \tilde{s}^2(\vec{y}) &= \tilde{s}^2(X\vec{\beta} + \vec{\varepsilon}) = \langle (X\vec{\beta} + \vec{\varepsilon}), A(X\vec{\beta} + \vec{\varepsilon}) \rangle + \langle \vec{b}, (X\vec{\beta} + \vec{\varepsilon}) \rangle + c \\ &= \langle X\vec{\beta}, AX\vec{\beta} \rangle + \langle \vec{\varepsilon}, AX\vec{\beta} \rangle + \langle X\vec{\beta}, A\vec{\varepsilon} \rangle + \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle \\ &\quad + \langle \vec{b}, X\vec{\beta} \rangle + \langle \vec{b}, \vec{\varepsilon} \rangle + c \end{aligned} \quad (16)$$

und damit

$$\sigma^2 \stackrel{!}{=} \mathbb{E}[\tilde{s}^2(\vec{y})] = \langle X\vec{\beta}, AX\vec{\beta} \rangle + \langle \vec{b}, X\vec{\beta} \rangle + c + \sigma^2 \text{Tr}(A) \quad (17)$$

Dabei haben wir die Identitäten

$$\begin{aligned} \mathbb{E}[\langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle] &= \sum_{i,j=1}^n a_{i,j} \mathbb{E}[\varepsilon_i \varepsilon_j] \\ &= \sum_{i,j=1}^n a_{i,j} \sigma^2 \delta_{i,j} \\ &= \sigma^2 \sum_{i=1}^n a_{ii} = \sigma^2 \text{Tr}(A) \end{aligned}$$

und

$$\mathbb{E}[\langle \vec{\varepsilon}, AX\vec{\beta} \rangle] = \mathbb{E}[\langle X\vec{\beta}, A\vec{\varepsilon} \rangle] = \mathbb{E}[\langle \vec{b}, \vec{\varepsilon} \rangle] = 0$$

benutzt. Für beliebige  $\vec{\beta}$  und beliebiges  $\sigma^2$  muss also gelten:

$$\langle \vec{\beta}, X^T AX\vec{\beta} \rangle + \langle X^T \vec{b}, \vec{\beta} \rangle + c + \sigma^2 \text{Tr}(A) \stackrel{!}{=} \sigma^2 \quad (18)$$

Aus (18) folgt

$$\begin{aligned} X^T AX &= 0 \\ X^T \vec{b} &= 0 \end{aligned} \quad (19)$$

und

$$c = 0 \quad (20)$$

$$\text{Tr}(A) = 1 \quad (21)$$

Wegen  $c = 0$  und der Positivität der Schätzer muss weiterhin gelten

$$\tilde{s}^2(\vec{y}) = \langle \vec{y}, A\vec{y} \rangle + \langle \vec{b}, \vec{y} \rangle \stackrel{!}{\geq} 0 = \tilde{s}^2(\vec{y} = \vec{0}) \quad (22)$$

Das heisst,  $\vec{y} = \vec{0}$  ist globales Minimum von  $\tilde{s}^2(\vec{y})$  und das wiederum bedeutet, dass der Gradient von  $\tilde{s}^2$  an der Stelle  $\vec{y} = \vec{0}$  verschwinden muss,

$$\nabla \tilde{s}^2(\vec{y}) \Big|_{\vec{y}=\vec{0}} = \vec{b} = \vec{0} \quad (23)$$

Also gilt für jedes  $\tilde{s}^2 \in \mathcal{Q}$

$$\tilde{s}^2(\vec{y}) = \langle \vec{y}, A\vec{y} \rangle \quad (24)$$

Wegen

$$\langle \vec{y}, A\vec{y} \rangle = \langle A\vec{y}, \vec{y} \rangle = \langle \vec{y}, A^T \vec{y} \rangle$$

können wir schreiben

$$\begin{aligned} \tilde{s}^2(\vec{y}) &= \langle \vec{y}, A\vec{y} \rangle \\ &= \frac{1}{2} \left\{ \langle \vec{y}, A\vec{y} \rangle + \langle \vec{y}, A^T \vec{y} \rangle \right\} \\ &= \left\langle \vec{y}, \frac{A+A^T}{2} \vec{y} \right\rangle \end{aligned}$$

so dass wir o.B.d.A. annehmen können, dass die Matrix  $A$  symmetrisch ist,

$$A = A^T$$

Aus Gleichung (16) folgt dann

$$\begin{aligned} \tilde{s}^2(\vec{y}) = \tilde{s}^2(X\vec{\beta} + \vec{\varepsilon}) &= \langle X\vec{\beta}, AX\vec{\beta} \rangle + \langle \vec{\varepsilon}, AX\vec{\beta} \rangle + \langle X\vec{\beta}, A\vec{\varepsilon} \rangle + \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle \\ &\quad + \langle \vec{b}, X\vec{\beta} \rangle + \langle \vec{b}, \vec{\varepsilon} \rangle + c \\ &\stackrel{(20,23)}{=} \langle \vec{\beta}, X^T AX\vec{\beta} \rangle + \langle \vec{\varepsilon}, AX\vec{\beta} \rangle + \langle AX\vec{\beta}, \vec{\varepsilon} \rangle + \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle \\ &\stackrel{(19)}{=} 2 \langle \vec{\varepsilon}, AX\vec{\beta} \rangle + \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle \\ &\stackrel{!}{\geq} 0 = \tilde{s}^2(X\vec{\beta} + \vec{\varepsilon}) \Big|_{\vec{\varepsilon}=\vec{0}} \end{aligned} \quad (25)$$

Die letzte Zeile folgt wieder aus der Positivität von  $\tilde{s}^2$  und bedeutet, dass bei  $\vec{\varepsilon} = \vec{0}$  ein globales Minimum ist und damit der Gradient nach  $\vec{\varepsilon}$  von  $\tilde{s}^2$  an der Stelle  $\vec{\varepsilon} = \vec{0}$  verschwinden muss:

$$\nabla_{\varepsilon} \tilde{s}^2(X\vec{\beta} + \vec{\varepsilon}) \Big|_{\vec{\varepsilon}=\vec{0}} = 2AX\vec{\beta} = 0$$

Da das für beliebige  $\vec{\beta} \in \mathbb{R}^{p+1}$  gelten muss, folgt

$$AX = 0$$

und mit (25) haben wir dann also die Darstellung

$$\tilde{s}^2(\vec{y}) = \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle$$

Damit ist die Behauptung 1 bewiesen. •

Eine äquivalente Formulierung von (15), das war die Gleichung  $AX = 0$ , ist

$$A\vec{x}_0 = A\vec{x}_1 = \dots = A\vec{x}_p = \vec{0} \quad (26)$$

Das heisst, die Regressoren  $\vec{x}_0, \vec{x}_1, \dots, \vec{x}_p$  sind Eigenvektoren von  $A$  zum Eigenwert  $\lambda = 0$ . Da wir das  $A$  als symmetrisch voraussetzen können, ist es diagonalisierbar. Es seien  $\lambda_1, \lambda_2, \dots, \lambda_n$  die Eigenwerte von  $A$ . Aus der Positivität von  $\tilde{s}^2$  folgt

$$\lambda_i \geq 0 \quad \forall i$$

Wir ordnen die Eigenwerte der Grösse nach von gross nach klein,

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

Wegen (26) gilt dann

$$\lambda_n = \lambda_{n-1} = \dots = \lambda_{n-p} = 0 \quad (27)$$

Wir zeigen jetzt die folgende

Behauptung 2: Für jedes  $\tilde{s}^2 \in \mathcal{Q}$  gilt

$$\mathbb{V}[\tilde{s}^2] = 2\sigma^4 \sum_{i=1}^n \lambda_i^2 \stackrel{(27)}{=} 2\sigma^4 \sum_{i=1}^{n-p-1} \lambda_i^2 \quad (28)$$

Beweis Behauptung 2: Es sei

$$\{ \vec{v}_1, \vec{v}_2, \dots, \vec{v}_n \}$$

eine ONB aus Eigenvektoren von  $A$ , also

$$\begin{aligned} A\vec{v}_i &= \lambda_i \vec{v}_i \\ \langle \vec{v}_i, \vec{v}_j \rangle &= \delta_{i,j} \end{aligned}$$

und

$$A = VDV^T$$

mit

$$V = \left( \begin{array}{c|ccc|c} & & & & \\ & \vec{v}_1 & \dots & \vec{v}_n & \\ & & & & \end{array} \right)$$

und

$$D = \text{diag}(\lambda_1, \dots, \lambda_n)$$

Dann gilt

$$\begin{aligned} \tilde{s}^2(\vec{y}) &\stackrel{(14)}{=} \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle \\ &= \langle \vec{\varepsilon}, VDV^T\vec{\varepsilon} \rangle \\ &= \langle V^T\vec{\varepsilon}, DV^T\vec{\varepsilon} \rangle \\ &= \sum_{i=1}^n \lambda_i (\vec{v}_i \cdot \vec{\varepsilon})^2 \end{aligned}$$

und wir bekommen

$$\begin{aligned} \mathbb{V}[\tilde{s}^2] &= \mathbb{E}[(\tilde{s}^2)^2] - (\mathbb{E}[\tilde{s}^2])^2 \\ &= \sum_{i,j=1}^n \lambda_i \lambda_j \mathbb{E}[(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2] - (\sigma^2)^2 \end{aligned} \quad (29)$$

Jetzt wenden wir den Hilfssatz 9.2 vom letzten Mal an. Die Aussage von Teil (b) war: Es gilt die folgende Formel

$$\mathbb{E}[(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2] = \begin{cases} 3\sigma^4 & \text{falls } i = j \\ \sigma^4 & \text{falls } i \neq j \end{cases} \quad (30)$$

Also,

$$\begin{aligned} \mathbb{V}[\tilde{s}^2] &= \sum_{i,j=1}^n \lambda_i \lambda_j \mathbb{E}[(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2] - \sigma^4 \\ &= \sum_{\substack{i,j=1 \\ i \neq j}}^n \lambda_i \lambda_j \mathbb{E}[(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2] + \sum_{\substack{i,j=1 \\ i=j}}^n \lambda_i \lambda_j \mathbb{E}[(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2] - \sigma^4 \\ &= \sum_{\substack{i,j=1 \\ i \neq j}}^n \lambda_i \lambda_j \sigma^4 + \sum_{i=1}^n \lambda_i^2 3\sigma^4 - \sigma^4 \\ &= \sum_{i,j=1}^n \lambda_i \lambda_j \sigma^4 + \sum_{i=1}^n \lambda_i^2 2\sigma^4 - \sigma^4 \\ &= \left( \sum_{i=1}^n \lambda_i \right)^2 \sigma^4 + \sum_{i=1}^n \lambda_i^2 2\sigma^4 - \sigma^4 \\ &= [\text{Tr}(A)]^2 \sigma^4 + \sum_{i=1}^n \lambda_i^2 2\sigma^4 - \sigma^4 \end{aligned}$$

Jetzt benutzen wir die Gleichung (21),

$$\text{Tr}(A) = 1$$

und erhalten

$$\begin{aligned} \mathbb{V}[\tilde{s}^2] &= [\text{Tr}(A)]^2 \sigma^4 + \sum_{i=1}^n \lambda_i^2 2\sigma^4 - \sigma^4 \\ &= 2\sigma^4 \sum_{i=1}^n \lambda_i^2 = 2\sigma^4 \sum_{i=1}^{n-p-1} \lambda_i^2 \end{aligned}$$

Damit ist die Behauptung 2 gezeigt. •

Die Aussage von Theorem 10.2 folgt nun aus der folgenden

Behauptung 3: Es gilt

$$\sum_{i=1}^{n-(p+1)} \lambda_i^2 \geq \frac{1}{n-(p+1)} \left( \sum_{i=1}^{n-(p+1)} \lambda_i \right)^2 \quad (31)$$

Beweis Behauptung 3: Übungsblatt 10. •

Also können wir jetzt schreiben

$$\begin{aligned} \mathbb{V}[\tilde{s}^2] &\stackrel{\text{Beh.2}}{=} 2\sigma^4 \sum_{i=1}^{n-p-1} \lambda_i^2 \\ &\stackrel{\text{Beh.3}}{\geq} 2\sigma^4 \frac{1}{n-(p+1)} \left( \sum_{i=1}^{n-(p+1)} \lambda_i \right)^2 \\ &= \frac{2\sigma^4}{n-(p+1)} \left( \sum_{i=1}^n \lambda_i \right)^2 \\ &= \frac{2\sigma^4}{n-(p+1)} [\text{Tr}(A)]^2 \\ &\stackrel{(21)}{=} \frac{2\sigma^4}{n-(p+1)} = \mathbb{V}[\hat{s}^2] \end{aligned}$$

Damit ist das Theorem 10.2 bewiesen. ■