VL11: Effizienz der Maximum-Likelihood-Schätzer

Es gelten dieselben Bezeichnungen wie in den VL's 9+10. Die Maximum-Likelihood-Schätzer waren durch die folgenden Formeln gegeben:

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

$$\widehat{\sigma^2}_{\text{ML}} \;=\; \tfrac{1}{n} \left[\, P_{X^\perp} \vec{y} \,\right]^2 \tag{2}$$

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

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

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

$$\widehat{s^2} \;=\; \tfrac{1}{n-(p+1)} \left[\, P_{X^\perp} \vec{y} \,\right]^2 \tag{3}$$

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 β_j und definieren die folgende Menge von linearen, erwartungstreuen Schätzern

$$\mathcal{L} \;:=\; \Big\{ \tilde{\beta} : \mathbb{R}^n \to \mathbb{R}^{p+1} \;\big|\; \tilde{\beta}(\vec{y}) = A\vec{y} + \vec{b},\; A \in \mathbb{R}^{(p+1)\times n},\, \vec{b} \in \mathbb{R}^{p+1},\; \mathsf{E}[\, \tilde{\beta}(\vec{y}) \,] = \vec{\beta} \;\Big\} \tag{4}$$

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

$$\vec{y} \;=\; X\vec{\beta} + \vec{\varepsilon} \tag{5}$$

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

Theorem 11.1: 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:

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

Beweis: Wegen

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

muss gelten

$$AX \;=\; Id \tag{7}$$

$$\vec{b} \;=\; 0 \tag{8}$$

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

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

wobei wir in der letzten Zeile von (9) das Lemma 9.2 verwendet haben. Wir müssen zeigen: Die rechte Seite von (9) wird minimiert durch

$$A \;=\; (X^TX)^{-1}X^T \tag{10}$$

Dazu schreiben wir

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

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

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

bekommen wir dann

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

Nun ist aber

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

und wir erhalten

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

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)

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

und das Theorem 11.1 ist bewiesen. ■

Jetzt wollen wir ein analoges Theorem für den Schätzer

$$\begin{aligned}
\widehat{s^2} \;=\; \widehat{s^2}(\vec{y}) \;&=\; \tfrac{1}{n-(p+1)}\,\big[\,P_{X^\perp}\vec{y}\,\big]^2 \\
&=\; \tfrac{1}{n-(p+1)}\,\big\langle P_{X^\perp}\vec{y}, P_{X^\perp}\vec{y}\,\big\rangle \\
&=\; \tfrac{1}{n-(p+1)}\,\big\langle \vec{y}, P_{X^\perp}^T P_{X^\perp}\vec{y}\,\big\rangle \\
&=\; \tfrac{1}{n-(p+1)}\,\big\langle \vec{y}, P_{X^\perp} P_{X^\perp}\vec{y}\,\big\rangle \\
&=\; \tfrac{1}{n-(p+1)}\,\big\langle \vec{y}, P_{X^\perp}\vec{y}\,\big\rangle \qquad (11)
\end{aligned}$$

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

$$\mathcal{Q} \;:=\; \Big\{ \widetilde{s^2} : \mathbb{R}^n \to \mathbb{R} \;\Big|\; \widetilde{s^2}(\vec{y}) \;=\; \langle \vec{y}, A\vec{y}\rangle \;+\; \langle \vec{b}, \vec{y}\rangle \;+\; c \;\geq\; 0\,, \quad A \in \mathbb{R}^{n\times n},\; \vec{b} \in \mathbb{R}^n,\; c \in \mathbb{R},\;\; \mathsf{E}[\,\widetilde{s^2}(\vec{y})\,] = \sigma^2 \Big\} \qquad (12)$$

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

Theorem 11.2: Der Schätzer $\widehat{s^2}$ 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 $\widetilde{s^2}$ aus $\mathcal{Q}$ gilt:

$$\mathsf{V}[\,\widetilde{s^2}\,] \;\geq\; \frac{2\sigma^4}{n-(p+1)} \;=\; \mathsf{V}[\,\widehat{s^2}\,] \qquad (13)$$

Beweis: Wir zeigen zunächst die folgende

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

$$\widetilde{s^2}(\vec{y}) \;=\; \langle \vec{\varepsilon}, A\vec{\varepsilon} \rangle \tag{14}$$

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

$$AX \;=\; 0 \;. \tag{15}$$

Beweis Behauptung 1: Wegen

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

ist

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

und damit

$$\sigma^2 \;\stackrel{!}{=}\; \mathsf{E}\big[\widetilde{s^2}(\vec{y})\big] \;=\; \langle X\vec{\beta}, AX\vec{\beta} \rangle \;+\; \langle \vec{b}, X\vec{\beta} \rangle \;+\; c \;+\; \sigma^2 \,\mathrm{Tr}(A) \tag{17}$$

Dabei haben wir die Identitäten

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

und

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

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

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

Aus (18) folgt

$$\begin{aligned}
X^T AX \;&=\; 0 \\
X^T\vec{b} \;&=\; 0
\end{aligned} \tag{19}$$

und

$$\begin{aligned} c &= 0 & (20)\\ \mathrm{Tr}(A) &= 1 & (21)\end{aligned}$$

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

$$\widetilde{s^2}(\vec{y}) \;=\; \langle \vec{y}, A\vec{y}\rangle \;+\; \langle \vec{b}, \vec{y}\rangle \;\overset{!}{\geq}\; 0 \;=\; \widetilde{s^2}(\vec{y} = \vec{0}) \qquad (22)$$

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

$$\nabla \widetilde{s^2}(\vec{y})\,\Big|_{\vec{y}=\vec{0}} \;=\; \vec{b} \;=\; \vec{0} \qquad (23)$$

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

$$\widetilde{s^2}(\vec{y}) \;=\; \langle \vec{y}, A\vec{y}\rangle \qquad (24)$$

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

$$\begin{aligned} \widetilde{s^2}(\vec{y}) &= \langle \vec{y}, A\vec{y}\rangle \\ &= \tfrac{1}{2}\Big\{ \langle \vec{y}, A\vec{y}\rangle \;+\; \langle \vec{y}, A^T\vec{y}\rangle \Big\} \\ &= \Big\langle\, \vec{y}\,,\; \tfrac{A+A^T}{2}\,\vec{y}\,\Big\rangle \end{aligned}$$

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

$$A \;=\; A^T$$

Aus Gleichung (16) folgt dann

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

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

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

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

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

Damit ist die Behauptung 1 bewiesen. $\bullet$

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

$$A\vec{x}_0 \;=\; A\vec{x}_1 \;=\; \cdots \;=\; A\vec{x}_p \;=\; \vec{0} \tag{26}$$

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

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

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

$$\lambda_1 \;\geq\; \lambda_2 \;\geq\; \cdots \;\geq\; \lambda_n \;\geq\; 0$$

Wegen (26) gilt dann

$$\lambda_n \;=\; \lambda_{n-1} \;=\; \cdots \;=\; \lambda_{n-p} \;=\; 0 \tag{27}$$

Wir zeigen jetzt die folgende

<u>Behauptung 2</u>: Für jedes $\widetilde{s^2} \in \mathcal{Q}$ gilt

$$\mathsf{V}\big[\,\widetilde{s^2}\,\big] \;=\; 2\sigma^4 \sum_{i=1}^{n} \lambda_i^2 \;\overset{(27)}{=}\; 2\sigma^4 \sum_{i=1}^{n-p-1} \lambda_i^2 \tag{28}$$

<u>Beweis Behauptung 2</u>: Es sei

$$\{\; \vec{v}_1 \;,\; \vec{v}_2 \;\cdots\;,\; \vec{v}_n \;\}$$

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

$$A \;=\; V D V^T$$

mit

$$V \;=\; \begin{pmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{pmatrix}$$

und

$$D \;=\; \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$$

Dann gilt

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

und wir bekommen

$$\begin{aligned}
\mathsf{V}\big[\,\widetilde{s^2}\,\big] &= \mathsf{E}\big[\,(\widetilde{s^2})^2\,\big] \;-\; \big(\,\mathsf{E}\big[\,\widetilde{s^2}\,\big]\,\big)^2 \\
&= \textstyle\sum_{i,j=1}^n \lambda_i \lambda_j\, \mathsf{E}\big[\,(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2\,\big] \;-\; \big(\,\sigma^2\,\big)^2 \qquad (29)
\end{aligned}$$

Jetzt wenden wir den Hilfssatz 10.2 vom letzten Mal an. Die Aussage von Teil (b) war: Es gilt die folgende Formel

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

Also,

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

Jetzt benutzen wir die Gleichung (21),

$$\mathrm{Tr}(A) \;=\; 1$$

und erhalten

$$\begin{aligned}
\mathsf{V}\big[\,\widetilde{s^2}\,\big] &= \big[\,\mathrm{Tr}(A)\,\big]^2\, \sigma^4 \;+\; \sum_{i=1}^n \lambda_i^2\, 2\sigma^4 \;-\; \sigma^4 \\
&= 2\sigma^4 \sum_{i=1}^n \lambda_i^2 \;=\; 2\sigma^4 \sum_{i=1}^{n-p-1} \lambda_i^2
\end{aligned}$$

Damit ist die Behauptung 2 gezeigt. $\bullet$

Die Aussage von Theorem 11.2 folgt nun aus der folgenden

Behauptung 3: Es gilt

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

Beweis Behauptung 3: Übungsblatt 11. $\bullet$

Also können wir jetzt schreiben

$$\begin{aligned}
\mathsf{V}\left[\, \widetilde{s^2} \,\right] \quad &\overset{\text{Beh.2}}{=} \quad 2\sigma^4 \sum_{i=1}^{n-p-1} \lambda_i^2 \\
&\overset{\text{Beh.3}}{\geq} \quad 2\sigma^4 \, \frac{1}{n-(p+1)} \left(\sum_{i=1}^{n-(p+1)} \lambda_i \right)^2 \\
&= \quad \frac{2\sigma^4}{n-(p+1)} \left(\sum_{i=1}^{n} \lambda_i \right)^2 \\
&= \quad \frac{2\sigma^4}{n-(p+1)} \left[\, \mathrm{Tr}(A) \,\right]^2 \\
&\overset{(21)}{=} \quad \frac{2\sigma^4}{n-(p+1)} \quad = \quad \mathsf{V}\left[\, \widehat{s^2} \,\right]
\end{aligned}$$

Damit ist das Theorem 11.2 bewiesen. $\blacksquare$