week9: Erwartungswert und Varianz der Maximum-Likelihood-Schätzer, Teil2

Es gelten dieselben Bezeichnungen wie in dem week8. Heute wollen wir die Varianzen der Maximum-Likelihood-Schätzer

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

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

oder im Falle von Gleichung (2) nehmen wir die erwartungstreue Version,

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

berechnen. Diese brauchen wir, wenn wir zeigen wollen, dass die Schätzer effizient sind. Es gilt das folgende

Theorem 9.1: a) Die Varianz des Maximum-Likelihood-Schätzers für den j-ten Regressionskoeffizienten ist gegeben durch

$$\mathsf{V}\big[\,\hat{\beta}_{j,\mathrm{ML}}\,\big] \;=\; \sigma^2 \left[(X^T X)^{-1} \right]_{j,j} \tag{4}$$

Allgemeiner gilt:

$$\mathsf{Cov}\big[\,\hat{\beta}_{j,\mathrm{ML}}\,,\, \hat{\beta}_{k,\mathrm{ML}}\,\big] \;=\; \sigma^2 \left[(X^T X)^{-1} \right]_{j,k} \tag{5}$$

b) Die Varianz für das $\widehat{s^2}$ ist gegeben durch die folgende Formel:

$$\mathsf{V}\big[\,\widehat{s^2}\,\big] \;=\; \tfrac{2\sigma^4}{n-(p+1)} \tag{6}$$

Beweis: a) Mit

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

bekommen wir

$$\begin{aligned}
\hat{\beta}_{\mathrm{ML}} \;&=\; (X^T X)^{-1} X^T \vec{y} \\
&=\; (X^T X)^{-1} X^T \big(\, X\vec{\beta} \,+\, \vec{\varepsilon}\,\big) \\
&=\; \vec{\beta} \,+\, (X^T X)^{-1} X^T \vec{\varepsilon} \\
&=:\; \vec{\beta} \,+\, A\vec{\varepsilon}
\end{aligned} \tag{7}$$

mit der Matrix

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

Damit bekommen wir (wir lassen das 'ML' an den $\hat{\beta}$'s jetzt mal weg)

$$\begin{aligned}
\mathsf{Cov}\big[\,\hat{\beta}_j\,,\,\hat{\beta}_k\,\big] \quad &= \quad \mathsf{E}\Big[\,\big(\hat{\beta}_j - \mathsf{E}[\hat{\beta}_j]\big)\big(\hat{\beta}_k - \mathsf{E}[\hat{\beta}_k]\big)\,\Big] \\
&= \quad \mathsf{E}\Big[\,\big(\hat{\beta}_j - \beta_j\big)\big(\hat{\beta}_k - \beta_k\big)\,\Big] \\
&\overset{(7)}{=} \quad \mathsf{E}\big[\,(A\vec{\varepsilon})_j (A\vec{\varepsilon})_k\,\big] \\
&\overset{\text{Lemma 8.2}}{=} \quad \sigma^2\,(AA^T)_{j,k}
\end{aligned}$$

wobei

$$\begin{aligned}
AA^T \quad &= \quad (X^T X)^{-1} X^T \big[(X^T X)^{-1} X^T\big]^T \\
&= \quad (X^T X)^{-1} X^T X \big[(X^T X)^{-1}\big]^T \\
&= \quad \big[(X^T X)^{-1}\big]^T \\
&= \quad \big[(X^T X)^T\big]^{-1} \\
&= \quad \big[X^T X\big]^{-1}
\end{aligned}$$

Damit ist der Teil (a) bewiesen. Zum Beweis von Teil (b) benötigen wir den folgenden

Hilfssatz 9.2: Es seien $\vec{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)$ normalverteilte, unabhängige Zufallszahlen mit Mittelwert 0 und Standardabweichung σ. Weiter seien n Vektoren

$$\vec{v}_1,\ \vec{v}_2,\ \cdots\ ,\ \vec{v}_n \ \in \ \mathbb{R}^n$$

gegeben, die eine Orthonormalbasis des $\mathbb{R}^n$ bilden. Es gelte also

$$\vec{v}_i \cdot \vec{v}_j \ = \ \delta_{i,j}$$

Dann gelten die folgenden Aussagen:

a) Es sei $F : \mathbb{R}^n \to \mathbb{R}$ eine beliebige Funktion. Dann gilt

$$\mathsf{E}\big[\,F(\vec{v}_1\vec{\varepsilon}, \cdots, \vec{v}_n\vec{\varepsilon})\,\big] \ = \ \mathsf{E}\big[\,F(\varepsilon_1, \cdots, \varepsilon_n)\,\big] \tag{8}$$

b) Es gilt die folgende Formel

$$\mathsf{E}\big[\,(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2\,\big] \ = \ \begin{cases} 3\sigma^4 & \text{falls } i = j \\ \sigma^4 & \text{falls } i \neq j \end{cases} \tag{9}$$

Beweis: Wenn wir die Vektoren $\vec{v}_1, \cdots, \vec{v}_n$ in der Matrix

$$V \;:=\; \begin{pmatrix} - & \vec{v}_1 & - \\ & \vdots & \\ - & \vec{v}_n & - \end{pmatrix} \;\in\; \mathbb{R}^{n\times n}$$

zusammenfassen, dann ist V eine orthogonale Matrix, $V^{-1} = V^T$ und es gilt $|\det V| = 1$. Die Gleichung (8) können wir dann auch folgendermassen schreiben:

$$\mathsf{E}\big[\,F(V\vec{\varepsilon})\,\big] \;=\; \mathsf{E}\big[\,F(\vec{\varepsilon})\,\big]$$

Mit der Variablensubstitution

$$\vec{\varepsilon} \;=\; V^T\vec{\phi}$$

erhalten wir dann

$$\begin{aligned}
\mathsf{E}\big[\,F(V\vec{\varepsilon})\,\big] \;&=\; \int_{\mathbb{R}^n} F(V\vec{\varepsilon}) \prod_{k=1}^{n}\Big\{\, e^{-\frac{\varepsilon_k^2}{2\sigma^2}} \,\tfrac{d\varepsilon_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&=\; \int_{\mathbb{R}^n} F(V\vec{\varepsilon})\; e^{-\frac{1}{2\sigma^2}\langle\vec{\varepsilon},\vec{\varepsilon}\rangle}\; \tfrac{d^n\varepsilon}{(2\pi\sigma^2)^{n/2}} \\
&=\; \int_{\mathbb{R}^n} F(VV^T\vec{\phi})\; e^{-\frac{1}{2\sigma^2}\langle V^T\vec{\phi},V^T\vec{\phi}\rangle}\; \tfrac{|\det V^T|\,d^n\phi}{(2\pi\sigma^2)^{n/2}} \\
&=\; \int_{\mathbb{R}^n} F(\vec{\phi})\; e^{-\frac{1}{2\sigma^2}\langle\vec{\phi},\vec{\phi}\rangle}\; \tfrac{d^n\phi}{(2\pi\sigma^2)^{n/2}} \\
&=\; \int_{\mathbb{R}^n} F(\vec{\varepsilon})\; e^{-\frac{1}{2\sigma^2}\langle\vec{\varepsilon},\vec{\varepsilon}\rangle}\; \tfrac{d^n\varepsilon}{(2\pi\sigma^2)^{n/2}} \;\;=\;\; \mathsf{E}\big[\,F(\vec{\varepsilon})\,\big]
\end{aligned}$$

und der Teil (a) ist bewiesen. Teil (b) bekommen wir dann folgendermassen:

$$\mathsf{E}\big[\,(\vec{v}_i\cdot\vec{\varepsilon})^2(\vec{v}_j\cdot\vec{\varepsilon})^2\,\big] \;\overset{(a)}{=}\; \mathsf{E}\big[\,\varepsilon_i^2\,\varepsilon_j^2\,\big] \;=\; \int_{\mathbb{R}^n} \varepsilon_i^2\,\varepsilon_j^2 \prod_{k=1}^{n}\Big\{\, e^{-\frac{\varepsilon_k^2}{2\sigma^2}} \,\tfrac{d\varepsilon_k}{\sqrt{2\pi\sigma^2}}\,\Big\}$$

Für $i \neq j$ liefert das

$$\int_{\mathbb{R}^2} \varepsilon_i^2\,\varepsilon_j^2\, e^{-\frac{\varepsilon_i^2}{2\sigma^2}}\, e^{-\frac{\varepsilon_i^2}{2\sigma^2}}\, \tfrac{d\varepsilon_i\,d\varepsilon_j}{\sqrt{2\pi\sigma^2}^{\,2}} \;=\; \sigma^2\times\sigma^2 \;=\; \sigma^4$$

und für $i = j$ bekommen wir

$$\int_{\mathbb{R}} \varepsilon_i^4\, e^{-\frac{\varepsilon_i^2}{2\sigma^2}}\, \tfrac{d\varepsilon_i}{\sqrt{2\pi\sigma^2}} \;=\; 3\,\sigma^4$$

Damit ist der Hilfssatz bewiesen. ■

Beweis Teil b Theorem 9.1: Es war

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

mit

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

und

$$P_{X^\perp} \;=\; Id - P_X \;\; .$$

Mit

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

bekommen wir dann

$$P_{X^\perp}\vec{y} \;=\; P_{X^\perp}X\vec{\beta} \,+\, P_{X^\perp}\vec{\varepsilon}$$

Wegen

$$\begin{aligned} P_{X^\perp}X &= [Id - P_X]X \\ &= X \,-\, X(X^TX)^{-1}X^TX \\ &= X \,-\, X \;=\; 0 \end{aligned}$$

haben wir also

$$P_{X^\perp}\vec{y} \;=\; P_{X^\perp}\vec{\varepsilon}$$

und damit

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

Wir wählen jetzt eine ONB

$$\mathcal{B} \;=\; \left\{\,\vec{v}_0,\; \vec{v}_1,\; \cdots\;,\; \vec{v}_{n-1}\,\right\}$$

vom $\mathbb{R}^n$ folgendermassen: Die ersten $p+1$ Vektoren

$$\vec{v}_0,\; \vec{v}_1,\; \cdots\;,\; \vec{v}_p$$

sind eine ONB von X, der Raum der von den Regressoren $\vec{x}_0, \cdots, \vec{x}_p$ aufgespannt wird, und die restlichen $n-(p+1)$ Vektoren

$$\vec{v}_{p+1},\; \vec{v}_{p+2},\; \cdots\;,\; \vec{v}_{n-1}$$

sind eine ONB von $X^\perp$. Dann können wir schreiben

$$\begin{aligned} \vec{\varepsilon} &= \sum_{k=0}^{n-1}\langle\vec{\varepsilon},\vec{v}_k\rangle\,\vec{v}_k \\ P_{X^\perp}\vec{\varepsilon} &= \sum_{k=p+1}^{n-1}\langle\vec{\varepsilon},\vec{v}_k\rangle\,\vec{v}_k \end{aligned}$$

und

$$\left[P_{X^\perp}\vec{\varepsilon}\right]^2 \;=\; \sum_{k=p+1}^{n-1}\langle\vec{\varepsilon},\vec{v}_k\rangle^2$$

Jetzt können wir die Varianz von $\widehat{s^2}$ berechnen:

$$
\begin{aligned}
\mathsf{V}\big[\,\widehat{s^2}\,\big] \;&=\; \mathsf{E}\big[\,(\widehat{s^2})^2\,\big] \;-\; \mathsf{E}\big[\,\widehat{s^2}\,\big]^2 \\
&=\; \tfrac{1}{[n-(p+1)]^2}\mathsf{E}\left[\,\left\{\textstyle\sum_{k=p+1}^{n-1}\langle\vec{\varepsilon},\vec{v}_k\rangle^2\right\}^2\,\right] \;-\; (\sigma^2)^2 \\
&=\; \tfrac{1}{[n-(p+1)]^2}\;\sum_{k,\ell=p+1}^{n-1}\mathsf{E}\big[\,\langle\vec{\varepsilon},\vec{v}_k\rangle^2\,\langle\vec{\varepsilon},\vec{v}_\ell\rangle^2\,\big] \;-\; \sigma^4 \\
&=\; \tfrac{1}{[n-(p+1)]^2}\left\{\;\sum_{k=p+1}^{n-1}\mathsf{E}\big[\,\langle\vec{\varepsilon},\vec{v}_k\rangle^2\,\langle\vec{\varepsilon},\vec{v}_k\rangle^2\,\big] \;+\; \sum_{\substack{k,\ell=p+1\\k\neq\ell}}^{n-1}\mathsf{E}\big[\,\langle\vec{\varepsilon},\vec{v}_k\rangle^2\,\langle\vec{\varepsilon},\vec{v}_\ell\rangle^2\,\big]\;\right\} \;-\; \sigma^4 \\
&=\; \tfrac{1}{[n-(p+1)]^2}\left\{\;\sum_{k=p+1}^{n-1}3\sigma^4 \;+\; \sum_{\substack{k,\ell=p+1\\k\neq\ell}}^{n-1}\sigma^4\;\right\} \;-\; \sigma^4 \\
&=\; \tfrac{\sigma^4}{[n-(p+1)]^2}\Big\{\;3[n-(p+1)] \;+\; [n-(p+1)]^2-[n-(p+1)]\;\Big\} \;-\; \sigma^4 \\
&=\; \tfrac{\sigma^4}{[n-(p+1)]^2}\Big\{\;2[n-(p+1)]\;\Big\} \;+\; \sigma^4 \;-\; \sigma^4 \\
&=\; \tfrac{2\sigma^4}{[n-(p+1)]}
\end{aligned}
$$

Damit ist das Theorem 9.1 bewiesen. ■