

### 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

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

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

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

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

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

$$\text{V}[\hat{\beta}_{j,\text{ML}}] = \sigma^2 [(X^T X)^{-1}]_{j,j} \quad (4)$$

Allgemeiner gilt:

$$\text{Cov}[\hat{\beta}_{j,\text{ML}}, \hat{\beta}_{k,\text{ML}}] = \sigma^2 [(X^T X)^{-1}]_{j,k} \quad (5)$$

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

$$\text{V}[\hat{s}^2] = \frac{2\sigma^4}{n-(p+1)} \quad (6)$$

**Beweis: a)** Mit

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

bekommen wir

$$\begin{aligned} \hat{\beta}_{\text{ML}} &= (X^T X)^{-1} X^T \vec{y} \\ &= (X^T X)^{-1} X^T (X\vec{\beta} + \vec{\varepsilon}) \\ &= \vec{\beta} + (X^T X)^{-1} X^T \vec{\varepsilon} \\ &=: \vec{\beta} + A\vec{\varepsilon} \end{aligned} \quad (7)$$

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)

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

wobei

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

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

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

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

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  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  eine beliebige Funktion. Dann gilt

$$\mathbb{E}[F(\vec{v}_1\vec{\varepsilon}, \dots, \vec{v}_n\vec{\varepsilon})] = \mathbb{E}[F(\varepsilon_1, \dots, \varepsilon_n)] \quad (8)$$

b) Es gilt die folgende Formel

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

**Beweis:** Wenn wir die Vektoren  $\vec{v}_1, \dots, \vec{v}_n$  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:

$$\mathbb{E}[F(V\vec{\varepsilon})] = \mathbb{E}[F(\vec{\varepsilon})]$$

Mit der Variablensubstitution

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

erhalten wir dann

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

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

$$\mathbb{E}[(\vec{v}_i \cdot \vec{\varepsilon})^2 (\vec{v}_j \cdot \vec{\varepsilon})^2] \stackrel{(a)}{=} \mathbb{E}[\varepsilon_i^2 \varepsilon_j^2] = \int_{\mathbb{R}^n} \varepsilon_i^2 \varepsilon_j^2 \prod_{k=1}^n \left\{ e^{-\frac{\varepsilon_k^2}{2\sigma^2}} \frac{d\varepsilon_k}{\sqrt{2\pi\sigma^2}} \right\}$$

Für  $i \neq j$  liefert das

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

und für  $i = j$  bekommen wir

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

Damit ist der Hilfssatz bewiesen. ■

**Beweis Teil b Theorem 9.1:** Es war

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

mit

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

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

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

haben wir also

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

und damit

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

Wir wählen jetzt eine ONB

$$\mathcal{B} = \{ \vec{v}_0, \vec{v}_1, \dots, \vec{v}_{n-1} \}$$

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

$$\vec{v}_0, \vec{v}_1, \dots, \vec{v}_p$$

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

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

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

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

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

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

Damit ist das Theorem 9.1 bewiesen. ■