week10a: Kapitel 6: Erwartungstreue und Verteilung von Schaetzern, Teil2

Wir wollen zunächst die Varianzen der Maximum Likelihood Schätzer aus dem Beispiel 1 berechnen. Die Schätzer waren gegeben durch (für das σ^2 nehmen wir die erwartungstreue Version $\widehat{s^2}$, die hatte ein $1/(n-1)$ als Vorfaktor),

$$\hat{\mu}_{\mathrm{ML}}(x_1,\cdots,x_n) \;=\; \frac{1}{n}\sum_{i=1}^n x_i \;=:\; \bar{x} \tag{1}$$

$$\widehat{s^2}(x_1,\cdots,x_n) \;=\; \frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2 \tag{2}$$

und Erwartungswerte waren wie folgt zu berechnen:

$$\mathsf{E}\big[\,F(x_1,\cdots,x_n)\,\big] \;=\; \int_{\mathbb{R}^n} F(x_1,\cdots,x_n)\;\prod_{i=1}^n \Big\{\, e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\;\frac{dx_i}{\sqrt{2\pi\sigma^2}}\,\Big\} \tag{3}$$

mit $F = \hat{\mu}_{\mathrm{ML}}$ oder $F = \widehat{s^2}$. Mit den Formeln für die Varianzen aus dem folgendem Theorem 6.2.1 werden wir dann in der nächsten Woche zeigen können, dass die Schätzer (1) und (2) effizient oder asymptotisch effizient sind. Das heisst, sie haben eine minimale Varianz innerhalb der Menge aller erwartungstreuen Schätzer. Das bedeutet also, dass das ganz besonders schöne Schätzer sind, und das ist eine Eigenschaft, die man häufiger bei Maximum Likelihood Schätzern findet.

Theorem 6.2.1: Wir betrachten die Maximum-Likelihood-Schätzer (1) und (2) aus dem Beispiel 1. Dann gilt:

a) Für die Varianz von $\hat{\mu}_{\mathrm{ML}}$ bekommen wir

$$\mathsf{V}\big[\,\hat{\mu}_{\mathrm{ML}}(x_1,\cdots,x_n)\,\big] \;=\; \frac{\sigma^2}{n} \tag{4}$$

b) Für die Varianz des modifizierten erwartungstreuen Schätzers

$$\widehat{s^2} \;:=\; \frac{n}{n-1}\,\hat{\sigma}^2_{\mathrm{ML}} \;=\; \frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2 \tag{5}$$

erhalten wir

$$\mathsf{V}\big[\,\widehat{s^2}(x_1,\cdots,x_n)\,\big] \;=\; \frac{2\sigma^4}{n-1} \tag{6}$$

Beweis: a) Wir haben wegen Theorem 6.1.2 vom letzten Mal

$$\begin{aligned} \mathsf{V}\big[\,\hat{\mu}_{\mathrm{ML}}\,\big] &= \mathsf{E}\big[\,\hat{\mu}^2_{\mathrm{ML}}\,\big] \;-\; \big(\,\mathsf{E}\big[\,\hat{\mu}_{\mathrm{ML}}\,\big]\,\big)^2 \\ &= \mathsf{E}\big[\,\hat{\mu}^2_{\mathrm{ML}}\,\big] \;-\; \mu^2 \end{aligned}$$

mit

$$\hat{\mu}^2_{\mathrm{ML}} \;=\; \left\{\,\frac{1}{n}\sum_{i=1}^{n} x_i\,\right\}^2 \;=\; \frac{1}{n^2}\sum_{i=1}^{n} x_i^2 \;+\; \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n} x_i\,x_j$$

Im Beweis von Theorem 6.1.2 hatten wir letztes Mal schon gezeigt:

$$\begin{aligned} \mathsf{E}[\,x_i^2\,] &= \sigma^2 \;+\; \mu^2 \\ \mathsf{E}[\,x_i x_j\,] &\overset{i\neq j}{=} \mu^2 \end{aligned}$$

Damit,

$$\begin{aligned} \mathsf{E}\big[\,\hat{\mu}^2_{\mathrm{ML}}\,\big] &= \frac{1}{n^2}\sum_{i=1}^{n}(\sigma^2+\mu^2) \;+\; \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\mu^2 \\ &= \frac{n}{n^2}\,\sigma^2 \;+\; \frac{n+(n^2-n)}{n^2}\,\mu^2 \\ &= \frac{\sigma^2}{n} \;+\; \mu^2 \end{aligned}$$

und der Teil (a) folgt dann sofort aus

$$\begin{aligned} \mathsf{V}\big[\,\hat{\mu}_{\mathrm{ML}}\,\big] &= \mathsf{E}\big[\,\hat{\mu}^2_{\mathrm{ML}}\,\big] \;-\; \mu^2 \\ &= \frac{\sigma^2}{n} \;+\; \mu^2 \;-\; \mu^2 \;=\; \frac{\sigma^2}{n} \end{aligned}$$

b) Mit der Erwartungstreue von dem $\widehat{s^2}$ folgt

$$\begin{aligned} \mathsf{V}\big[\,\widehat{s^2}\,\big] &= \mathsf{E}\big[\,(\widehat{s^2})^2\,\big] \;-\; \big(\,\mathsf{E}\big[\,\widehat{s^2}\,\big]\,\big)^2 \\ &= \mathsf{E}\big[\,(\widehat{s^2})^2\,\big] \;-\; (\sigma^2)^2 \end{aligned}$$

Zunächst mal gilt für beliebige Konstanten c und λ

$$\begin{aligned} \widehat{s^2}(x_1,\cdots,x_n) &= \widehat{s^2}(x_1+c,\cdots,x_n+c) \\ \widehat{s^2}(\lambda x_1,\cdots,\lambda x_n) &= \lambda^2\,\widehat{s^2}(x_1,\cdots,x_n) \end{aligned}$$

denn:

$$\begin{aligned}
\widehat{s^2}(x_1,\cdots,x_n) &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\Big\{\, x_i \,-\, \tfrac{1}{n}\textstyle\sum_{j=1}^n x_j \,\Big\}^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\Big\{\, x_i + c \,-\, \tfrac{1}{n}\textstyle\sum_{j=1}^n (x_j + c) \,\Big\}^2 \\
&= \widehat{s^2}(x_1+c,\cdots,x_n+c)
\end{aligned}$$

und

$$\begin{aligned}
\widehat{s^2}(\lambda x_1,\cdots,\lambda x_n) &= \frac{1}{n-1}\sum_{i=1}^{n}\Big\{\, \lambda x_i \,-\, \tfrac{1}{n}\textstyle\sum_{j=1}^n \lambda x_j \,\Big\}^2 \\
&= \lambda^2\,\frac{1}{n-1}\sum_{i=1}^{n}\Big\{\, x_i \,-\, \tfrac{1}{n}\textstyle\sum_{j=1}^n x_j \,\Big\}^2 \\
&= \lambda^2\;\widehat{s^2}(x_1,\cdots,x_n)
\end{aligned}$$

Wir wählen jetzt $c=-\mu$ und bekommen

$$\begin{aligned}
\mathsf{E}\big[\,(\widehat{s^2})^2\,\big] &= \int_{\mathbb{R}^n}\Big\{\,\widehat{s^2}(x_1,\cdots,x_n)\,\Big\}^2\,\prod_{k=1}^{n}\Big\{\,e^{-\frac{(x_k-\mu)^2}{2\sigma^2}}\;\tfrac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&= \int_{\mathbb{R}^n}\Big\{\,\widehat{s^2}(x_1-\mu,\cdots,x_n-\mu)\,\Big\}^2\,\prod_{k=1}^{n}\Big\{\,e^{-\frac{(x_k-\mu)^2}{2\sigma^2}}\;\tfrac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&= \int_{\mathbb{R}^n}\Big\{\,\widehat{s^2}(y_1,\cdots,y_n)\,\Big\}^2\,\prod_{k=1}^{n}\Big\{\,e^{-\frac{y_k^2}{2\sigma^2}}\;\tfrac{dy_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&= \int_{\mathbb{R}^n}\Big\{\,\sigma^2\,\widehat{s^2}\big(\,\tfrac{y_1}{\sigma}\,,\,\cdots\,,\,\tfrac{y_n}{\sigma}\,\big)\,\Big\}^2\,\prod_{k=1}^{n}\Big\{\,e^{-\frac{y_k^2}{2\sigma^2}}\;\tfrac{dy_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&= \sigma^4\,\times\,\int_{\mathbb{R}^n}\Big\{\,\widehat{s^2}(z_1,\cdots,z_n)\,\Big\}^2\,\prod_{k=1}^{n}\Big\{\,e^{-\frac{z_k^2}{2}}\;\tfrac{dz_k}{\sqrt{2\pi}}\,\Big\} \\
&= \sigma^4\,\times\,\int_{\mathbb{R}^n}\Big\{\,\widehat{s^2}(z_1,\cdots,z_n)\,\Big\}^2\,e^{-\frac{1}{2}\langle z,z\rangle}\;\tfrac{d^n z}{(2\pi)^{n/2}}
\end{aligned}$$

mit

$$\widehat{s^2}(z_1,\cdots,z_n) \;=\; \frac{1}{n-1}\sum_{i=1}^{n}\Big\{\, z_i \,-\, \tfrac{1}{n}\textstyle\sum_{j=1}^n z_j \,\Big\}^2$$

Wir definieren den Vektor

$$v_1 \;:=\; \tfrac{1}{\sqrt{n}}\,(1,1,\cdots,1) \;\in\; \mathbb{R}^n$$

mit $\|v_1\|^2=1$. Wir haben dann

$$\langle z, v_1\rangle \;=\; \tfrac{1}{\sqrt{n}}\,\textstyle\sum_{j=1}^n z_j$$

und können schreiben

$$z_i \; - \; \tfrac{1}{n}\sum_{j=1}^n z_j \quad = \quad \Big[\, z \; - \; \langle z, v_1\rangle v_1 \,\Big]_{i-\text{te Komponente}}$$

Weiterhin, da v_1 ein Einheitsvektor ist, können wir das v_1 mit geeigneten Vektoren $v_2, \cdots, v_n$ zu einer Orthonormalbasis

$$B \;=\; \{v_1, v_2, \cdots, v_n\}$$

des $\mathbb{R}^n$ erweitern. Wir haben dann also die Skalarprodukte

$$\langle v_j, v_k\rangle \;=\; \delta_{j,k}$$

und können das $z \in \mathbb{R}^n$ nach den $v_k \in \mathbb{R}^n$ entwickeln,

$$z \;=\; \sum_{k=1}^n \langle z, v_k\rangle\, v_k$$

Insbesondere,

$$\begin{aligned}
z_i \; - \; \tfrac{1}{n}\sum_{j=1}^n z_j \quad &= \quad \Big[\, z \; - \; \langle z, v_1\rangle v_1 \,\Big]_{i-\text{te Komponente}} \\
&= \quad \left[\, \sum_{k=1}^n \langle z, v_k\rangle\, v_k \; - \; \langle z, v_1\rangle v_1 \,\right]_{i-\text{te Komponente}} \\
&= \quad \left[\, \sum_{k=2}^n \langle z, v_k\rangle\, v_k \,\right]_{i-\text{te Komponente}}
\end{aligned}$$

und wir können schreiben

$$\begin{aligned}
s^2(z_1, \cdots, z_n) \quad &= \quad \frac{1}{n-1} \sum_{i=1}^n \Big\{\, z_i \; - \; \tfrac{1}{n}\sum_{j=1}^n z_j \,\Big\}^2 \\
&= \quad \frac{1}{n-1} \left\| \, \sum_{k=2}^n \langle z, v_k\rangle v_k \, \right\|^2 \\
&= \quad \frac{1}{n-1} \left\langle \, \sum_{k=2}^n \langle z, v_k\rangle v_k \, , \, \sum_{\ell=2}^n \langle z, v_\ell\rangle v_\ell \, \right\rangle \\
&= \quad \frac{1}{n-1} \sum_{k,\ell=2}^n \langle z, v_k\rangle \langle z, v_\ell\rangle \, \underbrace{\langle v_k, v_\ell\rangle}_{=\delta_{k,\ell}} \\
&= \quad \frac{1}{n-1} \sum_{k=2}^n \langle z, v_k\rangle \langle z, v_k\rangle \\
&= \quad \frac{1}{n-1} \sum_{k=2}^n \langle z, v_k\rangle^2
\end{aligned}$$

Jetzt machen wir die Variablen-Substitution

$$
\begin{aligned}
\phi_1 &= \langle v_1, z \rangle \\
\phi_2 &= \langle v_2, z \rangle \\
&\vdots \\
\phi_n &= \langle v_n, z \rangle
\end{aligned}
$$

oder in Vektor-Matrix-Schreibweise

$$\phi \;=\; Vz$$

mit der orthogonalen Matrix

$$V \;:=\; \begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{pmatrix} \;\in\; \mathbb{R}^{n\times n}$$

Wegen $\langle v_k, v_\ell \rangle = \delta_{k,\ell}$ ist $V^T = V^{-1}$ und $\det V = 1$. Damit bekommen wir

$$
\begin{aligned}
\langle z, z \rangle &= \langle V^{-1}\phi, V^{-1}\phi \rangle = \langle V^T\phi, V^T\phi \rangle = \langle \phi, VV^T\phi \rangle = \langle \phi, \phi \rangle \\
d^n z &= |\det[V^{-1}]|\, d^n\phi = |\det[V^T]|\, d^n\phi = |\det V|\, d^n\phi = d^n\phi
\end{aligned}
$$

und mit

$$
\begin{aligned}
s^2(z_1, \cdots, z_n) &= \frac{1}{n-1} \sum_{k=2}^{n} \langle z, v_k \rangle^2 \\
&= \frac{1}{n-1} \sum_{k=2}^{n} \phi_k^2
\end{aligned}
$$

ist dann

$$
\begin{aligned}
\mathsf{E}\big[\,(\widehat{s^2})^2\,\big] &= \sigma^4 \times \int_{\mathbb{R}^n} \big\{\, s^2(z_1, \cdots, z_n)\,\big\}^2\, e^{-\frac{1}{2}\langle z,z\rangle}\, \tfrac{d^n z}{(2\pi)^{n/2}} \\
&= \frac{\sigma^4}{(n-1)^2} \times \sum_{k,\ell=2}^{n} \int_{\mathbb{R}^n} \phi_k^2\, \phi_\ell^2\, e^{-\frac{1}{2}\langle \phi,\phi\rangle}\, \tfrac{d^n \phi}{(2\pi)^{n/2}}
\end{aligned}
$$

Mit den Gaus'schen Integralen aus dem week2b, das war das Theorem 2.1.3, folgt

$$\int_{\mathbb{R}^n} \phi_k^2\, \phi_\ell^2\, e^{-\frac{1}{2}\langle \phi,\phi\rangle}\, \tfrac{d^n \phi}{(2\pi)^{n/2}} \;=\; \begin{cases} 3 & \text{falls } k = \ell \\ 1 & \text{falls } k \neq \ell \end{cases}$$

Also,

$$
\begin{aligned}
\mathsf{E}\big[\,(\widehat{s^2})^2\,\big] \;&=\; \frac{\sigma^4}{(n-1)^2} \times \sum_{k,\ell=2}^{n} \int_{\mathbb{R}^n} \phi_k^2\,\phi_\ell^2 \; e^{-\frac{1}{2}\langle\phi,\phi\rangle} \, \tfrac{d^n\phi}{(2\pi)^{n/2}} \\
&=\; \frac{\sigma^4}{(n-1)^2} \times \left\{ \sum_{\substack{k,\ell=2\\k=\ell}}^{n} 3 \;+\; \sum_{\substack{k,\ell=2\\k\neq\ell}}^{n} 1 \right\} \\
&=\; \frac{\sigma^4}{(n-1)^2} \times \left\{ \sum_{\substack{k,\ell=2\\k=\ell}}^{n} 2 \;+\; \sum_{\substack{k,\ell=2\\k=\ell}}^{n} 1 \;+\; \sum_{\substack{k,\ell=2\\k\neq\ell}}^{n} 1 \right\} \\
&=\; \frac{\sigma^4}{(n-1)^2} \times \left\{ \sum_{\substack{k,\ell=2\\k=\ell}}^{n} 2 \;+\; \sum_{k,\ell=2}^{n} 1 \right\} \\
&=\; \frac{\sigma^4}{(n-1)^2} \times \Big\{\; 2(n-1) \;+\; (n-1)^2 \;\Big\} \\
&=\; \frac{2\sigma^4}{n-1} \;+\; \sigma^4
\end{aligned}
$$

und damit

$$
\begin{aligned}
\mathsf{V}\big[\,\widehat{s^2}\,\big] \;&=\; \mathsf{E}\big[\,(\widehat{s^2})^2\,\big] \;-\; \Big(\mathsf{E}\big[\,\widehat{s^2}\,\big]\Big)^2 \;=\; \mathsf{E}\big[\,(\widehat{s^2})^2\,\big] \;-\; (\sigma^2)^2 \\
&=\; \frac{2\sigma^4}{n-1} \;+\; \sigma^4 \;-\; (\sigma^2)^2 \;=\; \frac{2\sigma^4}{n-1}
\end{aligned}
$$

und das Theorem ist bewiesen. ■