week8b: Kapitel 5: Die Maximum Likelihood Methode
5.2: Erwartungswert, Varianz und Verteilung von Schaetzern, Teil1

Ein wichtiger Begriff in der Statistik ist der der Erwartungstreue. Bei Schätzern möchte man immer gerne haben, dass sie erwartungstreu sind. Was heisst das jetzt genau? Wir wollen uns das zunächst erstmal an den Beispielen 1 und 2 aus dem week7a.pdf anschauen:

Beispiel 1, Erwartungstreue: Gegeben seien n Zufallszahlen $x_1, x_2, \cdots, x_n$. Wir wissen, dass diese Zahlen mit Hilfe einer Normalverteilung generiert worden sind, kennen aber nicht den Mittelwert μ und die Standardabweichung σ dieser Normalverteilung. Mit Hilfe der Maximum Likelihood Methode waren wir dann in dem Theorem 5.1.1 auf die folgenden Schätzer für μ und σ gekommen, das waren dann also die Maxium Likelihood Schätzer:

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

$$\hat{\sigma}^2_{\mathrm{ML}}(x_1, \cdots, x_n) \;=\; \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2}$$

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

Wenn die x_i's jetzt nicht irgendwelche gegebenen Daten sind, sondern wirklich durch Simulation generierte normalverteilte Zufallszahlen, dann sollte, wenn wir jetzt etwa 10000 mal solche n Zufallszahlen $(x_1, \cdots, x_n)$ simulieren und dann jeweils die Grösse (1) berechnen (für ein festes n, etwa $n = 30$), dann sollte der Mittelwert dieser 10000 $\hat{\mu}_{\mathrm{ML}}$ dann auch wirklich gegen das tatsächliche μ konvergieren. Das ist genau dann der Fall, wenn der Erwartungswert der $\hat{\mu}_{\mathrm{ML}}$'s, gegeben durch

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

gleich dem tatsächlichen μ ist. Das führt dann also zu der folgenden

Definition 5.2.1: Die Maximum-Likelihood-Schätzer (1) und (2) aus dem Beispiel 1 heissen erwartungstreu, wenn sie die folgende Eigenschaft haben:

$$\mathsf{E}\big[\, \hat{\mu}_{\mathrm{ML}}(x_1, \cdots, x_n) \,\big] \;=\; \mu \tag{4}$$

$$\mathsf{E}\big[\, \hat{\sigma}^2_{\mathrm{ML}}(x_1, \cdots, x_n) \,\big] \;=\; \sigma^2 \tag{5}$$

Dabei sind die Erwartungswerte gemäss Gleichung (3) zu berechnen. Es gilt nun das folgende

Theorem 5.2.2: Wir betrachten die Maximum-Likelihood-Schätzer (1) und (2) aus dem Beispiel 1. Dann gilt:

a) Der Schätzer $\hat{\mu}_{\mathrm{ML}}$ ist erwartungstreu,

$$\mathsf{E}\big[\,\hat{\mu}_{\mathrm{ML}}(x_1,\cdots,x_n)\,\big] \;=\; \mu\;. \tag{6}$$

b) Der Schätzer $\hat{\sigma}^2_{\mathrm{ML}}$ ist nur asymptotisch, im Limes $n \to \infty$, erwartungstreu. Genauer gilt:

$$\mathsf{E}\big[\,\hat{\sigma}^2_{\mathrm{ML}}(x_1,\cdots,x_n)\,\big] \;=\; \frac{n-1}{n}\,\sigma^2 \tag{7}$$

und der modifizierte Schätzer

$$\widehat{s^2} \;:=\; \frac{n}{n-1}\hat{\sigma}^2_{\mathrm{ML}} \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{8}$$

ist erwartungstreu.

Beweis: Wir kürzen ab:

$$p_{\mu,\sigma}(x) \;:=\; \tfrac{1}{\sqrt{2\pi\sigma^2}}\;e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Das $p_{\mu,\sigma}(x)$ ist also die Dichte der Normalverteilung mit Mittelwert μ und Standardabweichung σ. Da das eine W'keitsdichte ist, gilt

$$\textstyle\int_{\mathbb{R}} p_{\mu,\sigma}(x)\,dx \;=\; 1 \tag{9}$$

Weiterhin haben wir die folgenden Integrale: Der Mittelwert einer $p_{\mu,\sigma}$-verteilten Zufallszahl ist μ meint:

$$\textstyle\int_{\mathbb{R}} x\,p_{\mu,\sigma}(x)\,dx \;=\; \mu \tag{10}$$

Die Varianz einer $p_{\mu,\sigma}$-verteilten Zufallszahl ist σ^2 meint:

$$\textstyle\int_{\mathbb{R}}(x-\mu)^2\,p_{\mu,\sigma}(x)\,dx \;=\; \sigma^2$$

oder

$$\begin{aligned}\textstyle\int_{\mathbb{R}} x^2\,p_{\mu,\sigma}(x)\,dx &= \textstyle\int_{\mathbb{R}}(x-\mu+\mu)^2\,p_{\mu,\sigma}(x)\,dx \\ &= \textstyle\int_{\mathbb{R}}\big\{\,(x-\mu)^2 \,+\, 2(x-\mu)\mu \,+\, \mu^2\,\big\}\,p_{\mu,\sigma}(x)\,dx \\ &= \sigma^2 \,+\, 0 \,+\, \mu^2 \;=\; \sigma^2 \,+\, \mu^2 \end{aligned} \tag{11}$$

Damit bekommen wir:

$$\begin{aligned}\mathsf{E}\big[\,\hat{\mu}_{\mathrm{ML}}(x_1,\cdots,x_n)\,\big] &= \int_{\mathbb{R}^n}\hat{\mu}_{\mathrm{ML}}(x_1,\cdots,x_n)\;\prod_{k=1}^{n}\Big\{\,e^{-\frac{(x_k-\mu)^2}{2\sigma^2}}\;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\ &= \int_{\mathbb{R}^n}\frac{1}{n}\sum_{i=1}^{n}x_i\;\prod_{k=1}^{n}\Big\{\,e^{-\frac{(x_k-\mu)^2}{2\sigma^2}}\;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^n}x_i\;\prod_{k=1}^{n}\Big\{\,e^{-\frac{(x_k-\mu)^2}{2\sigma^2}}\;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\}\end{aligned}$$

mit

$$\int_{\mathbb{R}^n} x_i \prod_{k=1}^{n} \Big\{ e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \frac{dx_k}{\sqrt{2\pi\sigma^2}} \Big\} \;=\; \int_{\mathbb{R}} p_{\mu,\sigma}(x_1)\,dx_1 \;\cdots\; \int_{\mathbb{R}} x_i\, p_{\mu,\sigma}(x_i)\,dx_i \;\cdots\; \int_{\mathbb{R}} p_{\mu,\sigma}(x_n)\,dx_n$$
$$\overset{(9,10)}{=} \; 1 \times \cdots \times 1 \times \mu \times 1 \times \cdots \times 1$$
$$= \; \mu$$

Also,

$$\mathsf{E}\big[\, \hat{\mu}_{\mathrm{ML}}(x_1, \cdots, x_n) \,\big] \;=\; \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^n} x_i \prod_{k=1}^{n} \Big\{ e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \frac{dx_k}{\sqrt{2\pi\sigma^2}} \Big\}$$
$$= \; \frac{1}{n} \sum_{i=1}^{n} \mu \;=\; \mu$$

und $\hat{\mu}_{\mathrm{ML}}$ ist erwartungstreu. Für die Berechnung von $\mathsf{E}[\hat{\sigma}^2_{\mathrm{ML}}]$ machen wir zunächst ein paar Umformungen:

$$\begin{aligned}
\hat{\sigma}^2_{\mathrm{ML}}(x_1, \cdots, x_n) &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \big\{ x_i^2 \;-\; 2x_i\bar{x} \;+\; \bar{x}^2 \big\} \\
&= \frac{1}{n} \sum_{i=1}^{n} x_i^2 \;-\; 2\bar{x}\,\bar{x} \;+\; \bar{x}^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} x_i^2 \;-\; \bar{x}^2 \qquad (12)
\end{aligned}$$

mit

$$\begin{aligned}
\bar{x}^2 &= \Big\{ \frac{1}{n} \sum_{i=1}^{n} x_i \Big\}^2 \;=\; \frac{1}{n} \sum_{i=1}^{n} x_i \, \frac{1}{n} \sum_{j=1}^{n} x_j \;=\; \frac{1}{n^2} \sum_{i,j=1}^{n} x_i\, x_j \\
&= \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i=j}}^{n} x_i\, x_j \;+\; \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} x_i\, x_j \\
&= \frac{1}{n^2} \sum_{i=1}^{n} x_i^2 \;+\; \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} x_i\, x_j \qquad (13)
\end{aligned}$$

Wir setzen (13) in (12) ein und bekommen

$$\begin{aligned}
\hat{\sigma}^2_{\mathrm{ML}}(x_1, \cdots, x_n) &= \frac{1}{n} \sum_{i=1}^{n} x_i^2 \;-\; \bar{x}^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} x_i^2 \;-\; \frac{1}{n^2} \sum_{i=1}^{n} x_i^2 \;-\; \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} x_i\, x_j \\
&= \frac{n-1}{n^2} \sum_{i=1}^{n} x_i^2 \;-\; \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} x_i\, x_j \qquad (14)
\end{aligned}$$

Damit erhalten wir

$$
\begin{aligned}
\mathsf{E}\big[\,\hat\sigma^2_{\mathrm{ML}}(x_1,\cdots,x_n)\big] \;&=\; \int_{\mathbb{R}^n} \hat\sigma_{\mathrm{ML}}(x_1,\cdots,x_n) \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&\overset{(14)}{=} \int_{\mathbb{R}^n} \Big\{\, \frac{n-1}{n^2}\sum_{i=1}^{n} x_i^2 \;-\; \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n} x_i\,x_j \,\Big\} \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&= \frac{n-1}{n^2}\sum_{i=1}^{n}\int_{\mathbb{R}^n} x_i^2 \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&\quad - \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\int_{\mathbb{R}^n} x_i\,x_j \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \qquad (15)
\end{aligned}
$$

mit den Integralen

$$
\begin{aligned}
\int_{\mathbb{R}^n} x_i^2 \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \;&=\; \textstyle\int_{\mathbb{R}} p_{\mu,\sigma}(x_1)\,dx_1 \;\cdots\; \int_{\mathbb{R}} x_i^2\; p_{\mu,\sigma}(x_i)\,dx_i \;\cdots\; \int_{\mathbb{R}} p_{\mu,\sigma}(x_n)\,dx_n \\
&\overset{(9,11)}{=} 1\times\cdots\times 1\times(\sigma^2+\mu^2)\times 1\times\cdots\times 1 \\
&= \sigma^2+\mu^2
\end{aligned}
$$

und für $i \neq j$

$$
\begin{aligned}
&\int_{\mathbb{R}^n} x_i\,x_j \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&\quad= \textstyle\int_{\mathbb{R}} p_{\mu,\sigma}(x_1)\,dx_1 \;\cdots\; \int_{\mathbb{R}} x_i\; p_{\mu,\sigma}(x_i)\,dx_i \;\cdots\; \int_{\mathbb{R}} x_j\; p_{\mu,\sigma}(x_j)\,dx_j \;\cdots\; \int_{\mathbb{R}} p_{\mu,\sigma}(x_n)\,dx_n \\
&\quad\overset{(9,10)}{=} 1\times\cdots\times\mu\times\cdots\times\mu\times\cdots\times 1 \\
&\quad= 1^{n-2}\;\times\;\mu^2 \;=\; \mu^2
\end{aligned}
$$

Das setzen wir in (15) ein und bekommen

$$
\begin{aligned}
\mathsf{E}\big[\,\hat\sigma^2_{\mathrm{ML}}(x_1,\cdots,x_n)\big] \;&=\; \frac{n-1}{n^2}\sum_{i=1}^{n}\int_{\mathbb{R}^n} x_i^2 \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&\quad - \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\int_{\mathbb{R}^n} x_i\,x_j \prod_{k=1}^{n}\Big\{\, e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} \;\frac{dx_k}{\sqrt{2\pi\sigma^2}}\,\Big\} \\
&= \frac{n-1}{n^2}\sum_{i=1}^{n}(\sigma^2+\mu^2) \;-\; \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\mu^2 \\
&= \frac{n-1}{n}(\sigma^2+\mu^2) \;-\; \frac{n^2-n}{n^2}\,\mu^2 \\
&= \frac{n-1}{n}\,\sigma^2 \;+\; \frac{n-1}{n}\,\mu^2 \;-\; \frac{n-1}{n}\,\mu^2 \;=\; \frac{n-1}{n}\,\sigma^2
\end{aligned}
$$

Also ist $\mathsf{E}[\hat\sigma^2_{\mathrm{ML}}]$ nur im Limes $n\to\infty$ erwartungstreu und das Theorem ist bewiesen. ■

Beispiel 2, Erwartungstreue: Gegeben seien n Zufallszahlen $x_1, x_2, \cdots, x_n$. Wir wissen, dass diese Zahlen mit Hilfe einer Poisson-Verteilung generiert worden sind, kennen aber nicht den Parameter λ der Poisson-Verteilung. Welcher Wert von λ passt am besten, im Sinne der Maximum Likelihood Methode, zu den gegebenen Zahlen $x_1, \cdots, x_n$? Im week7a.pdf hatten wir dazu den Maximum Likelihood Schätzer

$$\hat{\lambda}_{\mathrm{ML}}(x_1, \cdots, x_n) \;=\; \frac{x_1 + \cdots + x_n}{n}$$

hergeleitet. Der Erwartungswert davon ist dann gegeben durch, jetzt mal etwas sehr ausführlich hingeschrieben,

$$\begin{aligned}
\mathsf{E}\big[\,\hat{\lambda}_{\mathrm{ML}}(x_1, \cdots, x_n)\,\big] \;&=\; \sum_{k_1, \cdots, k_n = 0}^{\infty} \hat{\lambda}_{\mathrm{ML}}(k_1, \cdots, k_n) \times \mathsf{Prob}\Big[\,x_1 = k_1\,,\; x_2 = k_2\,,\; \cdots\,,\; x_n = k_n\Big] \\
&= \sum_{k_1, \cdots, k_n = 0}^{\infty} \frac{k_1 + \cdots + k_n}{n} \prod_{i=1}^{n} \left\{ \frac{\lambda^{k_i}}{k_i!}\, e^{-\lambda} \right\} \\
&= \frac{1}{n} \sum_{j=1}^{n} \sum_{k_1, \cdots, k_n = 0}^{\infty} k_j \prod_{i=1}^{n} \left\{ \frac{\lambda^{k_i}}{k_i!}\, e^{-\lambda} \right\} \\
&= \frac{1}{n} \sum_{j=1}^{n} \left\{ \sum_{k_1=0}^{\infty} \frac{\lambda^{k_1}}{k_1!}\, e^{-\lambda} \right\} \times \cdots \times \left\{ \sum_{k_j=0}^{\infty} k_j\, \frac{\lambda^{k_j}}{k_j!}\, e^{-\lambda} \right\} \times \cdots \times \left\{ \sum_{k_n=0}^{\infty} \frac{\lambda^{k_n}}{k_n!}\, e^{-\lambda} \right\} \\
&= \frac{1}{n} \sum_{j=1}^{n} 1 \times \cdots \times 1 \times \underbrace{\mathsf{E}[x_j]}_{=\lambda} \times 1 \times \cdots \times 1 \\
&= \frac{1}{n} \sum_{j=1}^{n} \lambda \;=\; \lambda
\end{aligned}$$

also ist der Maximum Likelihood Schätzer $\hat{\lambda}_{\mathrm{ML}}$ erwartungstreu.