Marina Grachyova, Alexander Müller

Maximum-Likelihood Schätzung

Quellen: Erwin Kreyszik, Statistische Methoden und ihre Anwendungen,
U. Kockelkorn, Einführung in die Statistik.

Eines der wichtigsten Verfahren zur Gewinnung brauchbarer Schätzfunktionen für Parameter einer Verteilung ist die sogenannte **Maximum-Likelihood-Methode**. Wie betrachten zuerst die Methode für die Verteilung einer Zufallsvariablen X mit der Wahrscheinlichkeit (diskreter Fall) $P(X = x|\theta)$ bzw. Dichte (stetiger Fall) $f(x|\theta)$, die von einem einzelnen Parameter θ abhängen.
Wir führen das Experiment n-mal aus und erhalten n völlig voneinander unabhängige Werte $x_1, x_2, ..., x_n$.
Im Falle einer diskreten Variablen ist dann die Wahrscheinlichkeit, eine Stichprobe zu erhalten, die gerade aus den obigen Werten besteht, durch das Produkt

$$L := P(x_1, x_2, ..., x_n|\theta) = P(X = x_1|\theta)P(X = x_2|\theta) \cdots P(X = x_n|\theta)$$

gegeben.
Ist X stetig verteilt, so ist die Dichte:

$$L := f(x_1, x_2, ...x_n|\theta) = f(x_1|\theta)f(x_2|\theta) \cdots f(x_n|\theta).$$

So eine Funktion L heißt **Likelihood-Funktion**. Diese Funktion besitzt als Argument den Parameter θ.
Die Maximum-Likelihood-Methode besteht nun darin, dass man als Näherung für den unbekannten Parameter θ einen Wert nimmt, für den L einen möglichst großen Wert besitzt. Eine notwendige Bedingung dafür, dass L ein Maximum besitzt, ist das Verschwinden der Ableitung nach θ,

$$\frac{\partial L}{\partial \theta} = 0.$$

Eine von $x_1, x_2, ..., x_n$ abhängige Lösung $\hat{\theta}$ dieser Gleichung heißt eine **Maximum-Likelihood-Schätzfunktion** für den betreffenden Parameter θ.
Im Falle einer Verteilung, die mehrere, sagen wir r Parameter $\theta_1, \theta_2, ...\theta_r$ enthält, erhält man r Gleichungen

$$\frac{\partial L}{\partial \theta_1} = 0, ..., \frac{\partial L}{\partial \theta_r} = 0$$

zur Bestimmung von Schätzfunktionen für diese Parameter.
Manchmal empfiehlt es sich aber, statt der Likelihood- die logarithmierte Likelihood-Funktion zu maximieren. Also

$$\frac{\partial ln(L)}{\partial \theta} = 0.$$

Da Logarithmieren eine streng monoton wachsende Transformation ist, hat $ln(L)$ genau dort ein Maximum, wo L ein Maximum hat.
Für mehrere Parameter erhalten wir dann:

$$\frac{\partial ln(L)}{\partial \theta_1} = 0, ..., \frac{\partial ln(L)}{\partial \theta_r} = 0.$$

Beispiel:

Bei einem medizinischen Experiment soll die krebsverursachende Wirkung einer Substanz geschätzt werden. Wird eine Maus t Stunden der Substanz ausgesetzt, so haben sich mit der Wahrscheinlichkeit π_t nach 14 Tagen an den Nieren Tumore entwickelt, dabei

$$\pi_i(\theta) = 1 - e^{-\theta t_i}$$

Hier ist der Koeffizient θ ein Maß für die **Gefährlichkeit** des Stoffes.
Dieser soll in zwei gekoppelten Experimenten bestimmt werden.

Dazu werden

- $\mathbf{n_i}$ Mäuse
- während $\mathbf{t_i}$ Stunden seziert
- und die Anzahl $\mathbf{k_i}$ erkrankter Tiere bestimmt.

Die Ergebnisse stehen in der folgenden Tabelle:

	n_i	t_i	k_i	$n_i - k_i$
Experiment E_1	10	4	4	6
Experiment E_2	8	8	7	1

Wir betrachten zuerst jedes Experiment für sich.
Für das i-te Experiment gilt:

$$P(Y = k_i) = \binom{n_i}{k_i} \pi_i^{k_i}(1 - \pi_i)^{n_i - k_i}$$

Für das erste Experiment E_1 ist:

$$\pi_1 = 1 - e^{-\theta t_1} = 1 - e^{-4\theta} \quad \text{und} \quad \mathcal{L}(\theta|E_1) = (1 - e^{-4\theta})^4 (e^{-4\theta})^6 \ ,$$

(da der Binomialkoeffizient lediglich ein konstanter Faktor ist, lassen wir ihn weg.)

Außerdem ergibt sich für Experiment E_2:

$$\pi_2 = 1 - e^{-\theta t_2} = 1 - e^{-8\theta} \quad \text{und} \quad \mathcal{L}(\theta|E_2) = (1 - e^{-8\theta})^7 e^{-8\theta} \ .$$

Die Gesamt-Likelihood ist nach dem Multiplikationssatz:

$$\begin{aligned}\mathcal{L}(\theta|E_1 \wedge E_2) &= \mathcal{L}(\theta|E_1)\mathcal{L}(\theta|E_2) = (1-e^{-4\theta})^4(e^{-4\theta})^6(1-e^{-8\theta})^7 e^{-4\theta} \\ &= (1-e^{-4\theta})^4(1-e^{-8\theta})^7 e^{-28\theta}\end{aligned}$$

Plausible Schätzer liegen im Intervall $[0,1;0,35]$.
Wir bestimmen den ML-Schätzer $\hat{\theta}$ graphisch mit Maple zu $\hat{\theta} \approx 0,2$ *(s.u.)*.