Stochastik

Barth, Friedrich

München, [20]03

18. 2. Das Maximum-Likelihood-Prinzip

Eine so weit gefaßte Definition gibt uns keine Hilfe, wie man zu zweckmäßigen Schätzfunktionen gelangt. Und sie sagt uns erst recht nicht, welcher Schätzfunktion wir den Vorzug geben sollen, falls wir gar mehrere Schätzfunktionen gefunden haben.

18.2. Das Maximum-Likelihood-Prinzip

Ein besonders brauchbares Verfahren zur Gewinnung von Schätzgrößen führte 1760 *Johann Heinrich Lambert* (1728–1777) und unabhängig davon 1777 *Daniel Bernoulli* (1700–1782) in die Wahrscheinlichkeitsrechnung ein. *Carl Friedrich Gauß* (1777–1855) benützte es mehrfach, so z.B. 1798 zu seinem ersten Beweis der Methode der kleinsten Quadrate. Verallgemeinert hat das Verfahren aber erst 1912 *Ronald Aylmer Fisher* (1890–1962) zum

Maximum-Likelihood-Prinzip oder **Prinzip der maximalen Mutmaßlichkeit:** Eine Zufallsstichprobe $(X_1 | X_2 | \ldots | X_n)$ aus der Zufallsgröße X, deren Verteilung vom Parameter ϑ abhängt, zeitigte das Stichprobenergebnis $(a_1 | a_2 | \ldots | a_n)$. Als Schätzwert für den Parameter ϑ dient dann jeder Wert ϑ, für den die Wahrscheinlichkeit

$$P_\vartheta(X_1 = a_1 \wedge X_2 = a_2 \wedge \ldots \wedge X_n = a_n)$$

des tatsächlich eingetretenen Stichprobenergebnisses maximal wird.

Jedem möglichen Wert des Parameters ϑ wird also bei bekanntem Stichprobenergebnis eine Wahrscheinlichkeit zugeordnet. Diese Zuordnung

$$L\colon\ \vartheta \mapsto P(X_1 = a_1 \wedge X_2 = a_2 \wedge \ldots \wedge X_n = a_n)$$

heißt **Likelihood-Funktion** L. Eine Maximumstelle dieser Funktion L muß nicht notwendig existieren; andererseits kann es auch mehrere solcher Stellen geben. Ein nach dem Maximum-Likelihood-Prinzip bestimmter Schätzwert $\hat{\vartheta}(a_1, a_2, \ldots, a_n)$ heißt **Maximum-Likelihood-Schätzwert**, die zugehörige Zufallsgröße $\hat{\vartheta}(X_1, X_2, \ldots, X_n)$ dann **Maximum-Likelihood-Schätzgröße.**

Betrachten wir zum besseren Verständnis den besonders einfachen Fall, daß wir als Parameter ϑ den Anteil p einer Kugelsorte in einer Urne schätzen wollen. (Es sei $0 < p < 1$.) Wir entnehmen der Urne eine Kugel und betrachten die Zufallsgröße

$$X := \begin{cases} 1, & \text{falls die Kugel der Sorte angehört,} \\ 0 & \text{sonst.} \end{cases}$$

X ist nach $B(1; p)$ verteilt. Eine Stichprobe $(X_1 | X_2 | \ldots | X_n)$ der Länge n aus der Zufallsgröße X besteht im n-maligen Ziehen einer Kugel aus der Urne mit Zurücklegen. Die Wahrscheinlichkeit, daß bei einem gegebenen Kugelanteil p sich das Stichprobenergebnis $(a_1 | a_2 | \ldots | a_n)$, $a_i \in \{0; 1\}$, einstellt, wird durch die Likelihood-Funktion L in Abhängigkeit von p ausgedrückt:

$$L(p) := P_p(X_1 = a_1 \wedge X_2 = a_2 \wedge \ldots \wedge X_n = a_n).$$

Da in der Stichprobe die X_i stochastisch unabhängig sind, erhält man

$$L(p) = P_p(X_1 = a_1) \cdot P_p(X_2 = a_2) \cdot \ldots \cdot P_p(X_n = a_n).$$

Betrachten wir nun ein spezielles Stichprobenergebnis mit genau k Einsen, dann gilt

$$L(p) = p^k(1-p)^{n-k}.$$

Das Maximum dieser Wahrscheinlichkeit finden wir durch Differenzieren von $L(p)$ nach p. Für $0 < k < n$ erhalten wir – die Fälle $k = 0$ und $k = n$ erledigt man analog –

$$\begin{aligned}\frac{\mathrm{d}L(p)}{\mathrm{d}p} &= kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1} = \\ &= -p^{k-1}(1-p)^{n-k-1}(np-k).\end{aligned}$$

Als Nullstelle ergibt sich $p = \frac{k}{n}$; der Vorzeichenwechsel von $\frac{\mathrm{d}L(p)}{\mathrm{d}p}$ zeigt, daß es sich um eine Maximumstelle handelt.

$\hat{p}(a_1, a_2, \ldots, a_n) = \frac{k}{n} = \frac{1}{n} \sum_{i=1}^{n} a_i$ ist der Maximum-Likelihood-Schätzwert für den Kugelanteil p in der Urne. Die zugehörige Maximum-Likelihood-Schätzgröße $\hat{p}(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$ ist aber nichts anderes als die uns längst bekannte Zufallsgröße relative Häufigkeit H_n.

Es ist erfreulich, daß auch das Maximum-Likelihood-Prinzip die relative Häufigkeit eines Ereignisses als brauchbare Schätzgröße für die Wahrscheinlichkeit eines Ereignisses liefert. Auf Grund der Interpretationsregel für Wahrscheinlichkeiten, die durch die Gesetze der großen Zahlen wissenschaftlich abgesichert ist, war die relative Häufigkeit immer schon ein brauchbarer »Meßwert« für die Wahrscheinlichkeit eines Ereignisses.

18.3. Beurteilungskriterien für Schätzfunktionen

Das Maximum-Likelihood-Prinzip ist ein Verfahren zur Gewinnung von Schätzfunktionen. Wie sollen wir uns aber entscheiden, wenn wir uns durch verschiedene Betrachtungsweisen mehrere Schätzfunktionen verschafft haben? Eine Festlegung auf eine Schätzfunktion ist nicht eindeutig möglich, da die Eignung einer Stichprobenfunktion zur Schätzung eines Parameters nach sehr unterschiedlichen Gesichtspunkten beurteilt werden kann. In den Jahren 1921 und 1925 hat *Ronald Aylmer Fisher* (1890–1962) vier Kriterien zur Beurteilung von Schätzfunktionen entwickelt.

1. Von einer Schätzgröße T_n für den Parameter ϑ wird man erwarten, daß ihre Werte, d. h. also die Schätzwerte, nach beiden Seiten um den unbekannten Wert