Stochastik

Barth, Friedrich

München, [20]03

18. 1. Problemstellung

*18. Parameterschätzung

18.1. Problemstellung

Im vorausgegangenen Kapitel haben wir dargestellt, wie man Testprobleme lösen kann. Wir wenden uns nun der anderen typischen Fragestellung der Mathematischen Statistik zu, dem Schätzproblem, das wir auch nur für einfache Fälle angehen wollen.

Im einfachsten Fall handelt es sich darum, auf Grund eines Stichprobenergebnisses die Wahrscheinlichkeit p eines Ereignisses zu schätzen. Im Urnenmodell bedeutet dies, den Anteil p einer Kugelsorte zu schätzen. Als erster hat *Thomas Bayes* (1702–1761) diese Aufgabe in seiner erst 1763 erschienenen berühmten Schrift *An Essay towards solving a problem in the Doctrine of Chances* gestellt und unter der Voraussetzung, daß alle Werte von p aus $[0; 1]$ gleichwahrscheinlich in Frage kommen, durch eine **Intervallschätzung** für p gelöst. Erst 1934 gelang es *Jerzy Neyman* (1894–1981), das Problem allgemein durch Einführung der Konfidenzintervalle, wie wir sie in **14.8.** beschrieben haben, zu lösen.

Allgemeiner geht es darum, auf Grund von Stichprobenergebnissen gewisse Parameter der Wahrscheinlichkeitsverteilung einer Zufallsgröße zu schätzen. Solche Parameter sind beispielsweise der Parameter p einer Binomialverteilung, der Erwartungswert μ, der Median und die Quantile, die Standardabweichung und die Schiefe einer irgendwie gearteten Verteilung, ja sogar der Umfang der Grundgesamtheit.

Sei nun ϑ ein solcher zu schätzender Parameter der Wahrscheinlichkeitsverteilung einer Zufallsgröße X. Zu seiner Schätzung ziehen wir aus X eine Zufallsstichprobe $(X_1 | X_2 | \ldots | X_n)$. Sie liefere das Stichprobenergebnis $(a_1 | a_2 | \ldots | a_n)$. Aus diesen Stichprobenwerten a_i soll nun kein Intervall für ϑ, sondern durch eine geeignete Formel ein Näherungswert $\hat{\vartheta}$, eben ein Schätzwert, für den unbekannten Parameter ϑ errechnet werden. Man spricht dann von einer **Punktschätzung** für ϑ. Dieser Schätzwert ist somit eine Funktion der a_i; es gilt also $\hat{\vartheta} = T_n(a_1, a_2, \ldots, a_n)$. Dabei soll der Index n anzeigen, daß eine Stichprobe der Länge n gezogen wurde. Der auf diese Weise errechnete Schätzwert $\hat{\vartheta}$ hängt natürlich vom Zufall ab; denn a_i ist ja der zufällige Wert aus der Wertemenge $\mathfrak{S} = \{x_1, x_2, \ldots, x_s\}$ von X, den die Zufallsgröße X_i angenommen hat. $\hat{\vartheta}$ ist also aufzufassen als ein beobachteter Wert der aus den n Zufallsgrößen X_i gebildeten Stichprobenfunktion $T_n(X_1, X_2, \ldots, X_n)$, die als Funktion von Zufallsgrößen selbst wieder eine Zufallsgröße ist. Sie heißt im Zusammenhang mit dem Schätzproblem daher »Schätzgröße« oder auch »Schätzfunktion«. Wir merken uns

Definition 376.1: Ist $(X_1 | X_2 | \ldots | X_n)$ eine Stichprobe aus der Zufallsgröße X, dann heißt jede reellwertige Funktion

$$T_n\colon \quad (X_1 | X_2 | \ldots | X_n) \mapsto T_n(X_1, X_2, \ldots, X_n)$$

Schätzfunktion oder auch **Schätzgröße** für den reellen Parameter ϑ der Wahrscheinlichkeitsverteilung der Zufallsgröße X.

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: