Vorlesung am 26.01.2009

In der **schließenden Statistik** gehen wir davon aus, dass die beobachteten Daten gemäß einem stochastischen Modell erzeugt wurden.

Genauer:
Die beobachteten Daten sind Realisierungen (also beobachtete Werte) von Zufallsvariablen, deren Werte unbeeinflusst voneinander (d.h. unabhängig) und nach dem gleichen Prinzip (d.h. mit identischer Verteilung) entstanden sind.

Noch genauer:
Unsere Stichprobe $x_1, \ldots, x_n$ ist Realisierung der ersten n-Glieder $X_1, \ldots, X_n$ einer Folge $(X_k)_{k\in\mathbb{N}}$ von reellen Zufallsvariablen definiert auf dem gleichen Wahrscheinlichkeitsraum $(\Omega, \mathbf{P})$, die unabhängig und identisch verteilt sind.

Ausgehend von diesen Daten möchten wir Aussagen über die zugrundeliegende Verteilung (z.B. wie groß ist der Wert, der sich im Mittel bei dieser Verteilung ergibt) machen.

Genauer:
Konstruiere eine Schätzung $T_n(x_1, \ldots, x_n)$ von einem "Parameter" der Verteilung von X_1, z.B. vom Erwartungswert oder von der Varianz von X_1.

Beispiel:
Schätze Erwartungswert und Varianz der Verteilung durch

$$T_n(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{und} \quad T_n(x_1, \ldots, x_n) = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2 .$$

Wünschenswerte Eigenschaften von Schätzfunktionen sind erstens, dass sich der geschätzte Wert asymptotisch, also bei immer größer werdenden Stichprobenumfang, dem gesuchten Wert immer mehr annähert (sog. **starke Konsistenz**), und dass sich zweitens im Mittel, also bei wiederholtem Schätzen und festem Stichprobenumfang, der richtige Wert ergibt (sog. **Erwartungstreue**).

Im Falle der Schätzung des Erwartungswertes durch das empirische Mittel der Stichprobe bedeutet dies:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbf{E}X_1 \tag{1}$$

mit Wahrscheinlichkeit Eins (also fast sicher, abgekürzt f.s.), und

$$\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mathbf{E}X_1. \tag{2}$$

Nach dem **starken Gesetz der großen Zahlen** gilt:
Sind die auf dem selben Wahrscheinlichkeitsraum definierten reellen Zufallsvariablen $X_1, X_2, \ldots$ unabhängig und identisch verteilt, und existiert $\mathbf{E}X_1$, so gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbf{E}X_1 \quad f.s.$$

Also ist bei obiger Schätzung die Eigenschaft (1) (und nach den Rechenregeln für Erwartungswerten auch (2)) erfüllt.

Man kann zeigen, dass auch die empirische Varianz stark konsistent und erwartungstreu ist. Der Vorfaktor $1/(n-1)$ stellt dabei die Erwartungstreue sicher.