Die Verteilung der Stichprobenvarianz einer normalverteilten Variablen

VORAUSSETZUNG

Wie im Lernpfad *Einführung in Wahrscheinlichkeitstheorie und Statistik für PsychologInnen* im Abschnitt *4.3. Aus der Normalverteilung abgeleitete Verteilungen* beschrieben, ist die Quadratsumme mehrerer unabhängig standardnormalverteilter Zufallsvariablen χ^2-verteilt. Einziger Parameter der χ^2-Verteilung ist die Anzahl der Freiheitsgrade df (degrees of freedom), die genau der Zahl der summierten Quadrate entspricht.

Beispiel

$$U \sim N(0,1)$$
$$V \sim N(0,1)$$
$$W \sim N(0,1)$$

$$X = U^2 + V^2 + W^2$$
$$X \sim \chi^2 \ (df = 3)$$

kurz:

$$X \sim \chi^2[3]$$

Die Summe unabhängig χ^2-verteilter Variablen

ist ihrerseits wieder χ^2-verteilt. Die Anzahl der Freiheitsgrade ist die Summe der Freiheitsgrade der einzelnen Verteilungen.

Formal:

$$A \sim \chi^2[3]$$
$$B \sim \chi^2[2]$$
$$C \sim \chi^2[5]$$
$$(A, B, C \; unabh.)$$

$$X = A + B + C$$

$$\Rightarrow X \sim \chi^2[3+2+5]$$
$$X \sim \chi^2[10]$$

Anmerkung

Da χ^2 bei $df \to \infty$ gegen $N(df, \sqrt{2df})$ strebt, können die Werte für $df > 100$ mittels NV-Approximation ermittelt werden.

VERTEILUNG DER STICHPROBENVARIANZEN

Als Schätzung für σ_X^2 verwendet man:

$$s_X^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

Durch Umformung zu

$$s_X^2 = \frac{\sigma_X^2}{n-1} \sum_i \left(\frac{x_i - \bar{x}}{\sigma_X}\right)^2$$

erhält man in der Klammer eine Standardnormalverteilung.
Durch Quadrieren und Summieren ergibt sich eine χ^2-Verteilung.

Vorsicht bei der Anzahl der Freiheitsgrade

Zwar werden n Quadrate normalverteilter Variablen summiert, diese sind aber nicht völlig unabhängig. Da sich $\bar{x}$ unmittelbar aus den n verschiedenen x_i ergibt, sind nur $n-1$ x_i-Werte frei. Der letzte x_i-Wert ist durch $\bar{x}$ und die anderen x_i-Werte bereits fixiert und demnach nicht mehr frei. Folglich weist obige Summe nicht die Verteilung $\chi^2[n]$, sondern $\chi^2[n-1]$ auf. Also gilt für S_X^2:

$$S_X^2 = \frac{\sigma_X^2}{n-1} \cdot \chi^2[n-1]$$