Zufallsstichproben

Definition

Eine *einfache Zufallsstichprobe* vom Umfang n ist eine Folge von stochastisch unabhängigen und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$, den sogenannten *Stichprobenvariablen*. Dabei bezeichnet X_i die Merkmalsausprägung des i-ten Elements in der Stichprobe. Die beobachteten Merkmalswerte $x_1, \ldots, x_n$ der n Elemente sind Realisierungen der Zufallsvariablen $X_1, \ldots, X_n$ und heissen *Stichprobenwerte*.

Parameterschätzungen

Definition

Allgemein ist eine *Stichprobenfunktion* eine Funktion, die von den Stichprobenvariablen $X_1, \ldots, X_n$ abhängt. Eine *Schätzfunktion* $\Theta = g(X_1, \ldots, X_n)$ ist eine spezielle Stichprobenfunktion, nämlich eine „Formel“, mit der man den Wert eines Parameters θ der Grundgesamtheit schätzen kann: Setzt man eine konkrete Stichprobe $x_1, \ldots x_n$ ein, so erhält man einen *Schätzwert* $\hat{\theta} = g(x_1, \ldots, x_n)$ für den Parameter θ.

Definition

Eine Schätzfunktion Θ eines Parameters θ heisst *erwartungstreu*, wenn gilt:

$$E(\Theta) = \theta$$

Gegeben sind zwei erwartungstreue Schätzfunktionen Θ_1 und Θ_2 desselben Parameters θ. Man nennt Θ_1 *effizienter* als Θ_2, falls gilt:

$$V(\Theta_1) < V(\Theta_2)$$

Eine Schätzfunktion Θ eines Parameters θ heisst *konsistent*, wenn gilt:

$$E(\Theta) \to \theta \text{ und } V(\Theta) \to 0 \text{ für } n \to \infty$$

Schätzfunktionen für die wichtigsten statistischen Parameter

	Schätzfunktion	**Schätzwert**
Erwartungswert Spezialfall: Anteilswert einer Bernoulli-Verteilung	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$	$\hat{\mu} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ $\hat{p} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{\text{Anzahl 1en}}{n}$
Varianz	$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$	$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$

Standardabweichung	$S = \sqrt{S^2}$	$\hat{\sigma} = s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$

Satz
(1) $\bar{X}$ und S^2 sind erwartungstreu und konsistent.
(2) S ist konsistent, aber nicht erwartungstreu.

Vertrauensintervalle

Man bestimmt zwei Stichprobenfunktionen Θ_u und Θ_o, die den wahren Wert des Parameters θ mit der vorgegebenen Wahrscheinlichkeit γ einschliessen:

$$P(\Theta_u \leq \theta \leq \Theta_o) = \gamma$$

Setzt man nun die Werte $x_1, x_2, \ldots, x_n$ einer konkreten Stichprobe in Θ_u und Θ_o ein, so erhält man die Zahlen c_u und c_o. Das Intervall $[c_u; c_o]$ ist dann ein *Vertrauensintervall* für den unbekannten Parameter θ.

Die Wahrscheinlichkeit γ heisst *Vertrauensniveau* oder *statistische Sicherheit* (übliche Werte: 95% oder 99%); $\alpha = 1 - \gamma$ wird *Irrtumswahrscheinlichkeit* genannt.

Wenn man hundertmal eine Stichprobe nehmen und zu jeder Stichprobe das Vertrauensintervall berechnen würde, so würden etwa $100 \cdot \gamma$ dieser Intervalle (bei $\gamma = 95\%$ also etwa 95) den wahren Wert des Parameters einschliessen: