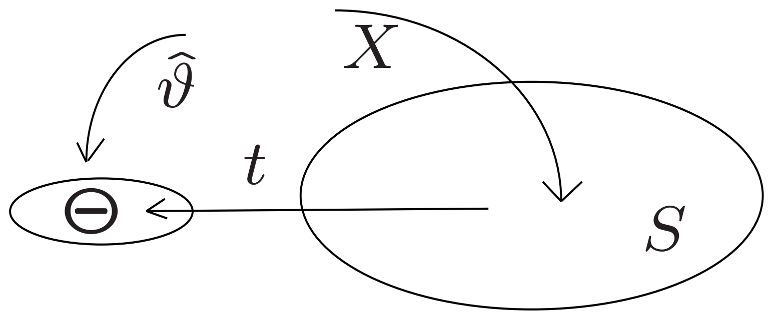
Vorlesung 12b

Konfidenzintervalle

Ein Logo der Statistik:



$$\mathbf{P}_\vartheta(X \in da) = \rho_\vartheta(da)\,, \quad \vartheta \in \Theta$$

Θ ... *Parameterraum*

S ... *Beobachtungsraum*

$\widehat{\vartheta} := t(X)$ ... *Schätzer* für den Parameter ϑ

Sei $m(\vartheta)$ ein reelles Parametermerkmal

und $I = I(X)$ ein aus den Daten konstruiertes Intervall.

Gilt für ein $\alpha \in (0, 1)$

$$\mathbf{P}_\vartheta(m(\vartheta) \in I) \geq 1 - \alpha \quad \text{für jedes } \vartheta \in \Theta$$

dann sagt man:

I ist ein *Konfidenzintervall* für $m(\vartheta)$ mit *Niveau* $1 - \alpha$,

es hält die *Überdeckungswahrscheinlichkeit* $1 - \alpha$ ein.

Eine Zahl ν heißt *Median* der Verteilung ρ auf $\mathbb{R}$,

wenn sowohl $\rho((-\infty, \nu]) \geq 1/2$

als auch $\rho([\nu, \infty)) \geq 1/2$ gilt.

Die *Ordnungsstatistiken* $X_{(1)} \leq \cdots \leq X_{(n)}$

sind die aufsteigend geordneten $X_1, \ldots, X_n$

Ein Kandidat für ein Konfidenzintervall für den Median ist

$$[X_{(1+j)}, X_{(n-j)}]$$

mit $0 \leq j < n/2$.

$$\mathbf{P}_\rho\big(\nu \notin [X_{(1)}, X_{(n)}]\big) = \mathbf{P}_\rho(X_{(1)} > \nu) + \mathbf{P}_\rho(X_{(n)} < \nu)\ .$$

$$\mathbf{P}_\rho(X_{(1)} > \nu) \le 2^{-n}$$

$$\mathbf{P}_\rho(X_{(n)} < \nu) \le 2^{-n}.$$

Also:

$$\mathbf{P}_\rho\big(\nu \in [X_{(1)}, X_{(n)}]\big) \ge 1 - \frac{1}{2^{n-1}}\ .$$

$$\mathbf{P}_\rho\big(\nu \notin [X_{(1+j)}, X_{(n-j)}]\big)$$
$$= \mathbf{P}_\rho(X_{(1+j)} > \nu) + \mathbf{P}_\rho(X_{(n-j)} < \nu)\ .$$

$$\mathbf{P}_\rho(X_{(j+1)} > \nu) = \mathbf{P}(\text{höchstens } j \text{ der } X_i \text{ sind } \leq \nu)$$
$$\leq \mathbf{P}(Y \leq j)$$

mit Y Bin$(n, 1/2)$-verteilt.

Denn:

$$\mathbf{P}(X_i \leq \nu) \geq 1/2$$

und für $k < 1/2$ ist die Funktion $p \mapsto p^k(1-p)^{n-k}$

monoton fallend auf $[\frac{1}{2}, 1]$.

$$\mathbf{P}_\rho\Big(\nu \notin [X_{(1+j)}, X_{(n-j)}]\Big)$$
$$= \mathbf{P}_\rho(X_{(1+j)} > \nu) + \mathbf{P}_\rho(X_{(n-j)} < \nu)\ .$$

$$\mathbf{P}_\rho(X_{(j+1)} > \nu) \leq \mathbf{P}(Y \leq j)$$

mit Y Bin$(n, 1/2)$-verteilt. Also:

$$\mathbf{P}_\rho\Big(\nu \in [X_{(1+j)}, X_{(n-j)}]\Big) \geq 1 - 2\mathbf{P}(Y \leq j)$$

Mit wachsendem j wird das Intervall kürzer

und die Überdeckungsw'keit nimmt ab!

Approximatives Konfidenzintervall für den Mittelwert

$$\mu := \int a\rho(da)$$

$$\widehat{\mu} := \frac{1}{n}(X_1 + \cdots + X_n)$$

$$\widehat{\sigma}^2 := \frac{1}{n}\Big((X_1 - \widehat{\mu})^2 + \cdots + (X_n - \widehat{\mu})^2\Big)$$

Aus dem Zentralen Grenzwertsatz folgt (vgl.Übung 29)

für $q > 0$ und N(0, 1)-verteiltes Z :

$$\mathbf{P}_\rho\Big(\mu \in \Big[\,\widehat{\mu} - q\frac{\widehat{\sigma}}{\sqrt{n}},\ \widehat{\mu} + q\frac{\widehat{\sigma}}{\sqrt{n}}\,\Big]\Big) \overset{n\to\infty}{\to} \mathbf{P}(-q \le Z \le q)$$

$$\mathbf{P}_\rho\Big(\mu \in \Big[\hat{\mu} - q\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{\mu} + q\frac{\hat{\sigma}}{\sqrt{n}}\Big]\Big) \overset{n\to\infty}{\to} \mathbf{P}(-q \le Z \le q)$$

Sei q so, dass

$$\mathbf{P}(-q \le Z \le q) = 1 - \alpha$$

(das $(1-\alpha/2)$-*Quantil* der Standard-Normalverteilung)

Dann ist

$$I := \left[\hat{\mu} - q\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{\mu} + q\frac{\hat{\sigma}}{\sqrt{n}}\right],$$

ein Konfidenzintervall für μ,

approximativ vom Niveau $1 - \alpha$ für große n.

Konfidenzintervalle für p:

(Anteil oder Erfolgswahrscheinlichkeit)

vgl. Vorlesung 12a,

Buch S. 122 und 129/130,

sowie *Woche 6* des Proseminars

“Exploratives Rechnen in der Stochastik”:

http://math.uni-frankfurt.de/ ferebee/explorativ/index.html