

Schätzung

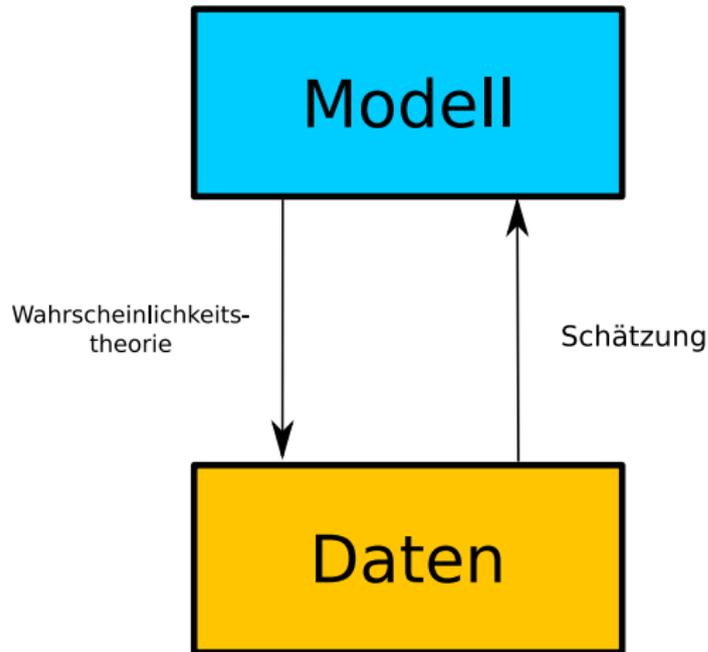
Raimar Sandner

Studentenseminar "Statistische Methoden in der Physik"

Gliederung

- 1 Theorie
 - Schätzfunktionen
 - Eigenschaften von Schätzern
- 2 Parameterschätzung
 - Momentenmethode
 - Maximum Likelihood
 - Methode der kleinsten Quadrate
- 3 Ausblick

Worum geht es hier?



Gliederung

- 1 Theorie
 - Schätzfunktionen
 - Eigenschaften von Schätzern
- 2 Parameterschätzung
 - Momentenmethode
 - Maximum Likelihood
 - Methode der kleinsten Quadrate
- 3 Ausblick

Stichproben

Gegeben eine Beobachtungsreihe $x = (x_1, x_2, \dots, x_n)$:

- Realisierung der n -dimensionalen Zufallsvariablen $X = (X_1, X_2, \dots, X_n)$
- **unabhängige** Stichprobe, falls X_1, \dots, X_n unabhängig sind
- **einfache** Stichprobe, falls zusätzlich X_1, \dots, X_n identisch verteilt

Beispiel

n unabhängige Realisierungen der gleichen Zufallsvariablen Z

Stichprobenfunktion

$$t_n = g_n(x_1, x_2, \dots, x_n)$$

Funktionswert ist Realisierung einer neuen Zufallsvariablen

Stichprobenfunktion

$$T_n = g_n(X_1, X_2, \dots, X_n)$$

- Verwendung zur Parameterschätzung \rightsquigarrow Schätzfunktion
- Verwendung für Tests \rightsquigarrow Testfunktion

Schätzfunktion

Gegeben Stichprobe $x = (x_1, x_2, \dots, x_n)$.

Problem

Wir kennen nicht die genaue Verteilung der X_i , haben aber eine Vermutung

Verteilung hängt von unbekanntem Parameter θ ab

Beispiel

Annahme: die Zufallsvariablen X_i haben Dichtefunktion

$$f(x) = \frac{1}{\theta_1 \sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{x - \theta_2}{\theta_1} \right)^2 \right)$$

Schätzfunktion

- Konstruiere Schätzfunktion:

$$T_n = g_n(X_1, X_2, \dots, X_n),$$

hängt auch von den Parametern ab.

- Aus der Menge der Schätzfunktionen zu einem Parameter:
finde „gute“ Schätzfunktionen

Gliederung

- 1 Theorie
 - Schätzfunktionen
 - **Eigenschaften von Schätzern**
- 2 Parameterschätzung
 - Momentenmethode
 - Maximum Likelihood
 - Methode der kleinsten Quadrate
- 3 Ausblick

einige Eigenschaften

erwartungstreu

$$E(T_n) = E(g_n(X_1, X_2, \dots, X_n)) = \theta$$

asymptotisch erwartungstreu

$$\lim_{n \rightarrow \infty} E(T_n) = \lim_{n \rightarrow \infty} E(g_n(X_1, X_2, \dots, X_n)) = \theta$$

konsistent

Folge von Schätzfunktionen $T_n = g_n(X_1, X_2, \dots, X_n)$ mit

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$$

heißt konsistent.

Tschebyschow-Ungleichung

Sei X Zufallsvariable mit Erwartungswert μ und endlicher Varianz σ^2 . Dann gilt für alle $\varepsilon > 0$

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Hinreichende Bedingung für Konsistenz

Alle T_n besitzen endliche Varianz σ_n^2 mit $\lim_{n \rightarrow \infty} \sigma_n^2 = 0$.

$$P(|T_n - \theta| \geq \varepsilon) \leq \frac{\sigma_n^2}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$$

wirksamste Schätzer

T_n^* (erwartungstreu) ist wirksamster Schätzer für Parameter θ ,
wenn

$$\text{var}(T_n^*) \leq \text{var}(T_n)$$

für jeden beliebigen erwartungstreuen Schätzer T_n von θ .

Wirksamkeit eines Schätzers T_n

$$W(T_n) = \frac{\text{var}(T_n^*)}{\text{var}(T_n)} \leq 1$$

untere Schranke der Varianz

Mit allgemeinen Regularitätsbedingungen gilt

Ungleichung von Rao-Cramér

$$\text{var}_\theta(T_n) \geq \frac{1}{n \cdot I(\theta)}, \quad T_n \text{ erwartungstreuer Schätzer für } \theta$$

Fisher-Information

$$I(\theta) := \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) dx$$

Gliederung

- 1 Theorie
 - Schätzfunktionen
 - Eigenschaften von Schätzern
- 2 **Parameterschätzung**
 - **Momentenmethode**
 - Maximum Likelihood
 - Methode der kleinsten Quadrate
- 3 Ausblick

Idee der Momentenmethode

Erinnerung

X Zufallsvariable, das k -te Moment (um Null) der Verteilung von X :

$$m_k = E_{\theta_1, \dots, \theta_j} (X^k)$$

- drücke die Parameter der Verteilung als Funktion der Momente aus:

$$\theta_i = h_i(m_1, \dots, m_r)$$

- setze für die Momente die empirischen Momente aus einer Stichprobe ein:

$$\hat{\theta}_i = h_i(\hat{m}_1, \dots, \hat{m}_r), \quad \hat{m}_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

Formal

einige Modellannahmen

- Parametervektor $\theta = (\theta_1, \dots, \theta_j)$
- für $k = 1, \dots, r$ soll sich das **theoretische** k -te Moment als Funktion von θ schreiben lassen: $m_k = g_k(\theta)$
- Das Gleichungssystem der **empirischen** Momente

$$\hat{m}_k(x_1, \dots, x_n) = g_k(\theta) \quad k \in \{1, \dots, r\}$$

für unbekanntes θ soll eindeutig lösbar sein, Lösung
 $\hat{\theta}(x_1, \dots, x_n)$

Zufallsvektor $\hat{\theta}(X_1, \dots, X_n)$ heißt **Momentenschätzer**

Beispiel

Betrachte normalverteilte Stichprobenvariablen, Mittelwert μ und Varianz σ^2 unbekannt, also $\theta = (\mu, \sigma^2)$.

$$m_1 = \mu \qquad m_2 = \sigma^2 + \mu^2$$

Ergibt Gleichungssystem

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n x_i = \mu \qquad \hat{m}_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = \sigma^2 + \mu^2$$

Lösung:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Eigenschaften der Momentenmethode

- unter bestimmten Stetigkeitsannahmen ist der Schätzer konsistent
- insbesondere asymptotisch erwartungstreu

Nachteile der Momentenmethode

- im Allgemeinen schlechte Wirksamkeit
- nicht immer erwartungstreu (wie gesehen)

Gliederung

- 1 Theorie
 - Schätzfunktionen
 - Eigenschaften von Schätzern
- 2 Parameterschätzung**
 - Momentenmethode
 - Maximum Likelihood**
 - Methode der kleinsten Quadrate
- 3 Ausblick

Idee der Maximum-Likelihood-Schätzung

Passe Parameter $\theta = (\theta_1, \dots, \theta_m)$ der Verteilung so an eine gegebene Stichprobe an, dass die Wahrscheinlichkeit (Dichte) für genau diese Stichprobe maximal wird.

Definition

Likelihood: Interpretiere Dichte als Funktion von θ bei konstanten Daten. Für einfache Stichprobe:

$$L(x_1, \dots, x_n; \theta_1, \dots, \theta_m) = \prod_{i=1}^m f(x_i; \theta_1, \dots, \theta_m)$$

Konstruktion eines Maximum-Likelihood-Schätzers

Maximierung von L führt auf Gleichungssystem

$$\frac{\partial L}{\partial \theta_1} = 0 \quad \frac{\partial L}{\partial \theta_2} = 0 \quad \dots \quad \frac{\partial L}{\partial \theta_m} = 0,$$

äquivalent zu

$$\frac{\partial \ln L}{\partial \theta_1} = 0 \quad \frac{\partial \ln L}{\partial \theta_2} = 0 \quad \dots \quad \frac{\partial \ln L}{\partial \theta_m} = 0.$$

Logarithmus oft einfacher, da

$$\ln L(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \ln f(x_i; \theta).$$

Beispiel

Betrachte Stichprobe (x_1, \dots, x_n) exponentiell verteilter Zufallsvariablen, Parameter μ :

$$f(x, \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$$

Likelihood-Funktion:

$$L(x_1, \dots, x_n; \mu) = \frac{1}{\mu^n} \prod_{i=1}^n e^{-\frac{x_i}{\mu}}$$

$$\ln L(x_1, \dots, x_n; \mu) = -n \ln \mu - \sum_{i=1}^n \frac{x_i}{\mu}$$

$$\frac{\partial L}{\partial \mu} = -\frac{n}{\mu} + \sum_{i=1}^n \frac{x_i}{\mu^2} \stackrel{!}{=} 0 \quad \Rightarrow \quad \mu = \sum_{i=1}^n \frac{x_i}{n}$$

Vor- und Nachteile der Maximum-Likelihood-Schätzung

Vorteile:

- unter bestimmten Regularitätsvoraussetzung existiert der Schätzer
- sofern sie existieren, sind sie asymptotisch effizient (schwache Konvergenz gegen eine Zufallsvariable mit minimaler Varianz) und konsistent

Nachteile:

- Manchmal ist Schätzer $\hat{\theta}(X_1, \dots, X_n)$ nicht eindeutig bestimmt
- Optimierungsproblem könnte analytisch nicht lösbar sein
↪ numerische Verfahren

Gliederung

- 1 Theorie
 - Schätzfunktionen
 - Eigenschaften von Schätzern
- 2 **Parameterschätzung**
 - Momentenmethode
 - Maximum Likelihood
 - **Methode der kleinsten Quadrate**
- 3 Ausblick

Idee der Methode

Problem

Finde Abhängigkeit einer Variable y von einer vorgegebenen Variable t bzw. von mehreren unabhängigen Variablen

$$\vec{t} = t_1, \dots, t_q.$$

$$y_m = f(t_1, \dots, t_q; \theta)$$

- f ist vermuteter Funktionstyp, darf von Parametern $\theta = (\theta_1, \dots, \theta_p)$ abhängen
- Beobachtungsdaten $((y_1, \vec{t}_1), \dots, (y_n, \vec{t}_n))$
- Daten sollten durch Modellfunktion „möglichst gut“ approximiert werden \rightsquigarrow minimiere die quadratische Abweichungen

Warum quadratische Abweichung minimieren?

Modellannahmen:

- Daten sind von der Form $f(\vec{t}_i, \theta) = y_i + v_i$
- Messfehler v_i normalverteilt mit Erwartungswert 0
- alle gleiche Varianz
- v_i stochastisch unabhängig von \vec{t}_i und von jedem anderen v_j .

Dann entspricht die Minimierung der quadratischen Abweichung dem Maximum-Likelihood-Schätzer für θ .

Weiterführende Themen

- Konstruktion von robusten Schätzern (robuste Statistik)
- Intervallschätzung, Konfidenzintervalle
- Testen