Schätzen und Testen I in 90 Minuten

Autorin: Christiane Fuchs
Bearbeitet von: Julia Sommer, Ludwig Bothmann, David Rügamer

2. Februar 2016

Kapitel 1:

Einführung in statistische Modelle und Inferenzkonzepte

Parametrische Modelle $X \sim f(x|\theta)$ für $\theta \in \Theta$

- Verteilungsfamilien
  - Darstellung über Standardverteilung mit Lokalisationsparameter a und Skalenparameter b

$$\frac{1}{b} f_0 \left(\frac{x-a}{b} \right)$$

  - Exponentialfamilien

$$f(x|\boldsymbol{\theta}) = h(x) \exp \left(b(\boldsymbol{\theta}) + \boldsymbol{\gamma}(\boldsymbol{\theta})' \mathbf{T}(x) \right)$$

    i.A. Fisher-regulär, schöne Eigenschaften

- übersicht über (generalisierte) lineare parametrische Modelle

Non- und semi-parametrische Modelle

- Schätzen & Testen II

Konzepte:

- Statistische Entscheidungstheorie
- Klassische Schätz- und Testtheorie
- Likelihood-Inferenz
- Bayes-Inferenz

Kapitel 2:

Klassische Schätz- und Testtheorie

Gewünschte Eigenschaften einer Statistik / eines Schätzers $\hat{\theta} = T(X)$

Suffizienz	„keine Info verschenken“	Faktorisierung (Neyman)
		Fisher-Info
Minimalsuffizienz	„suffizient & eindeutig“	Satz: Likelihoodquotient
		Lehmann-Scheffé: Suffizient + vollständig
MSE-optimal	„Bias+Varianz minimieren“	Zulässige Schätzer

=> Einschränkung auf erwartungstreue Schätzer

Erwartungstreue	„unverzerrt, kein Bias“	UMVU
Effizienz	„Varianz möglichst klein“	
		Fisher-Info
		Informationsungleichung & Cramer-Rao-Schranke
		Rao-Blackwellisierung

Asymptotische Eigenschaften einer Folge von Schätzern $\hat{\theta}_n = T(X_1, \ldots, X_n)$

- Asymptotische Erwartungstreue
- Konsistenz ($MSE_\theta(\hat{\theta}_n) \to 0$ für alle θ und $n \to \infty$)
- Asymptotische Normalität ($\sqrt{n}$-Normierung, Matrix-Normierung)
    - Delta-Methode
- Asymptotische Effizienz (BAN-Schätzer)
    - Asymptotische Cramer-Rao-Schranke
    - Fisher-Regularität

Gewünschte Eigenschaften eines Tests

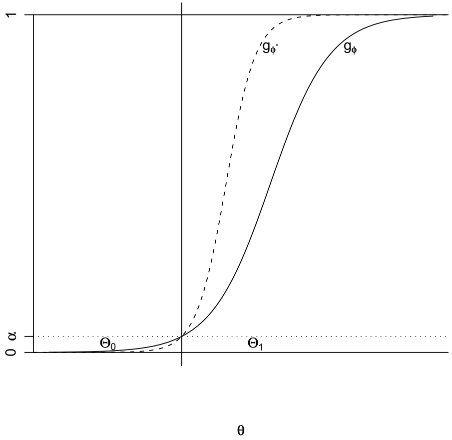
Testproblem:

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

Fehler 1. Art: $\alpha \rightarrow$ fest vorgegeben
Fehler 2. Art: $\beta \rightarrow$ minimieren (bzw. Macht maximieren)

Gesucht ist unter allen Tests ϕ der gleichmäßig beste Test ϕ^* zum Niveau α (UMP)

- Gütefunktion



UMP-Test für einfache Hypothesen

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1$$

- **Satz von Neyman-Pearson**:

  Der eindeutig bestimmte UMP-Test zum exakten Niveau α ist ein randomisierter LQ-Test, d.h. von der Form

$$\phi^*(x) = \begin{cases} 1 \;, \; \Lambda(x) > k & H_0 \text{ ablehnen} \\ \gamma^*, \; \Lambda(x) = k & \text{randomisieren} \\ 0 \;, \; \Lambda(x) < k & H_0 \text{ beibehalten} \end{cases}$$

  mit Konstante $k > 0$ und $0 \leq \gamma^* < 1$.

- **Niveaubedingung zur Bestimmung von γ^*:**

$$\boxed{\mathbb{P}_{\theta_0}(\Lambda(x) > k) + \gamma^* \mathbb{P}_{\theta_0}(\Lambda(x) = k) = \alpha}$$

UMP-Test für einseitige Hypothesen

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0$$

- **Satz (UMP-Test bei MLQ)**:

  Falls die Verteilungsfamilie MLQ in $T(x)$ (d.h. der Likelihoodquotient $\Lambda(\cdot)$ ist monoton steigend in $T(x)$), gibt es einen UMP-Test zum exakten Niveau α, nämlich

$$\phi^*(x) = \begin{cases} 1\ , & T(x) > c \qquad H_0 \text{ ablehnen} \\ \gamma^*, & T(x) = c \qquad \text{randomisieren} \\ 0\ , & T(x) < c \qquad H_0 \text{ beibehalten} \end{cases}$$

  mit Konstante $0 \leq \gamma^* < 1$.

- **Niveaubedingung zur Bestimmung von γ^* und c:**

$$\boxed{\mathbb{P}_{\theta_0}(T(x) > c) + \gamma^* \mathbb{P}_{\theta_0}(T(x) = c) = \alpha}$$

UMPU-Test für zweiseitige Hypothesen

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

- Zusätzliche Einschränkung auf unverfälschte Tests
- Entsprechende Bestimmung eines UMPU-Tests

Veranschaulichung der Randomisierung

Testproblem:

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

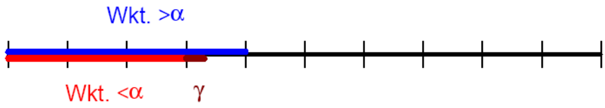
Randomisierter Test:

$$\phi(x) = \begin{cases} 1 \quad , & x \in B_1 \\ \gamma(x), & x \in B_{10} \\ 0 \quad , & x \in B_0 \end{cases}$$

B_1 strikter *Ablehnungsbereich*

B_0 strikter *Annahmebereich*

B_{10} *Randomisierungsbereich*, *Indifferenzbereich*



Klassische Schätz- und Testtheorie

was sonst noch behandelt wurde...

- Bereichsschätzung und Konfidenzintervalle
- Dualität von Konfidenzbereichen und Tests
- Multiples Testen

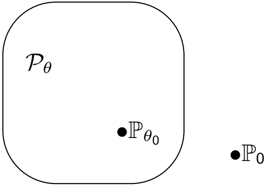
Kapitel 3:

Likelihood-Inferenz

Likelihood-Inferenz

- Schätzkonzept: Maximierung der Likelihood
- numerische Methoden zur Bestimmung des ML-Schätzers
- Asymptotische Eigenschaften
  - Konsistenz
  - Erwartungstreue
  - Normalität
  - ML-Schätzer sind BAN-Schätzer (unter Fisher-Regularität)
- Testen linearer Hypothesen
  - Konstruktion von Teststatistiken durch Ausnutzen von Verteilungsaussagen über Likelihood, Score-Funktion oder den ML-Schätzer selbst
- Asymptotische Konfidenzintervalle
- Inferenz nach Modellselektion

Fehlspezifikation



Gründe für Fehlspezifikation:

- falsche Verteilungsannahme
- falsche Erwartungswertstruktur
- falsche Kovarianzstruktur

Werkzeuge:

- Kullback-Leibler-Distanz
- Quasi-Likelihood

Kapitel 4:

Bayes-Inferenz

Basiskomponenten in der Bayes-Inferenz

- $p(\theta)$ Priori-Verteilung
- $f(x|\theta)$ Daten-Verteilung, Likelihood
- $f(\theta|x)$ Posteriori-Verteilung

$$f(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)} \propto f(x|\theta)p(\theta)$$

- $f(\widetilde{x}|x)$ prädiktive Verteilung

Modelle

- Modelle mit einem Parameter
- Multivariate Normalverteilung
- Dirichlet-Multinomial-Modell
- Bayesianisches LM und GLM

Dazu berechnet:

- gemeinsame Posteriori
- vollbedingte Dichten / Full Conditionals
- Posteriori-Randdichten / marginale Posterioridichten
- a posteriori prädiktive Verteilung

Wahl einer Priori Verteilung

- konjugiert, semi-konjugiert
  - Posteriori analytisch berechenbar, keine Simulation nötig
- beliebige Verteilung, die das Vorwissen gut ausdrückt
- nicht-informativ
  - 'Die Daten für sich sprechen lassen'
  - Jeffreys' Priori: Posterioriverteilung wird von den Daten dominiert, nicht vom Vorwissen. Erfüllt das Invarianzprinzip.
  - Alternativ: flache Priori, Priori mit maximaler Varianz

Punkt- und Intervallschätzer

Punktschätzer:

- Erwartungswert, Modus, Median der Posterioriverteilung

Intervallschätzer:

- Gleichendiges Kredibilitätsintervall: $\alpha/2$ und $(1-\alpha/2)$ – Quantil der Posterioriverteilung
- HPD-Kredibilitätsintervall: Kleinstes $(1-\alpha)$ – Kredibilitätsintervall

Algorithmen

Ziel: Simulation aus der analytisch nicht zugänglichen Posteriori-Verteilung.

- Direkte Simulation
- Markov Chain Monte Carlo Simulation
  - Gibbs
  - Metropolis-Hastings

$\Rightarrow$ **Schätzung** der theoretischen Momente der Posteriori-Verteilung durch empirische Analoga. Z.B. Erwartungswert durch arithmetisches Mittel oder theoretische Quantile durch empirische Quantile der simulierten Zufallszahlen.

Direkte Simulation von $\boldsymbol{\theta} = (\gamma, \tau)'$

Voraussetzung: Simulation aus $f(\gamma)$ und $f(\tau|\gamma)$ möglich.

Algorithmus:

- Wiederhole für $k = 1, \ldots, K$:

  Schritt 1: Ziehe $\gamma^{(k)}$ aus $f(\gamma)$.

  Schritt 2: Ziehe $\tau^{(k)}$ aus $f(\tau|\gamma^{(k)})$.
- Man erhält K Paare $\boldsymbol{\theta}^{(k)} = (\gamma^{(k)}, \tau^{(k)})'$ aus der Verteilung von $\boldsymbol{\theta}$.

Empirischer CDF-Sampler zur Simulation von θ

Voraussetzung: $f(\theta)$ muss bis auf Proportionalitätskonstante bekannt sein (und θ skalar).

Algorithmus:

- Diskretisiere den Träger von θ in eine Menge von N Punkten $x_1 \leq \ldots \leq x_N$.
- Evaluiere den Kern von $f(\theta)$ an $x_1 \leq \ldots \leq x_N$, um Werte $f_1, \ldots, f_N$ zu erhalten.
- Schätze die Proportionalitätskonstante c über $c = f_1 + \ldots + f_N$.
- Ziehe Zufallszahlen aus $x_1 \leq \ldots \leq x_N$ gemäß den Wahrscheinlichkeiten $f_1/c, \ldots, f_N/c$.

Gibbs-Sampler zur Simulation von $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$

$$\begin{aligned} \boldsymbol{\theta}_{-i} &= (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)' \\ \boldsymbol{\theta}_{-i}^{(k)} &= (\theta_1^{(k)}, \ldots, \theta_{i-1}^{(k)}, \theta_{i+1}^{(k-1)}, \ldots, \theta_n^{(k-1)})' \end{aligned}$$

Voraussetzung: Simulation von Zufallszahlen aus den vollbedingten Dichten $f(\theta_i|\boldsymbol{\theta}_{-i})$ muss möglich sein.

Algorithmus:

- Wähle einen Startwert $\boldsymbol{\theta}^{(0)}$.
- Wiederhole für $k = 1, \ldots, K$:
  - Für $i = 1, \ldots, n$: Ziehe $\theta_i^{(k)}$ aus $f(\theta_i|\boldsymbol{\theta}_{-i}^{(k)})$.
- Nach einer ausreichenden Zahl M von Wiederholungen können die Ziehungen $\boldsymbol{\theta}^{(M+1)}, \ldots, \boldsymbol{\theta}^{(K)}$ als Ziehungen aus $f(\boldsymbol{\theta})$ angesehen werden.

Metropolis-Hastings-Algorithmus zur Simulation von $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$

Voraussetzung: Der Kern der vollbedingten Dichten $f(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{x})$ oder der gemeinsamen Posteriori $f(\boldsymbol{\theta}|\mathbf{x})$ muss bekannt sein.

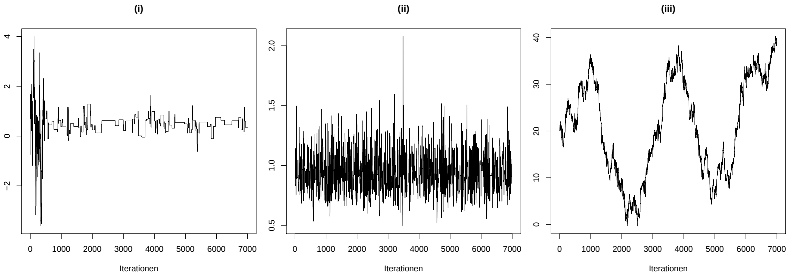
Algorithmus:

- Wähle einen Startwert $\boldsymbol{\theta}^{(0)}$.
- Wiederhole für $k = 1, \ldots, K$:
  - Für $i = 1, \ldots, n$:
    - Ziehe Y aus einer geeigneten Vorschlagsdichte $q(\cdot|\theta_i^{(k-1)}, \boldsymbol{\theta}_{-i}^{(k)})$.
    - Setze $\theta_i^{(k)} := Y$ mit Wahrscheinlichkeit

$$\alpha(Y, \theta_i^{(k-1)}) = \min\left\{1, \frac{f(Y|\boldsymbol{\theta}_{-i}^{(k)})}{f(\theta_i^{(k-1)}|\boldsymbol{\theta}_{-i}^{(k)})} \cdot \frac{q(\theta_i^{(k-1)}|Y, \boldsymbol{\theta}_{-i}^{(k)}))}{q(Y|\theta_i^{(k-1)}, \boldsymbol{\theta}_{-i}^{(k)}))}\right\} .$$

    - Ansonsten setze $\theta_i^{(k)} := \theta_i^{(k-1)}$.
- Nach einer ausreichenden Zahl M von Wiederholungen können die Ziehungen $\boldsymbol{\theta}^{(M+1)}, \ldots, \boldsymbol{\theta}^{(K)}$ als Ziehungen aus $f(\boldsymbol{\theta})$ angesehen werden.

Konvergenzuntersuchungen



Likelihood- und Bayes-Inferenz im Vergleich

	Likelihood	**Bayes**
Stützmaß für Inferenz	Likelihood-Funktion	Posteriori-Verteilung
a priori	$f(x\|\theta)$ Dichte von X	$f(x\|\theta)$ Dichte von X und $p(\theta)$ Dichte von θ
a posteriori	Likelihood: $L(\theta) = f(x\|\theta)$. Wie wahrscheinlich ist es gewesen, x zu beobachten, wenn θ der 'wahre' Parameter ist.	Posteriori: $f(\theta\|x)$ beschreibt die Verteilung des Parameters nach Beobachtung von x.
Parametrische Modellierung	Unbekannter Parameter θ ist fest, Wahrscheinlichkeitsaussagen betreffen nur den Schätzer $\hat{\theta}_n$.	Unbekannter Parameter θ ist stochastisch, hat eine Verteilung.
Berechnung & Werkzeuge	Likelihood maximieren ▶ analytisch ▶ numerisch (Newton Raphson, ...)	Posteriori berechnen ▶ analytisch ▶ numerisch über Simulation / MCMC