

3.2 Maximum-Likelihood-Schätzung

3.2.2 Iterative numerische Verfahren zur Berechnung des ML-Schätzers

Beispiel 3.7 (Gemischte Modelle)

Eine Herleitung für E- und M-Schritt in linearen gemischten Modellen findet sich in Pawitan, Kapitel 12.8.

3.2 Maximum-Likelihood-Schätzung

3.2.3 Asymptotische Eigenschaften

Satz 3.4

Seien X_1, \dots, X_n i.i.d. aus einer Dichte $f(x|\theta)$, die folgenden Annahmen genügt:

- ▶ $f(x|\theta)$ ist Fisher-regulär.
- ▶ Die Informationsmatrix $\mathbf{I}(\theta)$ ist positiv definit im Inneren von Θ .
- ▶ Es existieren Funktionen M_{jkl} derart, dass

$$\left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \log f(x|\theta) \right| \leq M_{jkl}(x)$$

und

$$\mathbb{E}_{\theta_0}[M_{jkl}(X)] < \infty$$

für alle j , k und l , wobei θ_0 den wahren Wert des Parameters bezeichnet.

3.2 Maximum-Likelihood-Schätzung

3.2.3 Asymptotische Eigenschaften

Satz 3.4 (fortgeführt)

Dann gilt unter weiteren, relativ schwachen Regularitätsannahmen:

- ▶ *Die Likelihood-(ML-)Gleichungen haben für $n \rightarrow \infty$ mit Wahrscheinlichkeit 1 eine Lösung $\hat{\theta}_n$ (d.h.*

$\mathbb{P}(\hat{\theta}_n \text{ existiert}) \rightarrow 1$) mit $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$; die konsistente Lösung $\hat{\theta}_n$ ist eindeutig und $\mathbb{P}(\hat{\theta}_n \text{ ist (lokales) Maximum}) \rightarrow 1$.

- ▶ $\hat{\theta}_n \stackrel{a}{\sim} N(\theta_0, \mathbf{I}_n^{-1}(\theta_0))$ bzw. $\mathbf{I}_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I)$,
- ▶ $\hat{\theta}_n \stackrel{a}{\sim} N(\theta_0, \mathbf{J}_n^{-1}(\theta_0))$ bzw. $\mathbf{J}_n^{1/2}(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I)$,

d.h. ML-Schätzer sind asymptotisch erwartungstreue BAN-Schätzer.

3.2 Maximum-Likelihood-Schätzung

3.2.3 Asymptotische Eigenschaften

Bemerkung.

1. Es sind auch andere Varianten von Regularitätsannahmen möglich.
2. Der Satz gilt unter stärkeren Regularitätsannahmen auch für i.n.i.d. und abhängige X_1, \dots, X_n
3. $\mathbf{I}(\theta_0)$ und $\mathbf{J}(\theta_0)$ können auch durch $\mathbf{I}(\hat{\theta}_n)$ bzw. $\mathbf{J}(\hat{\theta}_n)$ ersetzt werden.

3.2 Maximum-Likelihood-Schätzung

3.2.3 Asymptotische Eigenschaften

Beweis. Erfolgt lediglich skizzenhaft.

- *Existenz (für skalares θ):* Es gilt für alle $\theta \neq \theta_0$, dass

$$\mathbb{P}_{\theta_0} \left(\prod_{i=1}^n f(x_i|\theta_0) > \prod_{i=1}^n f(x_i|\theta) \right) \rightarrow 1 \text{ für } n \rightarrow \infty. \quad (2)$$

Beweis: Logarithmieren liefert

$$\frac{1}{n} \sum_{i=1}^n \log (f(x_i|\theta)/f(x_i|\theta_0)) < 0.$$

Nach dem Gesetz der großen Zahlen konvergiert die linke Seite in Wahrscheinlichkeit gegen die Kullback-Leibler-Distanz

$$\mathbb{E}_{\theta_0}[\log (f(x|\theta)/f(x|\theta_0))].$$

Anwendung der Ungleichung von Jensen liefert, dass

$$\mathbb{E}_{\theta_0}[\log (f(x|\theta)/f(x|\theta_0))] < \log E_{\theta_0}[f(x|\theta)/f(x|\theta_0)] = 0,$$

woraus die Behauptung (2) folgt.

3.2 Maximum-Likelihood-Schätzung

3.2.3 Asymptotische Eigenschaften

Wähle nun $a > 0$ klein genug, so dass $(\theta_0 - a; \theta_0 + a)$ vollständig in Θ enthalten ist. Setze

$$S_n = \{x : L(\theta_0; x) > L(\theta_0 - a; x) \text{ und } L(\theta_0; x) > L(\theta_0 + a; x)\}.$$

Für beliebige Stichproben $x \in S_n$ existiert somit ein Punkt $\hat{\theta}_n \in (\theta_0 - a; \theta_0 + a)$, der die Likelihood (lokal) maximiert, d.h. $s(\hat{\theta}_n) = 0$. Aus eben bewiesener Hilfsaussage folgt, dass $\mathbb{P}_{\theta_0}(S_n) \rightarrow 1$ für jedes beliebige a .

3.2 Maximum-Likelihood-Schätzung

3.2.3 Asymptotische Eigenschaften

- ▶ *Konsistenz und Eindeutigkeit (für skalares θ):* siehe Huzurbazar (1948).
- ▶ *Asymptotische Normalität der Score-Funktion:* Aus der Fisher-Regularität folgt, dass der Erwartungswert und die Kovarianzmatrix existieren und durch $\mathbb{E}[s_i(\theta)] = 0$ und $\text{Cov}(s_i(\theta)) = \mathbf{i}(\theta)$ gegeben sind. Der zentrale Grenzwertsatz liefert $s(\theta) \stackrel{a}{\sim} N(0, \mathbf{I}(\theta))$ bzw.

$$\mathbf{I}(\theta)^{-1/2} s(\theta) = (\mathbf{n}\mathbf{i}(\theta))^{-1/2} \left(\sum_{i=1}^n s_i(\theta) - 0 \right) \xrightarrow{d} N(0, I).$$

3.2 Maximum-Likelihood-Schätzung

3.2.3 Asymptotische Eigenschaften

- ▶ *Asymptotische Normalität von $\hat{\theta}_n$* : Eine Taylorentwicklung von $s(\hat{\theta}_n) = 0$ um θ führt zu

$$0 = s(\hat{\theta}_n) = s(\theta) - \mathbf{J}(\theta)(\hat{\theta}_n - \theta) + o(\hat{\theta}_n - \theta).$$

Wegen dem Satz von Slutsky können wir im Folgenden auch $\mathbf{J}(\theta)$ durch $\mathbf{I}(\theta) = \mathbb{E}[\mathbf{J}(\theta)]$ ersetzen, da

$\frac{1}{n}\mathbf{J}(\theta) = \frac{1}{n}\sum_{i=1}^n \mathbf{j}_i(\theta) \xrightarrow{\mathbb{P}} \mathbf{i}(\theta)$. Dies liefert

$$s(\theta) \stackrel{a}{\approx} \mathbf{I}(\theta)(\hat{\theta}_n - \theta) \quad \text{bzw.} \quad \hat{\theta}_n - \theta \stackrel{a}{\approx} \mathbf{I}^{-1}(\theta)s(\theta)$$

und somit

$$\hat{\theta}_n - \theta \stackrel{a}{\approx} N(0, \mathbf{I}^{-1}(\theta)\mathbf{I}(\theta)\mathbf{I}^{-1}(\theta)) = N(0, \mathbf{I}^{-1}(\theta)).$$

