Beispiel 3.7 (Gemischte Modelle)**.** *Eine Herleitung für E- und M-Schritt in linearen gemischten Modellen findet sich in Pawitan, Kapitel 12.8.*

3.2.3 Asymptotische Eigenschaften

Satz 3.4. *Seien $X_1, \ldots, X_n$ 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.

Dann gilt unter weiteren, relativ schwachen Regularitätsannahmen:

- *Die Likelihood-(ML-)Gleichungen haben für $n \to \infty$ mit Wahrscheinlichkeit 1 eine Lösung $\widehat{\theta}_n$ (d.h. $\mathbb{P}(\widehat{\theta}_n$ existiert$) \to 1$) mit $\widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$; die konsistente Lösung $\widehat{\theta}_n$ ist eindeutig und $\mathbb{P}(\widehat{\theta}_n$ ist (lokales) Maximum$) \to 1$.*
- *$\widehat{\theta}_n \overset{a}{\sim} N(\theta_0, \mathbf{I}_n^{-1}(\theta_0))$ bzw. $\mathbf{I}_n^{1/2}(\theta_0)(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I)$,*
- *$\widehat{\theta}_n \overset{a}{\sim} N(\theta_0, \mathbf{J}_n^{-1}(\theta_0))$ bzw. $\mathbf{J}_n^{1/2}(\theta_0)(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I)$,*

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

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 $X_1, \ldots, X_n$.*
3. *$\mathbf{I}(\theta_0)$ und $\mathbf{J}(\theta_0)$ können auch durch $\mathbf{I}(\widehat{\theta}_n)$ bzw. $\mathbf{J}(\widehat{\theta}_n)$ ersetzt werden.*

Beweis. Erfolgt lediglich skizzenhaft.

- *Existenz (für skalares θ):* Es gilt, dass

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

Beweis: Logarithmieren liefert

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

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

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

Anwendung der Ungleichung von Jensen liefert, dass

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

woraus die Behauptung folgt. 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 $\widehat{\theta}_n \in (\theta_0 - a; \theta_0 + a)$, der die Likelihood (lokal) maximiert, d.h. $s(\widehat{\theta}_n) = 0$. Aus eben bewiesener Hilfsaussage folgt, dass $\mathbb{P}_{\theta_0}(S_n) \rightarrow 1$ für jedes beliebige a.

- *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 $\mathrm{Cov}(s_i(\theta)) = \mathbf{i}(\theta)$ gegeben sind. Der zentrale Grenzwertsatz liefert $s(\theta) \overset{a}{\sim} N(0, \mathbf{I}(\theta))$ bzw.

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

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

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

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

$$s(\theta) \overset{a}{\sim} \mathbf{I}(\theta)(\widehat{\theta}_n - \theta) \quad \text{bzw.} \quad \widehat{\theta}_n - \theta \overset{a}{\sim} \mathbf{I}^{-1}(\theta)s(\theta)$$

und somit

$$\widehat{\theta}_n - \theta \overset{a}{\sim} N(0, \mathbf{I}^{-1}(\theta)\mathbf{I}(\theta)\mathbf{I}^{-1}(\theta)) = N(0, \mathbf{I}^{-1}(\theta)).$$

□