Kapitel VIII

Nichtlineare Modelle, Stochastische Prozesse

1 Nichtlineares Regressionsmodell

2 Verallgemeinertes lineares Modell (GLM)

3 Stochastische Prozesse

Unser Schätz- und Test-theoretischer Ansatz in V.3 und VI.4 ist allgemein genug, um parametrische statistische Analysen in stochastischen Prozessen durchführen zu können. In der Tat, wir haben zugelassen, dass die Zufallsvariablen $X_1, \ldots, X_n$ Abhängigkeiten aufweisen Allerdings beschränken wir uns hier ausschließlich auf den Fall des Liklihoods bzw. des *Quasi*-Likelihoods.

3.1 Likelihood und Scorevektor

Sei $f(x_1, \ldots, x_n, \vartheta)$ die gemeinsame Dichte der $X_1, \ldots, X_n$. Dann wird die bedingte Dichte $f_{(i-1)}(x_i, \vartheta)$ von X_i, gegeben $X_1, \ldots, X_{i-1}$ gemäß II 5.3 durch

$$f_{(i-1)}(x_i, \vartheta) \;=\; \frac{f(x_1, \ldots, x_i, \vartheta)}{f(x_1, \ldots, x_{i-1}, \vartheta)} \qquad [f_{(0)}(x_1, \vartheta) \equiv f(x_1, \vartheta)]$$

definiert. Die Bezeichnungen $\mathbb{E}_{(i-1),\vartheta}$, $\mathrm{Var}_{(i-1),\vartheta}$, $\mathbb{V}_{(i-1),\vartheta}$ beziehen sich dann auf diese Dichten, insbesondere gelten mit der Abkürzung $X^{(i-1)} = (X_1, \ldots, X_{i-1})$ die Notationen

$$\mathbb{E}_{(i-1),\vartheta}\,(\cdot) = \mathbb{E}_\vartheta\,(\,\cdot\,|X^{(i-1)}), \quad \mathrm{Var}_{(i-1),\vartheta}\,(\cdot) = \mathbb{E}_\vartheta\left((\cdot\, - \,\mathbb{E}_{(i-1),\vartheta}(\cdot))^2|X^{(i-1)}\right).$$

Wir führen Vertauschbarkeitsbedingungen wie in V 1.1 ein, jetzt aber für bedingte Dichten formuliert $(\int \equiv \int_{-\infty}^{\infty})$.

$$\frac{d}{d\vartheta}\int f_{(i-1)}(x, \vartheta)\,dx \;=\; \int \frac{d}{d\vartheta} f_{(i-1)}(x, \vartheta)\,dx\,, \qquad i = 1, 2, \ldots\,. \qquad V^{(1)}$$

$V^{(2)}$ möge die gleiche Bedingung mit zweiten anstatt ersten Ableitungen bezeichnen. Die linke Seite von $V^{(1)}$ und von $V^{(2)}$ ist gleich Null. Die log-Likelihoodfunktion $\ell_n(\vartheta)$ einer Beobachtung $X_1, \ldots, X_n$ lautet

$$\ell_n(\vartheta) \;=\; \sum_{i=1}^{n} \log f_{(i-1)}(X_i, \vartheta)\,.$$

Mit der Abkürzung

$$u_i(\vartheta) \;=\; \frac{d}{d\vartheta}\log f_{(i-1)}(X_i, \vartheta) \;=\; \frac{\frac{d}{d\vartheta} f_{(i-1)}(X_i, \vartheta)}{f_{(i-1)}(X_i, \vartheta)} \tag{3.1}$$

läßt sich der Scorevektor $U_n(\vartheta) = \frac{d}{d\vartheta}\ell_n(\vartheta)$ in der Form

$$U_n(\vartheta) \;=\; \sum_{i=1}^{n} u_i(\vartheta) \tag{3.2}$$

schreiben. Unter der stillschweigend gemachten Voraussetzung, dass $\mathbb{E}_\vartheta\,|u_i(\vartheta)|^2 < \infty$ für alle $i = 1, 2 \ldots$ gilt, haben wir auch $\mathbb{E}_\vartheta\,|U_n(\vartheta)|^2 < \infty$ für alle $n = 1, 2 \ldots$. Der Scorevektor U_n erweist sich als ein Martingal, die u_i's also als Martingaldifferenz-Folge, denn es gilt das

Lemma. *Unter der Bedingung $V^{(1)}$ besitzt die Folge $U_n(\vartheta)$, $n \geq 1$, von Zufallsvektoren die Martingaleigenschaft*

$$\mathbb{E}_{(n-1),\vartheta}\, U_n(\vartheta) \;=\; U_{n-1}(\vartheta) \qquad \mathbb{P}_\vartheta - f.s., \qquad n = 2, 3, \ldots . \tag{3.3}$$

Ferner gilt $\mathbb{E}_\vartheta\, U_n(\vartheta) \;=\; 0.$

Beweis. Aufgrund der Darstellung (3.2) ist (3.3) erfüllt, falls

$$\begin{aligned} \mathbb{E}_{(n-1),\vartheta}\, u_i(\vartheta) &= u_i(\vartheta) \quad \mathbb{P}_\vartheta - f.s. \quad \text{für } i = 1, \ldots, n-1 \\ \mathbb{E}_{(n-1),\vartheta}\, u_n(\vartheta) &= 0 \quad \mathbb{P}_\vartheta - f.s. \quad \text{für } n = 1, 2, \ldots . \end{aligned}$$

Während die erste Gleichung aus einer Eigenschaft der bedingten Erwartung folgt (für $i = 1, \ldots, n-1$ ist $u_i(\vartheta)$ messbare Funktion der $X_1, \ldots, X_{n-1}$), ergibt sich die zweite aus $V^{(1)}$, denn

$$\mathbb{E}_{(n-1),\vartheta}\, u_n(\vartheta) \;=\; \int \Big(\tfrac{d}{d\vartheta}\, \log f_{(n-1)}(x,\vartheta)\Big) \cdot f_{(n-1)}(x,\vartheta)\, dx \;=\; \int \tfrac{d}{d\vartheta}\, f_{(n-1)}(x,\vartheta)\, dx \;=\; 0.$$

Ferner $\mathbb{E}_\vartheta\, u_i(\vartheta) \;=\; \mathbb{E}_\vartheta\big(\mathbb{E}_{(i-1),\vartheta}\, u_i(\vartheta)\big) \;=\; 0$, $i = 1, 2, \ldots$, und deshalb über (3.2) auch $\mathbb{E}_\vartheta\, U_n(\vartheta) \;=\; 0.$ □

Die $d \times d$-Hessematrix von $\ell_n(\vartheta)$ ist

$$W_n(\vartheta) \;=\; \sum_{i=1}^n w_i(\vartheta), \qquad w_i(\vartheta) \;=\; \frac{d^2}{d\vartheta d\vartheta^\top} \log f_{(i-1)}(X_i, \vartheta)\,.$$

Unter $V^{(1)}$ und $V^{(2)}$ rechnet man ähnlich wie in V 1.1

$$\mathbb{E}_{(i-1),\vartheta}\, w_i(\vartheta) \;=\; -\mathbb{V}_{(i-1),\vartheta}\, \big(u_i(\vartheta)\big)\,. \tag{3.4}$$

3.2 Quasi-Likelihood, Bedingung U*

Auch ohne explizite Kenntnis der bedingten Dichten $f_{(i-1)}$ lassen sich asymptotische Ergebnisse für stochastische Prozesse gewinnen, wenn wir von einer *estimation function* $U_n(\vartheta)$ der Form (3.2) ausgehen, mit einer Martingaldifferenz-Folge $u_i(\vartheta)$, $i \geq 1$, bez. $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$. Das führt zur

Definition. *Quasi*-Likelihood (GEE-Typ): Für die *estimation function* $U_n(\vartheta)$ gilt

$$U_n(\vartheta) = \sum_{i=1}^n u_i(\vartheta), \quad u_i(\vartheta) \text{ ist } \mathcal{F}_i\text{-messbar, mit } E_{(i-1),\vartheta}(u_i(\vartheta)) = 0. \tag{M*}$$

Nach 3.1 ist M* im Likelihood-Fall, d. h. wenn (3.1) zusammen mit der Vertauschbarkeitsbedingung $V^{(1)}$ gelten, erfüllt.

Wir stellen einige Bedingungen auf, die in Hinblick auf U*, das ist die Verteilungskonvergenz von $\Gamma_n \cdot U_n(\vartheta)$ gegen die $N_d(0,\Sigma(\vartheta))$-Verteilung, von Bedeutung sind. Dazu setzen wir –mit einer Folge Γ_n, $n \geq 1$, von $d \times d$-Diagonalmatrizen mit positiven Diagonalelementen, die gegen Null konvergieren– zur Abkürzung

$$L_n(\vartheta,\varepsilon) = \sum_{i=1}^{n} \mathbb{E}_{(i-1),\vartheta}\left(1(|\Gamma_n u_i(\vartheta)| > \varepsilon) \cdot |\Gamma_n u_i(\vartheta)|^2\right)$$

$$V_n(\vartheta) = \sum_{i=1}^{n} v_i(\vartheta), \quad v_i(\vartheta) = \mathbb{V}_{(i-1),\vartheta}\,(u_i(\vartheta)) \qquad [d \times d\text{-Matrizen}].$$

Wir formulieren nun für jedes $\vartheta \in \Theta$ und für den Limes $n \to \infty$

$$L_n(\vartheta,\varepsilon) \xrightarrow{\mathbb{P}_\vartheta} 0 \qquad \text{für jedes } \varepsilon > 0 \qquad U_1{}^*$$

$$\Gamma_n \cdot V_n(\vartheta) \cdot \Gamma_n^\top \xrightarrow{\mathbb{P}_\vartheta} \Sigma(\vartheta) \qquad [\text{positiv-definite, symm. } d \times d\text{-Matrix}] \qquad U_2{}^*$$

Bemerkung. $U_1{}^*$ ist eine bedingte *Lindeberg*-Bedingung, $U_2{}^*$ stellt eine *ergodische* Bedingung dar, $V_n(\vartheta)$ heißt auch bedingte Fisher-Information(smatrix).

Satz 1. *Aus den Bedingungen M*, $U_1{}^*$ und $U_2{}^*$ folgt die Bedingung U*, das ist* $\Gamma_n \cdot U_n(\vartheta) \xrightarrow{\mathcal{D}_\vartheta} N_d(0,\Sigma(\vartheta))$.

Beweis. Mit der Festsetzung $Y_{n,i}(\vartheta) = \Gamma_n \cdot u_i(\vartheta)$ bilden wir das Dreiecksschema

$$Y_{n,1}(\vartheta), \ldots, Y_{n,n}(\vartheta), \qquad n = 1,2,\ldots .$$

Wegen $\mathbb{E}_{(i-1),\vartheta}\, Y_{n,i}(\vartheta) = 0$ gemäß M* stellt es ein *Martingal*differenz-Schema dar bezüglich $\mathcal{F}_{n,i} \equiv \mathcal{F}_i = \sigma(X_1,\ldots,X_i)$. Es gilt dann

$$\Gamma_n \cdot U_n(\vartheta) = \sum_{i=1}^{n} Y_{n,i}(\vartheta)$$

$$\sum_{i=1}^{n} \mathbb{V}_{(i-1),\vartheta}\,(Y_{n,i}(\vartheta)) = \Gamma_n \cdot V_n(\vartheta) \cdot \Gamma_n^\top .$$

Im Fall $d = 1$ sagt $U_2{}^*$ aus, dass $[\Sigma(\vartheta) \equiv \sigma^2(\vartheta)$ gesetzt$]$

$$\sum_{i=1}^{n} \mathrm{Var}_{(i-1),\vartheta}\,(Y_{n,i}(\vartheta)) \xrightarrow{\mathbb{P}_\vartheta} \sigma^2(\vartheta)\,,$$

das ist (A 3.5), während U_1* dann (A 3.6) entspricht. Die Voraussetzungen des ZGWS A 3.7 für Martingaldifferenz-Schemata sind also erfüllt, so dass

$$\sum_{i=1}^{n} Y_{n,i}(\vartheta) \xrightarrow{\mathcal{D}_\vartheta} N(0,\sigma^2(\vartheta))\,.$$

Im allgemeinen Fall $d \geq 1$ beweist man, dass für alle Vektoren $a \in \mathbb{R}^d$

$$\sum_{i=1}^{n} a^\top \cdot Y_{n,i}(\vartheta) \xrightarrow{\mathcal{D}_\vartheta} a^\top Z, \qquad Z \quad N_d(0,\Sigma(\vartheta)) - \text{verteilt},$$

so dass Satz 1 aus A 3.4, das ist der Cramér-Wold device, die Behauptung liefert. □

3.3 Autoregressiver Prozess AR(1)

Als einfaches Beispiel für einen stochastischen Prozess betrachten wir den autoregressiven Prozess der Ordnung 1 (kurz: AR(1)-Prozess) mit Erwartungswert 0. Erst in 3.4 werden wir einen Erwartungswert μ berücksichtigen.

Modell. Unter einem AR(1)-Prozess verstehen wir eine Folge $X_0, X_1, \ldots$ von Zufallsvariablen, welche der Gleichung

$$X_n = \vartheta \cdot X_{n-1} + e_n, \quad n = 1, 2 \ldots \tag{3.5}$$

genügt. Dabei ist $\vartheta \in (-1, 1)$ und

$$\begin{aligned} e_1, e_2, \ldots \quad & \text{ist eine Folge unabh., identisch verteilter Zufallsvariablen mit} \\ & \mathbb{E}(e_n) = 0, \ \mathrm{Var}(e_n) = \sigma_e^2\,, \\ & e_n \text{ unabhängig von } X_0, \ldots, X_{n-1}. \end{aligned} \tag{3.6}$$

Ein solcher Prozess ist *stationär* und hat die Darstellung

$$X_n = \sum_{i=0}^{\infty} \vartheta^i e_{n-i}, \quad n = 1, 2 \ldots . \qquad [L_2\text{–Limes}].$$

Es ergeben sich für $n = 1, 2 \ldots$ die folgenden unbedingten und bedingten Momente

$$\begin{aligned} \mathbb{E}_\vartheta\,(X_n) &= 0, & \mathbb{E}_\vartheta\,(X_n|X_{n-1}) &= \vartheta \cdot X_{n-1} \\ \mathrm{Var}_\vartheta(X_n) \equiv \sigma_x^2 &= \sum_{i=0}^{\infty} \vartheta^{2i}\,\sigma_e^2 = \frac{1}{1-\vartheta^2}\,\sigma_e^2, & \mathrm{Var}_\vartheta(X_n|X_{n-1}) &= \sigma_e^2\,. \end{aligned}$$

Der Prozess X_n, $n \geq 1$, unterliegt den folgenden ergodischen Gesetzen (Gesetze der großen Zahlen):

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}_\vartheta} \mathbb{E}_\vartheta\,(X_1) = 0\,, \tag{3.7}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{\mathbb{P}_\vartheta} \mathbb{E}_\vartheta\,(X_1^2) = \mathrm{Var}_\vartheta(X_1) \equiv \sigma_x^2\,, \tag{3.8}$$

vgl. Schlittgen & Streitberg (1984, sec. 4.1); es gilt sogar die $\mathbb{P}_\vartheta$ fast-sichere Konvergenz. In der Extremwerttheorie zeigt man, dass für eine ganze Reihe von Verteilungen der Fehlervariablen e_i, darunter auch die Normalverteilung,

$$\frac{1}{\sqrt{n}} \max(|X_1|, \ldots, |X_n|) \xrightarrow{\mathbb{P}_\vartheta} 0, \tag{3.9}$$

gilt, vergleiche dazu Embrechts et al (1997, Ex. 3.5.4, sec. 5.5.2).

Nun setzen wir zusätzlich voraus:

$$\text{Jedes } e_i \text{ ist } N(0, \sigma_e^2)\text{-verteilt.} \tag{3.10}$$

Die bedingte Verteilung von X_i, gegeben $X_0, \ldots, X_{i-1}$, ist dann die $N(\vartheta X_{i-1}, \sigma_e^2)$-Verteilung; zum Nachweis verwende man die letzte Gleichung in II 5.1, formuliert für bedingte Wahrscheinlichkeiten. Die zugehörige (bedingte) Dichte lautet

$$f_{(i-1)}(x_i, \vartheta) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\Big\{ -\frac{1}{2}\Big(\frac{x_i - \vartheta X_{i-1}}{\sigma_e}\Big)^2\Big\},$$

die gemeinsame Dichte der $X_1, \ldots, X_n$ also

$$f(x_1, \ldots, x_n, \vartheta) = f(x_1, \vartheta) \cdot f_{(1)}(x_2, \vartheta) \cdot \ldots \cdot f_{(n-1)}(x_n, \vartheta).$$

Statistische Analyse. Aus dem Likelihood $f(x_1, \ldots, x_n, \vartheta)$ ergeben sich die log Likelihoodfunktion $\ell_n(\vartheta)$, die Scorefunktion $U_n(\vartheta)$ und die zweite Ableitung $W_n(\vartheta)$, mit einem R, das nicht von ϑ abhängt,

$$\ell_n(\vartheta) = -\frac{1}{2\sigma_e^2} \sum_{i=1}^n (X_i - \vartheta X_{i-1})^2 + R, \tag{3.11}$$

$$U_n(\vartheta) = \frac{1}{\sigma_e^2} \sum_{i=1}^n X_{i-1} \cdot (X_i - \vartheta X_{i-1}) \equiv \sum_{i=1}^n u_i(\vartheta)$$

$$W_n(\vartheta) = -\frac{1}{\sigma_e^2} \sum_{i=1}^n X_{i-1}^2.$$

$U_n(\vartheta)$ ist eine 1-dimensionale –in ϑ lineare– *estimation* function, $u_i(\vartheta)$ ist ein Zufalls-gewichtetes Martingaldifferenz-Schema im Sinne von (A 3.7), $W_n(\vartheta)$ hängt nicht mehr von ϑ ab. Der ML-Schätzer für ϑ berechnet sich aus $U_n(\vartheta) = 0$ zu

$$\hat{\vartheta}_n = \frac{\sum_{i=1}^n X_{i-1} \cdot X_i}{\sum_{i=1}^n X_{i-1}^2}; \tag{3.12}$$

das ist eine Version $r(1)$ des Koeffizienten der Auto-Korrelation der Ordnung 1 für Prozesse mit Erwartungswert 0.

Wir überprüfen die Bedingungen U* und W*. Hier im Fall $d = 1$ setzen wir $\Gamma_n = \gamma_n$.

Satz. *Gegeben sei ein autoregressiver Prozess gemäß Gleichung (3.5) bis (3.10), mit* $\vartheta \in (-1,1)$. *Dann sind die Bedingungen* U^* *und* W^* *erfüllt, und zwar mit*

$$\gamma_n = \frac{1}{\sqrt{n}}, \qquad \Sigma(\vartheta) = B(\vartheta) = \frac{1}{1-\vartheta^2} \quad [\equiv \sigma^2(\vartheta)].$$

Beweis. (i) Wir zeigen zunächst mittels Lemma A 3.7, dass die bedingte Lindeberg-Bedingung $U_1{}^*$ erfüllt ist, wenn wir in (A 3.7)

$$A_{n,i} = \frac{1}{\sqrt{n}} \cdot \frac{X_{i-1}}{\sigma_e^2}, \qquad e_i = X_i - \vartheta \cdot X_{i-1}\,,$$

setzen (beachte, dass $\mathbb{E}_\vartheta(e_i|X_{i-1}) = \mathbb{E}_\vartheta(e_i) = 0$). In der Tat, die Momentenbedingungen (A 3.9) reduzieren sich hier auf

$$\mathrm{Var}_\vartheta(e_1) > 0\,, \qquad \mathbb{E}(|e_1|^\kappa) < \infty \quad \text{für ein } \kappa > 2\,; \tag{3.13}$$

sie sind wegen unserer Annahme (3.10) erfüllt. Auf Grund des ergodischen Gesetzes (3.8) ist (A 3.10) wegen A 3.4, Prop. 1 (i), erfüllt ist. Auf Grund der Extremwert-Eigenschaft (3.9) gilt auch (A 3.11).
(ii) Mit $u_i(\vartheta) = X_{i-1} \cdot e_i/\sigma_e^2$ erhalten wir für $v_i(\vartheta) = \mathrm{Var}_{(i-1),\vartheta}(u_i(\vartheta))$ die Gleichung

$$v_i(\vartheta) = \mathrm{Var}_\vartheta(e_i|X_{i-1})\,\frac{X_{i-1}^2}{\sigma_e^4} = \frac{X_{i-1}^2}{\sigma_e^2}\,,$$

so dass $U_2{}^*$ wegen (3.8) gültig ist, mit

$$\sigma^2(\vartheta) = \frac{\sigma_x^2}{\sigma_e^2} = \frac{1}{1-\vartheta^2}\,.$$

Nach Satz aus 4.1 ist damit die Bedingung U* erfüllt.
(iii) Die Bedingung W*, mit $B(\vartheta) = \sigma^2(\vartheta)$, ist wegen Aussage (3.8) erfüllt. □

Der ML-Schätzer (3.12) weist also $\sqrt{n}$-Konsistenz für ϑ auf; ferner asymptotische Normalität der Form

$$\sqrt{n}(\hat\vartheta_n - \vartheta) \xrightarrow{\mathcal{D}_\vartheta} N(0, 1-\vartheta^2) \ \text{ bzw. } \sqrt{\frac{n}{1-\hat\vartheta_n^2}}(\hat\vartheta_n - \vartheta) \xrightarrow{\mathcal{D}_\vartheta} N(0,1)\,. \tag{3.14}$$

Der obige Satz gilt auch im Quasi-Liklihood Fall, wenn wir also auf die Normalverteilungs-Annahme (3.10) verzichten und gleich mit (3.11) als Kriteriumsfunktion starten (R kann 0 gesetzt werden). Für die Verteilung von e_1 müssen dann nur noch die (schwachen) Annahmen (3.13) und (3.9) getroffen werden. (Tatsächlich ist auch die Gültigkeit von (3.9) durch die Verteilung von e_1 bestimmt). (3.12) ist dann ein Quasi ML-Schätzer (und zwar vom GEE-Typ).

Anwendung. Wir benutzen jetzt die in der Zeitreihenanalyse üblichen Bezeichnungen $\vartheta = \rho(1)$ und $\hat{\vartheta}_n = r(1)$. Der approximative Standardfehler des Schätzers $r(1)$ ist nach (3.14)

$$\operatorname{se}(r(1)) \;=\; \sqrt{\frac{1-(r(1))^2}{n}}\,.$$

Er ist angenähert gleich $\sqrt{-1/\hat{W}_n(\hat{\vartheta}_n)}$, wobei wir in die Formel für $W_n(\vartheta)$ den Schätzer $\hat{\sigma}_e^2 = (1/n)\sum_{i=1}^n (X_i - \hat{\vartheta}_n X_{i-1})^2$ für σ_e^2 eingesetzt haben. Ein Konfidenzintervall für $\rho(1)$ zum asymptotischen Niveau $1-\alpha$ lautet

$$r(1) \;-\; \operatorname{se}(r(1))\cdot u_{1-\alpha/2} \;\le\; \rho(1) \;\le\; r(1) \;+\; \operatorname{se}(r(1))\cdot u_{1-\alpha/2}\,. \tag{3.15}$$

Betrachte jetzt die Hypothese $H_0 : \rho(1) = 0$, das bedeutet die Unkorreliertheit der aufeinanderfolgenden Beobachtungen $X_1, X_2, \ldots$. Ein Test zum Prüfen von H_0 zum asymptotischen Signifikanzniveau α verwendet die Teststatistik

$$Z = \sqrt{n}\,\frac{r(1)}{\sqrt{1-r^2(1)}}\,,$$

die gemäß (3.14) unter H_0 asymptotisch $N(0,1)$-verteilt ist.

3.4 AR(1)-Prozess mit Mittelwert-Korrektur

Wir gehen jetzt zu dem realistischeren Fall über, dass die Variablen X_n einen Erwartungswert μ besitzt, der nicht notwedig –wie in 3.3– gleich 0 ist.

Modell. Ein AR(1)-Prozess mit *Mittelwert-Korrektur* gehorcht der Gleichung

$$X_n - \mu \;=\; \alpha\cdot(X_{n-1}-\mu) + e_n, \quad n = 1,2,\ldots\,, \tag{3.16}$$

bzw. $X_n = \alpha\cdot X_{n-1} + \mu\,(1-\alpha) + e_n$. Hierbei ist $\alpha \in (-1,1)$ und $\mu \in \mathbb{R}$, und es gelten wieder die Voraussetzungen (3.6) an die Fehlervariablen e_i (keine Normalverteilungs-Annahme). Es ergibt sich die Darstellung $X_n = \mu + \sum_{i=0}^{\infty} \alpha^i\, e_{n-i}$ und, mit dem zweidimensionalen Parametervektor $\vartheta = (\mu, \alpha)$,

$$\mathbb{E}_\vartheta(X_n) = \mu\,, \qquad \mathbb{E}_\vartheta(X_n|X_{n-1}) = \alpha\cdot X_{n-1} + \mu(1-\alpha),$$

während die Varianz-Formeln wie in 3.3 lauten, insbesondere $\sigma_e^2 = (1-\alpha^2)\sigma_x^2$. Ebenso gelten die ergodischen Gesetze (3.7), mit $\mathbb{E}_\vartheta(X_1) = \mu$, und (3.8), sowie für eine ganze Reihe von Verteilung der e_i die Extremwert-Eigenschaft (3.9).

Statistische Analyse (Quasi-Likelihood.) Ausgehend von der Kriteriumsfunktion

$$\ell_n(\vartheta) \;=\; -\frac{1}{2\sigma_e^2}\sum_{i=1}^n \big(X_i - \alpha X_{i-1} - \mu(1-\alpha)\big)^2\,, \qquad \vartheta = (\mu,\alpha)\,,$$

gelangen wir zum Gradientenvektor bzw. zur Hessematrix

$$U_n(\vartheta) \;=\; \frac{1}{\sigma_e^2}\sum_{i=1}^n \begin{pmatrix} 1-\alpha \\ X_{i-1}-\mu \end{pmatrix}\cdot\big(X_i - \alpha\cdot X_{i-1} - \mu(1-\alpha)\big)\,,$$

$$W_n(\vartheta) \;=\; -\frac{1}{\sigma_e^2}\sum_{i=1}^n \begin{pmatrix} v_{11}(\vartheta) & v_{12}(\vartheta) \\ v_{12}(\vartheta) & v_{22}(\vartheta) \end{pmatrix} \;+\; \begin{pmatrix} 0 & R_n' \\ R_n' & 0 \end{pmatrix},$$

($R_n' = c'(X_n - X_0)$ mit Konstante c'). Dabei hat die Matrix die Einträge

$$v_{11}(\vartheta) = (1-\alpha)^2\,, \quad v_{12}(\vartheta) = 2\,(X_{i-1}-\mu)(1-\alpha)\,, \quad v_{22}(\vartheta) = (X_{i-1}-\mu)^2\,.$$

Die Lösung $\hat\vartheta_n = (\hat\mu_n, \hat\alpha_n)$ der Schätzgleichung $U_n(\vartheta) = 0$ (vom GEE-Typ) lautet

$$\hat\mu_n \;=\; \frac{\sum_{i=1}^n (X_i - \alpha X_{i-1})}{n(1-\alpha)} \;=\; \frac{\sum_{i=1}^n X_i}{n} + \frac{1}{n}\,R_n'' \;\equiv\; \bar X_n + \frac{1}{n}\,R_n''$$

$$\hat\alpha \;=\; \frac{\sum_{i=1}^n (X_{i-1}-\hat\mu_n)\cdot(X_i - \hat\mu_n)}{\sum_{i=1}^n (X_{i-1}-\hat\mu_n)^2} \;=\; r(1)\,,$$

($R_n'' = c''(X_n - X_0)$ mit Konstante c''). Der Quasi ML-Schätzer $\hat\vartheta_n = (\hat\mu_n, \hat\alpha_n)$ für $\vartheta = (\mu,\alpha)$ besteht also (approximativ) aus dem Mittelwert und aus einer Version des Koeffizienten der Autokorrelation (Ordnung 1) der Stichprobe $X_1,\ldots,X_n$.

Proposition. *Gegeben sei ein autoregressiver Prozess mit Mittelwert-Korrektur gemäß Gleichung (3.16), mit $\alpha \in (-1,1)$. Ferner mögen (3.13) und (3.9) gelten. Dann sind die Bedingungen U^* und W^* erfüllt, und zwar mit*

$$\Gamma_n \;=\; Diag\Big(\frac{1}{\sqrt n}, \frac{1}{\sqrt n}\Big)\,, \qquad \Sigma(\vartheta) \;=\; B(\vartheta) \;=\; \begin{pmatrix} (1-\alpha)^2/\sigma_e^2 & 0 \\ 0 & 1/(1-\alpha^2) \end{pmatrix}\,.$$

Beweis. Der Beweis verläuft wie der zum Satz 3.3 oben, jedoch mit folgenden Erweiterungen.

ad U_2*: Es ist

$$U_n(\vartheta) = \sum_{i=1}^n u_i(\vartheta), \quad \text{mit} \quad u_i(\vartheta) \;=\; \frac{1}{\sigma_e^2}\begin{pmatrix} 1-\alpha \\ X_{i-1}-\mu \end{pmatrix}\cdot e_i\,.$$

Für $v_i(\vartheta) = \mathbb{V}_{(i-1),\vartheta}(u_i(\vartheta))$ erhalten wir wegen $\mathrm{Var}_{(i-1)}(e_i) = \sigma_e^2$

$$v_i(\vartheta) \;=\; \frac{1}{\sigma_e^2}\begin{pmatrix} (1-\alpha)^2 & (1-\alpha)(X_{i-1}-\mu) \\ (1-\alpha)(X_{i-1}-\mu) & (X_{i-1}-\mu)^2 \end{pmatrix}.$$

Die ergodischen Gesetze (3.7) und (3.8) liefern dann

$$\frac{1}{n}\sum_{i=1}^{n} v_i(\vartheta) \xrightarrow{\mathbb{P}_\vartheta} \frac{1}{\sigma_e^2}\begin{pmatrix} (1-\alpha)^2 & 0 \\ 0 & \sigma_x^2 \end{pmatrix} \;=\; \Sigma(\vartheta)\,,$$

Letzeres wegen $\sigma_x^2/\sigma_e^2 = 1/(1-\alpha^2)$.

ad W*: Aufgrund der ergodischen Gesetze (3.7) und (3.8) gilt

$$-\frac{1}{n}\,W_n(\vartheta) \xrightarrow{\mathbb{P}_\vartheta} \frac{1}{\sigma_e^2}\begin{pmatrix} (1-\alpha)^2 & 0 \\ 0 & \sigma_x^2 \end{pmatrix} \;=\; \Sigma(\vartheta)\,.$$

Diese Konvergenzaussage lässt sich auch mit $W_n(\vartheta_n^*)$, $\vartheta_n^* \xrightarrow{\mathbb{P}} \vartheta$, anstelle von $W_n(\vartheta)$ aufstellen. Für das Element $(2,2)$ etwa verwende man $\sum_{i=1}^n (X_{i-1} - \mu_n^*)^2/n = \sum_{i=1}^n (X_{i-1}-\mu)^2/n + \eta_n$, η_n stochastische Nullfolge. Insbesondere ist damit auch $B(\vartheta) = \Sigma(\vartheta)$ bewiesen. □

Anwendung. Test der (zusammengesetzten) Hypothese $H_0 : \mu = \mu_0$. Mit $\vartheta = (\mu, \alpha)$ schreiben wir

$$H_0 : r(\vartheta) \;=\; 0\,, \qquad r(\vartheta) = \mu - \mu_0\,.$$

Mit $\hat\vartheta_n = (\hat\mu_n, \hat\alpha_n)$, $\hat\mu_n = \bar X_n + (const/n)(X_n - X_0)$, gilt für die Wald-Statistik gemäß VI 4.4

$$T_n^{(W)} \;=\; \frac{\left(\hat\mu_n - \mu_0\right)^2}{\hat w_n}\,, \quad \text{wobei} \quad \hat w_n \;=\; \left[-\,W_n^{-1}(\hat\vartheta_n)\right]_{(1,1)}.$$

W* und $\sigma_e^2 = (1-\alpha^2)\sigma_x^2$ berücksichtigend, haben wir

$$\left[-\,n\,W_n^{-1}(\hat\vartheta_n)\right]_{(1,1)} \xrightarrow{\mathbb{P}_\vartheta} \left[\Sigma^{-1}(\vartheta)\right]_{(1,1)} \;=\; \frac{\sigma_e^2}{(1-\alpha)^2} \;=\; \frac{1+\alpha}{1-\alpha}\,\sigma_x^2\,.$$

Bedeutet $A_n \approx B_n$, dass A_n/B_n stochastisch gegen 1 konvergiert, so lässt sich

$$\hat w_n \;\approx\; \frac{1}{n}\,\frac{1+\hat\alpha_n}{1-\hat\alpha_n}\,\hat\sigma_x^2 \qquad [\;\hat\sigma_x^2 \text{ ein konsistenter Schätzer für } \sigma_x^2]$$

schreiben. Wir erhalten

$$T_n^{(W)} \;\approx\; n_\alpha\,\frac{(\bar X_n - \mu_0)^2}{\hat\sigma_x^2}\,, \qquad \text{mit} \quad n_\alpha \;=\; \frac{1-\hat\alpha_n}{1+\hat\alpha_n}\,n\,.$$

$T^{(W)}$ ist unter H_0 asymptotisch χ_1^2-verteilt. Anders formuliert,

$$Z \;=\; \sqrt{n_\alpha}\,\frac{\bar X - \mu_0}{\hat\sigma_x} \quad \text{ist unter } H_0 \text{ asymptotisch } N(0,1)\text{-verteilt.}$$

Dies ist der *asymptotische* t-Test mit *korrigiertem n*: Bei Vorhandensein einer Autokorrelation $\alpha \in (0,1)$ reduziert sich der „effektive" Stichprobenumfang um den Faktor $(1-\alpha)/(1+\alpha)$.

3.5 Endliche Markovketten (*)

Modell. Wir betrachten eine Markovkette X_n, $n \geq 0$, mit endlich vielen, nämlich I, Zuständen. Das bedeutet: Mit einem Wahrscheinlichkeitsvektor (*Startverteilung*) $p_0 = (p_{0,1}, \ldots, p_{0,I})$ und einer $I \times I$-Matrix P von *Ubergangswahrscheinlichkeiten*,

$$P = \big(p_{i,j},\ i,j = 1, \ldots, I\big), \qquad [\text{alle } p_{i,j} \geq 0, \quad \sum_{j=1}^{I} p_{i,j} = 1]$$

gilt $\mathbb{P}(X_0 = i) = p_{0,i}$, $i = 1, \ldots, I$, und

$$\mathbb{P}(X_n = j | X_{n-1} = i) = p_{i,j}\,, \qquad i,j = 1, \ldots, I\,.$$

Einfachhaltshalber setzen wir im Folgenden voraus, dass alle Elemente von P positiv sind, $p_{i,j} > 0$ für alle i,j. Alternativ läßt sich auch fordern, dass –mit einer festen Menge D von Doppelindizes– $p_{i,j} > 0$ genau für $(i,j) \in D$ gilt, wobei D so beschaffen ist, dass die Markovkette *irreduzubel* ist (vgl. Feller 1970, chap. XV). Dann gibt es einen (eindeutig bestimmten) Wahrscheinlichkeitsvektor $\pi = (\pi_1, \ldots, \pi_I)$, so dass

$$\pi \cdot P = \pi \qquad\qquad [\pi \text{ \textit{stationäre} Verteilung}]$$

gilt. Wählt man π als Startverteilung, so ist π auch die Verteilung für jede der Variablen $X_1, X_2, \ldots$. Unter Verwendung von π können wir ein Gesetz der großen Zahlen (*ergodisches* Gesetz) formulieren, vgl. Billingsley(1961). Für alle Startverteilungen und messbare Funktionen $g : I \times I \to \mathbb{R}$ gilt

$$\frac{1}{n} \sum_{k=1}^{n} g(X_{k-1}, X_k) \xrightarrow{\mathbb{P}} \mathbb{E}_\pi\big(g(X_1, X_2)\big) = \sum_i \sum_j \pi_i \cdot p_{i,j} \cdot g(i,j)\,, \tag{3.17}$$

$$\text{insbesondere} \quad \frac{1}{n} \sum_{k=1}^{n} 1(X_{k-1} = i, X_k = j) \quad \xrightarrow{\mathbb{P}} \pi_i \cdot p_{i,j}\,. \tag{3.18}$$

Statistische Analyse. Wir setzen die Startverteilung p_0 als gegeben voraus und betrachten

$$\vartheta \equiv P^- = \Big(p_{i,j},\ i = 1, \ldots, I,\ j = 1, \ldots, I-1 : \quad \text{alle } p_{i,j} > 0, \quad \sum_{j=1}^{I-1} p_{i,j} < 1\Big)$$

$[p_{i,I} = 1 - (p_{i,1} + \ldots + p_{i,I-1})]$ als (unbekannten) Parameter der Dimension $d = I \cdot (I-1)$. Die Likelihoodfunktion einer Stichprobe $X_0, X_1, \ldots, X_n$ lautet

$$L_n(\vartheta) = p_{0,X_0} \cdot \prod_{k=1}^{n} p_{X_{k-1},X_k} = p_{0,X_0} \cdot \prod_{i=1}^{I} \prod_{i=1}^{I} p_{i,j}^{n_{i,j}}\,,$$

wobei $n_{i,j}$ die Häufigkeit des Übergangs $i \to j$ bezeichnet,

$$n_{i,j} = \sum_{k=1}^{n} 1(X_{k-1} = i, X_k = j)\,.$$

Die log-Likelihoodfunktion heißt dann

$$\ell_n(\vartheta) = \sum_{i=1}^{I} \sum_{j=1}^{I} n_{i,j} \log p_{i,j}\,.$$

Wir erhalten den Scorevektor $U_n(\vartheta)$ und die Hessematrix $W_n(\vartheta)$, Letzteres eine Block-Diagonalmatrix mit Blöcken der Dimension $(I-1) \times (I-1)$, zu

$$U_{n,(i,j)}(\vartheta) = \frac{n_{i,j}}{p_{i,j}} - \frac{n_{i,I}}{p_{i,I}}, \quad i = 1,\ldots,I,\ j = 1,\ldots,I-1$$

$$W_n(\vartheta) = \begin{pmatrix} W_{n,1}(\vartheta) & & 0 \\ & \ddots & \\ 0 & & W_{n,I}(\vartheta) \end{pmatrix}, \qquad W_{n,i\,(j,j')}(\vartheta) = -\left(\frac{n_{i,j}}{p_{i,j}^2}\,\delta_{j,j'} + \frac{n_{i,I}}{p_{i,I}^2}\right).$$

Die ML-Schätzer für die Übergangswahrscheinlichkeiten $p_{i,j}$ sind die relativen Übergangshäufigkeiten

$$\hat{p}_{i,j} = \frac{n_{i,j}}{n_{i,\bullet}}, \quad \text{mit} \quad n_{i,\bullet} = \sum_{j=1}^{I} n_{i,j}\,.$$

Wir weisen nun die Voraussetzungen U* und W* nach, und zwar mit der Block-Diagonalmatrix

$$\Sigma(\vartheta) = \begin{pmatrix} \Sigma_1(\vartheta) & & 0 \\ & \ddots & \\ 0 & & \Sigma_I(\vartheta) \end{pmatrix}, \qquad \Sigma_{i\,(j,j')}(\vartheta) = \left(\frac{\pi_i}{p_{i,j}}\,\delta_{j,j'} + \frac{\pi_i}{p_{i,I}}\right). \tag{3.19}$$

<u>ad U*:</u> Wir schreiben

$$U_{n,(i,j)}(\vartheta) = \sum_{k=1}^{n} u_{k,(i,j)}(\vartheta)\,,$$
$$u_{k,(i,j)}(\vartheta) = 1(X_{k-1} = i, X_k = j)/p_{i,j} - 1(X_{k-1} = i, X_k = I)/p_{i,I}\,,$$

und setzen $\mathbb{E}_{(i')}(u_k(\vartheta))$, $\mathbb{V}_{(i')}(u_k(\vartheta))$, für die bedingte Erwartung bzw. bedingte Kovarianzmatrix von $u_k(\vartheta)$, und zwar gegeben $X_{k-1} = i'$. Wir erhalten wegen $\mathbb{E}\big(1(X_{k-1} = i, X_k = j)\big) = \mathbb{P}(X_{k-1} = i) \cdot p_{i,j}$

$$\begin{aligned} \mathbb{E}_{(i')}\big(u_{k,(i,j)}(\vartheta)\big) &= 0\,, \qquad \text{für alle } i', i, j \\ \mathbb{V}_{(i')}\big(u_{k,(i,\cdot)}(\vartheta)\big) &= \begin{cases} \left(\frac{1}{p_{ij}}\,\delta_{j,j'} + \frac{1}{p_{iI}}\right)_{j,j'=1,\ldots,I-1} & \text{falls } i = i' \\ 0 & \text{falls } i \neq i' \end{cases} \end{aligned}$$

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{V}_{(X_{k-1})}\big(u_{k,(i,\cdot)}(\vartheta)\big) \xrightarrow{\mathbb{P}} \Sigma_i(\vartheta)\,,$$

Letzteres mit einer beliebigen Startverteilung. Gemäß A 3.7 ist damit U* erfüllt [setze $X_{n,k} = A_{n,k}\, e_k$ mit $A_{n,k} = 1/\sqrt{n}$, $e_k = u_k(\vartheta)$ in (A 3.8) und Lemma A 3.7; die korrekte Behandlung des hier vorliegenden multivariaten Falles mittels Cramér-Wold device unterlassen wir].

ad W*: Gemäß (3.18) erhalten wir wegen $\mathbb{E}_\pi\left(n_{i,j}\right) = n \cdot \pi_i \cdot p_{i,j}$

$$-\frac{1}{n}\, W_{n,i\,(j,j')}(\vartheta) \;\xrightarrow{\mathbb{P}}\; \left(\frac{\pi_i\, p_{i,j}}{p_{i,j}^2}\,\delta_{j,j'} + \frac{\pi_i\, p_{i,I}}{p_{i,I}^2}\right) \;=\; \Sigma_{i\,(i,j')}(\vartheta)\,,$$

und dasselbe auch mit $W_n(\vartheta_n^*)$, $\vartheta_n^* \xrightarrow{\mathbb{P}} \vartheta$, anstatt $W_n(\vartheta)$. Es gilt insbesondere $B(\vartheta) = \Sigma(\vartheta)$. Wir haben damit die folgende Aussage bewiesen.

Proposition. *Gegeben sei eine irreduzible Markovkette mit Übergangswahrscheinlichkeiten $P = (p_{i,j})$ und stationärer Verteilung $\pi = (\pi_j)$. Dann sind die Bedingungen U^* und W^* erfüllt, mit $\vartheta = P^-$ und mit $\Sigma(\vartheta) = B(\vartheta)$ gemäß (3.19).*

Anwendung. Wir wollen die Hypothese einer Folge unabhängiger $X_0, X_1, \ldots$ prüfen. Mit einem Wahrscheinlichkeitsvektor $q = (q_1, \ldots, q_I)$ heißt das

$$H_0: \; p_{i,j} \;=\; q_j\,, \quad \text{bzw. } p_{i,j} \;=\; h_j((q_1, \ldots, q_{I-1}))\,, \; i = 1, \ldots, I, \; j = 1, \ldots I-1,$$

mit $h_j : \mathbb{R}^{I-1} \to \mathbb{R}$ als Projektion auf die j-te Komponente.

Aus $\ell_n(h(q)) = \sum_{j=1}^{I} n_{\bullet,j}\, \log q_j$ erhalten wir den ML-Schätzer $\hat{q}_j = n_{\bullet,j}/n$ für q_j, und daraus die log-LQ Statistik

$$T_n \;=\; 2\sum_{i=1}^{I}\sum_{j=1}^{I} n_{i,j} \log\frac{n_{i,j}}{e_{i,j}}\,, \qquad e_{i,j} \;=\; \frac{n_{i,\bullet}\cdot n_{\bullet,j}}{n} \qquad \text{[erwartete Häufigkeit].}$$

Unter H_0 ist T_n asymptotisch χ^2_{d-c}-verteilt, mit $d-c = I\cdot(I-1)-(I-1) = (I-1)^2$ Freiheitsgrade. Formal unterscheidet sich dieser Test nicht vom χ^2-Unabhängigkeitstest VII 4.8 in einer $I \times I$-Kontingenztafel.